

# PERMBLEDHJE PER STUDENTET

## MASTER PROFESIONAL NË MENAXHIM SHËNDETËSOR

SHTATOR 2018



## Statistikë

Me mbështetjen e



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Agjencia Zvicerane për Zhvillim  
dhe Bashkëpunim SDC





## PËRMBATJA E LËNDËS

1. Hyrje ne Statistikë: Qëllimi dhe fazat e analizës statistikore
2. Vrojtimi me Pyetësor dhe  zgjedhja ( Mostra) në Statistikë
3. Njohja me SPSS dhe përgatitja e të dhënave për analizë
4. Statistika përshkruese 2
5. Probabiliteti dhe Teorema e Bayes-it 2
6. Shpërndarjet probabilitare 2
7. Testimi i Hipotezave (rasti i një popullimi): testet 't' 2
8. Testimi i Hipotezave (rasti i dy popullimeve): testet 't' 2
9. Testimi i Hipotezave (rasti i 'k' popullimeve): Anova (një dhe dy faktoriale), Analiza e matjeve të përsëritura
10. Analiza e Fuqisë
11. Testimi i Hipotezave: Metodët Jo parametrike – testet ' $\chi^2$ '
12. Analiza e Regresionit mbështetur ne OLS (Ordinary Least square)
13. Analiza Korelative: Kovarianca, Korelacioni Pearson dhe të tjerë koeficientë korelacioni
14. Regresioni Logjistik Binar
15. Analiza Diskriminante







REPUBLIKA E SHQIPËRISË  
UNIVERSITETI I TIRANËS  
FAKULTETI I EKONOMISË



MIRATOI

Prof. Dr. Kozeta Sevrani

Përgjegjës i Departamentit Statistikë dhe Informatikë e Zbatuar

Programi i Studimit: **Master Profesional në Menaxhim Shëndetësor**

Viti Akademik 2016-2017

| Titulli i lëndës : “STATISTIKË ”   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                        |         |          |        |
|------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|---------|----------|--------|
| Lloji i Lëndës                     | ECTS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Orë mësimi në auditor  |         |          | Totali |
|                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Leksion                | Seminar | Ushtrime |        |
| B                                  | 5                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 30                     |         | 30       | 60     |
| Viti/Semestri kur zhvillohet lënda |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Viti I, Semestri I     |         |          |        |
| Lektori/Pedagogu i lëndës          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Prof. Dr. Besa SHAHINI |         |          |        |
| Përshkrimi i lëndës                | <p>Lënda synon të pajisë studentin me formimin e duhur statistikor për të kuptuar, analizuar dhe për të marrë vendimet e duhura në menaxhimin e një prej sektorëve më jetikë të realitetit, sic është ai i kujdesit shëndetësor.</p> <p>Njohuritë që studentët do të marrin në këtë modul, do t'u shërbejnë si guidë për të ndërmarrë vendime në fushën e menaxhimit shëndetësor, të bazuara në metoda dhe teknika shkencore statistikore.</p> <p>Temat e këtij moduli do të adresojnë koncepte dhe procedura të përshtatshme për aplikimin e metodave dhe teknikave të ndryshme statistikore. Kjo do të pasohet nga aplikimi i këtyre metodave ndaj problemeve reale të menaxhimit ekonomik dhe social të sektorit të shëndetësisë.</p> <p>Lënda ka një ndërthurje të aspekteve teorike me aspektin praktik duke bërë që të rritet aftësia e studentëve për një përdorim efikas të analizave statistikore. Software kompjuterik, kryesisht SPSS, do të përdoret për ilustrim në zgjidhjen e problemeve statistikore, ku cdo temë do të shoqërohet me një problem konkret dhe demonstrimin e aplikimit të metodës përkatëse statistikore në SPSS, për zgjidhjen e tij.</p> <p>Lënda fillon me konceptin e bazës së të dhënave, njohje me metodat e vrojtimit dhe pyetësorët, hartimin e përpunimit të tyre. Më pas bëhet njohja e studentëve me softin SPSS, duke vazhduar më tej me informacion të detajuar mbi statistikat përshkruese, probabilitetin dhe shpërndarjet probabilitare. Këto njohuri do të shërbejnë si bazë për të vazhduar më tutje me aftësimin e studentëve për të kryer teste dhe analiza të sakta statistikore për të dhënat cilësore dhe sasiore, mbi bazën e të cilave kuptohet realiteti multidimensional në të cilin jetojmë, duke mundësuar një vendimarrje sa më efektive dhe efciente, sidomos në kuadrin e menaxhimit në sektorin e shëndetit publik. Cdo kapitull përmban hap pas hapi një guidë të detajuar në lidhje me testet statistikore dhe</p> |                        |         |          |        |

|                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                     | <p>analizën e marrëdhënies midis variablave në shqyrtim, ndërthurur me instruksione praktike dhe aplikim në software për të gjeneruar dhe interpretuar rezultatet .</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <p><b>Objektivat e lëndës</b></p>   | <p>Lënda synon t'ju jap studentëve njohuritë e duhura statistikore , duke mundësuar ata, për përdorimin sa më drejtë dhe saktë të koncepteve dhe teknikave statistikore për të kuptuar dhe analizuar realitetin social dhe ekonomik në të cilin jetojmë, në funksion të vendimmarrjes sa më eficiente në sektorin e menaxhimit të sektorit të shëndetit publik. Gjithashtu, synohet që studentët të mësojnë si të aplikojnë dhe të përdorin teknikat kompjuterike për zgjidhjen e problemeve në fushën e menaxhimit duke implementuar modelet e duhura statistikore.</p> <p>Objektivat e këtij moduli janë:</p> <ul style="list-style-type: none"> <li>• Të kuptohet rëndësia e statistikës në fushën e menaxhimit</li> <li>• Të pajise studentet me njohuritë e duhura si të mbledhen, organizohen, përpunohen, analizohen dhe të interpretohen të dhënat</li> <li>• Të zhvillojë aftësitë e studentëve për të identifikuar teknikat e duhura statistikore për analizën e të dhënave.</li> <li>• Të përdoret një program kompjuterik (SPSS) për analizën e të dhënave.</li> <li>• Të kuptohet , ku, kur dhe si do të aplikohet metoda e përshtatshme statistikore në situata të ndryshme vendimmarrjeje menaxhuese në biznes, ekonomi dhe në sektorin e shëndetësisë</li> </ul> |
| <p><b>Rezultatet e pritshme</b></p> | <p>Në përfundim të këtij kursi, pritet të jetë realizuar:</p> <ul style="list-style-type: none"> <li>- Aftësimi i studentëve për të përdorur shifrat dhe faktet në procesin e analizës dhe të vendimmarrjes.</li> <li>- Aftësimi i studentëve për të ndërtuar dhe realizuar një pyetësor gjatë një procesi vërtetimi</li> <li>- Aftësimi i studentëve për të përdorur programin kompjuterik SPSS</li> <li>- Aftësimi i studentëve për të realizuar zgjedhjen e duhur nga popullimi në cilësi dhe në madhësi.</li> <li>- Aftësimi i studentëve për të përdorur një bazë të dhënash statistikore dhe për të përzgjedhur treguesit e duhur për të argumentuar analizat statistikore</li> <li>- Aftësimi i studentëve për një paraqitje të drejtë dhe lehtësisht të kuptueshme të treguesve dhe metodave të ndryshme statistikore</li> <li>- Aftësimi i studentëve për të testuar hipoteza dhe për të marrë vendime eficiente menaxheriale</li> <li>- Aftësimi i studentëve për të kuptuar lidhjen mes variablave dhe për të menaxhuar me sukses bashkëveprimin e faktoreve, mbështetur në teknika shkencore statistikore</li> </ul>                                                                                                                                                 |
| <p><b>Mënyra e vlerësimit</b></p>   | <ul style="list-style-type: none"> <li>• <u>Vlerësimi i vazhduar 30 %</u> <ul style="list-style-type: none"> <li>▪ Aktivizimi 10%,</li> <li>▪ Detyra e kursit 10%</li> <li>▪ Provimi i ndërmjetëm 10 %</li> </ul> </li> <li>• <u>Provimi përfundimtar 70 %.</u></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <p><b>Literatura e detyruar</b></p> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

|                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                 | <ul style="list-style-type: none"> <li>- Cikël leksionesh përgatitur nga lektori</li> <li>- “Statistika”- Grup Autoresh, Shtëpia botuese Pegi, (2016)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Literatura e rekomanduar</b> | <ul style="list-style-type: none"> <li>- “Statistical Methods in Healthcare”-Frederick Faltin (Editor), Ron S. Kenett (Editor), Fabrizio Ruggeri (Editor), August 2012</li> <li>- “Applied Statistics”, Warner A.R , 2008</li> <li>- “How to Conduct Surveys”, Fink A. 2009</li> <li>- “Intermediate Statistics- A modern Approach”, Stevens J, 2009</li> <li>- “Statistical for Business and Economics”, Newbold P, Carlson W, Thorne B, 2011</li> <li>- “Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014</li> <li>- “Statistika Mjekësore”, E. Petrela</li> <li>- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011</li> </ul> |

| <b>PËRMBAJTJA E LËNDËS</b> |                                                                                                                  |              |
|----------------------------|------------------------------------------------------------------------------------------------------------------|--------------|
| <b>Java</b>                | <b>Temat teorike: leksione</b>                                                                                   | <b>orë L</b> |
| <b>1</b>                   | Hyrje ne Statistikë: Qëllimi dhe fazat e analizës statistikore.                                                  | 2            |
| <b>2</b>                   | Vrojtimi me Pyetësor dhe Zgjedhja ( Mostra) në Statistikë                                                        | 2            |
| <b>3</b>                   | Njohja me SPSS dhe përgatitja e të dhënave për analizë                                                           | 2            |
| <b>4</b>                   | Statistika përshkruese                                                                                           | 2            |
| <b>5</b>                   | Probabiliteti dhe Teorema e Bayes-it                                                                             | 2            |
| <b>6</b>                   | Shpërndarjet probabilitare                                                                                       | 2            |
| <b>7</b>                   | Testimi i Hipotezave (rasti i një popullimi): testet ‘t’                                                         | 2            |
| <b>8</b>                   | Testimi i Hipotezave (rasti i dy popullimeve): testet ‘t’                                                        | 2            |
| <b>9</b>                   | Testimi i Hipotezave (rasti i ‘k’ popullimeve): Anova (një dhe dy faktoriale) , Analiza e matjeve të përsëritura | 2            |
| <b>10</b>                  | Analiza e Fuqisë                                                                                                 | 2            |
| <b>11</b>                  | Testimi i Hipotezave: Metodatat Jo parametrike – testet ‘ $\chi^2$ ’                                             | 2            |
| <b>12</b>                  | Analiza e Regresionit mbështetur ne OLS (Ordinary Least square)                                                  | 2            |

|             |                                                                                                                                                           |                |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| 13          | Analiza Korelative: Kovarianca, Korelacioni Pearson dhe të tjerë koeficientë korelacioni                                                                  | 2              |
| 14          | Regresioni Logjistik Binar                                                                                                                                | 2              |
| 15          | Analiza Diskriminante                                                                                                                                     | 2              |
| <b>Java</b> | <b>Temat praktike (seminare/ushtime)</b>                                                                                                                  | <b>orë S/U</b> |
| 1           | Hyrje ne Statistikë: Qëllimi dhe fazat e analizës statistikore.                                                                                           | 2              |
| 2           | Vrojtimi me Pyetësor dhe Zgjedhja ( Mostra) në Statistikë                                                                                                 | 2              |
| 3           | Njohja me SPSS dhe përgatitja e të dhënave për analizë                                                                                                    | 2              |
| 4           | Statistika përshkruese                                                                                                                                    | 2              |
| 5           | Probabiliteti dhe Teorema e Bayes-it                                                                                                                      | 2              |
| 6           | Shpërndarjet probabilitare                                                                                                                                | 2              |
| 7           | Testimi i Hipotezave (rasti i një popullimi): testet 't'                                                                                                  | 2              |
| 8           | Testimi i Hipotezave (rasti i dy popullimeve): testet 't'                                                                                                 | 2              |
| 9           | Testimi i Hipotezave (rasti i 'k' popullimeve): Anova (një dhe dy faktoriale) , Analiza e matjeve të përsëritura                                          | 2              |
| 10          | Analiza e Fuqisë                                                                                                                                          | 2              |
| 11          | Testimi i Hipotezave: Metodot Jo parametrike – testet ' $\chi^2$ '                                                                                        | 2              |
| 12          | Analiza e Regresionit mbështetur ne OLS (Ordinary Least square)                                                                                           | 2              |
| 13          | Analiza Korelative: Kovarianca, Korelacioni Pearson dhe të tjerë koeficientë korelacioni                                                                  | 2              |
| 14          | Regresioni Logjistik Binar                                                                                                                                | 2              |
| 15          | Analiza Diskriminante                                                                                                                                     | 2              |
| <b>Java</b> | <b>Detyrë Kursi - Detyrim</b>                                                                                                                             |                |
| <b>X, Y</b> | Tematika e detyrës është e lirë për cdo student, në ndërtimin e një baze të dhënash dhe analizën e të dhënave nëpërmjet teknikave të mësuara statistikore |                |

**Pedagogu i Lëndës**

**Prof. Dr. Besa SHAHINI**

## **TEMA 1:**

### **Hyrje në Statistikë: Qëllimi dhe fazat e analizës statistikore**

#### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje që :

- Të kuptojnë ç'është dhe përse shërben Statistika
- Të kuptojnë terminologjitë si " Variabël, zgjedhje, popullim, vërtetim, bazë të dhënash, gabim statistikor etj
- Të dinë tipet e variablave dhe shkallët e matjeve të tyre

Të dinë ç'është dhe përse shërben Statistika descriptive dhe Statistika inferenciale

■

#### **LITERATURA**

-“ Applied Statistics”, Warner A.R , 2008

- Statistika”- Grup Autoresh, Shtëpia botuese Pegi, (2016)

- Statistical for Business and Economics”, Newbold P, Carlson W, Thorne B, 2011

- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011

-“Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014

## TEMA 1

### HYRJE NË STATISTIKË: QËLLIMI DHE FAZAT

#### 1. STATISTIKA DHE PËRDORIMI I SAJ

Shpesh në jetën e përditshme me statistikë kuptojmë faktet dhe shifrat që karakterizojnë dukuri të ndryshme masive të jetës ekonomiko-shoqërore, të tilla sip s. nr. i popullsisë dhe ritmet e ndryshimit të saj, zgjatja mesatare e jetës, koeficientët e lindshmërisë, të vdekshmërisë, të migrimit etj, prodhimit të përgjithshëm dhe të ardhurat kombëtare, shpenzimet për mjeksinë, numri i punonjësve , niveli i papunësisë etj. Megjithatë fusha e statistikës përfshin shumë më tepër sesa llogaritjen dhe paraqitjen e fakteve numerike. Statistika është shkenca që studion se si mbledhen të dhënat, se si përpunohen, paraqiten , analizohen dhe interpretohen ato.

Sot ekziston një informacion shumë i pasur statistikor, kuptimi dhe përdorimi siç duhet i të cilit është kyçi i suksesit për drejtuesit dhe sipërmarrësit e vendimeve. Në disa raste një studim statistikor do të ndihmonte për të kuptuar problemin aktual dhe marrjen e vendimit që lidhet me të. Në raste të tjera statistika jep një informacion shumë të vlefshëm për ambientin e përgjithshëm ekonomik dhe të biznesit Brenda të cilit duhet të merren vendimet. Statistika gjen zbatim të gjerë Kështu ajo përdoret në prodhim për planifikimin e prodhimit, për analizën e faktorëve ekonomikë dhe pritjen apo tendencën e tyre, në mjeksi për matjen e treguesve të shëndetit të njerëzve, në arsim për matjen e nivelit të edukimit etj. Vendimmarrësit përdorin gjerësisht informacionin statistikor. Një përdorim të gjerë ka statistika në fushën e mjeksisë. Përpara hedhjes në qarkullim të një medikamenti, ai testohet me zgjedhje në gjallesa të tjera përveç njeriut dhe më pas, në varësi të rezultatit përshtatet për qënien njerëzore. Në shumë raste ekonomistëve u kërkohet të bëjnë prognoza për ekonominë në tërësi ose për tregues të veçantë dhe për këtë ata përdorin një informacion të pasur statistikor. Metodatat statistikore përdoren gjerësisht për analizën e prirjeve të mëparshme, të gjendjes aktuale dhe prognozën e aktivitet ekonomik në të ardhmen për një varg treguesish makroekonomikë si psh. burimet natyrore dhe ato të punës, për rritjen ekonomike, për inflacionin etj.

#### 2. TË DHËNAT, DHE DISA KUPTIME TË TJERA THEMELORE

**Të dhënat** janë faktet dhe shifrat që grumbullohen , analizohen dhe përmbledhen, ato janë lënda e parë për statistikën. Të gjitha të dhënat që grumbullohen në një studim të veçantë përbëjnë atë që quhet bashkësia e të dhënave të studimit.

**Elementët** janë njësitë për të cilat mbledhen të dhënat. Psh. Nëse jemi të interesuar për punonjësit e një institucioni, për të cilin analizojmë vetëm 1 prej tyre, themi se çdo punonjës është një element i cili identifikohet me emrin e tij. Meqenëse janë 12 të tillë, themi se bashkësia e të dhënave janë 12 elementë. **Variabli** është një karakteristikë e përbashkët e elementëve për të

cilin jemi të interesuar. Kështu të tilla mund të jenë psh. gjinia, mosha, funksioni, vjetërsia në punë dhe paga. Për këto variabla, kemi mbledhur informacion për secilin element. Bashkësia e matjeve të grumbulluara për një element të veçantë quhet vrojtim. Një studim ka aq vrojtime sa edhe elementë ka. Disa variabla janë numerikë ( psh. mosha , vjetërsia në punë dhe paga) dhe disa të tjera janë jo-numerike ( psh. Gjinia , funksioni). Variablat që marrin vlera numerike quhen sasiore , kurse variablat që nuk marrin vlera numerike, por që karakterizojnë kategori, emërtime ose që shprehen në simbole quhen variabla cilësore.

Në shumë raste ekziston një grup shumë i madh elementësh ( individë, aksione, qytete, kompani etj) për të cilët mblidhen të dhënat. Mirëpo për shkak të kohës, kostos dhe konsideratave të tjera, të dhënat grumbullohen vetëm nga një pjesë e vogël e grupit. Grupi i madh i elementëve në një studim quhet **popullim**, kurse grupi më i vogël për të cilin mblidhen të dhënat quhet **zgjedhje**. Kështu, popullimi është grumbulli i të gjithë elementëve për të cilët interesohemi në një studim të veçantë, kurse zgjedhja është një nënbashkësi e popullimit. Qëllimi i një studimi statistikor është studimi i popullimit. Në pamundësi të matjes së gjithë popullimit, ne realizojmë matjen e zgjedhjes, por me qëllim që karakteristikat e saj ti përdorim për t’ia “veshur” popullimit. Me këtë arësyetim, zgjedhja duhet të jetë sa më përfaqësuese e popullimit. Një nga mënyrat e seleksionit të elementëve të popullimit që do ti nënshtrohet studimit quhet zgjedhje e rastit ose zgjedhje probabilitare. Kjo mënyrë zgjedhjeje u krijon mundësi të gjithë elementëve të popullimit të marrë nga baza e zgjedhjes një shans të njohur (jo gjithnjë të barabartë) për tu përfshirë në zgjedhje.

### **3. SHKALLËT E MATJES DHE BURIMET E TË DHËNAVE**

Lloji i analizës statistikore që mund të përdoret për të dhënat e një variabli të veçantë varet nga shkalla e matjes e përdorur për këtë variabël. Shkallët e matjes së një variabli mund të klasifikohen në katër: Nominale, e zakonshme, intervale dhe në formë raporti. Shkalla e matjes përcakton sasinë e informacionit që përmbajnë të dhënat, në të njejtën kohë tregon llojin e përmbledhjes dhe të analizës statistikore që janë më të përshtatshme të zbatohen ndaj tyre.

Karakteristike për shkallët e matjes është që vetitë e tyre mbarten dhe kumulohen nga njëra shkallë te tjetra sipas renditjes së mësipërme.

**Shkalla e matjes nominale**, është niveli më i ulët i matjes së një variabli, përmban sasi të vogël infomacionit dhe si rrjedhojë ndaj të dhënave të kësaj shkalle mund të zbatohet një numër i kufizuar metodash statistikore. Shkalla e matjes së një variabli quhet nominale kur të dhënat që merren për variablin janë thjesht emërtime, simbole ose shenja që tregojnë për identifikimin e një vetie të elementit. Psh. Gjinia, qyteti, ngjyra e preferuar etj. Këto të dhëna vetëm mund të tabelohen. Pas kësaj për çdo kategori mund të llogariten përpjesat ose përqindjet e çdo kategorie. Të dhënat në shkallën nominale plotësojnë dy veti: Së pari kategoritë përjashtojnë njëra tjetrën, dmth çdo e dhënë për një element që është përfshirë në një kategori nuk mund të përfshihet në asnjërën nga kategoritë e tjera. Së dyti kategoritë përfshijnë të gjithë elementët , dmth çdo element përfshihet në njërën dhe vetëm njërën nga kategoritë.

**Shkalla e matjes e zakonshme**, shënon një nivel më të lartë të matjes së të dhënave se shkalla nominale. Shkalla e matjes së një variabli quhet e zakonshme , nqse të dhënat kanë vetitë e shkallës nominale dhe renditja e tyre në rendin rritës apo zbritës merr kuptim. Psh. Vlerësimi që keni për profesorin? (kategoritë e këtij variabli janë: Shumë e mirë (1) , Mirë (2) , Dobët(3) )Pavarësisht se kategoritë mund të kodohen me numra, themi se veprimet arithmetike në këtë rast, ashtu si edhe për shallën nominale, janë pa kuptim.

**Shkalla Intervalore**, e matjes së të dhënave përbën diçka më të sofistikuar krahasuar me dy shkallët e mëparshme. Ajo përmban më shumë informacion dhe lejon përdorimin e më shumë metodave statistikore për analizën e të dhënave. Shkalla e matjes së një variabli quhet në formë intervalore , nqse të dhënat kanë vetitë e shkallës së zakonshme dhe kur ndryshimi midis vlerave të të dhënave tregon se sa më shumë ose më pak vlerë të variablit ka një element kur krahasohet me një element tjetër. Të dhënat e një variabli të tillë shprehen gjithnjë në vlerë numerike , ku ka kuptim të kryhet vetëm veprimi arithmetik i mbledhjes dhe i zbritjes por jo i shumzimit dhe i pjestimit. Themelore për shkallën intervalore është se intervali njësi është madhësi konstante për çfarëdo dy vlera të të dhënave, kurse pika zero zgjidhet në mënyrë konvencionale dhe nuk shpreh faktin se nuk ekziston asgjë. Psh. variabli “ Temeperatura e qytetit tuaj “ është shembull tipik i variablit në shkallë matje intervalore. Ka kuptim të thuhet se në një qytet temperatura është x gradë më e ulët ose më e lartë, por jo x herë më ngrohtë apo më ftohtë. Dhe zero grade celcius tregon që aty ka një temperaturë dhe nuk do të thotë që aty nuk ekziston temperaturë.

**Shkalla në formë raporti**, është niveli më i lartë i matjes së të dhënave, dhe siguron sasinë më të madhe të informacionit. Shkalla e matjes së një variabli quhet në formë raporti, nqse të dhënat kanë vetitë e shkallës intervalore dhe raporti i vlerave të tyre ka kuptim të përcaktuar. Variabla të tilla si distancë, pesha, dhe koha janë variabla të matura në shkallën në formë raporti.

**Të dhënat mund të klasifikohen në të dhëna cilesore dhe të dhëna sasiore.** Të dhënat cilesore janë emërtime, shenja ose simbole për kategori të tilla që karakterizojnë gjendjen, llojin, tipin etj. Kategoritë për të dhënat cilesore mund të identifikohen edhe me shifra në raste të caktuara, por ato janë vetëm simbole ose kode. Të dhënat cilesore rezultojnë ose nga shkalla e matjes nominale ose të zakonshme.

Të dhënat sasiore tregojnë sasinë , se sa shumë ose më pak të një diçkaje. Të dhënat sasiore janë gjithmonë numerike dhe merren ose nga shkalla e matjes intervalore ose nga shkalla e matjes në formë raporti.

**Burimet e të dhënave janë primare** kur informacioni mblidhet nga pyetësorët, intervistat , pra sigurohen nga vetë studiuesi. **Burimet e të dhënave janë sekondare** , kur të dhënat i referohen burimeve ekzistuese, si psh. internet, proces verbaleve të takimeve, dokumentacioneve ekzistuese etj.

Studiuesit gjithnjë duhet të jenë të ndërgjegjshëm për **gabimet e mundshme** që ekzistojne në të dhënat që përdoren në studimet statistikore, të cilat reflektohen edhe në konkluzionet e analizës përfundimtare të tyre. Një gabim në të dhëna ndodh sa herë që vlera e të dhënës që merret nuk është e barabartë me vlerën e vërtetë ose më mirë me vlerën reale që do të mund të merrej, nqse do të zbatohej një procedurë e rregullt. Të tilla gabime mund të ndodhin në mënyra nga më të



ndryshmet: psh. disa gabime që mund të vijnë nga shënimi gabim i informacionit nga intervistuesi, gabime që vijnë për shkak të hedhjes gabim të të dhënave nga operatori, gabime që mund të bëhen nga përpunimi i të dhënave apo interpretimi i tyre. Këto janë gabime që mund të kontrollohen nga një sjellje individuale dhe profesionale e saktë. Por gabimi që vjen për shkak të zgjedhjes ( zgjedhje së “n” elementëve nga “N” të tillë që ka popullimi) quhet **gabim i zgjedhjes**, shpesh është i pashmangshëm por mund të kontrollohet duke realizuar një zgjedhje “n” sa më afër asaj të popullimit “N”.

Pjesa më e madhe e informacionit në gazeta, revista, raporte dhe botime të tjera ka rezultuar nga përmbledhja dhe paraqitja e të dhënave në një formë të thjeshtë dhe të kuptueshme për lexuesit. Kjo përmbledhje e të dhënave , e cila mund të jetë tabelare, grafike ose numerike quhet **statistikë përshkruese**.

**Statistika analitike**, quhet ajo degë e statistikës që ka të bëjë me teknikat dhe metodat e analizës shkencore të të dhënave. Statistika analitike na jep mundësi që edhe duke përdorur të dhënat e një zgjedhjeje të bëjmë vlerësime dhe të kontrollojmë pohime për një ose disa karakteristika të popullimit që studionet. Këto vlerësime dhe pohime për karakteristikat e një popullimi në bazë të studimit të zgjedhjes quhen gjykime statistikore.

#### **4. STRATEGJITË E NJË KËRKIMI SHKENCOR**

**1. Eksperimenti** : Është një formë klasike e kërimit. Përfshin:

- Prezantimin e hipotezës teorike
- Zgjedhjen e kampioneve apo mostrave nga popullime të njohura
- Alokimi i mostrave për kushte të ndryshme eksperimentale
- Paraqitja e ndryshimit të planifikuar të një apo shumë faktorëve
- Matja e disa variablave
- Kontrolli i variablave të tjerë.

**2. Vëzhgimi/Survejimi/Vrojtimi** Lidhet me metodën deduktive. Lejojnë mbledhjen e një sasi të madhe të të dhënave, nga një popullim i madh, në mënyrë ekonomike.

Të përfutur kryesisht nga pyetësi, të dhënat janë të standartizuara dhe të krahasueshme. Strategjia e vëzhgimit siguron kontroll të hollësishëm mbi procesin e kërimit.

**3. Rasti Studimor** Robinson (1999) e përkufizon: *“një strategji kërkimi që përfshin investigim empirik të një fenomeni të caktuar në kontekstin e jetës reale të tij, duke përdorur burime të shumta evidence*

Kjo strategji përdoret për ti dhënë përgjigje pyetjeve: Pse? Çfarë? Si?

**4. Teoria e bazuar (me bazë)** Këtu mbledhja e të dhënave fillon pa krijuar një kuadër të plotë teorik. Këto çojnë në parashikime të cilat më vonë testohen në observime të ardhme që mund të konfirmojnë apo hedhin poshtë ato parashikime.

Kështu ndërtohet një teori përmes kombinimit të induksionit (nxjerr një ligj nga faktet) dhe deduksionit (nxjerr një fakt nga ligjet)

5. **Etnografia** Bazohet në mënyrën induktive. Qëllimi është të interpretojë subjektet e kërimit të botës sociale në mënyrën që ata vetë e interpretojnë këtë shoqëri. Si strategji kërimit kërkon kohë të gjatë. Procesi i kërimit është fleksibël dhe reagues ndaj ndryshimit, meqë kërkuesi në mënyrë konstante zhvillon mënyra të reja të menduari rreth asaj që observohet. Kryesisht është për çështjet sociale.

6. **Kërkimi i veprimit** Kërkimi i veprimit dallon nga të tjera të kërimit, pikërisht nga fokusi eksplicit, duke promovuar ndryshimet në organizatë.

Kërkuesi është i përfshirë në veprimin për ndryshim. **Fuqia e kësaj strategjie** qëndron në fokusin e saj të ndryshimi; pranimi i kohës së nevojshme për tu përkushtuar në monitorim, vlerësim dhe përfshirje punonjësish në proces.

Strategjia kombinon mbledhjen e informacionit dhe lehtësimin e ndryshimit.

Strategjia ka dy qëllime:

1. Të përmbushë axhendën e kërkuesit dhe jo të sponsorit
2. Fillon me nevojën e një sponsori dhe përfshin ndërmarrjen e kërimit për çështje të sponsorit . Këto quhen aktivitete këshillimi, konsultimi.

Ndërhyrje të ndryshme bashkohen dhe ndërveprojnë përmes sponsorit dhe studiuesit.

**Procesi i këshillimit** ndihmon sponsorin të fitojnë aftësitë e duhura të diagnostifikimit dhe zgjidhjes së problemeve organizacionale, kështu që ai mund të zhvillojë autonominë në përmirësimin e organizatës.

### 7. **Kërkimi arkivor**

Kjo strategji përdor regjistrime administrative dhe dokumentacionin si burim kryesor të dhënash. Megjithëse termi “arkivor” ka sens historik, i referohet dokumentacioneve të shtrira në kohë më parë. Kjo strategji, lejon çështjet e kërimit të fokusuara në të shkuarën, të ndryshojnë me kohën dhe të gjejnë përgjigje nga kërkit nëpërmjet kërkimeve eksploruese, përshkruese ose shpjeguese. Megjithëse burimi kryesor është arkiva, shpesh mund të mungojnë aty të dhëna, ose të dhënat që gjenden nuk janë shumë relevante për kërkitin. Nëse zgjidhet kjo strategji, duhet duhet të dizajnohet kërkiti që të merret më e mira nga arkiva ekzistuese.

8. **Strategjitë mikse** Në një studim të vetëm mund të kombinohen metodat sasore dhe cilësore; si edhe të përdoren të dhëna primare dhe sekondare. Ka dy avantazhe përdorimi i strategjisë mikse:

1. Metoda të ndryshme mund të përdoren për qëllime të ndryshme . Psh. Mund të përdoren intervistat para se të formuloni pyetësorë; kjo ju jep besim në adresimin e çështjeve të rëndësishme.
2. Mundësohet funksionimi i “trekëndeshi” që është: Përdorimi i metodave të ndryshme të mbledhjes së të dhënave brenda një studimit, që të siguroheni që të dhënat tregojnë atë që ju mendoni.

## **5. TIPET E KËRKIMIT SHKENCOR :**

### ***A) SIPAS QËLLIMIT***

Një tjetër element që merret në konsideratë, që në fillim është edhe klasifikimi i kërkimit në varësi të qëllimit që ka ai. **Mbledhja e të dhënave fillon mbi bazën e Strategjisë dhe Klasifikimit të kërkimit** Sipas qëllimit kemi:

#### **1. Kërkim eksplorues**

Shërben për të zbuluar çfarë po ndodh dhe për tu orientuar në drejtime të reja, pra qartëson kuptimin tuaj ndaj një problemi. Tre rruge parimore për këtë kërkim:

1. Kërkim në literaturë
2. Bisedat me ekspertë
3. Intervista me grupe fokusi

Është fleksibël, dhe i adaptueshëm ndaj ndryshimeve. Fokusi i gjerë përcaktohet që në fillim, pastaj bëhet më i ngushtë gjatë procesit kërkimor.

2. **Kërkim Përshkrues** Objekti i këtij kërkimi është të japë një portret të profilit të personave, ngjarjeve apo organizatave. Është e domosdoshme të kemi një pamje të qartë të fenomenit për të cilin ju do mbledhni të dhëna, para se të filloni punën për mbledhjen e tyre. Kjo ju inkurajon në thellimin e kërkimit nëpërmjet analizave të sintetizuara. Ai mund të mendohet si mënyra për të kuptuar fundin sesa një fund në vetvete.

3. **Studimi shpjegues** Studime të tilla janë studimet që krijojnë marrëdhëniet shkakësore ndërmjet variabla. Theksi vendoset tek studimi i situatës me qëllim shpjegimin e lidhjes mes variabla.

### **B. SIPAS KOHËS SË MATJES**

Në varësi të kohës së cilës i kushtohet kërkimi do të përcaktohet edhe strategjia e kërkimit:

1. **Cross sectional:** Nëse duhet që kërkimi të jetë një fotografi e çastit, e marrë në një moment të veçantë.
2. **Longitudinal- studime gjatësore:** nëse duhet që kërkimi të jetë i ngjashëm me një ditar; që paraqet ngjarjet të shtrira në kohë. Kryesisht kur bëhen vëzhgime të fenomeneve në një hark të gjerë kohor, duke bërë të mundur kontrollin mbi variablat.

## **6. KREDIBILITETI I PROCESIT TË KËRKIMIT**

### **1. Besueshmëria**

I referohet shkallës për të cilën teknikat e mbledhjes së të dhënave ose procedurat e analizave do të çojnë në rezultate të duhura.

Besueshmëria vlerësohet nga tre pyetje kryesore:

1. A do të japin matjet rezultate të njejta për zgjedhje të tjera?
2. Kërkues të tjerë a do të arrijnë në të njejtat rezultate?
3. A ka kuptim transparenca e kryer nga të dhënat e pa përpunuara?

#### **Kërcënimet e besueshmërisë:**

1. Çështje psikologjike që lidhen me pjesëmarrësit
2. Mund të ketë subjektivitet apo anësi nga pjesëmarrësi
3. Gabim i observuesit
4. Mund të ketë subjektivitet apo anësi nga observuesi.

### **2.Vlefshmëria (validiteti)**

Lidhet me faktin nëse gjetjet e kërkimit janë apo jo reale. Lidhet me faktin nëse marrëdhënia shkak-pasojë vërtetohet për çdo matje të kryer.

#### **Kërcënimet e vlefshmërisë:**

1. *Historia* (nëse matjet do bëhen fill pas një ngjarjeje, përgjigjet mund të jenë çorientuese)
2. *Testimi* (nëse dikush punon në presionin se po matet, do të reagojë ndryshe)
3. *Instrumentalizmi* (Nëse dikush punon në presionin se po matet, mund ta detyrojnë shefat të punojë ndryshe)
4. *Vdekshmëria* (i referohet pjesëmarrësve që largohen nga vërtetimi)
5. *Maturimi* (ngjarja që po matet ndikohet nga diçka tjetër që ka ndodhur në organizatë)
6. *Dykuptimësia rreth drejtimit shkakësor* (psh në je studim u gjet se

Vlerësimi i performancës së dobët të punonjësve është e lidhur me qëndrimin negativ të menaxherit ndaj tyre. Ajo çfarë s'është e qartë është se performanca e dobët çon në qëndrim negativ, apo qëndrimi negativ çon në performancë të dobët?

### **3.Përgjithësimi**

Kjo quhet vlefshmëri e jashtme. Lidhet me faktin sa rezultatet e kërkimit janë përgjithësuese, sa rezultatet janë po aq të rëndësishme për kërkime të tjera në organizata të tjera. Nëse deklaroni që rezultatet tuaja janë të vlefshme edhe për organizata të ngjashme, atëherë kërkimi juaj duhet të këtë teknika që të kenë përgjithësim të rezultatit. Nëse kërkimi nuk dekalron përgjithësim atëherë nuk ka problem.

### **4.Hapat logjike dhe supozimet e rreme**

Pasi kemi arritur në këtë fazë të kërkimit, më tutje vazhdohet me:

1. *Identifikimin e popullimit*: nëse synimi është gjenerimi i konkluzioneve për popullimin, atëherë duhet që zgjedhja e popullimit të jetë logjike. Kjo do ta bënte kërkimin të vlefshëm.

2. **Mbledhja e të dhënave:** mënyra e duhur e mbledhjes së të dhënave rrit vlefshmërinë e kërimit.
3. **Interpretimi i të dhënave:** Ky është ndoshta rreziku më i madh i hapave logjike dhe supozimeve të rreme. Do tju duhet të lëvizni nga një pozicion ku keni një mal me të dhëna, në një tjetër pozicion që do tju duhet të jepni konkluzione. Ky është një proces intelektualisht sfidues, si dhe shumë kreativ.
4. **Zhvillimi i konkluzioneve:** Pyetja që bëhet në fund është: A janë konkluzionet e kërimit tim të besueshme dhe të vlefshme?

## **TEMA : VROJTIMI ME PYETËSOR DHE APLIKIMI NË SPSS**

- 1. QËLLIMI DHE FAZAT E STUDIMEVE ME ANË TË ANKETIMEVE**
- 2. NDËRTIMI I TREGUESVE DHE KRYERJA E ZGJDHJEVE**
- 3. VLERËSIMI I TREGUESVE**
- 4. REALIZIMI I NJË ZGJEDHJEJE**
- 5. INTERVISTIMI**
- 6. NDËRTIMI I PYETËSORIT**

### **1. QËLLIMI DHE FAZAT E STUDIMEVE ME ANË TË ANKETIMEVE**

Vrojtimet me anketim janë një metodë e vërtetimit statistikor që përdoret mjaft në fushën e studimeve ekonomiko sociale. Kjo metodë ashtu si çdo metodë tjetër ka avantazhet dhe disavantazhet e veta.

*Çfarë është një vërtetim me anë të anketimit?* Shumë studjues insistojnë se një vërtetim me anketim nuk është sinonim me një teknikë të veçantë të mbledhjes së informacionit: këtu pyetësoret përdoren gjerësisht por gjithashtu janë të përshtatshme edhe teknika të tjera si intervistat e strukturuar dhe të thëlla, vëzhgimet analize të shkallës së “kënaqësisë” etj, etj. Elementët më të vështirë të anketimeve janë mënyra e mbledhjes së të dhënave dhe metoda e analizës.

Mënyra e mbledhjes së të dhënave Vërtetimi me anketim karakterizohet prej një stukture ose një komplet sistematik të dhënash të cilat do t'i quajmë variabla të rastit të databazës së të dhënave.

E gjitha kjo do të thotë se ne mbledhim informacion rreth të njëjtëve variabla ose karakteristika prej të paktën 2 (normalisht me shumë) rastesh dhe përftojme një databazë të të dhënave të formës së paraqitur në tabelën 1.1.

*Tabela 1.1 Databaza e të dhënave të disa variablave të rastit*

| Rastet    |                 | Personi 1             | Personi 2 | Personi 3 | Personi 4 | Personi 5 |
|-----------|-----------------|-----------------------|-----------|-----------|-----------|-----------|
| Variablat | Seksi           | Mashkull              | Mashkull  | Femer     | Mashkull  | Femer     |
|           | Mosha           | 36 vjeç               | 19 vjeç   | 30 vjeç   | 55 vjeç   | 42 vjeç   |
|           | Bindja pol      | I majtë               | I qendrës | I djathtë | I majtë   | I djathtë |
|           | Statusi punësim | Punëtor i kualifikuar | Nëpunës   | I Papunë  | I papunë  | Biznesmen |

Problemet që lidhen me qëllimin e studimit duhet të zërthehen përpara se sa të mbledhen të dhënat. Që në fillim të studimit duhet të qartësohet problemi se a do të fokusohet studimi ynë në një fenomen të veçantë dhe ne duhet të përpiqemi të gjejmë sa më shumë shkaqe të mundshme për këtë fenomen, apo studimi ynë ka për qëllim të realizojë shpjegime të pjesëshme të disa ngjarjeve. Përcaktimi i qëllimit të studimit ka një rëndësi themelore sepse në varësi të qëllimit të studimit do të jenë të gjitha fazat e tjera të studimit që do të ndërmerret.

Qëllimi i studimit kushtëzon gjithashtu faktin në se do të kryejmë një studim deskriptiv (përshkruar) apo një studim më të gjerë dhe shpjegues.

#### *a. Fazat e studimit me anë të anketimeve*

*Faza përgatitore*, është një fazë shumë e rëndësishme. Në këtë fazë të studimit ndërtohen tabelat paraprake të rezultateve (databaza) që mendohet se do të merren në përfundim të studimit, përcaktohet njësia e vërtetimit, si dhe realizohet studimi pilot për të testuar jo vetëm pyetësorin por edhe elementë të tjerë të vërtetimit.

*Veprimet njohëse për grumbullimin e të dhënave* : Në këtë fazë ndërtohet pyetsori (i cili mund të përdoret për një vërtetim direkt, me postë, me telefon, me internet) bëhet kontrolli i pyetsorit, duke insistuar në faktin se informacioni i grumbulluar përmes pyetsorit është konfidencial, dhe përcaktohen llojet e gabimeve që janë të mundshme gjatë grumbullimit të të dhënave.

*Grumbullimi i të dhënave.* Veprimet në vazhdim të grumbullimit të të dhënave konsistojnë në rishikimin e pyetsorit në bazë të vërejtjeve të marra nga studimi pilot etj, kodifikimin e pyetjeve, regjistrimin (hedhjen e të dhënave), si dhe ndërtimin e maketeve të tabelave statistikore. Nëse do të përpiqemi të japim një strukturë tip të një studimi të tillë atëherë do mendojmë se ajo do të ketë formën e mëposhtme:

1. Identifikimi i problemit
2. Njohja e studime dhe të dhëna të mëparshme për problemin
3. Formulimi i hipotezave
4. Skicimi i studimit
5. Grumbullimi i të dhënave
6. Analiza e të dhënave
7. Interpretimi i rezultateve

Hipoteza është një artikullim i pjesëshëm i teorisë. Teoria është “e përgjithshme”, ndërsa hipoteza është “specifike”. P.sh. “ekziston një korelacion pozitiv midis pjesmarrjes në jetën politike dhe qendërsisë social-moshore” është një teori që mund të shtrohet midis të tjerave edhe me hipotezën e pjesmarrjes në votime më tepër të moshave të mesme në krahasim me moshat më të reja dhe më të vjetra.

## **2. NDËRTIMI I TREGUESVE DHE KRYERJA E ZGJIDHJEVE**

Kthimi i koncepteve në tregues mund të bëhet në tre hapa: qartësimi i koncepteve, krijimi i treguesve fillestar, vlerësimi i treguesve

### **2.1 Konceptet**

Konceptet janë thjeshtë përmbledhje abstrakte të sjelljeve, qëndrimeve dhe karakteristikave të cilat kanë diçka të përbashkët.

Skicimi i dimensionit të konceptit: shumë koncepte kanë aspekte apo dimensione të ndryshme. Kur qartësohen konceptet duhet të dallohen këto dimensione. Kjo mund të çojë në përdorimin në studim vetëm të një dimensionit ose mund të çojë në ndërtimin e treguesve për çdo dimension. Dallimet midis dimensioneve mund të çojnë në teori më të sofistikuar dhe në analiza më të dobishme.

Deprivimi është shembull i një koncepti shumë dimensional: mund të dallohen dimensione sociale, ekonomike, politike dhe psikologjike. Dallimi midis tyre mund të na detyrojë që të qartësojmë se me çfarë ka të bëjë studimi ynë.

### **2.2 Ndërtimi dhe vlerësimi i treguesve**

Proçesi i zhvendosjes prej aspekteve abstrakte në pikën në të cilën ne duhet të ndërtojmë seksionet e pyetësorëve për të prekur konceptet quhet “*zbritja e shkallës së abstraksionit*”. Ajo përfshin lëvizjen prej të përgjithshmes në specifikën, prej abstraktes tek konkretja. Kur specifikohen dimensionet atëherë bëhet edhe një hap tjetër. Kur arrijmë në pikën në të cilën mund të ndërtojmë tregues të ndryshëm atëherë mund të merremi me tre çështje: sa tregues duhet të përdorim, si t'i ndërtojmë këta tregues, si të formojmë seksionet në pyetësor.

#### **2.2.1. Sa tregues duhet të përdorim?**

Nuk ka një përgjigje të vetme për këtë problem por, mund të ndjekim udhëzimet e mëposhtme.

1. Kur nuk bini dakort për mënyrën për të matur konceptet atëherë duhet të bëhet ndërtimi i treguesve për një serë përshkrimesh dhe të shihet ndryshimi që kjo gjë do të sillte në rezultate dhe në interpretime.
2. Nëse koncepti është shumëdimensional vendosni vete nëse jeni vërtet të interesuar për të gjitha dimensionet.

3. Sigurohuni se keni matur konceptet bazë. Sjellja dhe zakonet që po përpigemi të shpjegojmë dhe shkakët e fenomenit duhet që të maten me kujdes duke përdorur disa tregues.
4. Zakonet tipike dhe opinionet janë komplekse dhe maten më mirë me disa pyetje.
5. Treguesit në testimin pilot janë një mënyrë e eliminimit të pyetjeve të panevojshme. Fillimisht mund të kemi 50 pyetje për të matur një tregues por zbulojmë se na duhen vetëm 10 prej tyre, 40 të tjerat mund të mos japin asgjë të re për treguesin tonë.
6. Numri i seksioneve të pyetesorit ndikohet prej konsideratave praktike si gjatësia totale e pyetesorit dhe mënyra e administrimit të tij.

### 2.2. 3. Si t'i ndërtojmë këta tregues?

Për shumë koncepte ndërtimi i një treguesi është i thjeshtë, por për koncepte të tjera më abstrakte kjo gjë është më e vështirë. Ekzistojnë tre këshilla për ndërtimin e treguesve fillestarë në pyetjet e një pyetesori. Në çdo rast këto tregues fillestarë do të përpunohen përgjatë fazës së vlerësimit.

- Së pari mund të përdoren treguesit e ndërtuar në studimet e mëparshme. Ka shumë raporte të testuara të cilat duhet t'i vlerësojmë. Ato mund të kenë nevojë të rivlerësohen që të përshtaten plotësisht me një kuptim të veçantë apo me një zgjedhje të veçantë. Përdorimi i treguesve të stabilizuar nga studimet e mëparshme është avantazhi i madh i mundësisë së krahasimit të rezultateve me studime të tjera.
- Së dyti për disa pika studimi në mënyrë të veçantë ato të cilat i përkasin një grupi të veçantë (migrantët, të rinjt) është mirë që për mbledhjen e të dhënave fillimisht të përdoren metoda pak të strukturuar (vëzhgimi etj). Kjo gjë mund që të na ndihmojë që të kuptojmë gjërat sipas këtyre njerëzve, të mësojmë mënyrën e tyre të të menduarit dhe kjo mund të na ndihmojë për ndërtimin e pyetjeve të rëndësishme për atë grup.

- Alternativa e tretë është përdorimi i “informatorëve” prej grupit që do t'i bëhen pyetjet. Në fund duhet që të vendosim se cilët tregues të përdorim. Për të bërë këtë ne duhet që të jemi sa më shumë të informuar rreth popullimit, të jemi të qartë rreth asaj që duam të masim, të shohim përpjekjet e njerëzve dhe të vlerësojmë treguesit tanë.

## 3. VLERËSIMI I TREGUESVE

Duke ndërtuar tregues tregues duhet që të jemi të sigurtë se ata masin konceptet për të cilët ne i kemi përdorur (*vlefshmëria*) dhe të sigurohemi se mund të mbështetemi në përgjigjet që njerëzit iu japin pyetjeve tona. Një pyetje është me pak vlerë nëse njerëzit i përgjigjen ndryshe sot dhe ndryshe nesër (*besueshmëria e informacionit*).

Patjetër duhet të vlerësohet besueshmëria dhe vlefshmëria e treguesve përpara se sa ata të përdoren në studimin aktual. Kjo fazë quhet *testim pilot* dhe bëhet duke ju drejtuar pyetjet një zgjedhje mjaft të vogël të popullimit, që është e ngjashme me atë të studimit.

Disa herë studjuesit prapëqen ta evitojnë testimin pilot duke futur një numër të madh treguesish në studim dhe duke përdorur ata tregues të cilët duken të vlefshëm dhe besueshëm. Kjo është e gabuar. Nëse bëjmë këtë gjë, atëherë treguesit mund të mos përfaqësojnë konceptet apo idenë që ne donim të testonim dhe kështu studimi mund të ketë një vlerë të vogël në krahasim me pyetjet e studimit origjinal.

### a. Besueshmëria

Një matje e besueshme është ajo prej së cilës ne përfitojmë të njëjtin rezultat kur përsëritim zgjedhjen.

### b. Vlefshmëria



Tregues i vlefshëm është ai tregues që mat atë gjë për të cilën është ndërtuar. Në fakt i vlefshëm apo i pavlefshëm nuk është treguesi por, e tillë është mënyra në të cilën përdoret ai. Ne mund të përdorim nivelin e arsimit për të matur statusin social. Problemi nuk qëndron në faktin nëse e kemi matur arsimin si duhet por nëse është niveli i arsimit një tregues i statusit social. Vlefshmëria e një treguesi varet prej asaj se si e kemi ndërtuar konceptin të cili ai duhet të mati.

#### **4. REALIZIMI I NJË ZGJEDHJEJE**

Një mënyrë e zgjedhjes së një grupi njerëzish është mbledhja e informacionit prej çdo pjestari të grupit. Për grupe të mëdha kjo gjë është e kushtueshme dhe jopraktike. Prandaj duhet të mblidhet informacion vetëm prej një pjese të grupit në mënyrë të tillë që përgjigjet dhe karakteristikat e tyre të reflektojnë karakteristikat e grupit prej të cilit ata janë marrë. Kjo procedurë është shumë më e lehtë, më pak e kushtueshme, më e shpejtë dhe më praktike. Ky është parimi bazë i zgjedhjes. Një *zgjedhje* përftohet duke mbledhur informacion për disa njësi të vrojtimit. Zgjedhje të ndryshme mund të përfaqsojnë me shkallë të ndryshme përpikmërie popullimin prej të cilit ato janë kryer. Një *zgjedhje* e cili përfaqëson me përpikmëri popullimin quhet *zgjedhje përfaqësuese*.

Për tu siguruar se një *zgjedhje* është përfaqësuese e popullimit është e rëndësishme që disa grupe të popullimit të mos përjashtohen në mënyrë sistematike prej kësaj *zgjedhjeje*. P.sh. nëqoftëse ne duam të përftojmë një *zgjedhje* të popullsisë së periferisë së qytetit duke shkuar gjatë ditës dhe duke trokitur në një në çdo 20 dyer në këtë zonë, ne do të kemi në mënyrë sistematike një nën-përfaqësim e atyre grupeve të njerëzve të cilët nuk janë në shtëpi gjatë ditës. Një *zgjedhje* e tillë ka *anësi* dhe ajo nuk mund të përdoret për të përgjithësuar popullimin.

Ka dy lloje *zgjedhjesh*: *probabilitare* dhe *joprobabilitare*.

##### A )Zgjedhje porbabilitare

Zgjedhje probabilitare është ajo *zgjedhje* në të cilin çdo njësi e popullimit ka një mundësi të barabartë, ose të paktën të njohur, për të qënë e zgjedhur ndërsa në një *zgjedhje joprobabilitare* disa njësi kanë një mundësi më të madh por të panjohur për tu zgjedhur. Rruga më e mirë për të pasur një probabilitet të njëjtë *zgjedhjeje* është përdorimi i *zgjedhjes* së rastit. Kjo *zgjedhje* ka të bëjë me listimin e të gjithë anëtarëve të popullimit dhe më pas nga lista zgjidhen disa prej emrave. Është e pamundur që *zgjedhja* të jetë plotësisht përfaqësuese. Vetëm prej ndikimit të faktorëve të rastit do të ketë ndryshime midis popullimit dhe *zgjedhjes*. Këto ndryshime pjesërisht do të jenë si pasojë e *gabimeve të zgjedhjes*. E rëndësishme është se karakteristikat e pjesës më të madhe të *zgjedhjeve* të rastit do të jenë afër atyre të popullimit. Meqë pjesa më e madhe e *zgjedhjeve* japin vlerësime të përafërta me të dhënat e popullimit, ne mund të përdorim teorinë e probabilitetit për të vlerësuar se sa afër me të dhënat e popullimit do të jenë të dhënat e *zgjedhjes*. Për këtë qëllim përdoret një statistikë që quhet *gabimi standart*. Zgjedhjet probabilitare janë më të preferuara sepse ato ka më shumë mundësi që të jenë përfaqësuese dhe mundësojnë vlerësimin e saktësisë së *zgjedhjes*.

Vëllimi i *zgjedhjes* varen prej dy faktorëve: saktësia që kërkojmë për *zgjedhjen* dhe gjerësia e variacionit në popullim sipas variablave kryesor në studim. Ne duhet që të vendosim shkallën e gabimit që mund të tolerojmë. Ne relacionin midis vëllimit të *zgjedhjes* dhe saktësisë së saj duhen theksuar disa çështje kryesore.

- Së pari kur kemi të bëjmë me *zgjedhje* të vogla një rritje e vogël e vëllimit të *zgjedhjes* mund të çojë në një rritje të konsiderueshme të saktësisë. Në *zgjedhjet* e mëdha rritja e vëllimit nuk na jep të njëjtin rezultat. Rregulli është se për të përgjysmuar gabimin e *zgjedhjes* në duhet që të katërfishojmë vëllimin e *zgjedhjes*.
- Së dyti numri i përgjithshëm i njësive të popullimit, prej të cilit nxjerrim *zgjedhjen*, është disi i parëndësishëm për saktësinë e *zgjedhjes*. E rëndësishme është vlera absolute e vëllimit të *zgjedhjes*.
- Së treti për popullim homogjen, ku pjesa më e madhe e njerëzve do t'i përgjigjen një pyetjeje në mënyrë të njëjtë, rezultat e duhura mund t'i arrijmë dhe me një *zgjedhje* të vogël.

Ka mjaft vështirësi për të aplikuar këto teknika për të përcaktuar vëllimin e zgjedhjes. Përveç specifikimit të nivelit të saktësisë që duam ne duhet që të kemi një ide të përafërt se si do të përgjigjen njerëzit.

Për arsye të ndryshme njerëzit e zgjedhur për tu anketuar mund të mos përfshihen në anketim sepse disa refuzojnë, për disa është i pamundur kontaktimi dhe disa do të jenë të paintervistueshëm. *Mospërgjigjja* mund të krijojë dy probleme të mëdha: reduktimin e papranueshëm të përmasave të zgjedhjes dhe anshmërinë.

Problemi i përmasave të zgjedhjes mund të zgjidhet në dy mënyra. Së pari përdorimi i teknikave të cilat reduktojnë mospërgjigjen. Së dyti mund të marrim një vëllim zgjedhjeje më të madhe se sa nevojitet. Duke supozuar se do të kemi mospërgjigje rreth 20% atëherë bëjmë një zgjedhje me vëllim 20% më i madh se ai që nevojitet.

Shpesh është e vështirë që të përftojmë zgjedhje masive dhe përfaqësuese që të lejojnë një analizë adequate: nuk kemi kohën, paratë apo ekspertët për të përftuar një zgjedhje me të vërtetë të mirë. Por në shumë raste ka një alternativë: të dhënat e mbledhura nga persona apo agjensi të tjera shpesh mund të përdoren që të marrë informacion për pyetjet e studimit tonë. Kjo quhet *analizë dytësore*. Kërkimi kryesor kryhet prej agjensive shtetërore, organizatave të mëdha dhe grupeve të cilët realizojnë zgjedhje mjaft të mira dhe shpesh përfshijnë të dhëna për shumë probleme, sjellje dhe pikëpamje personale që i bëjnë ato të përshtatshme për projekte kërkimi përtej qëllimit të përcaktuar në anketimin fillestar të tyre. Më tej këto të dhëna bazuar në masa sigurie që garantojnë anonimin, depozitohen në një arshivë të dhënash dhe pasi studjuesit e parë kanë analizuar të dhënat ato janë të hapura për publikun. Rezultatet janë të disponueshme për tu blerë me çmime shumë të ulta ose ka raste që jepen falas.

Ka databaze të dhënash që janë të gatshme për analiza me shumë qëllime. Shembull i qartë është rregjistrimi i popullsisë.

Analiza dytësore ka edhe disavantazhe. Të dhënat që mund të duhen për një studim të veçantë mund të mos jenë të disponueshme. Shpesh pyetjet që bëhen në anketimet e të tjerëve nuk janë pikërisht ato çfarë ne do të kishim pyetur vetë. Ose në anketimet që bëhen në vende të tjera apo në vite të ndryshme struktura e pyetjeve mund të jetë e ndryshme. Për këtë analiza dytësore nuk na jep një zgjidhje.

Por me pak imagjinatë, krijimtari dhe kompromis shpesh është e mundur që të përftohen të dhëna të kënaqshme për konceptet që kërkojmë në studimin tonë. Me një vështrim të parë disa variabla mund të mos sigurohen tek të dhënat por shpesh blloqet e ndërtimit të këtyre variablave janë të disponueshme, gjë që u bën të mundur studjuesve që të krijojnë variablat e kërkuar.

#### b. Zgjedhjet jo-probabilitare

Shpesh ka situata ku probabiliteti i teknikave të zgjedhjes është jopraktik ose edhe jo i nevojshëm. Në situata të tilla përdoren teknika jo-probabilitare si teknika me kosto më të ulët. Këto teknika janë më të mira kur listat e njësive të vrojtimit janë të padisponueshme ose popullimi është kaq i shpërndarë sa që zgjedhja shumshkallëshe do të ishte joeficiente. Në fazat paraprake të studimit, të tilla si testimi i pyetësorëve, modelet jo të rastit janë të kënaqshëm. Në raste të tilla ne thjeshtë duhet të përpiqemi të kemi një varietet me njerëz në zgjedhje pa u merakosur nëse çdo grup i popullsisë paraqitet në zgjedhje në proporcion korrekt.

c. *Zgjedhja e qëllimshme* është formë e zgjedhjes joprobabilare ku rastet vrojtohen si raste tipike të disa grupeve që i interesojnë studjuesve. Ato nuk zgjidhen rastësisht. Anketat në rastin e votimeve politike shpesh përdorin zgjedhjen e qëllimshme. Këtu zgjidhen lagje të caktuara sepse më të kaluarën zgjedhja e tyre ka siguruar ide të mira rreth rezultateve të të gjithë elektoratit. Megjithatë nuk përdoren teknika të zgjedhjes probabilitare kjo metodë mund të na sigurojë një informacion me kosto të ulët dhe çuditërisht edhe parashikime efikente.

b. *Zgjedhja me kuota* është një teknikë joprobabilare që na jep zgjedhje përfaqësuese. Intervistuesit kërkojnë të gjejnë rastet me karakteristika të veçanta, që japin kuotat e tipeve të njerëzve të veçantë. Kuotat organizohen në mënyrë të tillë që në kushtet e kuotave karakteristike zgjedhja përfundimtare

do të jetë përfaqësuese. Që të ndërtojmë kuotat duhet të vendosim se në cilat karakteristika do bazohemi me qëllim që zgjedhja përfundimtare të jetë përfaqësuese, të gjejmë shpërndarjen e këtyre variablave në popullim dhe të vendosim kuotat sipas tyre. Teknika e kuotave nuk është probabilitare sepse intervistuesit mund të zgjedhin çdo rast që mund të përmbush disa kritere. Kjo do të çojë në anësi pasi intervistuesit do të tentojnë të zgjedhin ata persona që është më e lehtë për t'i intervistuar dhe ata tek të cilët ata ndjehen më rehat (miqtë). Një vështirësi tjetër është se mund të mos jenë në dispozicion proporcione të sakta popullimi. Së fundi, meqë nuk përdoret zgjedhja e rastit, atëherë është e pamundur që të vlerësojmë saktësinë e çdo zgjedhjeje me kuota.

c. *Zgjedhjet në grupe të kufizuara* janë të zakonshme por duhet që të përdoren me kujdes dhe vetëm për qëllime specifike sepse janë teknika që nxjerrin zgjedhjet më pak përfaqësuese. Anketimet ku gazetatat u kërkojnë lexuesve që të plotësojnë dhe të kthejnë pyetësorët e printuar në gazeta apo stacionet e TV që presin telefonata për votimet janë shembuj të këtyre zgjedhjeve. Këto teknika mund të na japin zgjedhje me vëllim të madh me kosto të ulët, por vëllimi i tyre nuk kompeson jo përfaqësueshmërinë e tyre. Këto lloj zgjedhjesh janë të dobishme për testime pilot të pyetësorëve apo për kërkime eksploruese për të përfutur një rang të madh idesh dhe për të zhvilluar tipologjitë, por nuk duhet të përdoren me pretendimin për të paraqitur diçka më tepër se sa zgjedhja e bërë.

## **5. INTERVISTIMI**

Anketimet ballë për ballë janë më të mirë në përfitimin e zgjedhjeve përfaqësuese dhe krijojnë më pak vështirësi gjatë ndërtimit të pyetësorit dhe të hartimit të pyetjeve, ndërsa anketimi me postë është anketimi që krijon më shumë vështirësi.

Metoda e përdorur në praktikë për intervistim ndikon në cilësinë e rezultateve, llojin e pyetjeve që do të bëhen, mënyrën e ndërtimit të pyetjeve si dhe në paraqitjen e pyetësorëve. Ky kapitull fokusohet në dy probleme kryesore:

- Së pari krahason tre mënyra të intervistimit (kontakti direkt ose ballë për ballë, me postë dhe me telefon) në termat e normës së tyre të refuzimit, cilësisë së zgjedhjes, problemeve të strukturës, cilësinë e përgjigjeve, praktikitetin e tyre etj.
- Së dyti më pas jepën disa këshilla të cilat ndihmojnë në zbatimin në praktikë të pyetësorëve.

Ndërsa intervistat ballë për ballë që mund të realizohet në mjaftë abjente të ndryshme (në rrugë, dyqan, derë më derë ose kudo që dy njerëz mund të bisedojnë), kanë pasur dhe kanë një reputacion shumë të mirë, janë një serë arsyesh të tjera, të cilat bëjnë të mundur që të merren në konsideratë edhe alternativa të tjera. Këto vitet e fundit teknikat e intervistimit me postë dhe telefon janë përmirësuar shumë, ndërsa anketimet ballë për ballë po ballafaqohen me probleme të reja.

Një prej avantazheve të teknikës ballë për ballë ka qënë norma e tyre e lartë e përgjigjeve (norma e ulët e refuzimit), por kjo normë ka rrënë me 15% si pasojë e rritjes së niveleve të refuzimit dhe rritjes së vështirësive së gjetjes së të intervistuarve për shkak të shpërndarjes gjeografike. Këto probleme kanë çuar si në rritjen e kostos ashtu edhe në rrëniën e cilësisë së zgjedhjeve.

## **6. NDËRTIMI I PYETËSORIT**

Pyetsori është i ndërtuar si një kamjon transportues i të dhënave dhe si një pasqyrë e objektivave të ndërmara dhe të refleksioneve kritike të përpunuara në momentin e konceptimit të projektit.

### ***1. Formulimi dhe qartësia e pyetjeve të një anketimi***

Pyetjet e anketimit duhet që të ndërtohen në mënyrë të tillë që të jenë sa më të fokusuara rreth problemit që studjohet. P.sh. nuk është e mjaftueshme të thuhet “jemi të interesuar që të kemi disa përgjigje rreth pabarazisë sociale”. Çfarë duam të dimë? Duam të dimë shtrirjen e pabarazisë, shpërndarjen e saj, shkaqet, efektet apo diçka tjetër? Për çfarë lloj pabarazie jemi të interesuar? Për

cilën periudhë? Është diçka tunduese por jo efiçente të mblidhen të dhëna përpara se të jetë formuluar qartë fokusi i studimit. Kjo është shumë inefiçente sidomos kur mblidhen të dhëna të gabuara. Duhet të kemi disa linja udhëzuese se si të ndërtohen pyetjet me qëllim që ato të jenë të qarta. Kjo nuk do të thotë se pyetjet fillestare që ndërtohen gjat studimit duhet të jenë përfundimtare, por ato mund të riformuohen për të qenë më të qarta.

### 1.1. Llojet e pyetjeve

Këtu mendojmë se është e rëndësishme të japim përkufizimet e disa termave. Një *variabël* është një karakteristikë e cila mund të marrë më shumë se një vlerë. Moshë është një variabël që merr vlera të ndryshme (1 vjeç, 2 vjeç, etj.). Megjithatë një person i përket një moshe të caktuar. Një variabël pra është një karakteristikë prej së cilës njerëzit mund të ndryshojnë prej njëri tjetrit. Dallojmë variabla të pavarur dhe të varur. Pasoja quhet një variabël *i varur* (shënohet me Y) dhe është variabël që varet nga diçka tjetër. Shkaku është variabël *i pavarur* (shënohet me X). Psh nëqoftëse duam të shohim nëse niveli i arsimit ndikon në të ardhura atëhere e ardhura do të jetë një variabël *i varur* dhe arsimit një variabël *i pavarur*. Një variabël quhet *ndërrhyrës* (shënohet me Z) dhe ai shpjegon mënyrën se si variabli *i pavarur* ndikon mbi variablin *e varur*. Në shembullin tonë mund të themi se arsimit ndikon tek lloji i punës e cila nga ana e vet është burimi *i të ardhurave*. Këtu puna është variabli ndërrhyrës.

### 1.2. Kërkimet deskriptive.

Le të marrim psh temën e divorcit. Pyetjet e mëposhtëme mund të na ndihmojnë që të fokusohemi qart në këtë temë.

1. Cili është *kufiri kohor* i interesit tonë? Duam të dimë rreth divorcit për periudhën e tanishme apo të shkuarën apo duam që të shikojmë trendin e 10 viteve të fundit?
2. Cila është *hapsira gjeografike* që na intereson të studjojmë? Duam të dimë rreth divorcit për një pjesë të vendit, për të gjithë vendin apo për disa vende të ndryshme? Apo duam të krahasojmë vendin tonë me vende të tjera?
3. Kemi interes për një përshkrim të gjerë apo për të bërë *krahasime* dhe *specifikime* të modeleve për nëngrupet? P.sh. duam të dimë normën e divorcit të një vendi apo duam të shikojmë se si ndryshon kjo normë në varësi të moshës së martesës, punësimit, zgjatjes së martesës, arsimit etj.
4. Në çfarë *aspekti* jemi të interesuar? A ka të bëjë norma e divorcit me mënyrën që njerëzit vendosin të divorcohen, ligjet rreth divorcit, problemeve etj?
5. Sa *abstrakt* është interesi ynë? Jemi të interesuar rreth fakteve apo rreth asaj se çfarë ato tregojnë në nivele abstrakte. P.sh. interesohemi për normën e divorcit si e tillë apo në atë se çfarë ajo mund të tregojë ajo rreth nivelit të konflikteve sociale.

Llojet e pyetjeve që bëhen gjat studimit përfshijnë kërkime shpjeguese. Ato variojnë në bazë të fokusit dhe kompleksitetit. Gjatë viteve të fundit norma e divorcit në Shqipëri është rritur ndjeshëm. Ne mund të interesohemi të dimë përse ndodh kjo. Por, interesi ynë mund të jetë edhe zbulimi i pasojave të rritjes së normës së divorcit.

Hapi i parë është që të vendoset nëse do të studjojmë shkaqet apo pasojat. Së dyti duhet të qartësojmë se çfarë jemi duke studjuar tek shkaqet apo pasojat. Hapi tjetër është listimi i shkaqeve apo pasojave të mundëshme dhe pastaj mbledhja e të dhënave të nevojshme. Ka disa mënyra se si mund ta hartojmë këtë listë.

1. *Studime të mëparshme*: Është e rëndësishme që të përdoren studimet e mëparshme rreth problemit. Por, shpesh ndodh që ka boshllëk në studimet e mëparshme, si p.sh. mungesa të të dhënave dhe probleme ende të pazgjidhura. Gjithsesi këto studime mund të ndihmojnë të fokusohemi më mirë në studimin tonë.
2. *Faktet*: përshkrimet e detajuara të një fenomeni përmes të dhënave të plota mund të stimulojnë ide rreth shkaqeve të mundëshme të tij.
3. *Parandjenjat tona*: idetë tona, besimi dhe eksperiencia jonë rreth problemit janë burime të vlefshme idesh për sa kohë që ne i testojmë ato përkundrejt fakteve.
4. *Rahja e mendimeve me specialistë të tjerë*: shpesh ka njerëz të cilët kanë njohuri të gjera për problemin dhe janë në gjendje për të na dhënë ide për tu mbështetur në studimin tonë.

Si përfundim për studimet deskriptive ne duhet të përpiqemi të jemi të qartë dhe specifik rreth asaj që duam të përshkruajm. Për këto studime duhet të pyesim

1. Çfarë po përpiqemi të shpjegojmë?
2. Cilat janë shkaqet e mundshme?
3. Cilat shkaqe do të eksplorojmë?
4. Cilat janë mekanizmat e mundshme?

Shpesh këto tipe të ndryshme të pyetjeve të studimit të përfaqësojnë stadi të ndryshme të tij. Fillimisht mund të interesohemi thjeshtë në përshkrimin e fenomeneve dhe më vonë të bëjmë pyetjet “pse” rreth shkaqeve të zhvillimit. Kështu, ne mund të fokusohemi në një shkak të veçantë dhe pastaj të bëjmë pyetje më komplekse rreth një faktori të veçantë.

## **2. Ndërtimi i pyetsorit**

Metoda e përdorur për të realizuar intervistimin (me postë, me telefon, direkt) luan një rol të rëndësishëm për faktin se çfarë pyetje do të bëhen, sa pyetje do të bëhen dhe në çfarë renditje do të bëhen pyetjet.

### **2.1. Renditja e pyetjeve.**

Mënyra e paraqitjes së pyetjeve është një moment mjaft i rëndësishëm në vërtetim. Pyetjet e një intervistimi duhet të jenë të lehta dhe në të njëjtën kohë interesante për tu përgjigjur. Në rastin e intervistës me postë, sa më e shkurtër që të jetë intervista aq më e madhe është edhe përqindja e përgjigjeve të kthyer. Kjo sepse në rast se në një familje vjen një pyetsor mjaft i gjatë atëherë ka shumë mundësi që personat e kësaj familjeje të mos e marrin fare në konsideratë dhe as të mos e lexojnë atë. Në këtë mënyrë ata nuk e plotësojnë pyetsorin dhe as që e ripostojnë atë drejt grupit të studimit. Duhet të kihet parasysh se intervista direkte është forma ku të intervistuarit janë më tolerantë ndaj vërtetimeve të gjata dhe mjaft voluminoze.

### **2.2. Formatimi i pyetsorit**

Në një pyetsor mund të përdoren një seri mënyrash të formatimit të pjesëve të ndryshme. Kjo gjë bëhet me qëllim që ta lehtësojë intervistuesin gjat procesit të intervistimit. Por gjithnjë ajo është marveshje mes të gjithë antarëve të grupit të studimit. Eksperienca tregon se në mjaft raste përdoren rekomandimet e mëposhme:

- Tekst i shkruar me gërma të vogla. Frazat në formë pyetjesh ose thëniesh të shkruara me gërma të vogla janë pjesë e intervistës dhe duhet që t'i lexohen të intervistuarit.
- Tekst me gërma të mëdha (kapitale). Frazat e shkruara me gërma të mëdha (kapitale) nuk duhet që t'i lexohen të intervistuarit. Frazat e tilla zakonisht janë tre llojsh:
- a. Instruksiione për intervistuesin. Në këtë rast ato tregojnë mënyrën se si duhet të bëhet pyetja ose të shkruhet informacioni për një pyetje specifike.
- b. Sondazhe Ato përdoren për të futur informacionin që intervistuesi ka grumbulluar. P.sh. një intervistë mund të fillojë kështu: “Në këtë intervistë ne do tju pyesim ju për mendimet tuaja rreth shumë problemeve për njerzit që jetojnë në (EMRI I QYTETIT TE TE INTERVISTUARIT) ...”. Pra, aty do të vendoset emri i qytetit në të cilin banon i intervistuari.
- c. Pyetje që intervistuesi duhet tju përgjigjet vetë. Kjo gjë mund të bëhet në pyetjet filtër të cilat intervistuesi i plotëson sipas informacionit që ka mbledhur.
  - Tekst në kllapa. Teksti i dhënë në formën e deklaratave në kllapa ka tre qëllime dhe përdoret në tre mënyra:
- *Deklarata që janë opsionale.* Këto janë fjalë ose deklarime që në vartësi të situatave të veçanta të intervistës mund të aplikohen ose jo. P.sh. “A bëni pjesë ju (ose ndonjë anëtar tjetër i familjes



suaj) në ndonjë sindikatë?” Pjesa e fjalisë “ndonjë anëtar tjetër i familjes suaj” lexohet vetëm në se në atë familje ka persona të tjerë të rritur.

- *Informacion që duhet të gjendet nga ana e intervistuesit.* Siç u theksua edhe në pikën b më sipër, fjalët që shkruhen me gërma kapitale dhe janë në kllapa nuk duhet të lexohen, por duhet të shkruhet informacioni i marrë më parë nga intervistuesi.
- *Shembuj që intervistuesi duhet të zgjedh midis dy frazave.* Frazat futen në tekstin e pyetsorit, në bazë të informacionit që intervistuesi ka marrë më parë. Psh “më thoni statusin e punës së gruas/partneres suaj?” Këtu intervistuesi duhet të zgjedh midis përdorimit të fjalës grua dhe partnere në bazë të informacionit që ai ka marrë më parë nga i intervistuari.

### 2.3. Kapërcimet

Ka shumë raste kur një pyetje ose një sekuençë pyetjeje nuk mund të aplikohet tek një i intervistuar i veçantë që intervistuesi po interviston. Në këtë rast intervistuesi duhet ta kapërcej pyetjen ose sekuençën përkatëse. Ka tre lloj kapërcimesh që mund të ndeshen gjat një intervistimi.

“KALO TEK PYETJA 10”. Këtu bëhet kalimi tek një pyetje tjetër, por në të njejtën faqe.

“KALO TEK PYETJA 10 NE FAQEN TJETER”. Këtu bëhet kalimi tek një pyetje tjetër, që ndodhet në një faqen pasardhëse.

“KALO TEK PYETJA 10 NE FAQEN 5”. Këtu bëhet kalimi tek një pyetje tjetër, por që ndodhet në faqen me numër 5.

Në këtë rast intervistuesi nuk duhet të bëjë asnjë shënim në pyetjet ndërmjetëse që kapërcehen.

### 2.4. Llojet e pyetjeve

Pyetje demografike. Këto pyetje janë pyetje deiskriptive të cilat japin një informacion të gjerë rreth statusit social ekonomik të të intervistuarit psh, mosha, gjinia, arsimi, kualifikimet, gjendja civile, numri i antarve në familje etj. Këto pyetje në një studim me anë të intervistimit janë po aq të rëndësishme sa edhe pyetjet e tjera që përbëjnë një intervistë. Por, zakonisht të intervistuarit mund të ndihen të lodhur prej këtyre pyetjeve ose atyre mund t'i duken si shumë personale. Në këto kushte shpesh rekomandohet që këto lloj pyetjesh të vihen në fund të pyetsorit sepse i intervistuari është përgjigjur deri në atë pikë (pranë fundit të intervistës) dhe si të thuash ka një marveshje të heshtur për ta përfunduar intervistën. Por, kjo gjë në asnjë mënyrë nuk duhet marrë si një regull strikt. Shpesh këto pyetje vendosen në fillim të pyetsorit, duke shërbyer si një thyerje akulli dhe njohje reciproke mes dy partnerëve.

Fjalët e pyetjeve. Formulimi i pyetjeve duhet të jetë mjaft i kujdesëshëm. Pyetjet e dykuptimta krijojnë mundësi për keqkuptime. Fillimisht duhet të vendoset se çfarë informacioni do të kërkohet dhe pastaj duhet të ndërtohet pyetja. Kështu pyetja se “Qysh kur jetoni në Tiranë?” mund të marrë përgjigje të tilla si “Që fëmijë”, “Mbasi jam martuar” etj. Pra për të shmangur këto lloj keqkuptimesh pyetja duhet formuluar më drejtë. Ajo mund të ketë një formë të tillë “Prej sa vitesh jetoni në Tiranë?”

Pyetjet e hapura. Në këto lloj pyetjesh ka liri të plotë për të dhënë një informacion të plotë. Shpesh këto pyetje paraqesin vështirësi statistikore për kodim të të dhënave. Si një shembull të një pyetje të tillë mund të përmendim: “Cili është shkaku kryesor i korupsionit në Shqipëri?”, Në këtë rast i intervistuari është i lirë të jap një përgjigje që ai mendon se është shkaku kryesor i korupsionit. Duke qenë se ekzistojnë mendime, eksperiencia dhe botkuptime të ndryshme mes njerzve, atëherë është e kuptueshme që edhe përgjigjet për këtë pyetje mund të jenë të shumta.

Pyetjet e mbyllura. Në këtë rast të intervistuarit i jepen një numër i caktuar alternativash të përgjigjes dhe ai duhet të zgjedh njerën prej tyre. Numri i alternativave zakonisht është i kufizuar (mund të jetë edhe relativisht i madh). Në këto kushte duhet patur kujdes me alternativat që jepen për këto pyetje sepse mund të ndodhemi në kushtet kur të intervistuarit nuk spiegojnë dot pse (alternativat nuk janë të plota). Në pyetjet e mbyllura duhet mbajtur parasysh se informacioni që është i nevojshëm të jepet duhet të jetë në fillim të pyetjes. Kështu p.sh. nuk duhet thënë “A besoni

se duhet të jetë i lejuar aborti nëse jeta e gruas është në rrezik?” por pyetja duhet ndërtuar në formë të tillë “Nëse jeta e gruas është në rrezik, a duhet të jetë i lejuar aborti?”

Pyetjet me ngarkesë emocionale. Këto janë pyetje që nuk janë neutrale por që kanë terma emocional. Një pyetje e tillë është p.sh. “A mendoni se disa segmente ekstermistësh radikal duhet që të lejohen që të djegin ndërtesat publike?” Këtu nuk duhet përdorur fjala ekstremistësh radikal sepse e sygjeron vetvetiu përgjigjen për problemin që trajtohet. Në këto kushte duhet të kihet kujdes në ndërtimin e pyetjeve për të shmangur specifikimet e ndryshme (p.sh. ekstremistët radikal) të cilat e bëjnë të vështirë dhënjen e një përgjigje të ndershme për mendimin e vetë të intervistuarit.

Pyetjet udhëheqëse. Këto pyetje kanë anshmëri e cila vihet re nga drejtimi që pyetja përpiket t’i jap të intervistuarit për tu përgjigjur. Një pyetje e tillë është p.sh. “Shumë njerëz besojnë se riciklimi ndihmon në mbrojtjen e rezervave, ju a jeni dakort?”. Me formulimin “shumë njerëz besojnë” formuluesit e intervistës i japin një drejtim përgjigjes duke i pozicionuar në një drejtim të caktuar përgjigjet që duhet të jepen.

Pyetjet dyfishe. Në këto raste bëhet më tepër se një pyetje në të njëjtën kohë (brenda të njëjtës pyetje). Në raste të tilla shpesh përdoren fjalët “ose, apo”. Kështu p.sh. “Puna që ju bëni, për ju është interesante apo ajo bëhet për ekzistencë?” Në këtë formulim, kjo pyetje duket sikur nxit përgjigjet ekstreme dhe mohon mundësinë për përgjigje ndërmjetëse të tipit “edhe edhe”. Pra, duhet të ketë më tepër shkallëzime për mundësitë e përgjigjeve ose të ndahen pyetjet nga njëra tjetra.

Pyetjet filtër. Janë pyetje mjaft të rëndësishme në një pyetsor intervistimi. Kur ndërton një pyetsor qëllimi është që të ndërtosh pyetje të drejtëpërdrejta, jo dykuptimshme dhe me paanshmëri që duhet të të çojnë në përgjigje të vlefshme, të përdorëshme dhe të besueshme. Jo vetëm renditja e fjalëve në një pyetje, por edhe renditja e pyetjeve në intervistë ka rëndësi të madhe. Për këtë qëllim përdoren pyetjet filter, që e instruktajnë si intervistuesin dhe të intervistuarin se cila do të jetë pyetja tjetër që do të bëhet dhe kjo në varësi të përgjigjes që jepet për pyetjen filter. P.sh. “A pini duhan, cigare, puro?” është një pyetje që ka dy alternativa për përgjigje (Po ose Jo). Nëse përgjigja është Jo atëherë shmangen disa pyetje të tjera që vijnë menjëherë mbas kësaj pyetjeje dhe që kërkojnë të marrin informacion vetëm nga ata të intervistuar që janë përdorues të duhanit. Mbas pyetjes filter në intervistë theksohet “kalo tek pyetja Nr. x”, pra shmang bllokun e pyetjeve të përmendur më sipër.

Pyetjet hinkë (funnel). Në vërtetimin me anë të intervistimit mund të përdoren edhe pyetjet hinkë. Këto lloje pyetjesh janë një grup pyetjesh që janë të renditura nga më e përgjithshmeja deri në më pyetje që kërkojnë informacion më në detaje dhe më specifike. Këto lloje pyetjesh zakonisht shërbejnë për të paraqitur një problem dhe gradualisht për tu fokusuar në të, pra për tu zhvendosur në informacionin e kërkuar në formën e një hinkë.

Me përdorim të kujdesëshëm dhe me një sistemim të mirë të pyetjeve, një vërtetim mund të sjell informacion të vlefshëm dhe të interpretueshëm. Por, fatkeqësisht shpesh është shumë më e lehtë që të ndërtohen pyetje konfuze ose të paraqiten pyetjet me një renditje jologjike, veprime këto që çojnë në dykuptimshmëri të pyetjeve dhe në të dhëna që interpretohen me vështirësi ose edhe në të dhëna të gati të painterpertueshme nga ana e stafit të studimit.

## 2.5. Pyetsori

Pyetsori në shumë vërtetime ka rol thelbësor, sepse është i ndërtuar si kamjon transportues i të dhënave dhe si pasqyrë e objektivave të ndërmara dhe të refleksioneve kritike të përpunuara në momentin e konceptimit të projektit. Pyetsori duhet të jetë në gjendje që përmes intervistimit të japë përgjigje të saktë dhe paekujvoke në shumëllojshmërinë e pyetjeve, ku secila prej tyre vjen nga njohja e një ose më tepër karakteristikave të njësishë së vërtetuar dhe përfaqëson aspektin teknik-njohës që i korespondon një hipoteze të saktësuar mirë të punës. Përgatitja e pyetsorit përfshin një seri problemesh, metodash dhe teknikash në lidhje me natyrën dhe llojin e kërkesës, në mënyrë që atë ta paraqesësh dhe ta rendisësh në vazhdueshmërinë koherente dhe që në çdo rast të bëjë një paraqitje të plotë të fenomeneve që trajton. Suksesi i një vërtetimi varet në një masë të madhe nga cilësia e pyetsorit.

Në krijimin e modelit të vërtetimit para së gjithash duhet të arihet të disponohet dhe analizohet një model konceptual i formuar nga bashkimi i të gjitha strukturave të përfaqësimeve të nevojshme për

të marrë aspektet e fenomenit të vrojtuar. Ndërtimi i një pyetësoi nuk është një ndërmarrje e lehtë ashtu siç mund të duket në momentin e parë: duke mbajtur parasysh veç të tjerave se defektet e ndërtimit të pyetësoit mund të të shpien me lehtësi në shumëfishin e gabimeve (midis të cilave edhe mungesë përgjigjesh, ose përgjigje të gabuara për shumë lloje motivesh). I intervistuari përballë pyetësoit duhet të kuptojë pyetjen, të arrijë të njohë përgjigjet e mundëshme dhe të formulojë përgjigjen e tij.

Për këtë qëllim duhet:

- a) Të formulohen pyetje të tilla që të mos kushtëzojnë përgjigjet.
- b) të evitohen pyetjet e dykuptimta
- c) Të përdoret një gjuhë e kuptueshme për të gjithë për të arritur një interpretim të njëjtë dhe të menjëhershëm të pyetjes nga ana e të intervistuarve.
- d) Të organizohen pyetjet ose grup-pyetjet në një mënyrë të tillë që të mos influencojnë në përgjigjet e pyetjeve në vazhdim.
- e) Të adoptohet formulimi i pyetjeve në mënyrë të përshtatshme për llojin e intervistimit (me postë, me telefon, me vetplotësim)
- f) Të evitohen pyetësorët shumë të gjatë
- g) Të preçizohet perioda e kohës të cilës i referohet pyetja.
- h) Të tregohet njësia e matjes së përgjigjeve numerike.
- i) Të trajtohet me kujdes paraqitja e pyetësoit (forma, titulli, pozicioni i vendosjes së kutijave të kodeve, hapësirat, gërmat, kushtëzimet etj.) jo vetëm për një paraqitje sa më të mirë estetike, por edhe për të realizuar një komunikim sa më të mirë midis intervistuesit dhe të intervistuarit.

## 2.6. Tipologjia e pyetjeve.

Përgatitja dhe formulimi i një pyetjeje nuk është një detyrë e thjeshtë që kërkon që t'i jepet përgjigje një serë problemesh të tilla si dhënia lirisht të intervistuarit për tu përgjigjur cilat dhe sa janë përgjigjet e mundshme, cila duhet të jetë renditja e përgjigjeve etj. Në realitet ekzistojnë një serë kërkesash për pyetjet që duhen bërë, midis të cilave po përmendim ato më kryesoret:

a) Pyetje e hapur ose përgjigje e lirë, e cila i jep liri përgjigjeje të intervistuarit. P.sh. Ç'farë aktivitete bën gjatë kohës së lirë? Avantazhi që kanë këto pyetje nga lehtësia e ndërtimit ka rezik të kushtëzojë përgjigjen dhe prandaj duhet trajtuar me mjaft delikatesë. Kjo sepse ka mjaft disavantazhe gjatë ndërtimit të këtyre llojeve pyetjeje sidomos për faktin se nuk paraqiten në një formë lehtësisht të sintetizueshme nga studjuesi dhe janë pyetje që kanë një vështirësi mjaft të madhe për t'u kodifikuar.

b) Pyetje me përgjigje të paracaktuara ose pyetje të strukturuar në të cilat jepen përgjigje të paracaktuara më parë dhe i intervistuari duhet të zgjedhë mes tyre vetëm njërin ose duhet të zgjedhë sipas renditjes disa prej tyre.

P.sh. Cilët janë sportet që preferoni?

futboll , tenis , basketboll , boks , not , golf , të tjera .

Natyrisht që modalitetet e përgjigjeve nuk duhet që të krijojnë boshllëqe. Duke ditur që mundësitë e përgjigjeve janë të shumta atëherë patjetër duhet që të vendoset edhe mundësia e përgjigjes të tjera të cilat përdoren për mundësitë që preferohen më pak se të tjerat dhe që nuk përdoren në opsionet e dhëna.

Nëse nuk njihen të gjitha përgjigjet e mundshme, atëherë është e këshillueshme të bëhet një provë pilot (kampion) ku përgjigja të jetë një pyetje e hapur ose përgjigje e lirë dhe të studjohen përgjigjet e dhëna ose të bëhet konsultimi i përgjigjeve të mundëshme me ekspert të fushës përkatëse për të cilën bëhet studimi. Avantazhet e pyetjeve të strukturuar janë mjaft të qarta dhe mund të përmendim: plotësimi me lehtësi nga intervistuesi dhe kodifikimi i lehtë, krahasueshmëria e përgjigjeve të cilat nuk janë në një numër mjaft të madh etj. Një rast i veçantë i ndërtimit të pyetjeve të mbyllura është paraqitja e 2 (në realitet 3) alternativave për përgjigje. Pra në këtë rast kemi pyetje dhe përgjigje dikotomike (po ose jo e favorshme ose e pafavorshme, pozitive ose negative etj). Sugjerohet që në grupin e përgjigjeve të vendoset dhe përgjigja “nuk e di” (përgjigja e tretë).



c) Pyetjet filtër përdoren me qëllim që nga përgjigja e dhënë nga të intervistuarit, ata të ndahen në nëngrupe të veçanta që duhet ose nuk duhet t'u përgjigjen një grupi pyetjesh në vazhdim. P.sh. për studentët "A banoni në konvikt?" Nëse jo kalo tek pyetja 10.

d) Pyetje për përgjigje hierarkike të strukturuar gjatë të cilave kërkohet preferenca ose klasifikimi në rendin e rëndësisë të një numri të caktuar të përgjigjeve. P.sh. Renditni sipas rëndësisë nga 1 deri në 6 (nga më i rëndësishmi tek më pak i rëndësishmi) faktorët e mëposhtëm që ndikojnë në vështirësitë për të marrë kredi nga bankat?

| Faktori                                                        | Renditja |
|----------------------------------------------------------------|----------|
| Norma e interesit                                              |          |
| Procedurat komplekse të aplikimit                              |          |
| Kolaterali i kërkuar                                           |          |
| Lidhjet me menaxherët e bankës                                 |          |
| Kërkesa e bankës për të ditur gjenjen e kompanisë sipas viteve |          |
| Zgjatja e kohës nga kërkesa deri në marrjen e kredisë          |          |

## 2.7. Gjuha e përdorur

Përgjigjet e dhëna nga të intervistuarit varen në një masë të madhe nga mënyra e formulimit të pyetjeve në një pyetësor. Duke qënë se përgjigjet janë homogjene dhe të konfrontueshme, pyetjet duhet të shprehen në mënyrë të tillë që të japin të njëjtin kuptim për të gjithë subjektet që duhet të përgjigjen.

Në këtë kuptim, gjuha e përdorur luan një rol thelbësor, prandaj mbi të gjitha duhet përdorur një fjalor dhe një ndërtim fjalie që lehtëson komunikimin, duke përdorur fjalë të thjeshta jo dykuptimshme, të kuptueshme nga të gjithë të intervistuarit, preçize, pa dyshueshmëri përdorimi dhe gjithashtu duhet të ndërtohen fjali në funksion, në funksion të kompetencës së intervistimit, të evitohen pyetjet e gjata, të përbëra nga shumë periudha kohore. Në përgjithësi duhet të mos përdoren terma teknike, përveçse në raste të pakta kur vërtetimi do të bëhet në grupe të caktuara tëspecializuara (p.sh mjekët). Në këtë rast përdorimi i shprehjeve teknike është i domosdoshëm (p.sh përshkrimi i sëmundjes)

## 2.8. Sekuenca e pyetjeve

Një rregull i përgjithshëm është që të krijohet një zonë e pyetësorit që i lihet kodeve të identifikimit dhe kontrollit eventual të cilësisë. Pyetjet duhet të formulohen dhe të vendosen në një vazhdimësi llogjike dhe në mënyrë që të jetë e lehtë për të intervistuarin që të kalojë nga një argument tek tjetri dhe gjithashtu të jetë e mundshme që të dallohen seksionet e destinuara për tematika homogjene dhe për ndërlidhje. Në qoftë se pyetja kërkon sforcim të memories është e këshillueshme që ajo të vendoset në mes të pyetësorit me qëllim që të mos trëmbet i intervistuari që në fillim të punës. Në fillim duhet të vendosen ato pyetje që japin pak informacion me qëllim që të jepet inerci për vazhdimin e intervistës pa e shqetësuar ose emocionuar të intervistuarin ose për të mos i krijuar mendimin se ai është në pozicionin e një të hetuari.

Në rastet kur studimi i fenomenit kërkon bërjen e shumë pyetjeve në vazhdim të njëra tjetrës është e këshillueshme të realizohet në fillim një njohje e përgjithshme dhe pastaj të kalohet në detaje. Në këtë rast është e këshillueshme kalimi me pyetje hink nga pyetjet e përgjithshme në ato më specifike, por duke e ditur se përgjigjet më komplekse janë frut i refleksioneve ose i ndërgjegjësimit atëherë mund të përdoret edhe teknika e pyetjeve suksesive ose hinka e përmbysur në të cilat fillohet me pyetje specifike dhe kalohet në pyetje më të përgjithshme.

## 2.9. Pyetjet retrospektive

Këto përdoren kur duam informacion për një periudhë të kaluar kohe pak a shumë të gjatë. Këto pyetje paraqesin informacion me karakteristika më delikate sidomos kur i referohen një periudhe preçize të kohës të specifikuar nga rregullat e studimit. P.sh. informacion kur personi ose familjarë të tij kanë qënë viktimë e agresionit fizik, vjedhje portofoli, vjedhje të tjera etj.) gjatë 3 muajve të

parë të vitit. Ai mund të harrojë ngjarje të periudhës ose mund të përfshijë edhe ngjarje të mëparshme (efekti i teleskopimit të mëparshëm) ose edhe ngjarje të mëvonëshme (efekti i teleskopimit të mëvonshëm) duke na dhënë informacion të gabuar. Rekomandohet që pyetjet të mos bëhen për periudha shumë të largëta të kohës. Në se pyetësi përmban periudha referuese (viti i fundit, muaji i fundit, mesatarisht gjatë një viti etj.) numri duhet të saktësohet për të mos krijuar situata të përzierjes së periudhave.

#### *2.10. Shkallët e vlerësimit të përdorura*

Shumë shpesh përfundimi i një vërtetimi kërkon njohjen dhe gjykimin nga ana e të intervistuarit të çështjeve dhe problemeve të vërtetimit, pra është e nevojshme të konvertohen gjykimet në matje të tyre nominale të përcaktuara në renditje me vazhdimësi joarbitrare dhe numerike. Kjo gjë realizohet kryesisht në shkallë vlerësimi ku përcaktohen raportet përkatëse (A është më e madhe se B) për të cilat i intervistuari të japë gjykimin e tij. P.sh. e shkëlqyer, shumë e mirë, e mirë. . . shumë, mjaftueshëm, pak, aspak ose shumë e favorshme, e favorshme...) Shkallët duhet të jenë shumë të balancuara midis attributeve pozitive dhe negative dhe të mos krijojnë distanca të ndryshme mes tyre dhe të kenë një element neutral midis tyre për të qënë në ekuilibër dhe me qëllim që të jepet mundësia edhe për përgjigje natyrale dhe jo detyruese. Shpesh preferohen shkallët numerike të shprehura me numra vlerësues (p.sh. nga 1 deri në 5) mes të cilave i intervistuari duhet të zgjedhë atë që është më pranë gjykimit ose shkallës së tij të vlerësimit.

#### *2.11. Pyetjet delikate ose të kondicionuara.*

Shumë shpesh në vërtetime ndodhemi në situata shumë delikate të cilat indukojnë tek i intervistuari mekanizma psikologjik dhe emocional që mund të përcaktojnë refuzimin e bashkëpunimit ose që të kompromentojnë përgjigjet e disa pyetjeve: rezistencë për rregullin e pagesave të tatimeve fiskale, vënia në pozitë sociale të pafavorëshme ose shkelje e privatësisë së jetës së tij etj. Të tilla janë të ardhurat, pasuria, konsumi sidomos i alkoolit, drogës, përdorimi i kontraceptivëve, sëmundshmëria etj. Por në situata të tilla mund të viresh edhe me pyetje më pak delikate kur kërkohen përgjigje pozitive ose negative në terma preçize ose detyrues ose që krijojnë emocione si ndalim të një aktiviteti, abrogim të një ligji, aprovim të një orientimi politik etj.

Për këto lloj pyetjesh me argumenta delikatë situata është e gjithnjë specifike dhe e ndryshme nga rastet e tjera: në këto raste është e domosdoshme të afroheni gradualisht në pyetjen thelbësore duke kaluar përmes pyetjeve më të përgjithshme ose duke përdorur teknikën e pyetjeve indirekte duke vendosur të intervistuarin në pozicionin e një personi të tretë me karakteristika të njëjta me të tijat. P.sh. për të ardhurat nuk këshillohet të bëhet pyetje direkte dhe në një farë mënyre edhe brutale për të ardhurat personale ose edhe familjare por duke bërë dy pyetjet në vazhdim:

1. Sipas jush sa janë nevojat e një familje në kushte si tuajat për të jetuar në këtë qytet pa luks dhe pa i munguar gjërat e nevojshme?

2. Sa para të duhen për të jetuar me familjen ...

Në realitet testet verifikuese kanë treguar se edhe me këto lloj formulimesh ka një fshehje të të ardhurave që ka një person ose një familje e caktuar.

### **3. Broshura e të intervistuarit dhe studimi pilot**

#### *3.1. Broshura e të intervistuarit*

Broshura e të intervistuarit është një mjet që duhet të përdoret e ndërlidhur me pyetësin. Ajo publikon kategoritë e përgjigjeve për pyetjet ku janë të mundshme shumë zgjedhje dhe qëllimi i përdorimit të saj është që të ndihmojë të intervistuarin që të mbaj mend kategoritë e ndryshme prej të cilave ai duhet të zgjedh përgjigjet. Kur broshura e të intervistuarit duhet të përdoret për një pyetje, atëherë kjo gjë duhet të theksohet në pyetësor që në fillim të pyetjes, psh (broshura e të intervistuarit faqe 1.).

Ka raste kur një i intervistuar mund ta përdor vetë broshurën e të intervistuarit. Në këtë rast intervistuesi duhet që t'i tregoj atij vetëm faqen që duhet të hapi dhe të lexojë instruksionin. Në këtë rast intervistuesi duhet të sigurohet që i intervistuari e ka hapur broshurën e të intervistuarit në faqen e duhur, përpara se sa të bëjë pyetjen përkatëse. Mbasi i intervistuari e ka parë faqen e broshurës së të intervistuarit, atëherë intervistuesi me takt duhet tja marrë broshurën të intervistuarit,

me qëllim që i intervistuari të mos lexoj asnjë faqe tjetër, përpara se sa kjo gje të theksohet në pyetsor.

### *3.2. Verifikimi dhe kolaudimi i pyetësorit*

Pyetësori i ndërtuar duhet të kalojë në një serë verifikimesh dhe revizionimesh sidomos përmes kontributit të ekspertëve të fushës së studimit duke arritur në fund në përfundimin e kontrollit se hipotezat e punës janë përfshirë në model në mënyrë korrekte dhe nëse i korespondojnë njohurive të paraqitura. Në këtë fazë duhet të verifikohet nëse mungon ose është nënvleftësuar ndonjë çështje, nëse modalitetet e përgjigjeve të marra në konsideratë për ndonjë pyetje janë të kompletuara, nëse pyetjet retrospektive paraqesin vështirësi. Vërejtjet e bëra nga ekspertët shërbejnë si bazë për modifikimet dhe revizionimet e nevojshme të pyetësorit. Kjo procedurë duhet të respektohet me rigorozitet.

Në vazhdim duhet të bëhet një kolaudim përfundimtar i pyetësorit duke bërë një paratestim me një numër të vogël intervistuarish që përfshijnë të gjithë spektrin në mënyrë që të kemi të gjithë gamën e situatave operative dhe të vrojtueshme sipas nivelit kulturor dhe shpërndarjes territoriale. Këshillohet që paratestimi të bëhet në 5-10% numrit të intervistave që do të bëhen në studim. Në përfundim të tij merren në konsideratë të gjitha problemet e shfaqura (renditja e pyetjeve, periudha e gjatë kohore, asistencë për tu përgjigjur etj.) Në disa raste për një pyetje përdoren dy versione për të vendosur më vonë se cili është më i miri.

Kontrolli i fundit bëhet përmes një vrojtimi pilot ku testohen të gjitha aspektet operative dhe organizative (funksionimi i skemës së vrojtimit, bashkëpunimi, koha, lista e të intervistuarve etj.) Në vrojtimin pilot stimulohen situata të ndryshme operative, ku marrin pjesë strukturat e përfshira në vrojtimit dhe ku verifikohet jo vetëm struktura formale e pyetësorit por gjithashtu edhe të gjitha aspektet e revizionit, paravrojtimit, konfidencës, regjistrimit të mikro të dhënave në programin kompjuterik, korrektimi automatik etj.

## **4. Kodimi**

Gjat ndërtimit të një pyetsori duhet të theksojmë se kemi dy lloj kodimesh. Kodimi i parë është kodimi i përgjigjeve të mundëshme për pyetjet e ndryshme që përmban një pyetsor, ndërsa kodimi i dytë ka të bëjë me kodet që përcaktohen sipas pyetjeve të pyetësorit për të ndërtuar tabelën e të dhënave (databazën e pyetësorit). Kodimi i përgjigjeve të mundëshme për pyetjet e ndryshme që përmban një pyetsor bëhet sepse kryerja e analizës kërkon që përgjigjet e dhëna prej të intervistuarve rreth pyetjeve të ndryshme të konvertohen në numura. Ky proces quhet kodim. Kodimi përfshin tre hapa kryesorë: përcaktimin e kodeve për përgjigjet e çdo pyetjeje (apo variabli); mënyra e hedhjes së kodeve në kompjuter; dhe kontrollin e kodeve.

### *4.1. Përcaktimi i kodeve për çdo variabël*

Secili prej variablave ka të paktën dy kategori dhe çdo person duhet që të përfshihet vetëm në njërin prej kategorive. Thelbi i kodimit është që t'i caktohet çdo përgjigjeje të pyetjejeve të pyetësorit një numur i veçantë. Pra, çdo përgjigje e një pyetjeje të veçantë duhet që t'i vendoset një kod i veçantë. Në këtë mënyrë krijohet një korespondencë e plotë ku çdo numur përfaqëson një përgjigje të përcaktuar në pyetjen në shqyrtim dhe nga ana tjetër çdo përgjigje e mundëshme e pyetjes në shqyrtim ka të përcaktuar një numër të vetëm. Kodet mund të vendosen si përpara se të jenë dhënë përgjigjet (përpara se të jetë plotësuar pyetsori), pra kemi parakodim ose pasi janë dhënë përgjigjet, pra kemi paskodim.

*Parakodimi.* Tek pyetjet që janë me zgjedhje të detyruar të përgjigjes është e mundur që kodet për përgjigjet të vendosen përpara se të intervistuarit të plotësojnë pyetësorin. Kjo është e mundur pasi kemi një rang përgjigjesh të përcaktuara që më përpara. Për ta bërë më të lehtë kodimin e përgjigjeve aktuale të dikujt, kodet që përfaqësojnë çdo përgjigje janë të printuara në pyetësor. Kjo mund që të bëhet në mënyra të ndryshme.

*Paskodimi.* Tek pyetjet me fund të hapur kodimi nuk mund që të bëhet derisa pyetsori nuk është plotësuar nga të intervistuarit, përderisa nuk i kemi akoma përgjigjet. Ka dy mënyra për të bërë paskodimin. Mund që të përdoret një skemë e paracaktuar kodimi. Kur kemi idenë e plotë se si ne dëshirojmë që t'i kodojmë përgjigjet, ne thjeshtë kodojmë në bazë të kësaj skeme. Një shembull i

tillë është ai i kodimit të punësimit. Punësimi aktual mund që të vendoset në kategori të gjera, të paracaktuara sipas njohjes që ne kemi. Kur bëhet kodimi në këtë mënyrë duhet të kihet parasysh që nga analiza që do të bëhet më vonë të nxirret sa më shumë informacion i detajuar.

Për shembull, në vend që të kodohet punësimi thjeshtë në grupe si nëpunës dhe punëtorë, ne mund që të kodojmë punësimin në një mënyrë më të arsyeshme sipas llojeve të punësimeve më specifike (p.sh. menaxherë, elektrikist, murator etj) apo edhe në kategori më pak të gjëra (p.sh. nëpunës, punëtorë, tregtarë etj.). Këto kode specifike me vonë mund të ndahen në kategori në të vogla. E kundërta nuk është e mundur: nuk mund që të krijohen katagori specifike nëse kodimi origjinal nuk përfshin detaje të plota.

Ndërtimi i kategorive të kodimit në bazë të përgjigjeve të të intervistuarve. Kjo procedurë ka të bëjë me faktin se ne krijojmë mundësi që kategoritë e kodimit të rrjedhin prej të dhënave në vend që t'i vendosim që më parë (apriori) këto kategori tek të dhënat. Për shembull ne mund që të kemi ndërtuar pyetjen “Çfarë mendoni mbi politikën e vendeve europiane mbi emigracionin?” Për të ndërtuar një komplet kodesh ne duhet që të shikojmë 50 –100 përgjigjet e para dhe duke u bazuar në to bëjmë një listë me përgjigjet e dhëna dhe shohim nëse ato mund që të grupohen në grupe të ndryshme më konçize.

Disa prej përgjigjeve mund të klasifikohen si aprovuese dhe disa si mosaprovuese. Përgjigjet që hyjnë në kategorinë e mosaprovimit mund që t'i ndajmë në nën kategori si “jo të kënaqur sepse lejojnë shumë pak emigrantë”, apo “sepse krijojnë bariera artificiale ndaj emigrantëve”. Pasi të kemi ndërtuar një komplet kategorish, të tillë që të gjitha përgjigjet e marra të mund të klasifikohin (të hyjnë) në njërën prej tyre, çdo kategorie i përcaktohet një kod dhe çdo përgjigje kodohet. Duhet patur kujdes që çdo kategori është ekskluzivisht reciproke, ku një person duhet që të kodohet vetëm në njërën prej kategorive për variabël.

#### *4.2. Kodimi i të dhënave numerike*

Për disa variabla të tillë si mosha, të ardhurat, numuri i fëmijëve etj, përgjigjet janë vetë më formë numerike, kështu që nuk është e nevojshme që përgjigjet të kodohen. Ne mund t'i konsiderojmë përgjigjen e tyre si kodime në mënyrë mjaft egzakthe. Kështu nëse dikush është 39 vjeç, ne mund që ta kodojmë ato thjesht 39. Shpesh njerëzve u kërkohet që të zgjedhin njërën prej kategorive të të ardhurave apo kategorive të moshës (p.sh. 25-35vjeç, 36-45 vjeç). Në këto raste është më mirë që këtyre përgjigjeve tu vendoset një kod i cili përfaqëson mesin e kategorisë. Kështu, nëse dikush bën pjesë në kategorinë 36-45vjeç, ne mund që ta kodojmë atë si 40. Kur njerëzve u kërkohet që të përgjigjen në këtë mënyrë atëherë ja vlen që të vendoset pika e mesit përbri secilës kategori për ta bërë kodim më të lehtë.

#### *4.3. Si të veprojmë me të dhënat që mungojnë*

Çdo variabël për secilin rast duhet që të ketë një kod edhe kur personi nuk i përgjigjet një pyetjeje. Këto kode quhen kodet e të dhënave që mungojnë dhe bëhen në të njëjtën mënyrë si në kodimin normal. Gjëja kryesore që duhet që të sigurohemi kur mungojnë të dhëna është që të ndërtohen kode të dallueshme për të dhënat që mungojnë. Duhet që kodi të jetë i dallueshëm prej kodeve të tjera të përgjigjeve të vlefshme. Meqë ka mjaft arsye të ndryshme që spjegojnë faktin se pse të intervistuarit nuk i përgjigjen pyetjeve, shpesh tipeve të ndryshme të mungesës së përgjigjeve u caktohen kode të ndryshme. Ka katër tipe kryesore të arsyeve pse pyetjet mbeten pa përgjigje: Të intervistuarit nuk i kërkohet që t'i përgjigjet pyetjes (p.sh. një pyetje rreth moshës së fëmijëve kur i intervistuari ka thënë se nuk ka fëmijë).

E pa verifikuar (e pa kapur): në rastet kur intervistuesi apo i intervistuari e kanë kaluar pa e parë pyetjen, apo në rastet kur pyetja nuk ishte e qartë. I intervistuari ka refuzuar që të përgjigjet në një ose disa pyetje. I intervistuari nuk dinte si të përgjigjej apo nuk kishte ndonjë opinion për problemin për të cilin u pyet. Ndonjëherë kjo mund që të trajtohet edhe si përgjigje e vlefshme ndërsa në disa raste ajo trajtohet si mungesë të dhënash.

Në këto kushte, këtyre tipeve të ndryshme të mospërgjigjeve, mund të nevojitet që tu caktohen kode të ndryshme sipas llojit të të dhënave që mungojnë. Nuk ka rregulla të përcaktuara qartë rreth asaj

se çfarë kodesh të veçanta duhet që tu vendosen të dhënave që mungojnë për sa kohë që kodi i të dhënave që na mungojnë nuk ngatërrohet me kodet e vlefshme. Kodet të cilat janë “të lira” për të dhënat që mungojnë varen prej asaj se sa kode të vlefshme kemi për pyetjet. Nëse janë “të lira” përdoren shpesh edhe kodet nga 0 në 9. Në mjaft raste sygjerohet që të përdoren kode me numra njëshifrorë negativë.

Është e këshillueshme që të përpiqemi që t’i japim të njëjtin lloj kodi të dhënave që mungojnë në sa më shumë pyetje në mënyrë që të shmangim konfuzionin dhe që ta bëjmë programin në kompjuter sa më të thjeshtë. Ideale është që për të dhënat që mungojnë të përdorim të njëjtat kode gjatë gjithë pyetsorit dhe kjo realizohet duke përdorur numrat njëshifrorë negativë. Për shembull, shpesh numrat pozitiv nga 0 deri në 9, por dhe dyshifrorë mund që të jenë kode të vlefshme (p.sh. në rastin e moshës), në këto kushte të dhënave që mungojnë mund tu vendosim kode të tjera me numra negativë suke filluar nga -1 deri në -9.

#### *4.4. Mënyra e hedhjes së kodeve në kompjuter*

Që të vendosim kodet në kompjuter ato duhet që të vendosen në një dokument (arshivë). Kodet për të gjithë të intervistuarit duhet të vendosen së bashku në një dokument në mënyrë të tillë që çdo rresht t’i përkasi një të intervistuari dhe kështu në vazhdim. Vendosja në këtë mënyrë e të dhënave sipas rreshtave krijon dosjen e të dhënave (data file).

Kur hedhim kodet prej pyetësorit në kompjuter duhet që kodet të cilat përfaqësojnë përgjigjet për pyetjet e veçanta duhet të vendosen në të njëjtin pozicion, pra në të njëjtën shtyllë për secilin rast. Për shembull, le të themi se kodin për opsionet e përgjigjes “gjinia” duhet t’i vendosim në shtyllën me numër x (të themi 5) dhe për opsionet e përgjigjes “arsimin” i vendosim në shtyllën y (le të themi me numër 6) dhe kështu me rradhë. Për të ndihmuar operatorin gjatë hedhjes së të dhënave është normale që të përcaktohet një kolonë në të djathtë të fletës së printuar të pyetësorit për të vendosur kodet e përcaktuara. Kjo ndihmon dhe lejon operatorin që të mund që të hedh të dhënat në kompjuter direkt prej pyetësorit. Zakonisht, marxhina e anës së djathtë të pyetësorit i rezervohet këtij qëllimi. Kështu, në anën e djathtë të pyetjeve krijohet një vend në të cilin mund vendosen kodet e pyetjeve.

Operatori kur hedh të dhënat mund që përqëndrohet vetëm në këtë “kolonë” dhe të hedh lehtësisht të dhënat në kompjuter, duke bërë edhe kontroll të vazhdueshëm nëse i ka vendosur kodet e duhura në shtyllat përkatëse. Pyetsori në ankesin I në fund të këtij libri paraqet shembullin e një modeli ilustrues.

#### *4.5. Kontrolli për gabimet në kodim*

Pavarësisht se mund të tregojmë shumë kujdes gjithmonë është e mundur që të bëhen gabime, si në kodim ashtu edhe në hedhjen e të dhënave në kompjuter. Shumë prej problemeve të hedhjes së të dhënave në kompjuter janë minimizuar prej programeve kompjuterike të sofistikuara. Megjithatë, në vetvehte programet nuk mjaftojnë dhe për këtë qëllim duhet bërë kontrolle të herëpashershme. Kjo duhet që të bëhet me ndihmën e programeve kompjuterike dhe arkivohen lehtësisht me SPSS. Në këtë rast bëhen tre lloje kontrollesh kryesore.

*Kontrolli i rangut të vlefshëm:* për çdo variabël ka vetëm disa kode që janë të vlefshme si kode të përgjigjeve të vlefshme ashtu edhe për të dhënat që mungojnë. Çdo kod jashtë këtij rangu tregon një kod të gabuar dhe duhet që të bëhen korrigjimet e duhura.

*Kontrollet me filtër:* nëse përdoren pyetjet filter (disa pyetje duhet që të marrin përgjigje vetëm prej një pjese të të intervistuarve) atëherë kontrollohet nëse ka të intervistuar që janë përgjigjur megjithëse ata nuk duhet ta bënë këtë gjë ose edhe e anasjelltë.

*Kontrollet logjike:* disa përgjigje mund të jenë jologjike. Kështu për ilustrim po japim rastin e mëposhtëm. Për shembull, nëse një i intervistuar përgjigjet se ai është 50 vjeç është jologjike që përgjigja për moshën e fëmijës së tij më të madh të jetë 48 vjeç.

#### *4.6. Kodet sipas pyetjeve të pyetësorit*

Një masë tjetër që shërben për të ndihmuar operatorin për të hedhur sa më lehtësisht të dhënat është edhe kodimi i pyetjeve të pyetësorit. Kjo procedurë ka të bëjë me vendosja e numrave për çdo pyetje apo të dhënë tjetër të pyetësorit duke filluar nga numri 1. Këta numra do të shërbejnë edhe si



shoqërues (ose si zëvendësues të fjalëve) në emërtimin e kokës së çdo shtylle të tabelës së të dhënave duke përcaktuar kështu edhe se në cilën shtyllë do të hidhen të dhënat e secilës pyetje. Në këtë mënyrë kemi ndërtuar strukturën e tabelës së të dhënave (databaza). Kjo gjë lehtëson punën dhe lejon operatorin që të hedh të dhënat në kompjuter direkt prej pyetësorit. Zakonisht, për këtë qëllim rezervohet ekstremi i djathtë, jashtë marxhinave të fletës së pyetësorit.

Pra, në ekstremin e djathtë të fletës së printuar, anash pyetjeve dhe kollonave të përcaktuara për kodet e përgjigjeve të pyetjeve vendosen numrat e kodeve të vetë pyetjeve. Në varësi të informacionit që jep, çdo pyetje mund të ketë një ose disa kode.

## **5. Ndërtimi i shkallëve të matjes**

Në studimet me anë të pyetësorëve ndërtimi i shkallëzimeve është thjeshtë një version më i formalizuar dhe më sistematik dhe një masë e përbërë e një koncepti, një masë e përbërë prej informacionit që rrjedh prej një sërë pyetjesh apo indikatorësh. Për të krijuar një shkallë ne thjeshtë konvertojmë shumë variabla relativisht specifik në një apo më shumë variabla abstrakt. Kështu, në vend që të masim konservatorizmin thjesht duke pyetur se për kë parti politike voton dikush, ne mund që ta pyesim për një varg çështjesh që ne mendojmë se përshkruajnë konservatorizmin.

### **5.1. Pse i ndërtojmë shkallët**

Ka shumë arsye pse është e nevojshme që të masim një koncept duke përdorur disa indikatorë dhe jo vetëm një indikatorë të vetëm. **Së pari** kjo gjë na ndihmon që të përftojmë një kompleksitet të konceptit. Nëse nuk e kemi përkufizuar një koncept shumë ngushtë atëherë ne na duhen shumë indikatorë që të arrijmë kompleksitetin e shumicës së koncepteve. Për të matur besimin fetar ne mund që të pyesim se sa shpesh dikush shkon në kishë/xhami, por nëse e shikojmë besim si më kompleks se sa vajtja në institucionet fetare, ne duhet që të pyesim mbi besim apo ritualet, dhe të nxjerrim se sa i rëndësishëm është besimi në jetën e dikujt. Duke i marrë së bashku këto pyetje përftojmë një masë më të mirë matjeje mbi besimin se sa do të kishim vetëm nga një matje e vetme.

**Së dyti**, kjo gjë ka të bëjë me ndërtimin e matjeve më të vlefshme. Shpesh një vëzhgim në vetëvete mund që të jetë çorirentues dhe ne na duhet që të shohim të gjithë kontekstin e vëzhgimeve të tjera që të mënjanojmë çorirentimin. Ndërtimi i shkallëve me shumë njësi mund të na ndihmojë që të shmangim çorirentimet dhe gabimet të cilat mund që të na dalin duke përdorur vetëm një masë me një njësi për konceptet komplekse.

**Së treti**, na ndihmon që të rrisim besueshmërinë. Mënyra në të cilën ndërtohet një fjali mund që të ndikojë shumë në mënyrën se si përgjigjen të intervistuarve. Nëse do të mbështetemi vetëm në një pyetje, përgjigjet e të intervistuarve do të varen shumë nga mënyra e ndërtimit të fjalisë. Duke përdorur shumë pyetje minimizohen gabimet në gjetjen e fjalëve të duhura.

**Së katërti**, kemi më shumë saktësi. Një pyetje e vetme nuk na lejon që të bëjmë diferencimet e duhura me shumë saktësi. Për shembull, në se ne do të përdorim opsionin e vendbanimit të njerëzve, vendbanimi në periferi i përdorur si masë për të matur statusin e njerëzve do të na çonte në një klasifikim të padrejtë. Ne do t'i klasifikonim banorët e periferisë në zona të pasura me një status më të lartë se ata të periferisë në zona të varfra, por kemi edhe shumë njerëz të cilët variojnë në vendbanim duke patur të njëjtin status. Në këtë kushtë duhen disa pyetje shitesë, le të themi, mbi punësimin, edukimin, të ardhurat, pyetje të cilat do të na ndihmonin që të bënim dallimin midis nivele të statusit. Vetëm kështu ne do t'i klasifikojmë më drejt njerëzit në varësi të statusit të tyre. Ndërsa vetëm me një indikator kjo saktësi nuk mund që të arrihet në asnjë mënyrë.

**Së pesti**, duke përmbledhur informacionin që shprehet prej një numuri pyetjesh rreth një variabli analiza thjeshtohet në mënyrë të konsiderueshme. Në vend që të analizajomë çdo pyetje në mënyrë të veçantë ne mund ta bëjmë këtë gjë në tërësi vetëm me një përpjekje.

### **5.2. Llogjika e shumimit të shkallëve**

Një shkallë konsiston në përgjigjet e dhëna ndaj një numuri pyetjesh. Për secilën pyetje të intervistuarit japin përgjigje që vlerësohet me pikë që varen prej përgjigjeve të tyre. Pikët shpërndahen sipas përgjigjeve të veçanta në varësi të asaj se sa favorizuese është përgjigja lidhur me atë që do të matet. Në fund të këtij procesi, për çdo të intervistuar pikët e secilës pyetje mbledhen së bashku për të përftuar pikët gjithsej. Këto pikë totale përdoren që të tregojmë

“pozicionin” në dimensionin abstrakt, pozicionim për të cilin ndërtohen edhe pyetjet individuale. Kështu për shembull:

**Si i vlerësoni vështirësitë e mëposhtme në marrëdhëniet me gjykatat:**

*Lutemi përgjigjuni në shkallën nga 0 deri në 4 ku: 0=aspak problem dhe 5= shumë problem*

|                                                    |                     |
|----------------------------------------------------|---------------------|
| Shpenzimet gjyqësore jozyrtare janë shumë të larta | (0) (1) (2) (3) (4) |
| Vendimet e gjyqtarëve janë të pandershme           | (1) (2) (3) (4) (5) |
| Gjykatësit nuk janë kompetentë                     | (1) (2) (3) (4) (5) |
| Proçeset gjyqësore zgjasin shumë                   | (1) (2) (3) (4) (5) |
| Përmbartimi nuk zbaton vendimet                    | (1) (2) (3) (4) (5) |

Në vazhdim analiza bëhet si për çdo pyetje në veçanti dhe për përmblendhjen e të gjitha pyetjeve së bashku. Pikët gjithsej të marra nga përmblendhja e të dhënave të të pesë pyetjeve mund të interpretohen si rezultante e problemeve që të intervistuarit kanë në momentet që ata kanë patur punë me gjykatat.

| I intervistuari | Pyetja |   |   |   |   | Pikët gjithsej |
|-----------------|--------|---|---|---|---|----------------|
|                 | a      | b | c | D | e |                |
| 1               | 4      | 4 | 4 | 4 | 4 | 20             |
| 2               | 4      | 2 | 1 | 0 | 3 | 10             |
| 3               | 2      | 1 | 2 | 1 | 1 | 7              |
| 4               | 0      | 0 | 0 | 0 | 0 | 0              |

Megjithëse mënyra e dhënë më sipër është mënyra më e zakonshme, ajo nuk është e vetmja mundësi për të përmblendhur shkallëzimin. Ka mjaftë raste kur mund të përdoren dhe metodat e mëposhtme. Jepet një numur i ndryshëm përgjigjesh për kategori. Ndonjëherë përdoren vetëm dy përgjigje (po dhe jo ose dakort dhe jodakort) dhe në raste të tjera mund të përdoren më shumë kategori si p.sh. katër, pesë, shtatë apo edhe dhjetë mundësi përgjigjesh.

Shpesh më mirë se sa të pyeten të intervistuarit në se janë apo jo dakort me një fjali, atyre ju kërkohet që të japin një vlerësim (të vendosin një sasi pikësh) si përgjigje për pyetjen e bërë. Për shembull, ne mund na nevojitet që të vlerësojmë mardhënjet e dikujt me nënën e vet. Në këtë rast këtë gjë mund ta bëjmë në mënyra të ndryshme si më poshtë:

Ju lutem vlerësoni mardhënjet tuaja me nënën tuaj, a janë ato:

Të ngrohta 0 1 2 3 4 Të ftohta  
Të lehta 0 1 2 3 4 Të vështira  
Të hapura 0 1 2 3 4 Të mbyllura  
Sipërfaqësore 0 1 2 3 4 Të thella

Duhet të theksojmë se shumimi i shkallëve thjeshtë kërkon që të masim fortësinë e disa opinioneve mbi disa çështje dhe kjo gjë bëhet duke vlerësuar “pozicionin” mbi një numër opinionesh. Problemi i të gjitha teknikave të shkallëzimit qëndron në atë se si duhet të përcaktojmë pyetjet që do të përfshihen në shkallëzim. Ne nuk mund që të mbledhim së bashku pikët prej çdo lloj grup pyetjesh, por ne duhet të jemi të sigurtë se ato të gjitha na çojnë në të njëjtin koncept. Kjo gjë mund të bëhet përmes ndërtimit të një *shkalle jo të rregullt*, e cila mund të realizohet me katër hapa:

*Ndërtohen një grup pyetjesh* të cilat në pamje të parë masin një koncept të rëndësishëm. Këto pyetje zakonisht kanë një përbërje përkufizimesh të favorshme dhe të pafavorshme për të cilat të intervistuarit duhet të shprehin nëse janë dakort apo jo, ose për nivelin deri ku janë dakort ose jo. Meqë një nga qëllimet e shkallëzimit është që të jetë në gjendje që të bëhet dallimi midis të intervistuarve të cilët shprehin për nivele të ndryshme (japin vlerësime me pikë në nivele të ndryshme), nuk ka kuptimi ndërtimi i njësisë ekstreme me të cilat të gjithë duhet që të bien dakort



(apo të mos jenë dakort). Për shembull, nesë ne do të donim që të dinim mendimin e njerëzve për nënat që punojnë, njësitë e mëposhtëme nuk do të ishin shumë të goditura. Nuk ka asnjë rrethanë në të cilën nënat do linin fëmijët e tyre vetëm me qëllim që të gjenin një punë me pagesë. Më mirë do të ishte që të ndërtonim pyetje më pak ekstreme si: Normalish do të ishte më mirë për nënat me fëmijë të vegjël që të rrinin në shtëpi se sa të punonin.

*Më pas këtyre pyetjeve ju japin përgjigje një grupi njerëzish me një strukturë të ngjashëm me objektin e studimit që ne duam që tu përgjigjen pyetjeve të ndërtuara.*

*Vendosen pikët e përgjigjeve për çdo person sipas çdo pyetjeje. Është shumë e rëndësishme që renditja e pikëve të bëhet në mënyrë të vazhdueshme (në drejtimin rritës) sepse një përgjigje “e favorshme” gjithmonë të ketë më shumë pikë. Kështu që për përgjigjet “pozitive” (p.sh. të orientuar nga e ardhmja) përgjigja plotësisht dakort të ketë më shumë pikë.*

*Sipas pyetjeve, mbledhen pikët e çdo personi për të nxjerrë pikët totale. Në këtë rast pikët maksimale do të kenë një vlerë të caktuar e cila është e pamundur të kapërcehet. Gjithashtu theksojmë se pikët minimale që mund të arrihen janë zero.*

### 5.3. Probleme për ndërtimin dhe përdorimin e shkallëve

Ka disa probleme kryesore që mund ta komplikojnë ndërtimin dhe përdorimin e shkallëve.

#### *Interpretimi i pikëve të niveleve*

Meqë pikët e një niveli janë një përmbledhje e përgjigjeve të një sërë pyetjesh, atëherë është e kuptueshme se ka humbje në detaje. Ky përgjithësim çon në një numër vështirësish.

**Së pari**, ne nuk e dimë se si është përgjigjur dikush në një pyetje të veçantë pra humbasin detaje të informacionit.

**Së dyti**, megjithëse dy njerëz mund të kenë të njëjtat pikë nuk është e thënë që u janë përgjigjur pyetjeve në mënyrë identike. Të njëjtat pikë mund që të arrihen duke ju përgjigjur grup pyetjeve në mënyrë thuajse të ndryshme. Megjithëse dy njerëz mund që të jenë të ngjashëm më sjelljen e tyre në përgjithësi ata mund që të ndryshojnë në pyetje të veçanta.

**Së treti**, pikët e nivelit duhet që të interpretohen në terma relative. Kjo gjë do të thotë së pari se p.sh. dikush që ka marrë 30 pikë në një shkallë konservatorizmi me maksimum 40 nuk mund që të përkufizohet si 75% konservator apo dy herë më shumë se një që ka marrë 15 pikë. Meqë pikët e nivelit nuk kanë një interpretim preciz, gjithçka që mund të themi është se sa më shumë pikë të ketë marrë dikush aq më shumë konservator është ai.

Ka edhe një kuptim të dytë në të cilin pikët janë relative. Shpesh ne duam të dimë nesë pikët e dikujt duhen konsideruar në nivele “të larta” apo “të ulta”. Kjo gjë mund që të nxirret prej shpërndarjes relative të pikëve të personave të tjerë në model. Për shembull, në një shkallë me pikë nga 0 në 40, si do t’i interpretonim pikët 30? Nëse të gjithë kanë marrë midis 30 dhe 40 atëherë do të themi se vlera 30 është për pikë relativisht të ulëta. Nëse të tjerët kanë marrë nën 30 pikë atëherë vlera 30 do të konsiderohet si pikë relativisht të larta. Shpesh studjuesit theksojnë se i intervistuari i tretë në model me pikët më të larta do të quhet me nivel të lartë dhe i intervistuari i tretë nga fundi me pikët më të ulta do të quhet me nivel të ulët. Pikët e tre persoave të intervistuar të cilët ndodhen në mes të serisë do të quhen pikë të moderuara.

#### *Njësitë ekuivalente*

Siç kemi theksuar deri tani, çdo njësi në shkallë kontribuon njësoj në pikët totale. Në shembujt e dhënë më sipër çdo njësi ka pasur një rang pikësh nga 0 në 4. Shpesh kërkuesit ndjejnë se disa njësi janë më të rëndësishme se sa të tjerat dhe dëshirojnë që ato të kontribujnë më shumë te pikët totale. Për këtë arsye njësitë mund të vlerësohen. Për ta bërë një njësi dy herë në të rëndësishme se sa të tjerat, ne mund që t’i vendosim 8 pikë “plotësisht dakort”. Duke lënë mënjëanë vështirësitë në përpunimin e të dhënave, lind pyetja se sa më shumë ekstra duhet që të vlerësohet njësia e veçantë? Studjuesit zakonisht nuk e vlerësojnë shumë këtë procedurë dhe i trajtojnë të gjitha njësitë si të shkallëzimi Likert sikur të jenë njësoj të rëndësishme dhe ndërtojnë të ashtuquajturat shkallë të vlerësuar njësoj.

Kur përdorim shkallët e vlerësuar njësoj dalin dy probleme kryesore. Këto probleme shkaktohen si prej *variablave që kanë shpërndarje të ndryshme* dhe nga *variabla që kanë numur të ndryshëm*

*përgjigjesh.* Kombinimi i njësive që kanë shpërndarje të ndryshme në shkallë do të thotë se tek pikët totale, njësi të ndryshme do të jenë të vlerësuara në mënyrë të ndryshme. Ky problem sjell edhe një herë në fokus të kuptuarit e pikëve të veçanta apo përgjigjeve në njësi të ndryshme të përdorura për të ndërtuar një shkallë.

Supozoni se duam që të ndërtojmë një indeks të kënaqësisë së jetës të bazuar në 5 njësi, secili prej tyre është ndërtuar për të matur ndienjën e njerëzve për kënaqësi në aspekte të ndryshme të jetës së tyre si puna, martesë, edukimi, standarti i jetës dhe jeta sociale. Për secilën prej tyre të intervistuarit duhet që të zgjedhin midis pikëve 0 dhe 10, për atë se sa të kënaqur janë me atë aspekt të jetës. Sa më shumë pikë aq më të kënaqur janë njerëzit me jetën e tyre.

Pyetja qëndron në atë se nëse një vlerë me 5 pikë në pyetjen e martesës është e njëjtë me një vlerë me 5 pikë në standartin e jetesës. A duhet që të dyja të kenë të njëjtën peshë? A kanë ato të njëjtën sasi satisfaksioni nëse ne po masim një masë të përgjithëshme kënaqësie? Megjithatë është i njëjti numur, ato mund të kenë një kuptim thuajse të ndryshëm kur shihen në kontekstin e pjesës tjetër të të intervistuarve që japin përgjigje. Kështu p.sh. për sa i përket marrtesës mund të tregojnë një pakënaqësi relative, ndërsa te standarti i jetesës mund të tregojnë një kënaqësi relative. Thelbi është se pikët që njerëzit marrin nuk janë absolute në kuptim por ato janë relative.

Problemi i dytë në ndërtimin e shkallëve njësor të vlefshme ngrihet kur ne do të dëshironim që të kombinonim numër të ndryshëm kategori përgjigjesh. Nëse nuk janë bërë transformime, gjatësia e ndryshme e shkallëve do të thotë se disa njësi janë bërë automatikisht më të rëndësishme. Për shembull, një njësi mund që të ketë dy kategori (dakort/jo dakort) me kode 0 dhe 1, dhe një tjetër me pesë kategori (plotësisht dakort deri te plotësisht jo dakort) me kode 0, 1, 2, 3 dhe 4 ndërsa një tjetër mund që të ketë 10 kategori me një shkallëzim më të imët me kode nga 1 në 10. Në këtë rast një “plotësisht dakort” në kategorinë me 10 njësi do të merrte 10 pikë, në kategorinë me katër do të merrte 4 pikë, dhe në kategorinë me dy njësi do të merrte 1 pikë. Është e qartë se “plotësisht dakort” në kategorinë me 10 njësi kontribuon më shumë te pikët përfundimtare sesa te njësitë e tjera.

Në këtë rast një mënyrë zgjidhje është që të rivlerësohen njësitë në mënyrë të tillë që të jenë të barabarta. Në këtë rast ne mund të shumëzojmë pikët e shkallës me dy njësi me 10, kështu që e bëjmë “plotësisht dakort” në atë njësi të barabartë me 10, dhe “dakort” do të mbetet 0. Te njësitë me pikë nga 0 te 4 do t'i shumëzojmë me 2.5, duke e bërë “plotësisht dakort” të barabartë me 10. Kështu qenia jo dakort do të ketë të njëjtën vlerë në të tre rastet.

### **(Bashkengjitur 3 Aneke)**

Data: \_\_\_\_\_/\_\_\_\_\_/2002

Dita Muaji

Koha e fillimit: \_\_\_\_\_

Koha e mbarimit: \_\_\_\_\_

ID e intervistës \_\_\_\_\_

ID e Intervistuesit: \_\_\_\_\_

## PYETËSOR MBI INVESTIMET E HUAJA DIREKTE NË PREFEKTURËN XXX

Mirëdita!

Më quajnë \_\_\_\_\_ dhe do të dëshiroja të flisja me Ju si përfaqësues/menaxher i firmës.

Unë jam bashkëpunëtor i *Institutit XXXX*, i cili po punon për një projekt për mjedisin e investimeve të huaja në prefekturën e XXX. Në bazën e kësaj analize është edhe kryerja e vrojtimit të kompanive me kapital të huaj. Qëllimi i këtij vrojtimi është të vlerësojë pikëpamjet e investorëve të huaj në lidhje me ambjentin politik, ekonomik dhe ligjor në Shqipëri, avantazhet dhe pengesat me të cilat janë përballur gjatë periudhës së veprimtarisë si dhe mendimet e tyre për përmirësimin e klimës së investimeve në Shqipëri. Kompania juaj

A.

V

**ËSHTRI  
M I  
PËRGJI  
THSHË  
M**

- Cila është veprimtaria kryesore e biznesit tuaj në Shqipëri (ju lutem zgjidhni sektorin kryesor të biznesit nëse jeni në më shumë se 1 sektor):** 1= Prodhim, 2= Tregti, 3= Ndërtim, 4= Banka, 5= Shërbime, 6= Transport, 7= Bujqësi 8=Të tjera (specifikoni)\_\_\_\_\_ (1) (2) (3) (4) (5) (6) (7) (8) (1)
- Muaji dhe viti i fillimit të veprimtarisë** Muaji \_\_\_\_/\_\_\_\_ Viti (2)
- Cila është forma ligjore e firmës tuaj?** (1) (2) (3) (4) (5) (6) (3)  
1= ShPK, 2= Shoqëri anonime (aksionere), 3= Shoqëri Kolektive, 4= Shoqëri komandite, 6= Të tjera (specifikoni)\_\_\_\_\_
- Cili është produkti/shërbimi juaj kryesor?** \_\_\_\_\_ (4)
- Sa punëtorë me kohë të plotë janë të punësuar në kompaninë tuaj:**  
1= Aktualisht \_\_\_\_\_ (5)  
2= Në vitin 1996 \_\_\_\_\_ (6)  
3= Në vitin e krijimit \_\_\_\_\_ (7)
- Ç'pjesë e firmës suaj zotërohet nga:**  
1= Të huaj \_\_\_\_\_% (8)  
2= Persona fizike/juridike vendas \_\_\_\_\_% (9)

3= Shteti\_\_\_\_% (10)

7. Sa është numri i aksionerëve në kompaninë tuaj? \_\_\_\_\_  
(11)(11a)

8. Cila është origjina e kapitalit të huaj në kompaninë tuaj? 1=Greqi, 2=Itali, 3=Gjermani, 4=Turqi, 5=ShBA, 6=Tjetër Specifikoni) \_\_\_\_\_ (1) (2) (3) (4) (5) (6)  
(12)(12a)

9. Pronari kryesor i huaj i kompanisë suaj është i përfshirë në ndonjë biznes tjetër jashtë Shqipërisë? 0=Jo, 1=Po (0) (1) (13)

**Shenim për intervistuesin: Nëse 0=Jo kalo tek pyetja 11**

10. Në ç'farë shteti \_\_\_\_\_ (14)

11. A ka qenë shteti pronar 100% i kësaj firme? 0=Jo, 1=Po (0) (1)  
(15)

12. A i perkiste më parë ndonjë nga mjetet (pasuritë) e firmës tuaj, një ndërmarrjeje shtetërore? 0=Jo, 1=Po (0) (1) (16)

**B. SHITJA**

13. Gjatë dy viteve të fundit a ka importuar firma juaj ndonjë produkt/ lëndë të parë? 0=Jo, 1=Po (0) (1) (17)

**Shënim për intervistuesin: Nëse përgjigja është 0=Jo, kalo tek pyetja 15**

14. Nëse PO, nga cilat vende? 0=Jo 1=Po

|                                 |        |               |
|---------------------------------|--------|---------------|
| Greqia                          | (0 (1) | (18)          |
| Italia                          | (0 (1) | (19)          |
| Europa Lindore                  | (0 (1) | (20)          |
| Vendet e Ish Bashkimit Sovjetik | (0 (1) | (21)          |
| Vendet e ish Jugosllavisë       | (0 (1) | (22)          |
| Tjetër (Specifikoni)_____       | (0 (1) | (23)<br>(23a) |

15. Gjatë dy viteve të fundit a ka eksportuar firma juaj ndonjë produkt? 0=Jo, 1=Po (0) (1) (24)

**Nëse përgjigja është 0=Jo kalo tek pyetja 17**

16. Nëse Po, në cilat vende? 0=Jo, 1=Po

|                                 |        |               |
|---------------------------------|--------|---------------|
| Greqia                          | (0 (1) | (25)          |
| Italia                          | (0 (1) | (26)          |
| Europa Lindore                  | (0 (1) | (27)          |
| Vendet e Ish Bashkimit Sovjetik | (0 (1) | (28)          |
| Vendet e ish Jugosllavisë       | (0 (1) | (29)          |
| Tjetër (Specifikoni)_____       | (0 (1) | (30)<br>(30a) |

**17. Afërsisht, ç'pjesë e xhiros vjetore realizohet në Shqipëri? \_\_\_\_\_**  
**(31)**

**18. Tregu juaj është kryesisht (zgjidhni vetëm një alternativë)**

1=Ndërkombëtar, 2= Kombëtar, 3= Rajonal, 4=Lokal (1) (2) (3) (4)  
**(32)**

**19. Për produktin apo shërbimin tuaj kryesor, ç'pjesë të tregut tuaj zotëroni?** 1= 0-5%  
2=6-25% 3=26-50%, 4=51-75%, 5=76-100% (1) (2) (3) (4) (5)  
**(33)**

## C. MJEDISI I BIZNESIT

### 20. Sa problematike janë çështjet e mëposhtme për biznesin tuaj.

*Ju lutemi përgjigjuni në shkallën nga 1 deri në 5, ku 1= nuk është pengesë dhe 5= është pengesë shumë e madhe*

|                                                       |                             |
|-------------------------------------------------------|-----------------------------|
| Aktet ligjore dhe nënligjore për fillimin e biznesit  | (1) (2) (3) (4) (5)<br>[34] |
| Aktet ligjore dhe nënligjore për investimet e huaja   | (1) (2) (3) (4) (5)<br>[35] |
| Paparashikueshmëria e ligjeve/vendimeve juridike      | (1) (2) (3) (4) (5)<br>[36] |
| Probleme me regjistrimin e pasurisë                   | (1) (2) (3) (4) (5)<br>[37] |
| Niveli i taksave                                      | (1) (2) (3) (4) (5)<br>[38] |
| Kosto e lartë e korrupsionit                          | (1) (2) (3) (4) (5)<br>[39] |
| Burokracia shtetërore                                 | (1) (2) (3) (4) (5)<br>[40] |
| Probleme me madhësinë e tregut të brendshëm           | (1) (2) (3) (4) (5)<br>[41] |
| Konkurenca e pandershme/ monopoli                     | (1) (2) (3) (4) (5)<br>[42] |
| Financimi                                             | (1) (2) (3) (4) (5)<br>[43] |
| Inflacioni                                            | (1) (2) (3) (4) (5)<br>[44] |
| Sigurimi i burimeve të kualifikuara të punës          | (1) (2) (3) (4) (5)<br>[45] |
| Niveli i zhvillimit të infrastrukturës fizike/teknike | (1) (2) (3) (4) (5)<br>[46] |
| Probleme lidhur me prodhimin/teknologjinë e biznesit  | (1) (2) (3) (4) (5)<br>[47] |
| Krimi dhe vjedhja                                     | (1) (2) (3) (4) (5)<br>[48] |
| Paqëndrueshmëria politike                             | (1) (2) (3) (4) (5)<br>[49] |

21. Deri në çfarë mase ndiheni i mbrojtur nga legjislacioni shqiptar? 1=shumë i mbrojtur, 2=deri diku i mbrojtur, 3=deri diku i pambrojtur, 4=krejtësisht i pambrojtur (1) (2) (3) (4)  
[50]

22. Në rast moszbatimi/thyerje kontratash, a keni frikë nga gjykimi i jo i drejtë i gjykatave?  
0=Jo 1=Po (0) (1)  
[51]

23. Gjatë 2 viteve të fundit, sa herë ka qenë firma juaj në gjykatë:

23a. Si akuzuese \_\_\_\_\_  
[52]

23b. Si e akuzuar \_\_\_\_\_  
[53]



24. Gjatë dy viteve të fundit a ka apeluar firma juaj për ndonjë vendim juridik në një gjykatë më të lartë? 0=Jo 1=Po (0) (1)  
[54]

**Shënim për intervistuesin: Nëse përgjigja është 0=JO, kalo tek pyetja 26**

25. Është kthyer vendimi përfundimtar në favorin tuaj?  
0=Jo, 1=Po (0) (1)  
(55)

**26. Si i vlerësoni vështirësitë e mëposhtme në marrëdhëniet me gjykatat:**

*Lutemi përgjigjuni në shkallën nga 0 deri në 5 ku: 0=nuk kam mendim, 1= aspak problem dhe 5= shumë problem*

|                                                       |                             |
|-------------------------------------------------------|-----------------------------|
| Shpenzimet gjyqësore zyrtare janë shumë të larta      | (1) (2) (3) (4) (5)<br>[56] |
| Shpenzimet gjyqësore jozyrtare janë shumë të larta    | (1) (2) (3) (4) (5)<br>[57] |
| Mungesa e një avokati të zot                          | (1) (2) (3) (4) (5)<br>[58] |
| Vendimet e gjyqtarëve janë të pandershme              | (1) (2) (3) (4) (5)<br>[59] |
| Gjykatësit nuk janë kompetentë                        | (1) (2) (3) (4) (5)<br>[60] |
| Proçeset gjyqësore zgjasin shumë                      | (1) (2) (3) (4) (5)<br>[61] |
| Përmbartimi nuk zbaton vendimet                       | (1) (2) (3) (4) (5)<br>[62] |
| Rreziku i nxjerrjes së sekretit tregtar është i lartë | (1) (2) (3) (4) (5)<br>[63] |

27. A ju është dashur që për ushtrimin e biznesit të paguani rryshfete zyrtarëve shtetërorë për probleme të tilla si:

27a. Sigurim sipërfaqe toke për zgjerimin/ ushtrimin e biznesit  
0=Jo 1=Po 2= E pa aplikueshme (0) (1) (2)  
[64]

27b. Zhdoganimet  
0=Jo 1=Po 2= E pa aplikueshme (0) (1) (2)  
[65]

27c. Blerje lëndë e parë nga ndërmarrje shtetërore, të nevojshme për biznesin tuaj  
0=Jo 1=Po 2= E pa aplikueshme (0) (1) (2)  
[66]

27d. Inspektime të ndryshme shtetërore (punësim /taksa/ higjiena/ ndërtim/ liçensa, etj)  
0=Jo 1=Po 2= E pa aplikueshme (0) (1) (2)  
[67]

28. Afërsisht, sa kohë shpenzon pronari kryesor (menaxheri) me administratën shtetërore për interpretimin apo sqarimin e vendimeve të ndryshme qeveritare \_\_\_\_ % e kohës së punës  
[68]

29. Nëse ju e llogarisni të gjithën, sa kohë shpenzohet nga Ju dhe punonjësit e administratës suaj për të trajtuar aktet juridike të qeverisë dhe për të punuar me punonjësit e administratës shtetërore? \_\_\_\_ % e kohës së punës  
[69]

## D. FINANCAT

### 30. Sa problematike janë elementët e mëposhtëm të sistemit bankar në Shqipëri?

*Ju lutem përgjigjuni në shkallën nga 1 deri në 5 ku 1=shumë i pazhvilluar, 5=shumë i zhvilluar*

|                                            |                             |
|--------------------------------------------|-----------------------------|
| Format e pagesave                          | (1) (2) (3) (4) (5)<br>[70] |
| Shpejtësia e kryerjes së veprimeve bankare | (1) (2) (3) (4) (5)<br>[71] |
| Niveli i komisioneve bankare               | (1) (2) (3) (4) (5)<br>[72] |
| Niveli i interesave bankare                | (1) (2) (3) (4) (5)<br>[73] |

### 31. Gjatë gjashtë muajve të fundit, sa përqind e të ardhurave të firmës suaj është realizuar:

31a. Përmes bankës \_\_\_\_\_% [74]

31b. Në të holla në dorë \_\_\_\_\_% [75]

31c. Në mallra \_\_\_\_\_% [76]

### 32. Që nga krijimi, a ka marrë firma juaj kredi? 0=Jo, 1=Po (0) (1) (77)

### Shënim për intervistuesin: Nëse përgjigja është 0=Jo, kalo tek pyetja 34

33. Nëse Po, cili ka qenë burimi i kredisë? 1= Banka, 2= Burime private, 3=Të dyja  
(1) (2) (3) [78]

34. Gjatë 12 muajve të kaluar, a ka ndërmarrë firma juaj ndonjë investim në tokë, ndërtesa, makineri etj? 0=Jo 1=Po (0) (1) [79]

35. A planifikon firma juaj të bëjë ndonjë investim në 12 muaj e ardhshëm në tokë, ndërtesa makineri etj? 0=Jo 1=Po, (0) (1) (80)

### 36. Si e vlerësoni nivelin e detyrimeve të mëposhtme financiare në Shqipëri:

*Lutemi përgjigjuni në shakllën nga 1 deri në pesë ku 1=shumë i ulët dhe 5=shumë i lartë*

|                                                             |                             |
|-------------------------------------------------------------|-----------------------------|
| TVSH                                                        | (1) (2) (3) (4) (5)<br>[81] |
| Tatimi mbi fitimin                                          | (1) (2) (3) (4) (5)<br>[82] |
| Tatimin mbi të ardhurat personale                           | (1) (2) (3) (4) (5)<br>[83] |
| Taksa mbi pasurinë e paluajtshme                            | (1) (2) (3) (4) (5)<br>[84] |
| Sigurimet Shoqërore dhe Shëndetësore (pjesa e punëdhënësit) | (1) (2) (3) (4) (5)<br>[85] |
| Tarifat doganore (nëse e aplikueshme)                       | (1) (2) (3) (4) (5)<br>[86] |

## E. PYETJE PËRMBLEDHËSE

37. Sa do të jetë rritja e kompanisë suaj vitin e ardhshëm? \_\_\_\_\_%  
[87]

38. Sa ishte vëllimi i qarkullimit në muajin e fundit? \_\_\_\_\_lekë (88)

39. Gjatë vitit të fundit, ç'pjesë e të ardhurave tuaja të përgjithshme në %, janë shpenzuar për:

|                                               |        |      |
|-----------------------------------------------|--------|------|
| 39a. Kosto e inputeve (përfshirë amortizimin) | _____% | [89] |
| 39b. Kosto e punës                            | _____% | [90] |
| 39c. Kosto e shërbimit të huasë/interetit     | _____% | [91] |
| 39d. Taksa                                    | _____% | [92] |
| 39e. Kosto të tjera (vendimet, korrupsioni)   | _____% | [93] |

40. Gjatë 2 viteve të fundit, a ka ndodhur që të mos realizonit një investim të cilin e kishit planifikuar? 0=Jo 1=Po (0) (1) (94)

**Shënim: Nëse përgjigja është 0=J0, kalo tek pyetja 42**

**41. Vendimi juaj për të mos bërë këtë investim ishte si pasojë e:**

|                                                         |         |                 |
|---------------------------------------------------------|---------|-----------------|
| Kërkesës së pamjaftueshme për produkt/shërbim që ofroni | (1) (2) | [95]            |
| Pasigurisë së pronës                                    | (1) (2) | [96]            |
| Taksave të larta                                        | (1) (2) | [97]            |
| Mossigurim i kredisë                                    | (1) (2) | [98]            |
| Kosto e lartë e zbatimit të vendimeve qeveritare        | (1) (2) | [99]            |
| Tjetër (specifikoni)_____                               | (1) (2) | [100]<br>[100a] |

42. Do të dëshironim të dinim opinionin tuaj për disa probleme të renditura më poshtë. (Ju lutem jepni mendimin tuaj në shkallë nga 0 deri në 10, ku "0" jeni plotësisht dakort me thënien në të majtë dhe "10" jeni plotësisht dakort me thënien në të djathtë.)

Ne Shqipëri nuk mund t'i besosh shumë njerezve  Ne Shqipëri mund t'i besosh shumë njerezve [101]

Njeriu duhet të jete shumë i kujdesshem për të bërë ndryshime në jete  Njeriu nuk mund të arrijë kurrë në jete nëse nuk vebron verberisht [102]

[103]

## F. TË DHËNA RRETH TË INTERVISTUARIT

43. Gjinia 1=Femër 2= Mashkull (0) (1) [104]

44. Mosha \_\_\_\_\_ [105]

45. Kombësia juaj? \_\_\_\_\_ [106]

46. Nënshtetësia juaj? \_\_\_\_\_ [107]

47. Cili është pozicioni juaj në firmë? (1) (2) (3) (4) [108] [108a]

1=Pronar, 2=Bashkëpronar, 3=Menaxher kryesor, 4=Tjetër (specifiko)\_\_\_\_\_

48. Niveli juaj i arsimimit? (1) (2) (3) (4) (5)

1= tetëvjeçar, 2=e mesme/kolegj, 3= e lartë (univesitet) e pambaruar, 4= Universitet/Institut, 5= Pasunivesitarë

[109]

49. A e zotëroni gjuhën vendase? 0=Jo 1= Po (0) (1)  
[110]

50. A dëshironi të bëni një koment të shkurtër  
a. rreth këtij studimi/pyetësoi

[111]

---

b. Pengesave që ndeshni në biznesin tuaj

[112]

---

**JU FALEMINDERIT PËR BASHKËPUNIMIN!**

## ANEKSI II

### Sondazhe standarte

Shkurtime

Përsërit pyetjen (PP)

*Përsërit strukturën e referencës*

Përsërit zgjedhjet

Çfardo që të thotë (ÇT)

Përkufizime që lejohen në pyetje pas pyetje (Përk)

Çfarë kuptoni? (ÇK)

Si e kuptoni?

A mund të më thoni diçka më tepër rreth mendimeve tuaja për këtë gjë? (TM)

A mund të më thoni mendimin tuaj?

Çfarë mendoni? (ÇM)

Çfarë prisni?

Cila është më afër mendimit tuaj (CA)

A ka arsye të tjera që ju e mendoni kështu? (AT)

A keni ndonjë gjë tjetër për të shtuar? (GJT)

### **Aneks III**

#### **Frazat pohuese nxitëse për sjellje nxitëse**

*Të shkurtra*

Ju kuptoj ...  
Ehë, ehë ...  
Faleminderit

*Të gjata*

Ky është një informacion i vlefshëm.  
Keni ide të mira për këtë problem.  
Faleminderit, opinioni juaj është i rëndësishëm për studimin.  
Ka mjaft rëndësi të kuptosh se çfarë mendojnë njerëzit për këtë gjë.  
Kjo gjë është mjaft e vlefshme për studimin tonë.

*Komente lidhur me detyrat e intervistuesit*

Të lutem më jep pak kohë që ta shkruaj këtë gjë.  
Unë duhet që t'i shkruaj të gjitha ato që ju thoni.  
Unë dua që të jem i sigurt që i kam shkruar mendimet tuaja siç duhet (Intervistuesi i lexon ato që ka shkruar).  
Ne e kemi diskutuar këtë gjë edhe më parë, por mua më duhet që tju pyes për çdo pyetje që është në pyetsor



## **Hyrje në Statistikë, Metodat Sasiore të Kërkimit dhe Pyetësorët**

### **DIZAJNET E KËRKIMIT**

- **KLASIFIKIMI:**
  - **EKSPLORUES**
    - Qëllimi kryesor është në zbulimin e ideve dhe detajeve të problemit.
  - **PËRSHKRUES (DESKRIPTIVE)**
  - **SHKAKËSOR (Causal)**
  - **KONKLUDUES**

### **DIZAJNI EKSPLORUES**

- QËLLIMET**
- Për të siguruar më shumë informacion në lidhje me problemin
  - Për të formuluar /definuar më saktë problemin
  - Për të identifikuar alternativa
  - Për të zhvilluar hipoteza
  - Për të sqaruar koncepte



## DIZAJNI EKSPLORUES

### KARAKTERISTIKAT

- Më fleksibël
- Më pak formal
  - të dhënat dytësore
  - Anketimin e eksperteve
  - Case studies (rastet studimore)

---

---

---

---

---

---

---

## DIZAJNI DESKRIPTIV

- Qëllimi: përshkrim, psh. Karakteristikat e tregut
- Përcaktohet nga formulimi paraparak i hipotezave
- Dizajn i paraplanifikuar dhe i strukturuar
- Kërkon të përcaktohet:

- Kush do ta japë informacionin
- Çfarë informacioni dhe si të sigurohet
- Kur duhet të sigurohet informacioni
- Ku do të sigurohet informacioni
- Pse nevojitet ky informacion

---

---

---

---

---

---

---

## KLASIFIKIMI I DIZAJNIT DESKRIPTIV

Dizajni ndërsektorial-**(Cross-sectional design)**:  
Mbledhje e informacionit nga një zgjedhje vetëm një herë

- single ( i vetëm) cross-sectional design: Informacioni mblidhet nga një zgjedhje një herë
- multiple ( shumëfishtë) cross-sectional design: Ka dy ose me shumë zgjedhje dhe informacioni mblidhet vetëm një herë

**Longitudinal** (ose panel): Një zgjedhje e caktuar e elementëve matet vazhdimisht ;  
zgjedhja mbetet e njejte gjatë gjithë kohës.

---

---

---

---

---

---

---

### **KËRKIMI SHKAKËSOR (CAUSAL)**

- Objektivi kryesor është të sigurohet evidencë për lidhjen Shkak- pasojë të variablaeve
  - Të kuptohet cilët variabla janë shkaku (variabla të pavarur) dhe cilët variabla janë efekti (variabla të varur) të fenomenit).
  - Të përcaktohet natyra e marrëdhënieve ndërmjet variablaeve shkakësore dhe të parashikohet efekti.
- Kërkon një strukturë të planifikuar të dizajnit
- Metoda kryesore është eksperimenti

### **DALLIMET NDËRMJET KËRKIMIT EKSPLORUES DHE KONKLUZIV**

#### • EKSPLORUES

Objektivi: Siguron detaje dhe arësytim Informacioni i nevojshëm është i gjerë

Karakteristika: Procesi i kërkimit është fleksibël dhe i pa strukturuar Zgjedhja është e vogël dhe jo përfaqësuese

Analiza e të dhënave primare është kualitative

Gjetjet: Tentativë

Rezultati: Zakonisht pasohet nga kërkime të tjera eksploruese ose konkludive

#### • KONKLUZIV

Objektivi: Teston hipoteza specifike dhe ekzaminon marrëdhënie Informacioni i nevojshëm përcaktohet qartë

Karakteristika: Procesi i kërkimit është formal dhe i strukturuar Zgjedhja është e madhe dhe përfaqësuese

Analiza e të dhënave është sasore

Gjetjet: Konkludues

Rezultati: Gjetjet përdoren si input për vendimmarrje

### **BURIMET E INFORMACIONIT**

- Të dhëna Dytësore
  - Të dhënat mblidhen për disa qëllime përveçse atij që po trajton
  - Gjenden lehtë dhe pak të kushtueshme
- Të dhëna Primare
  - Të dhëna që mblidhen nga kërkuesi me qëllim specifik dhe që adresojnë problemin
  - Të kushtueshme dhe kërkojnë kohë

### **KRITERI I EVALUIMIT TË TË DHËNAVE** **DYTËSORE**

- Metodologjia për mbledhjen e të dhënave
- Saktësia e të dhënave
- Sa aktuale janë të dhënat?
- Qëllimi për të cilin janë mbledhur të dhënat
- Natyra e përmbajtjes së të dhënave
- Sa të varuara janë të dhënat?

### **TË DHENAT PRIMARE**

- Kërkim Kuantitativ
  - Një metodologji kërkimi që kërkon të kuantifikojë të dhënat, duke aplikuar analiza statistikore
- Kërkimi Kualitativ
  - Një metodologji kërkimi e pastrukturuar bazuar në zgjedhje të vogla që sigurojnë detajet dhe të kuptuarit e problemit që trajtohet.

### **KËRKIMI KUALITATIV VS.** **KUANTITATIVE**

#### **Kërkimi Kualitativ**

**Objektivi:** Të sigurohet të kuptuarit cilësor të arsyeve dhe motiveve të problemit

**Zgjedhja:** Numër i vogël i rasteve jo përfaqësuese

**Mbledhja e të dhënave:** E pastrukturuar

**Anbaliza e të dhënave:** Jostatistikore

**Rezultati:** Zhvillimi i të kuptuarit

#### **Kërkimi Kuantitativ**

Të kuantifikohen të dhënat dhe të përgjithësohen rezultatet nga zgjedhja në interes të popullimit

Numër i madh i rasteve përfaqësuese

E Strukturuar

Statistikore

Rekomandohet një kurs final i veprimit

## **KLASIFIKIMI I TEKNIKAVE KUALITATIVE** **TË KËRKIMIT**

- Grup Fokus
- Intervista
- Teknika Projektive

## **KLASIFIKIMI I TEKNIKAVE KUANTITATIVE TË KËRKIMIT**

### • Anketimi

1. *Intervistë Personale* : (të besueshme, të lehta, ka normë të lartë përgjigjeje, kanë kosto të lartë)
2. *Interviste telefonike*: (e besueshme, e lehtë, pak e kushtueshme, hidhen direkt të dhënat në kompjuter nga intervistuesi...)
3. *Intervistë me postë*: (normë e ulët përgjigjeje, por bëhen përpjekje për ta rritur këtë normë...)
4. *Intervistë me email*: (s'ka kosto, ka normë të ulët përgjigjeje, por mund të korrigjohet...)

### • Vrojtimi

### • Eksperimenti

## **PYETËSORI**

Një teknikë e strukturuar për mbledhjen e të dhënave që konsiston në një seri pyetjesh të shkruara ose të thëna, të cilat kërkojnë përgjigje.

### • Procesi i dizajnit të pyetësorit:

1. Specifikohet informacioni i kërkuar
2. Specifikohet tipi i pyetësorit dhe metodat e administrimit
3. Përcaktohet forma e përgjigjes për çdo pyetje
4. Përcaktohet përmbajtja e pyetjeve individuale (në mënyrë të tillë që të kapërcehet paafësi dhe jo vullneti i atyre që përgjigjen)
5. Përcaktohet struktura e pyetjes dhe formëzohet me fjalë ajo
6. Përcaktohet sekuenca e pyetjeve
7. Përcaktohet forma dhe paraqitja fizike e pyetësorit
8. Para testohet pyetësori

### Dizenjimi i pyetesoreve

- Shumë pyetësorë dështojnë se ose bëjnë pyetje të gabuara ose pyesin nëpërmjet një pyetje të saktë por të ndërtuar në mënyrë të gabuar.

Çfarë duhet të kihet parasysh kur ndërtohet një pyetësor:

1. Duhet të ketë pyetje që lidhen me njëra tjetrën dhe të ketë sekuençë logjike
2. Të jetë sa më i shkurtër i mundshëm.

3. Pyetjet duhet të jenë të shkurtër, të thjeshta, pa keqkuptime, të lehta për tu kuptuar etj.
4. Përdorni shprehje sa më neutrale psh jo: A e pëlqeni këtë revistë? Por Si e vlerëson këtë revistë nga 1 deri te 5?
5. Formuloni me kujdes pyetjet personale: psh më mirë: A keni dalë në pension? Sesa A jeni një pensionist në moshë?
6. Mos jep paralajmërim që në fillim të pyetjes psh Jo: Ne e kuptojmë që ti s'do ti përgjigjesh kësaj pyetjeje por per kë do të votoni?
7. Shmangni pyetjet hipotetike psh jo: Sa do të shpenzoje për produktin nqse...
8. Shmangi pyetjet e hapura psh: "Keni ndonjë tjetër koment"
9. Gjithmonë kryej një "survejim pilot" para se të fillosh vrojtimin...
10. Shmangni pyetjet që trajtojnë dy çështje brenda tyre Psh jo: Sa të kënaqur jeni me pagesën dhe kushtet e punës?
11. Shmangni pyetje që përfshijnë negative brenda tyre. Mund të ndodhë që i pyetri të humbasë fjalën 'nuk' në pyetje Psh jo: A jeni dakort me idenë që studentët nuk duhet të marrin borxhe për të financuar arësimin e lartë? (Gabim)

12. Shmangni termat teknike Psh jo. A mendoni se influenza TUC e menaxherit ndikon në efektivitetin tuaj? (gabim- çfarë është TUC???)

13. Mos bëni pyetje që i stërmundojnë të intervistuarit në memorizimin e gjërave, shpesh memorja e tyre mund edhe ti tradhetojë ato...

Edhe pyetësi më i mirë nuk do të marrë përgjigje nga çdo i intervistuar sepse:

1. Nuk i kuptojnë pyetjet
2. Kanë arsye personale (janë sëmurë, nuk gjenden aty për momentin, janë të zënë me punë)
3. Nuk dinë të përgjigjen
4. Thjesht refuzojnë të përgjigjen.

### VËZHIMI (OBSERVIM)

• Regjistrohen tipet e sjelljes së njerëzve, objekteve dhe ngjarjeve në mënyrë sistematike për të siguruar informacionin në lidhje me fenomenin e interesit

• Observuesi nuk pyet apo komunikon me njerëzit që observohen.

• Informacioni mund të regjistrohet gjatë kohës që ndodh ngjarja ose për ngjarje të ndodhura më parë

• Kush observohet, Çfarë aspekti Observohet, Kur observohet, Ku observohet?

### EKSPERIMENTI

• Eksperimenti përdoret për studimin e marrëdhënieve shkak- pasojë

• Kuptimi shkencor i shkakësisë (I fjalisë “X shkakton Y”):

– X është një numër i mundshëm që shkakton Y

– Ndodhja e X bën ndodhjen e Y më të mundshme (x është një shkak probabilitar i Y)

– Nuk mund të provojmë që X definitivisht është shkak i Y. Më së shumti mund të dëshmojmë që X është shkak i Y.

### PËRKUFIZIMI I KONCEPTEVE

**Variabël i pavarur:** Variabla që mund të manipulohet, efektet e të cilëve maten dhe krahasohen.

**Njësitë që testohen:** individë, organizata, ose të tjera entitete, përgjigjet e të cilëve si variabla të pavarur ekzaminohen

**Variabla të varur:** janë variabla që masin efektin e variablave të pavarur në njësinë e testimit.

**Variabla të jashtëm:** janë variabla të ndryshme nga ato të pavarurit, që ndikojnë përgjigjen e njësive të testimit (keto variabla mund të ngatërrojnë variablin e varur i cili matur në këtë ndikim mund të dobësojë ose të zhvlerësojë vlefshmërinë e rezultateve të eksperimentit)

**Experiment:** kemi të bëjmë me eksperiment kur kërkuesi manipulon një ose më shumë variabla dhe mat efektin e tyre ndërsa kontrollojnë efektin e variablave të jashtëm.

**Dizajni eksperimental:** është seti i procedurave që specifikojnë njësitë e testimit dhe si këto njësi ndahen në nënzgjedhje homogjene, çfarë variablash të pavarur duhet të manipulohen, çfarë variablash të varur duhet të maten, dhe si mund të kontrollohen variablat e jashtëm.

## **PËRKUFIZIMI I KONCEPTEVE**

**Variabël i pavarur:** Variabla që mund të manipulohet , efektet e të cilëve maten dhe krahasohen.

**Njësitë që testohen:** individë, organizata, ose të tjera entitete, përgjigjet e të cilëve si variabla të pavarur ekzaminohen

**Variabla të varur:** janë variabla që masin efektin e variablave të pavarur në njësinë e testimit.

**Variabla të jashtëm:** janë variabla të ndryshme nga ato të pavarurit, që ndokojnë përgjigjen e njësive të testimit (keto variabla mund të ngatërrojnë variablin e varur i cili matur në këtë ndikim mund të dobësojë ose të zhvlerësojë vlefshmërinë e rezultateve të eksperimentit)

**Experiment:** kemi të bëjmë me eksperiment kur kërkuesi manipulon një ose më shumë variabla dhe mat efektin e tyre ndërsa kontrollojnë efektin e variablave të jashtëm.

**Dizajni eksperimental:** është seti i procedurave që specifikojnë njësitë e testimit dhe si këto njësi ndahen në nënsgjedhje homogjene, çfarë variablash të pavarur duhet të manipulohet, çfarë variablash të varur duhet të maten, dhe si mund të kontrollohen variablat e jashtëm .

## **VLEFSHMËRIA DHE BESUESHMËRIA**

**VLEFSHMËRIA:** Një matje është e vlefshme për atë kohë sa variabiliteti në rezultate ndërmjet objekteve reflekton ndryshimin e vërtetë të karakteristikave që maten.

E shohim saktësinë e matjes duke parë tre tipe validitetesh:

1. **Vlefshmëri parashikuese** (ne vlerësojmë vlefshmërinë e matjes si parashikues i disa karakteristikave të individit psh. Provimi GMAT) Përcaktohet nga korelacioni ndërmjet matjes dhe kriterit (kur korelacioni është i madh thuhet se ka vlefshmëri parashikuese, psh. Notat në MBA)
2. **Vlefshmëria e përmbajtjes**, fokusohet në përshtatshmërinë me të cilën karakteristika kapet nga matja. (vështirë kjo sepse është çështje gjykimi), por duhet përcaktuar mirë çështje variabli. (psh “kënaqësia në punë” të ketë vlefshmëri në përmbajtje duhet të përfshijë çdo aspekt)

3. **Vlefshmëria e konstruktit** lidhet me pyetjen: “ a mat në realitet instrumenti atë çfarë ne synojmë të matim? ”

Çfarë mënyre formulimi, konceptesh ose trajtimesh përkufizojnë rezultatin e arritur në test?

Një matje me vlefshmëri të fortë në konstrukt duhet të jetë e përshtatshme, por një matje e përshtatshme jo gjithmonë ka vlefshmëri të konstruktit. Psh një matje vërtet mund të përdoret për të matur diçka, por kujdes mos matësh diçka të ndryshme nga ajo që ti shpresoje të masje.



**BESUESHMËRIA:** Ngjashmëria e rezultateve të siguruara nga matje të pavarura por të krahasueshme të të njejtit objekt quhet Besueshmëri.

Besueshmëria është indeks përrputshmërie( korelacioni ndërmjet dy matjeve të të njejtit koncept kur bëjmë të njejtën matje ) Vlefshmëria është ( korelacioni ndërmjet dy matjeve të të njejtit koncept edhe kur metodat që përdoren janë të ndryshme)

Besueshmëria është kusht i nevojshëmepor jo i mjaftueshëm për vlefshmërinë. Besueshmëria është më e lehtë për t'u matur se vlefshmëria, prandaj dhe kërkuesit e theksojnë vlerën e besueshmërisë.

### **ZGJEDHJA**

Qëllimi i kërkuesit është të arrijë në konkluzione për popullimin.

**Zgjedhja** e cila duhet të jetë sa më përfaqësuese e popullimit, për të cilin kërkuesi synon përgjithësimin e rezultateve.

(“Validiteti i jashtëm” lidhet me faktin se sa gjetjet e studimit mund të përgjithësohen për elementët përtej nivelit të studimit. )

**Permasa e zgjedhjes:** është e lidhur me përmasën e efektit dhe me fuqinë e gjykimit statistikor. Zgjedhjet e mëdha kanë fuqi të mjaftueshme statistikore për të testuar në mënyrë sasiore marrëdhëniet ndërmjet variablave; ndërkohë që zgjedhjet e vogla nuk e kanë këtë fuqi.

1. **Zgjedhjet probabilitare** (Zgjedhje e rastit e thjeshtë, sistematike, me strata, ne seri)
2. **Zgjedhjet Jo-probabilitare** (Metodat e përshtatshme, Zgjedhja me kuota, Zgjedhja e qëllimshme (gjykuese), Zgjedhja Top-bore)

### **BURIMET E MUNDSHME TË GABIMIT**

#### **GABIMI TOTAL**

Ndryshimi midis vlerës reale të mesatares variablit të interesit në popullim dhe vlerës mesatare të vërtetuar në projektin e kërkimit.

### GABIMI TOTAL

#### GABIMI I RASTIT I ZGJEDHJES

Gabimi që ndodh kur zgjedhja e realizuar është një përfaqësim jo perfekt i popullimit në lidhje me çështjen që trajtohet.

#### GABIMI JO I ZGJEDHJES

Gabime që i atribuohen burimeve të ndryshme nga zgjedhja. Mund të jenë të rastit ose jo të rastit.

### GABIMI TOTAL

#### GABIMI I RASTIT I ZGJEDHJES

#### GABIMI JO I ZGJEDHJES

#### GABIMI I PËRGJIGJEVE

Gabime që vijnë nga të intervistuarit që përgjigjen, por japin përgjigje të pasakta, dhe përgjigjet e tyre janë keqkoduara ose keq analizuar.

#### GABIMI JO I PËRGJIGJEVE

Disa nga elementët e zgjedhjes nuk përgjigjen.

### MBLEDHJA DHE PËRPUNIMI I TË DHËNAVE:

1. Qëllimi i të dhënave
2. Përshkrimi i të dhënave që duhen për realizimin e qëllimit
3. Konsultohen të dhënat sekondare sa të vlefshme janë ato.
4. Përcaktohet popullimi dhe kuadri i zgjedhjes për të dhënat primare
5. Zgjidhet metoda e zgjedhjes dhe përmasat e saj
6. Identifikohet zgjedhja e përshtatshme
7. Dizenjohet pyetësi ose metodat e tjera të mbledhjes së të dhënave
8. Realizohet studimi pilot nga ku dalin problemet
9. Trajtohen intervistuesit, vrojtuesit ose eksperimentuesit
10. Mblidhen të dhënat
11. Realizohet follow-up, rikontaktimi me ata që nuk përgjigjen
12. Përgatiten të dhënat dhe hidhen në soft
13. Përpunohen, analizohen dhe interpretohen rezultatet.

## UNIVARIATE ANALYSIS- ANALIZAT E THJESHTA

### LLOJET E STATISTIKAVE

- **Statistika pershkruese - deskriptive:** përdor metodat grafike dhe numerike për të përmbledhur dhe paraqitur informacionin që prmban një seri të dhënash. Qëllimi nuk është përgjithësimi.
- **Statistika inferenciale:** përdor të dhënat e zgjedhjes për të bërë vlerësime, parashikime, ose përgjithësime të tjera për një seri më të madhe të dhënash (popullimi), që ndihmojnë në vendimmarrje. Për shkak të karakterit dëshmue, ekziston gabimi, ndaj flasim për interval besimi, një matje e sigurisë së dëshmisë.
- **Statistika e marrëdhënieve:** përdor të dhënat e zgjedhjes për të matur shkallën e marrëdhënies ndërmjet dy apo më shumë variabla. Edhe këtu ka karakter dëshmue, ekziston gabimi, ndaj flasim për interval besimi.

## MULTIVARIATE ANALYSIS- ANALIZAT SHUMËPËRMASORE

Teknikat multivariabël lejojnë të analizohen dy apo më shumë variabla në të njëjtën kohë .

Analiza regressive  
Analiza faktoriale  
MANOVA etj

## SHKALLËT E MATJES SË VARIABLIT

**Popullimi:** Tërësia e elementëve që i intereson studiuesit

**Zgjedhja:** Një pjesë e elementëve të popullimit.

**Variabli:** Karakteristikë e përbashkët e elementeve (sasiorë e cilësorë)

**Vrojtimi:** Tërësia e të dhënave për një element.

**Shkalla e matjes:** Shkalla e informacionit që mbart variabli, dhe lidhet me analizën e mëtejshme të tij

1. **Shkalla e matjes nominale:** Shrehet në numër, simbole, gërma etj që tregojnë grupin në të cilin bën pjesë elementi
2. **Shkalla e matjes e zakonshme (ordinale) :** ka vetitë e shkallës nominale por ka kuptim renditja e elementëve
3. **Shkalla e matjes intervalore:** Shprehet në numra që mbartin sasi, ka kuptim veprimi i mbledhjes dhe zbritjes por jo i shumëzimit dhe pjesimit
4. **Shkallë e matjes ne forme raporti:** Ka vetitë e shkallës së matjes intervalore, por ka kuptim edhe veprimi i shumëzimit dhe i pjesimit.

---

---

---

---

---

---

---

### **TEMA 3:**

## **Njohja me SPSS dhe përgatitja e të dhënave për analizë**

### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje që :

- Të kuptojnë ç'është dhe përse shërben programi kompjuterik SPSS
  - Të dinë si të hedhin të dhënat në SPSS
  - Të dinë si të kodojnë të dhënat dhe të kuptojnë çdo fushë në SPSS
- Të dinë si të pastrojnë të dhënat dhe ti bëjnë gati për analizë

■

### **LITERATURA**

- “ Applied Statistics”, Warner A.R , 2008
- Statistical for Business and Economics”, Newbold P, Carlson W, Thorne B, 2011
- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011
- “Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014

## PYETËSOR I PUNONJËSVE TË NJË INSTITUCIONI

Q1: Mosha në vite: \_\_\_\_\_

Q2: Gjendja civile : I pamartuar (0) \_\_\_\_\_ I martuar(1) \_\_\_\_\_

Q3: Sa vite keni që banoni në adresën aktuale: \_\_\_\_\_

Q4: Të ardhurat mujore të familjes në ( mijë Dollarë): \_\_\_\_\_

Q5: Shpenzimet mujore për karbutant: (në Euro): \_\_\_\_\_

Q6: Niveli I arësimit:

- a) Tetëvjeçare (1)
- b) Gjimnaz(2)
- c) Kolegj(3)
- d) Universitet (4)
- e) Pasuniversitar (5)

Q7: Sa vite keni në punën aktuale: \_\_\_\_\_

Q8: Sa I kënaqur jeni në punë:

- a) Shumë I pakënaqur (1)\_\_\_\_\_
- b) Disi I pakënaqur (2)
- c) Neutral (3) \_\_\_\_\_
- d) Disi I kënaqur (4) \_\_\_\_\_
- e) Shumë I kënaqur (5) \_\_\_\_\_

Q9: Gjinia : Femër (f)\_\_\_\_\_ Mashkull (m) \_\_\_\_\_

Q10: Sa anëtarë jeni në familjen tuaj: \_\_\_\_\_

Q11: A keni shërbim Wireles në telefonin tuaj: Jo (0) \_\_\_\_\_ Po(1) \_\_\_\_\_

Q12: A keni të aktivizuar në telefon : Voice mail: : Jo (0) \_\_\_\_\_ Po(1) \_\_\_\_\_

Q13: A keni të aktivizuar në telefon : Internetin:

Jo (0) \_\_\_\_\_ Po(1) \_\_\_\_\_ Nuk e di (8) \_\_\_\_\_ Pa përgjigje (9)\_\_\_\_\_

Q14: A keni të aktivizuar në telefon Caller ID : Jo (0) \_\_\_\_\_ Po(1) \_\_\_\_\_

Q15: A keni të aktivizuar në telefon : Call waiting: : Jo (0) \_\_\_\_\_ Po(1) \_\_\_\_\_

Q16: A zotëroni TV: : Jo (0) \_\_\_\_\_ Po(1) \_\_\_\_\_

Q17: A zotëroni/ DVD player : Jo (0) \_\_\_\_ Po(1) \_\_\_\_

Q18: A zotëroni Kompjuter: : Jo (0) \_\_\_\_ Po(1) \_\_\_\_

Q19: A zotëroni fax : Jo (0) \_\_\_\_ Po(1) \_\_\_\_

Q20: A jeni abonura në ndonjë gazetë: Jo (0) \_\_\_\_ Po(1) \_\_\_\_

**TEMA 3:**

**Njohja me SPSS dhe përgatitja e të dhënave për analizë**

**Ushtrime**

**Ushtrim 1:** Mbështetur në pyetësin që keni ndërtuar nga tema 2, kodoni të dhënat, hidhini në SPSS data base-in dhe bëni pastrimin e tyre, duke i bërë gati për analizën e të dhënave .



## Statistika deskriptive

### ORGANIZIMI I TE DHENAVE

Të gjitha institucionet apo organizatat që ushtrojnë veprimtari të ndryshme, kanë respektivisht zyrat e informacionit dhe të statistikes, nga ku grumbullohen një pafundësi të dhënash, të cilat në formën e tyre origjinale janë të pakuptueshme për shumë prej nesh. Statistika deskriptive përbën teknikën, e cila na ndihmon në kondensimin e të dhënave të shumta përmes përdorimit të tabelave, grafikëve dhe përmbledhjes së të dhënave.

Kështu, statistika përshkruese (deskriptive) është një ndër përbërësit më të rëndësishëm, që jep metoda efektive për grumbullimin dhe analizën e të dhënave.

Në shumicën e rasteve, të dhënat duhet të përmbledhen në mënyrë që të jenë të dobishme.

Metoda më e zakonshme e përmbledhjes së të dhënave është ajo me anë të statistikës përshkruese dhe grafikëve. Edhe në qoftë se jeni duke planifikuar për të analizuar të dhënat tuaja duke përdorur një teknikë statistikore të tillë si: t-test, analizën e variancës, ose regresionin logjistik, ju duhet të filloni duke shqyrtuar paraprakisht të dhënat tuaja. Ky hap paraprak ju ndihmon për të përcaktuar se cilat teknika të analizës statistikore duhet të përdoren për t'ju përgjigjur pyetjeve tuaja kërkimore.

Në fakt, ky proces i shqyrtimit të dhënave tuaja, shpesh zbulon informacione që mund t'ju habisin, teksa ju informojnë për shpërndarjen e të dhënave. Ju mund të zbuloni vlera jashtëzakonisht të larta ose të ulëta në të dhënat tuaja. Ndoshta këto "outliers" janë shkaktuar nga të dhënat e koduara gabimisht, ose ato mund të zbulojnë informacione të rëndësishme në lidhje me të dhënat tuaja (ose subjekte) që ju nuk i keni parashikuar. Ju mund të vëreni se të dhënat tuaja nuk janë shpërndara normalisht.

Po ashtu, ju mund të vëreni se një histogram me të dhënat tuaja, tregon dy kulme të dallueshëm, duke ju bërë të kuptoni se të dhënat tuaja tregojnë një dallim në mes të grupeve.

Gjetje të tilla, shpesh janë rezultat i përdorimit të duhur të teknikave përshkruese.

Pjesët në vijim të këtij kapitulli diskutojnë për taktikat më të përdorura, me qëllim që të kuptohen drejt të dhënat, për të raportuar informacion apo për të përgatitur të dhënat për një analizë të mëtejshme.

Ky kapitull, do të na ndihmojë të mësojmë se si do t'i organizojmë dhe paraqesim të dhënat duke përdorur tabelat dhe grafikët. Ne do të mësojmë si të përgatitim tabelat e shpërndarjes së frekuencave për të dhëna cilësore dhe sasiore; si të ndërtojmë një bar-diagram, diagram me sipërfaqe (pie chart), histogramë dhe poligonin e frekuencave për këto të dhëna dhe si të përgatitim shpërndarjen sipas trungut dhe gjethes.

### 1.1 Të dhënat në rradhë

Kur mbledhen të dhënat, informacioni që mbledhet për çdo element të popullatës ose të mostrës, shënohet në një sekuencë dhe konsiderohet i vlefshëm. Kjo sekuencë e të dhënave është e rastësishme dhe e parangëzuar. Këto të dhëna, para se të grupohen ose të rangëzohen quhen të dhëna në rradhë.

**Përkufizim :**

*Të dhënat e mbledhura, para se të grupohen ose të rangëzohen quhen të dhëna në rradhë.*

Supozojmë se kemi mbledhur të dhëna mbi peshën e 50 nxënësve të një shkolle të fillore, të cilët i kemi zgjedhur në mënyrë të rastësishme dhe peshat janë regjistruar sipas rregullit se cili student është peshuar i pari, cili i dyti, e kështu me rradhë. Të dhënat paraqiten në tabelën 6. Të dhënat në tabelën 1 janë të dhëna sasiore.

Tabela 1 Pesha e 50 nxënësve

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 21 | 19 | 24 | 25 | 29 | 34 | 26 | 27 | 37 | 33 |
| 18 | 20 | 19 | 22 | 19 | 19 | 25 | 22 | 25 | 23 |
| 25 | 19 | 31 | 19 | 23 | 18 | 23 | 19 | 23 | 26 |
| 22 | 28 | 21 | 20 | 22 | 22 | 21 | 20 | 19 | 21 |
| 25 | 23 | 18 | 37 | 27 | 23 | 21 | 25 | 21 | 24 |

Po këta 50 nxënës, u pyetën se me cilin nga prindërit jetonin (me të dy prindërit –D, vetëm me nënën –N apo vetëm me babain-B) dhe u regjistruan përgjigjet e tyre, të cilat u paraqitën në tabelën 7. Të dhënat në tabelën 2 janë të dhëna cilësore (ose kategorike).

Tabela 2 Statusi i 50 nxënësve

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| N | B | D | D | N | N | D | N | N | N |
| D | D | N | B | B | B | D | D | D | N |
| N | B | D | D | D | D | N | B | D | D |
| D | D | N | D | D | N | N | D | D | D |
| D | D | B | D | N | D | D | N | D | D |

Të dhënat e paraqitura në tabelën 7 quhen të dhëna të pa grupuara dhe përmbajnë informacion për secilin element të popullatës respektivisht.

## 1.2 ORGANIZIMI DHE PARAQITJA E TE DHENAVE CILESORE

### Shpërndarja e frekuencave.

Për të sqaruar sa më mirë se ku synon shpërndarja e frekuencave, i referohemi një shembulli të thjeshtë, i cili paraqitet si më poshtë.

Shembull: 100 studentë të Fakultetit të Mjekësisë u pyetën se çfarë do të bënin (tipi i punësimit) kur të mbaronin shkollën, nga ku u morën të dhënat e mëposhtme, si në tabelën 3.

Tabela 3 Tipi i punësimit

|                                    |                                                |
|------------------------------------|------------------------------------------------|
| <b>Variabli</b> → Tipi i punësimit | Nr. i studentëve → <b>Kolona e frekuencave</b> |
|------------------------------------|------------------------------------------------|

|                                      |                       |
|--------------------------------------|-----------------------|
| Kompani private                      | 44                    |
| <b>Kategori</b> → Qeverisje qendrore | 16 → <b>Frekuenca</b> |
| Qeverisje lokale                     | 23                    |
| Biznes te vetin                      | 17                    |
| Shuma                                | 100                   |

**Përkufizim:**

*Shpërndarja e frekuencave për të dhëna cilësore, **rendit** të gjitha kategoritë dhe numrat e elementëve që i përkasin secilës kategori.*

Shembull:

Tridhjetë punonjës të një kompanie u zgjodhën në mënyrë të rastësishme dhe u pyetën për shkallën e stresit të pranishëm në punë. Përgjigjet e marra u grupuan sipas tri kategorive; shumë stres, kur kishin shkallë të madhe të stresit në punë; diçka, kur kishin një shkallë të konsiderueshme të stresit në punë dhe mungon, kur nuk kishin stres. Informacioni i marrë, paraqitet në tabelën 4:

Tabela 4. Paraqitja e stresit ne punë

|       |        |        |       |        |        |
|-------|--------|--------|-------|--------|--------|
| diçka | mungon | diçka  | shumë | shumë  | mungon |
| shumë | diçka  | diçka  | shumë | diçka  | diçka  |
| shumë | diçka  | mungon | shumë | mungon | diçka  |
| diçka | shumë  | diçka  | diçka | shumë  | mungon |
| diçka | shumë  | shumë  | diçka | mungon | diçka. |

Në rast se duam të ndërtojmë tabelën e shpërndarjes së frekuencave për këto të dhëna, veprojmë si më poshtë:

Theksojmë se variabli në shembullin tonë është stresi me shkallët e tij. Ky variabël klasifikohet në tri kategori: *shumë*, *diçka* dhe *mungon*, të cilat ne i grupojmë në një kollonë (shih tab. 5). Në kollonën e dytë vendosim frekuencat, të cilat i shënojmë me (fr). Shuma e të gjitha frekuencave për secilën kategori duhet të japë numrin total të individëve të mostrës sonë (në rastin tonë 30).

Tabela 5. Shpërndarja e frekuencave për stresin në punë

| Stresi në punë | Frekuenca (fr) |
|----------------|----------------|
| Shumë          | 10             |
| Diçka          | 14             |
| Mungon         | 6              |
| Shuma          | 30             |

### 1.3 SHPERNDARJA E FREKUENCAVE RELATIVE DHE E PERQINDJEVE

Frekuenca relative për një kategori, merret duke pjesëtuar frekuencën për këtë kategori me numrin total të individëve të mostrës. Kështu, frekuenca relative tregon se cili fraksion apo proporcion i frekuencave totale i përket secilës kategori. Shpërndarja e frekuencave relative përcakton frekuencën relative për çdo kategori.

$$\text{Frekuenca relative për një kategori (FR)} = \frac{\text{Frekuencën për këtë kategori}}{\text{Shuma e të gjithë frekuencave}}$$

Përqindja e frekuencave llogaritet duke shumëzuar frekuencën relative të çdo kategorie me 100. Shpërndarja e përqindjeve shënon përqindjet për të gjitha kategoritë

$$\text{Përqindja} = (\text{frekuenca relative}) \times 100$$

Duke kryer veprimet përmes aplikimit të dy formulave më lart, plotësojmë si në tabelën 6.

Tabela 6. Shpërndarja e frekuencave për stresin në punë

| Stresi në punë | Frekuenca (FR) | Përqindja          |
|----------------|----------------|--------------------|
| Shumë          | 10/30 = 0.333  | 0.333 x 100 = 33.3 |
| Diçka          | 14/30 = 0.467  | 0.467 x 100 = 46.7 |
| Mungon         | 6/30 = 0.200   | 0.200 x 100 = 20.0 |
| Shuma          | 1.000          | 100.0              |

Për shkallën e stresit “shumë” kemi dhjetë raste, të cilat pjesëtuar me totalin 30 (që janë gjithsej raste të përfshira në studim), jep frekuencën relative (FR) = 0,0333. Kështu veprojmë dhe për kategoritë “diçka” dhe “mungon”. Shuma e frekuencave relative është, sigurisht, e barabartë me një.

Për të llogaritur përqindjen, shumëzojmë frekuencën relative të çdo kategorie me 100%. P.sh.: për kategorinë “shumë”, shumëzojmë 0,333 me 100% dhe rezultati është 33,3%. Kështu veprojmë dhe për kategoritë “diçka” dhe “mungon”. Shuma e përqindjeve është e barabartë me 100%.

#### Ushtrime

1. Të dhënat në tabelën 7, paraqesin rezultatet e një testi diagnostik, ku me A shënohen rezultatet pozitiv, B rezultatet negativ dhe C rezultati i dyshimtë për këtë lloj testi.

Tabela 7 Rezultatet e testit diagnostik

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| A | B | B | A | C | B | C | C | C | A |
| C | B | C | A | C | C | B | C | C | A |
| B | C | B | B | A | C | A | A | C | A |

- a) Ndërtoni tabelën e shpërndarjes së frekuencave.
  - b) Llogarisni frekuencën relative dhe përqindjen për cdo kategori.
  - c) Llogarisni frekuencën kumulative.
  - d) C’përqindje e elementëve të kësaj mostre i përkasin kategorisë C?
  - e) C’përqindje e elementëve të kësaj mostre i përkasin kategorisë A ose B?
  - f) Ndërtoni grafikun për këtë shpërndarje frekuencash.
2. Të dhënat në tabelën 8 paraqesin rezultatet e një pyetsori të realizuar në disa fëmijë, të cilët u pyetën se me kë jetonin. Me N shënohen rastet që jetojnë vetëm me nënën, B rastet që jetojnë vetëm me babain dhe D rastet që jetojnë me të dy prindërit së bashku.

Tabela 8 Rezultatet e pyetsorit

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| N | N | D | D | B | N | N | D | D |
| B | N | N | D | D | D | D | D | D |
| B | N | B | B | N | N | N | N | N |

- a) Ndërtoni tabelën e shpërndarjes së frekuencave.
  - b) Llogarisni frekuencën relative dhe përqindjen për cdo kategori.
  - c) Llogarisni frekuencën kumulative.
  - d) C’përqindje e elementëve të kësaj mostre jetojnë me të dy prindërit?
  - e) C’përqindje e elementëve të kësaj mostre vetëm me nënën ose vetëm me babain?
  - f) Ndërtoni grafikun për këtë shpërndarje frekuencash.
3. Në një spital rajonal u zgjodhën rastësisht 50 punonjës të cilët u pyetën për profesionin e tyre. Me M u shënuan mjekët, I infermieret dhe me N personeli ndihmës. Të dhënat janë pasqyruar në tabelën 9.

Tabela 9 Llojet e profesionit

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| D | N | N | I | I | I | N | I | D | I |
| I | I | I | I | N | I | I | N | N | I |
| N | I | I | N | D | N | I | I | I | I |
| I | I | N | N | I | I | N | D | D | D |
| I | D | D | D | I | I | I | I | I | I |

- a) Ndërtoni tabelën e shpërndarjes së frekuencave.
- b) Llogarisni frekuencën relative dhe përqindjen për cdo kategori.
- c) Llogarisni frekuencën kumulative.
- d) C'përqindje e elementëve të kësaj mostre janë mjekë?
- e) Ndërtoni grafikun me sipërfaqe për shpërndarjen e përqindjeve.

## PARAQITJA GRAFIKE E TË DHËNAVE

Paraqitjet statistikore realizohen përmes grafikëve dhe tabelave. Ato e bëjnë më të lexueshëm fenomenin dhe janë më të thjeshtë për komunikimin dhe transmetimin e mesazhit.

**Përkufizim. Grafiku është paraqitja e të dhënave me anë të pikave, vijave, figurave gjeometrike, hartave të gatshme apo dhe simboleve.**

Çdo grafik ka të domosdoshëm tre elementë:

- a) titullin;
- b) shkallën dhe;
- c) legjendën.

Grafikët klasifikohen sipas qëllimit të ndërtimit dhe sipas mënyrës së ndërtimit.

Sipas **qëllimit të ndërtimit** ata klasifikohen në grafikë që shprehin:

- ✓ Dinamikën e dukurisë;
- ✓ Strukturën e dukurisë;
- ✓ Sezonalitetin e dukurisë;
- ✓ Shtrirjen gjeografike;
- ✓ Grafikë që bëjnë të mundur krahasimin mes dukurive të ndryshme, ose pjesëve përbërëse të të njëjtës dukuri.

Sipas **mënyrës së ndërtimit**, grafikët klasifikohen në grafikë të tipit:

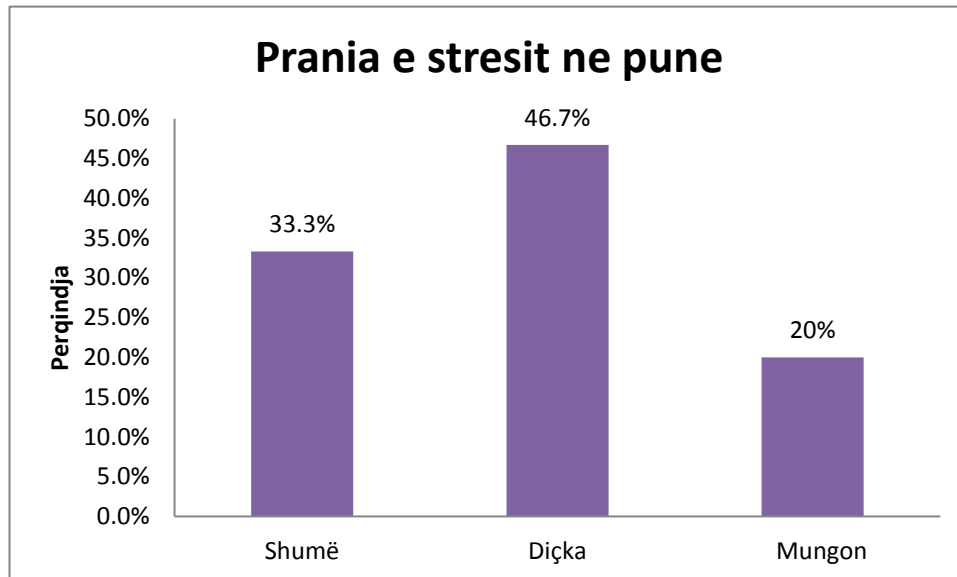
- ✓ Diagramë vijor;
- ✓ Histogramë;
- ✓ Bar-diagramë;
- ✓ Grafik me sipërfaqe;
- ✓ Diagramë radial, polar, spider-diagramë;
- ✓ Hartogramë;
- ✓ Hartëdiagramë;
- ✓ Poligoni i frekuencave;
- ✓ Diagramë me pika;
- ✓ Box-plott.

Është e rëndësishme të theksohet që jo cdo lloj grafiku, mund të përdoret cdo lloj të dhëne.

### 2.1 PARAQITJA GRAFIKE E TË DHËNAVE CILËSORE

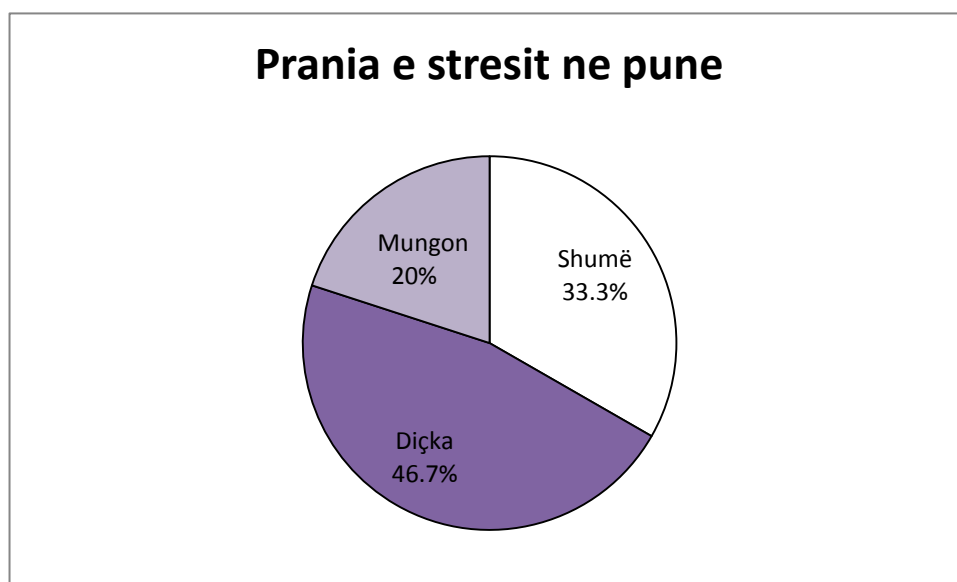
**Bar-diagrami.** Shërben për të paraqitur dinamikën e dukurisë. Bazë e ndërtimit të këtij grafiku, janë kolonat me bazë të njëjtë dhe lartësi të ndryshme. Për të ndërtuar një bar diagram, kategoritë e ndryshme i vendosim në boshtin horizontal dhe frekuencat në boshtin vertikal. Pasi

kemi bërë këtë, ndërtojmë grafikun me kollona me bazë të njëjtë dhe lartësi të ndryshme, të cilat tregojnë dinamikën e fenomenit. (Graf.1)



Grafiku1. Stresi në punë.

**2.1.2 Diagrami me sipërfaqe** (pie-chart) shërben për të paraqitur strukturën e fenomenit dhe zakonisht përdoret për të shprehur përqindjet, frekuencat ose frekuencat relative. Bazë e ndërtimit të këtij grafiku, është rrethi me sipërfaqe qarku  $360^\circ$ , në të cilin do të paraqitet 100% e dukurisë. Kështu që, 1% të dukurisë, i takojnë  $3,6^\circ$  në rreth. Duke shumëzuar përqindjet me  $3,6^\circ$ , marrim sipërfaqen e qarkut që i korrespondon secilës kategori. Rrethi i plotë, prezanton mostrën ose popullatën në tërësi, kurse pjesë të tij (pjesët me ngjyra të ndryshme), përfaqësojnë proporcionet që paraqet secila kategori. (Graf. 2)



Grafiku 2. Stresi ne pune



**2.1.3 Hartogrami.** Bazë e ndërtimit të këtij grafiku janë hartat e gatshme politike të vendeve në studim. Janë grafikë që tregojnë dinamikën e dukurisë. Për këtë përdoren ngjyra të ndryshme ose intensitete të ndryshme të të njëjtës ngjyre (të cilat sqarohen në legjendë). Zakonisht, në rastet kur përdoren intensitete të ndryshme të të njëjtës ngjyre, ngjyra me intensitet më të fortë i korrespondon rasteve me shpeshtësi më të madhe të hasjes së dukurisë, të cilat detyrimisht duhet të sqarohen në legjendë.



Grafiku 3. Hartogrami

## 2.1.4 Diagrami radial/polar/spider

Përdoret për të treguar shtrirjen e dukurisë ose sezonalitetin e dukurisë.

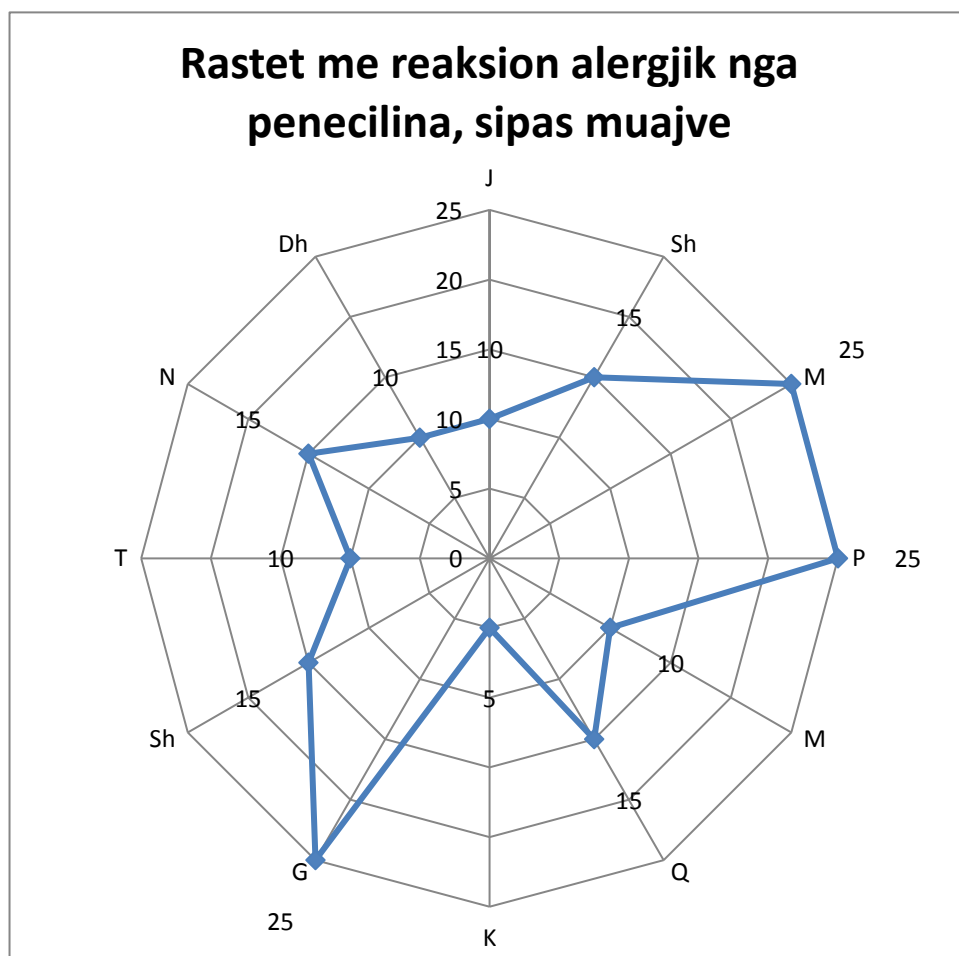
Bazë e ndërtimi të këtij grafiku është rrethi me rreze “r” të barabartë me numrin mesatar të rasteve (ne ratin tone  $180/12 = 15$ ) ose me një numër të plotë-pjestueshëm me të. Ndahet rrethi në aq pjesë të barabarta sa kategori të variablilit kemi (në rastin tonë do të ndahet në dymbëdhjetë pjesë), ku secila kategori (muajt e vitit) do të prezantohet me një bosht, që e merr origjinën nga qendra e rrethit. Në secilin nga këto boshte vendosen vlerat përkatëse sipas muajve, të cilat prezantohen me një pikë në bosht. Bashkimi i këtyre pikave jep grafikun e kërkuar. (Ky grafik, quhet spider, sepse imiton rrjetën e merimangës).

Shembull. Në tabelën 10, jepet numri i rasteve me reaksion alergjik nga penicilina, të paraqitur në një qendër shëndetësore, në një vit të caktuar. Kërkohet të bëhet paraqitja grafike.

Tabela10. Rastet me reaksion alergjik nga penicilina

| Muaj            | J  | Sh | M  | P  | M  | Q  | K | G  | Sh | T  | N  | Dh | Σ   |
|-----------------|----|----|----|----|----|----|---|----|----|----|----|----|-----|
| Numri i rasteve | 10 | 15 | 25 | 25 | 10 | 15 | 5 | 25 | 15 | 10 | 15 | 10 | 180 |

Në këtë rast bëjmë paraqitjen grafike me anë të spider diagramë, që paraqitet në grafikun 4.



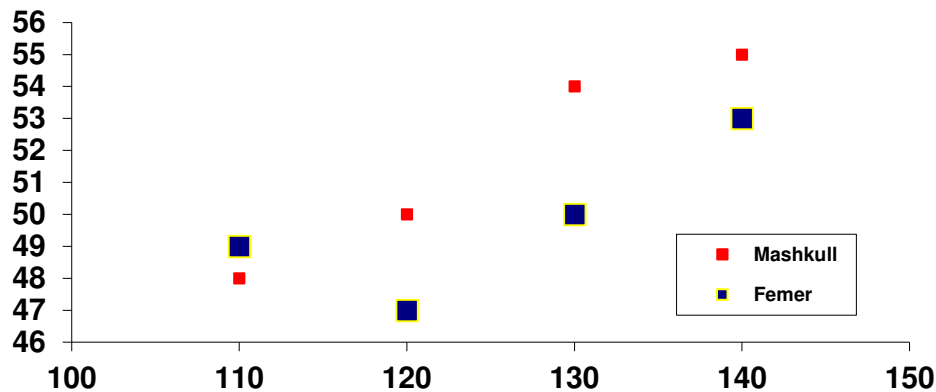
Grafiku 4. Diagrami radial

## 3.2 ORGANIZIMI DHE PARAQITJA GRAFIKE E TE DHENAVE SASIORE

Në këtë kapitull do të flasim për paraqitjen grafike të të dhënave sasiore.

**3.2.1 Diagrami me pika** (scatter-diagrame). Është tipik për paraqitjen e lidhjes mes dy dukurive. Diagrami me pika paraqitet në formën e një bashkësie pikash. Për këtë grafik do të flitet tek kapitulli mbi korrelacionin, por paraprakisht mund të themi në lidhjen mes dy dukurive (korrelacioni), paraqitja grafike është një bashkësi pikash në formën e topit të rrekbisë, i cili orientohet në varësi të vlerave që merr koeficienti i korrelacionit.

**Graf.5 Lidhja mes peshes dhe gjatesise**



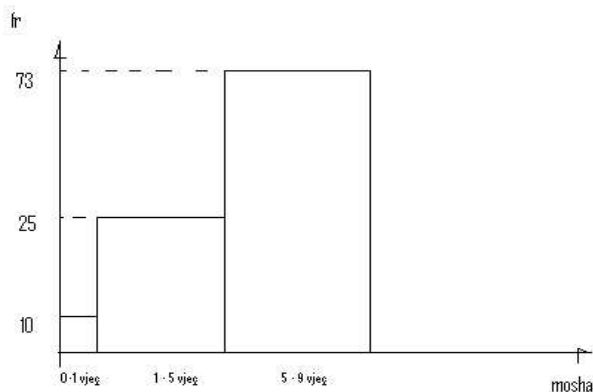
**3.2.2 Histogrami.** Ky lloj grafiku përdoret për paraqitjen e të dhënave sasiore të vazhdueshme. Bazë e ndërtimit të këtij grafiku janë kallonat me bazë dhe lartësi të ndryshme, që shërbejnë për të paraqitur dinamikën e dukurisë.

Supozojmë se në tabelën më poshtë paraqitet mosha në vite dhe frekuenca.

Tab.. Mosha në vite

| Mosha<br>(vite) | Numri i<br>rasteve |
|-----------------|--------------------|
| 0-1 vjeç        | 10                 |
| 1-5 vjeç        | 25                 |
| 5-9 vjeç        | 73                 |

Paraqitja grafike e të cilave realizohet me anë të histogramit, që paraqitet në grafikun 6. Siç shihet në këtë grafik kolonat janë me bazë të ndryshme, pasi gjerësia e klasës që ato paraqesin është e ndryshme (në kolonën e parë përfshihet vetëm një vit; në kolonën e dytë përfshihen 4 vite (mosha 1, 2, 3 dhe 4 vjeç dhe në kolonën e tretë përfshihen po katër vite (moshat 5, 6, 7 dhe 8 vjeç). Lartësia e ndryshme e kolonave paraqet shpeshësinë e hasjes, pra dinamikën e dukurisë, që respektivisht është dhjetë për klasën 0-1 vjeç, njëzet e pesë për klasën 1-5 vjeç dhe shtatëdhjetë e tre për klasën 5-9 vjeç. (Shih grafikun më poshtë).



Graf. 6. Histogrami, mosha në vite

### i. Shpërndarja e frekuencave.

Në tabelën e mëposhtme (tabela 18), paraqiten të ardhurat mujore të 100 punonjësve që punojnë në QSUT “Nënë TEREZA”. Në kolonën e parë paraqitet variabli sasior (kuantitativ), sasia e parave të marra nga punonjësit, e quajtur klasë. Për variabelët sasiore, intervali i cili përfshin të gjithë vlerat midis dy numrave limite, nga më i vogli tek më i madhi, quhet klasë. Kujdes, çdo klasë paraqet një variabël. Siç e shihni nga tabela më poshtë, klasat nuk mbivendosen tek njëra-tjetra (klasa e parë përfshin vlerat 120-140, klasa e dytë vlerat 141-170, e kështu me rradhë). Kolona e dytë në tabelë, paraqet numrin e punonjësve për secilën klasë, pra frekuencën për çdo kategori page.

Tab. Të ardhurat mujore të 100 punonjësve që punojnë në QSUT “Nënë TEREZA”

| Variabli →  | Paga mujore | Numri i punonjësve (Frekuenca) |
|-------------|-------------|--------------------------------|
|             | 121-140     | 31                             |
|             | 141-170     | 39                             |
| klasa III → | 171-200     | 15 ← { frek. per kl. trete     |
|             | 201-300     | 9                              |
|             | 301-500     | 6                              |
|             |             | 100                            |

Vlera më e vogël e klasës së tretë

Vlera më e madhe e klasës së tretë

Për të dhënat sasiore, frekuenca për secilën klasë paraqet numrin vlerave të përfshira në këtë klasë. Cdo klase ka vlerën e saj më të vogël dhe më të madhe. Vlerat 121, 141, 171, 201 dhe 301 përfaqësojnë vlerat më të vogla (limitin e poshtëm) dhe vlerat 140, 170, 200, 300, dhe 500 përfaqësojnë vlerat më të mëdha (limitin e sipërm).

Të dhënat e paraqitura në tabelën 13 janë një ilustrim i tabelës së shpërndarjes së frekuencave për të dhëna sasiore. Të dhënat e paraqitura në tabelën e shpërndarjes së frekuencave, quhen të dhëna të grupuara.

## PERKUFIZIM.

Shpërndarja e frekuencave për të dhëna sasiore, paraqet të gjitha klasat dhe numrin e vlerave për secilën klasë. Të dhënat e paraqitura në formën e shpërndarjes së frekuencave quhen të dhëna të grupuara.

Për të gjetur pikën e mesit (midpoint) të limitit të sipërm për klasën e parë dhe atij të poshtëm, për klasën e dytë, ne i mbledhim ato dhe i pjesëtojmë me dy, p.sh., për të gjetur pikën e mesit mes klasës së parë dhe të dytë kemi:

$$m = \frac{140 + 141}{2} = 140,5$$

Numri 140.5 quhet kufiri i sipërm për klasën e parë dhe kufiri i poshtëm për klasën e dytë. Duke përdorur këtë teknikë, mund të transformojmë çdo klasë limitesh në klasë kufijsh, që nuk është gjë tjetër veçse një klasë e vërtetë limitesh. Diferenca mes dy kufijve për secilën klasë, jep **gjerësinë e klasës** që shënohet me  $\Delta i$ . **Gjerësia e klasës** quhet gjithashtu madhësia e klasës.

$$Gjerësia\ e\ klases = \Delta i = kufiri\ i\ sipërm - kufiri\ i\ poshtëm\ i\ klases$$

$$\Delta i = gjerësia\ per\ klasen\ e\ pare = 140,5 - 120,5 = 20$$

Kështu, tabela merr këtë formë (shih tabelën më poshtë):

Tabela

| Klasat Limit | Klasa kufij                         | Gjerësia e klases | Klasa midpoint |
|--------------|-------------------------------------|-------------------|----------------|
| 121-140      | 120.5 deri tek me te vogel se 140.5 | 20                | 130.5          |
| 141-170      | 140.5 deri tek me te vogla se 170.5 | 30                | 155.5          |
| 171-200      | 170.5 deri tek me te vogla se 200.5 | 30                | 185.5          |
| 201-300      | 200.5 deri tek me te vogla se 300.5 | 100               | 250.5          |
| 301-500      | 300.5 deri tek me te vogla se 500.5 | 200               | 400.5          |

Klasa midpoint merret duke mbledhur dy vlerat e limiteve ose dy vlerat kufi të secilës klase dhe duke i pjesëtuar për dy. Pra, për klasën e parë kemi:

$$Klasa\ midpoint\ (per\ klasen\ e\ pare) = \frac{121 + 140}{2} = 130,5\ ose\ \frac{120,5 + 140,5}{2} = 130,5$$

Siç shihni në tabelën 19, në kolonën e dytë, tek klasat kufi, është shkruar shprehja “**deri tek më të vogla se**” për t’u siguruar se çdo vlerë i përket një klase dhe vetëm një, sepse ju e shihni, kufiri i sipërm i njëjës klase dhe kufiri i poshtëm i klasës pasardhëse, janë të njëjtë.

Shpërndarja e frekuencave për të dhëna sasiore paraqet të gjitha klasat dhe numrin e vlerave për secilën klasë. Të dhënat e paraqitura në formën e shpërndarjes së frekuencave quhen **të dhëna të grupuara**.

Pikën mesatare (midpoint) për klasën e dytë e llogarisim si shuma e limitit të sipërm për klasën e dytë dhe atij të poshtëm për klasën e dytë, pjesëtuar me dy:

$$m_2 = \frac{170 + 141}{2} = 155,5$$

Dhe këtu, gjërësia e klasës gjendet si:

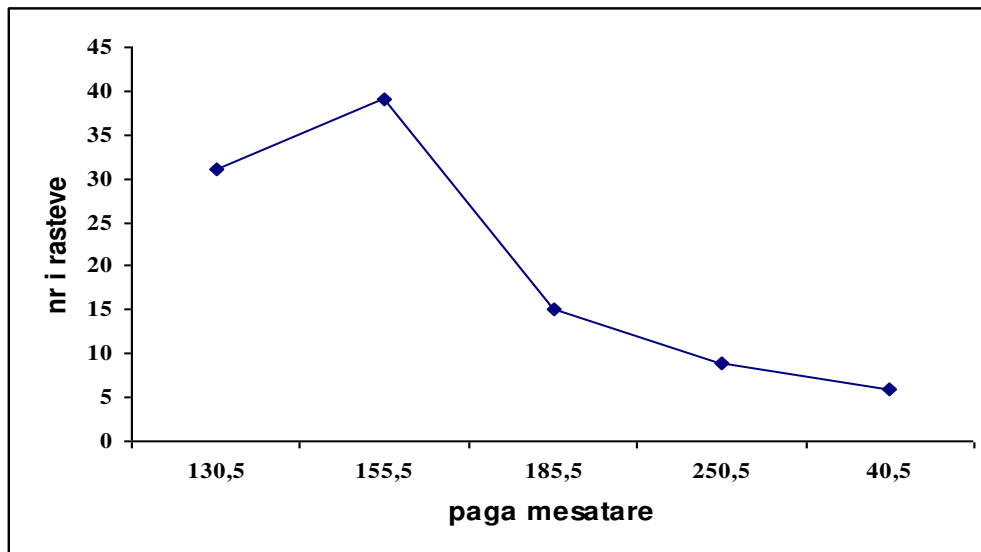
$$\text{Gjerësia e klases} = \Delta i = \text{kufiri i sipërm} - \text{kufiri i poshtëm i klases}$$

$$\Delta 2 = \text{gjerësia për klasën e dytë} = 170,5 - 140,5 = 30$$

$$\text{Klasa midpoint (për klasën e dytë)} = \frac{141 + 170}{2} = 155,5 \text{ ose } \frac{140,5 + 170,5}{2} = 155,5$$

Kështu me rradhë, në të njëjtën mënyrë veprojmë për të gjitha klasat, duke përcaktuar limitin e sipërm dhe të poshtëm të klasave dhe më pas duke llogaritur gjerësinë e klasës dhe klasën midpoint. Pasi kemi llogaritur të gjitha këto, bëjmë paraqitjen grafike të të dhënave në tabelë, duke vendosur në grafik pikat midpoint për secilën klasë përkatësisht.

Bashkimi i të gjitha pikave të mesit (midpoint), jep atë që quhet poligoni i frekuencave, i cili paraqitet në grafikun 7.



Grafiku 7. Poligoni i frekuencave

### 3.3 MASAT E TENDENCËS/PRERJES QËNDRORE

Shpërndarjet e frekuencës së të dhënave vazhdojnë të përkufizohen në dy tipe masash ose parametrash:

- masat e tendencës qendrore;
- masat e dispersionit.

Masat e tendencës qendrore vendosin një shkallë observimesh në hapësirë dhe janë dicka e ngjashme si “adresa” me variablin fqinjë. Masat e dispersionit sugjerojnë/tregojnë se sa shumë janë të përhapura observimet dhe mund të mendohen si rrethimi i një shtepie, që në rastin tonë është variabli.

### 3.3.1 MASAT E PRERJES QENDRORE

Në një ekzaminim të dhënash hapi i parë që bëhet është gjetja e tendencës qendrore së të dhënave. Pjesa më e madhe e të dhënave mjekësore tenton të grupohet në mënyrë të tillë, që densiteti i vlerave të vezhguara të jetë më i madh në qendër të shpërndarjes.

Hapi tjetër është përcaktimi i modës, medianës dhe mesatareve.

### 3.3.2 MODA

Në një seri të dhënash, modë quhet varianti me frekuencë më të lartë. Moda, megjithëse nuk ka ndonjë vlerë të madhe nga ana statistikore, përsëri paraqet interes . Në një varg vlerash mund të kemi një modë (fig.4). Në raste të tjera shpërndarja mund të jetë bimodale, sepse popullata mund të përbëhet nga dy nëngrupe, ku secili ka një shpërndarje të ndryshme, që paraqet modë në pika të ndryshme. Gjithashtu, mund të prodhohen artificialisht më shumë se dy moda, përmes asaj që quhet preferenca e shifrës, e cila ndodh atëherë kur vëzhguesit tentojnë të favorizojnë disa numra kundrejt të tjerëve. P.sh. personat që masin tensionin arterial tentojnë të favorizojnë numrat që përfundojnë me 0, psh; nëse tensioni arterial është 127/88 mmHg, personi që mat tensionin arterial e deklaron atë si 130/90 mmHg .

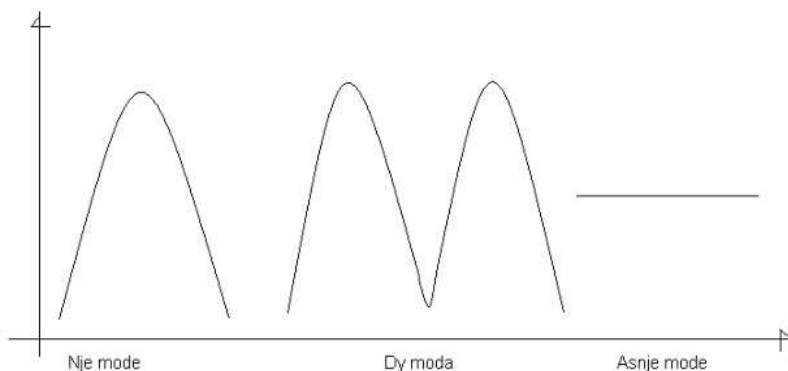


Fig 4. Moda dhe variacionet e saj

Gjithashtu mund të kemi rastin kur të gjithë variabellet hasen me të njëjtën frekuencë, atëherë themi se nuk kemi asnjë modë.

### 3.3.3.a MEDIANA

Mediana i korrespondon variablit që ze pozicionin e mesit në një varg vlerash, kur vlerat janë të renditura nga vlera më e madhe tek më e vogla, ose ansjelltas. Mediana llogaritet me formulën :

$$Me = \frac{n+1}{2}$$

ku n është numri i variableve në studim.

Kur numri i observacioneve është tek, mediana është një numër që ekziston realisht brenda vargut të vlerave; kur numri i observimeve është çift, mediana, ndoshta, është një vlerë që nuk ekziston objektivisht brenda vargut. Ajo gjendet si mesatare e dy vlerave qëndrore në varg.

Mediana ka disa karakteristika:

Mediana është e vetme në një varg vlerash.

Mediana nuk ndikohet nga vlerat skajore (ekstreme) të vargut të vlerave, për arsyen e thjeshtë se mediana lidhet me pozicionin e variablit në varg dhe jo me vlerën e tij.

**Ushtrim:** Është dhënë vargu i vlerave si me poshtë:

0, 16, 3, 9, 24, 45, 33.

Gjeni medianën.

Zgjidhje:

Fillimisht i rendisim vlerat nga më e vogla tek më e madhja.

0, 3, 9, **16**, 24, 33, 45.

Më pas bëjmë zëvendësimet, duke ditur që mediana gjendet me formulën:

$$Me = \frac{n+1}{2}$$

Shohim se kemi 7 elementë në varg, pra  $n = 7$

$$Me = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

Vlera  $Me = 4$ , tregon pozicionin në të cilin ndodhet mediana (duke qenë se është një mesatare pozicioni). Vlera e medianës është **16**. (Sic e shihni në këtë rast, mediana është një vlerë që ekziston realisht brenda vargut).

### II. Ushtrim:

Është dhënë vargu i vlerave si më poshtë:

0, 16, 3, 9, 24, 11, 45, 33.

Gjeni medianën.

Zgjidhje:



Fillimisht i rendisim vlerat nga më e vogla tek më e madhja.

0, 3, 9, **11, 16**, 24, 33, 45.  $n = 8$

Më pas bëjmë zëvendësimet në formulën e medianës:

$$Me = \frac{n+1}{2}$$

Pra mediana është në pozicionin

$$Me = \frac{n+1}{2} = \frac{8+1}{2} = 4,5$$

$$Me = (n+1) / 2 = (8 + 1) / 2 = 4,5$$

Mediana ndodhet mes pozicionit katër dhe pesë në varg. Kështu, vlera e medianës llogaritet si shuma e vlerës në pozicionin 4 (që është 11) me vlerën e pozicionit 5 (që është 16) dhe pjesëtohen me dy, si më poshtë:

$$Vlera e Me = \frac{11+16}{2} = \frac{27}{2} = 13,5$$

Në këtë rast, sic shihet, mediana është një vlerë që nuk ekziston realisht brenda vargut.

Vlera e medianes është quajtur si vlera e percentilit të 50° në varg, sepse 50% e observimeve bien nën këtë vlerë dhe 50% e vlerave bien mbi vlerën e medianës.

### 3.3.3.b Mediana për të dhëna të grupuara

Për të dhëna të grupuara vlera e medianës jepet me formulën:

$$Me = x_i + \Delta_i \frac{0,5 - F_{(cR)}}{F_R}$$

ku  $x_i$  është limiti inferior i klasës që përmban medianën;

$\Delta_i$  është gjerësia e klasës që përmban medianën;

$F_{(cR)}$  është frekuenca kumulative relative e klasës mediane

$F_R$  është frekuenca relative e klasës mediane.

Le ta sqarojmë këtë me një shembull:

Supozojmë se është matur niveli i glicemisë në një grup të sëmurësh si në tabelën më poshtë:

Tab. Niveli i glicemisë

| X<br>(glicemia mg/dl) | Fr | F <sub>R</sub> | F <sub>CR</sub> | $\Delta_i$  |
|-----------------------|----|----------------|-----------------|-------------|
| 124-145               | 10 | 10/42          | 10/42           | 145-124=21  |
| 146-167               | 15 | 15/42          | 25/42           | 167-146=21  |
| 168-189               | 15 | 15/42          | 40/42           | 189-168= 21 |
| 190-211               | 2  | 2/42           | 42/42           | 211-190=21  |

Për të gjetur medianën, arsyetojmë kështu: kemi gjithsej dyzetë e dy vlera, pra  $n=42$ , mediana zë pozicionin qendror në një varg vlerash, kur ato janë të renditura nga më e vogla tek më e madhja, dhe gjendet me formulën :

$$Me = \frac{n+1}{2} = \frac{42+1}{2} = 21,5$$

Kështu, mediana ndodhet mes pozicioneve 21-22 në varg.  
Si pasojë klase mediane është ajo **146-167**.

Vlera e medianës është:

$$Me = x_i + \Delta_i \frac{0,5 - F_{(cR)}}{F_R}$$

$$Me = 146 + 21 \frac{\left| 0,5 - \frac{25}{42} \right|}{\frac{15}{42}} = 146 + 21 \frac{|0,5 - 0,59|}{0,35} = 146 + 5,4 = 151,4 \text{ mg / dl}$$

Vlera e medianës është 151,4 mg/dl.

### 3.4. Mesataret Analitike

Në mesataret analitike bëjnë pjesë mesatarja aritmetike, mesatarja gjeometrike dhe mesatarja harmonike.

#### 3.4.1.a Mesatarja aritmetike

Mesatarja aritmetike është një nga treguesit më të rëndësishëm të të dhënave. Ajo konsiderohet si “zonja e madhe” në statistikë.

Mesatarja aritmetike është raporti mes shumës së variabileve dhe numrin total të variabileve (obervimeve) të marra (te kryera) në një studim.

$$\bar{x} = \frac{\sum x}{n}$$

ku:

$\bar{x}$  variablet;

$n$  numri i të gjithë variableve ( $\Sigma fr$ ).

Si masë e prerjes qendrore, mesatarja ka avantazhe si teorike ashtu dhe praktike. Meqënëse ka dhe vecori matematikore, kjo e favorizon të marrë pjesë aktive në statistikën e avancuar.

Vecoritë e mesatares:

- ♦ mesatarja e nje mostre është një vleresim i shkëlqyer i mesatares së popullatës nga e cila ajo rrjedh (mostra)

- ♦ Mesatarja është një vlerësim matematik që ndikohet nga vlerat skajeve të vargut të vlerave.
- ♦ Shuma e devijacioneve kuadratike nga mesatarja është fikse për një grup të caktuar observimesh.
- ♦ Mesatarja aritmetike është e vetme në një varg vlerash.

**Shembull:**

Supozojmë se numrat 2, 3, 3.5, 4, dhe 5.5 japin peshat në kilogramë të pesë bebeve. Për të gjetur peshën mesatare të tyre, i mbledhim të pesë vlerat dhe i pjesëtojmë me pesë.

$$\bar{x} = \frac{\sum x}{n} = \frac{(2+3+3,5+4+5,5)}{5} = 3,6kg$$

Në këtë rast themi se pesha mesatare e këtyre bebeve është 3,6 kg.

Kur kemi të dhëna (variable) të cilat përsëriten disa herë, mesatarja aritmetike, gjendet me formulën:

$$\bar{x} = \frac{\sum (xf)}{n}$$

dhe interpretohet si raporti i shumës së prodhimit të variableve me frekuencën përkatëse, përmbi numrin total të variableve në varg.

**Shembull:**

Supozojmë se është matur niveli i glicemisë në një mostër me 16 individë, si në poshtë:  
80, 80, 85, 85, 85, 85, 90, 90, 90, 95, 95, 100, 100, 100, 110, 110.

Gjeni vlerën mesatare të glicemisë në këtë grup individësh.

**Zgjidhje.**

Fillimisht i vendosim të dhënat në tabelë duke vendosur dhe shpeshtësinë e hasjes së variableve përkatësiht si në tabelën më poshtë. Më pas, bëjmë prodhimin xf (vlera e variablit shumëzuar me frekuencën përkatëse), të cilat pasqyrohen në kollonën e tretë të tabelës 24. Shumën e të gjithë prodhimeve (1480) e pjesëtojmë me numrin total të individëve (që është 16) dhe gjejmë mesataren aritmetike për këtë grup individësh.

Tab. Niveli i glicemisë

| Glicemia (mg/dl) (x) | fr          | xf            |
|----------------------|-------------|---------------|
| 80                   | 2           | 160           |
| 85                   | 4           | 340           |
| 90                   | 3           | 270           |
| 95                   | 2           | 190           |
| 100                  | 3           | 300           |
| 110                  | 2           | 220           |
|                      | $\Sigma=16$ | $\Sigma=1480$ |

$$\bar{x} = \frac{\sum(xf)}{n} = \frac{1480}{16} = 92,5 \text{ mg / dl}$$

Vlera mesatare e glicemisë në këtë grup individësh është 92,5 mg/dl

### 3.4.1.b Mesatarja aritmetike për të dhëna të grupuara

Kur të dhënat janë të grupuara, mesatarja aritmetike jepet me formulën:

$$\bar{x} = \frac{\sum(m_i f_i)}{n}$$

Ku:

$m_i$  është vlera e mesit për klasën në fjalë;  
 $f_i$  shpeshtësia e hajsës (frekuenca);  
 $n$  numri i rasteve

Po t'i kthehemi dhe një herë shembullit në tabelën 25, do të gjejmë fillimisht vlerat  $m$  për secilë klasë (kolona e tretë në tabelën më poshtë). Më pas do t'i shumëzojmë me frekuencën përkatëse ( $mf$ ). Prodhimet e gjetura do t'i mbledhim dhe shumën e gjetur (17542,5) do ta pjesëtojmë me numrin total të individëve në mostër, që në rastin tonë është 100.

Tab. Paga mujore e punonjësve

| Paga mujore | Numri I punonjesve (Frekuenca) | m     | mf                 |
|-------------|--------------------------------|-------|--------------------|
| 121-140     | 31                             | 130,5 | 4045,5             |
| 141-170     | 39                             | 155,5 | 6064,5             |
| 171-200     | 15                             | 185,5 | 2775,0             |
| 201-300     | 9                              | 250,5 | 2254,5             |
| 301-500     | 6                              | 400,5 | 2403,0             |
|             | 100                            |       | $\Sigma = 17542,5$ |

Nisur nga të dhënat e tabelës 25, bëjmë raportin e shumës së prodhimit  $mf$  me numrin “ $n$ ” të rasteve.

$$\bar{x} = \frac{\sum(m_i f_i)}{n} = \frac{17542,5}{100} = 175,425 \text{ lek}$$

Mesatarja aritmetike pagës mujore, në grupin tonë të individëve, është 175,425 lek.

### 3.4.5. MESATARJA harmonike.

Mesatarja harmonike shpreh raportin ndërmjet numrit total të variableve dhe shumës së inversit të variableve të veçanta të vargut.

$$H = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \frac{n_i}{x_i}}$$

Kjo lloj mesatare përdoret në rastet kur kemi vërtetime që zgjatin në kohë.

### 3.4.6 MESATARJA gjeometrike

Mesatarja gjeometrike është vlera e rrënjës me tregues "n" të prodhimit të një numri "n" madhësish.

Mesatarja gjeometrike përdoret kryesisht për të gjetur koeficientin mesatar, ose ritmin mesatar të zhvillimit të një sëmundjeje. Kjo është më pak e ndjeshme se mesatarja aritmetike ndaj ndryshimeve të madhësisë së serisë dhe niveli i saj është nën atë të mesatares aritmetike.

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}} = \sqrt[n]{\prod_{i=1}^k x_i^{n_i}}$$

### Ushtrime

1. Në tabelën më poshtë jepen numri i rasteve dhe incidenca (për 100.000 banorë) e rubeolës në vite.

Tabela Incidenca e rubeolës në vite.

| Viti | Numri i rasteve | Incidenca /100.000 |
|------|-----------------|--------------------|
| 1984 | 21              | 0.7                |
| 1985 | 78594           | 2653.2             |
| 1986 | 7               | 0.2                |
| 1987 | 10              | 0.3                |
| 1988 | 5               | 0.15               |
| 1989 | 8               | 0.25               |
| 1990 | 15              | 0.46               |
| 1991 | 9               | 0.3                |
| 1992 | 12              | 0.4                |
| 1993 | 111             | 3.4                |
| 1994 | 3432            | 104.4              |
| 1995 | 10              | 0.3                |
| 1996 | 180             | 5.5                |
| 1997 | 66              | 2                  |

Epidemiologjia e sëmundjeve infektive në Shqipëri (1960-2001), E. Kakarriqi, Tiranë, 2003.

Përcaktoni modën.

Bëni paraqitjen grafike për incidencën dhe numrin e rasteve me rubeolë.

1. Në tabelën më poshtë jepet numri i rasteve me malarie sipas grup-moshave.

Tabela Rastet me malarie sipas grupmoshave

| Grup-mosha<br>(në vite) | Numri i<br>rasteve |
|-------------------------|--------------------|
| 0-5                     | 4650               |
| 5-14                    | 12019              |
| 14-49                   | 49075              |
| >49                     | 5514               |

World malaria report 2005, EHO, UNICEF, f. 104.

Gjeni modën.

Gjeni medianën dhe vlerën e saj.

Gjeni mesataren aritmetike.

Bëni paraqitjen grafike.

3. Është matur niveli i tensionit arterial sistolik në disa pacintë me sëmundje ishemike të zemrës (tabela më poshtë).

Tabela paraqitja e velerave te tensionit arterial

| Tensioni arterial<br>(mmHg) | Numri i<br>rasteve |
|-----------------------------|--------------------|
| 115-124                     | 4                  |
| 125-134                     | 4                  |
| 135-144                     | 5                  |
| 145-154                     | 8                  |
| 155-164                     | 5                  |
| 165-174                     | 4                  |
| 175-184                     | 5                  |

Gjeni modën.

Gjeni medianën dhe vlerën e saj.

Gjeni mesataren aritmetike.

Bëni paraqitjen grafike.

4. Dy kardiologë analizuan lidhjen mes atakut kardiak në 1500 meshkuj (grupmosha 55-75 vjec) dhe aktivitetit fizik që paraprin atakun në 24 orë. Aktiviteti fizik klasifikohet si i rëndë, i moderuar, i mesëm dhe i lehtë. 52% referuan se kishin bërë sforcim të rëndë 24 orë para takut, 24 % të moderuar, 35% aktivitet aktivitetet të mesëm dhe 15% aktivitetet të lehtë.

Gjeni cili nga grafikët do ta prezantonte më mirë këtë situatë:

- Bar-diagrami
- Histogrami
- Poligoni i frekuencave me frekuencat absolute në boshtin Y
- Poligoni i frekuencave me përqindjet e frekuencave në boshtin Y

5. Në tabelën 20 jepet gjendja e punësimit për 30 persona të pyetur rastësisht: P- i punësuar, PP- i papunë, PJ– me punë të përkohshme.  
Ndërtoni tabelën e shpërndarjes së frekuencave për të dhënat më lart. Gjeni frekuencen relative, përqindjet për secilin rast dhe ndërtoni grafikun e duhur.

Tabela 20 Gjendja e punësimit

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| P  | PP | PJ | PP | P  | P  | P  | PJ | PJ | P  |
| P  | P  | P  | P  | PJ | P  | P  | P  | P  | PJ |
| PP | PP | PJ | P  | P  | PP | PJ | P  | P  | P  |

6. Në një studim të kryer mbi statistikat për kolagenozat në QSUT për vitet 1995-1999, (Studim epidemiologjik mbi statistikat për kolagenozat në QSUT “Nënë Tereza”, E. Petrela, Y. Sarolli), u morën të dhënat si në tabelën më poshtë:

Tabela Paraqitja e kolagenozave në QSUT, 1995-1999)

| Kolagenozat                    | Shpeshësia e hasjes (%) |
|--------------------------------|-------------------------|
| Arthrit Rheumatoid             | 48                      |
| Lupus erithematos i diseminuar | 32                      |
| Sklerodermi                    | 20                      |
| Totali                         | 100.0                   |

a) Ndërtoni grafikun për këto të dhëna.

7. Në një studim mbi vlerësimin e cilësisë në spitalet universitare Tiranë, njera nga pikaqë pikat që u morr mendimi i pacientëve ishte mbi vlerësimin e “sjelljes së mjekëve” dhe u morën të dhënat si në tabelën më poshtë:

Tabela. Vlerësimi i “sjelljes së mjekëve” nga këndvështrimi pacientit

| Vlerësimi i sjelljes së mjekëve | Përqindja |
|---------------------------------|-----------|
| Shkëlqyer                       | 21.7      |
| Mirë                            | 40.1      |
| Dobët                           | 31.8      |
| Keq                             | 6.0       |
| Shumë keq                       | 0.5       |

Burimi: Vlerësimi i cilësisë në spitalet universitare – Tiranë, Hoxha A., Sarolli Y., Dhjetor, 2006

- Llogarisni frekuencën relative për çdo kategori.
- Ndërtoni grafikun me sipërfaqe për shpërndarjen e përqindjeve.

#### 4. PËRSHKRIMI I PËRHAPJES SË VLERAVE PËRRETH QENDRËS

Përmes analizës së masave të tendendës qëndrore, ne mund të marrim informacion lokalizimit në “qendër të vlerave, por është shumë e rëndësishme që ne të dimë se si përhapën vlerat kundrejt qendrës, në një varg vlerash.

Pikërisht, treguesit që përdoren për të treguar përhapjen e vlerave kundrejt vlerave mesatare quhen masa të dispersionit. Këtu përfshihen: rangi, varianca dhe devijacioni standard.

##### 4.1 RANGU

Rangu është diferenca mes vlerës më të madhe dhe më të vogël në një varg vlerash. (Daniel, 1991, p 24). Psh në tabelën më poshtë, paraqiten gjatësitë në cm të pesë bebeve të porsalindura. Siç e shihni rangi është 8. ( $52 - 44 = 8$ )

Tabela. Gjatësia e bebeve

| Bebet | Gjatësia (në cm) |
|-------|------------------|
| A     | 48               |
| B     | 47.5             |
| C     | 44               |
| D     | 52               |
| E     | 50               |

##### 4.2 MASAT E DISPERSIONIT

Masat e dispersionit i kemi a) masa të dispersionit të bazuara në percentile, b) masa të dispersionit të bazuara në mesatare

**4.2.1 Masat e dispersionit të bazuara në percentile.** Percentili ose kuartili është përqindja e observacioneve poshtë pikes së referimit, kur të gjithë observacionet janë klasifikuar sipas një shkalle rritëse. Mediana zë percentilin e 50°. Percentili 75° përfaqëson pikën nën të cilën ndodhen 75% e vlerave të observuara dhe percentili 25° përfaqëson pikën nën të cilën ndodhen 25% e vlerave të observuara.

Rangu kompleksiv është diferenca mes vlerës më të madhe dhe asaj më të vogël të të dhënave.

Rangu interkuartil është diferenca mes kuartilit të 75° dhe atij të 25°.

Avantazhet e përdorimit të percentileve, qëndrojnë se mund të përdoren për të gjithë tipet e të dhënave, pra edhe në rastet kur nuk dihet shpërndarja e të dhënave.

##### 4.2.2 Masat e dispersionit të bazuara në mesatare.

$$\text{Devijacioni i mesatares "s"} = \frac{\sum (x - \bar{x})}{n}$$

Varianca në një varg të dhënash është shuma e devijacionit kuadratik nga mesatarja, pjesëtuar me numrin "n" të observacioneve minus 1.



Ngritja ne katror e variancës "s" është bërë për të shmangur faktin që shuma  $(x-x)$  del 0. Pjesëtohet me  $n-1$  (që quhen shkalle lirie) dhe jo me "n" meqënëse variaca e thjeshte nuk është një vlerësim i shkëlqyer i variancës së popullatës. Shuma  $(x-\bar{x})^2$  shpreh një entitet shumë të rëndësishëm në statistikë, që quhet shuma e katrorëve "SS" ose shuma totale e katrorëve "TSS". TSS-ja mat sasinë totale të variacioneve në një grup observimesh.

### 4.3. DEVIJACIONI STANDART

Meqënëse varianca tenton të jetë një numër i madh, vlera e saj mund të bjerë jashtë rangjeve të vlerave të observuara të një grupi të dhënash. Për këtë arsye, përdoret devijacioni standart, që shënohet me "SD".

Kur numri i rasteve është më i vogël ose i barabartë me 30 ( $n \leq 30$ ), devijacioni standart llogaritet me formulën

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

Kur numri i rasteve është më i madh se 30 ( $n > 30$ ), devijacioni standart llogaritet me formulën:

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Në një grup të dhënash të observuara, termi  $\bar{x} \pm SD$ , paraqet një devijacion standart nën dhe mbi mesataren aritmetike.

Për të dhëna të grupuara, kur numri i rasteve është më i vogël ose i barabartë me 30 ( $n \leq 30$ ), devijacioni standart llogaritet me formulën:

$$SD = \sqrt{\frac{\sum (m - \bar{x})^2 fr}{n-1}}$$

ku:

- $m_i$  - është vlera e mesit për klasën në fjalë;
- $f_i$  - shpeshtësia e hajsës (frekuenca);
- $n$  - numri i rasteve

Kur numri i rasteve është më i madh se 30 ( $n > 30$ ), devijacioni standart për të dhëna të grupuara, llogaritet me formulën:

$$SD = \sqrt{\frac{\sum (m - \bar{x})^2 fr}{n}}$$

### Ushtrime.

Në tabelën më poshtë, paraqitet shpërndarja e familjeve sipas numrit të anëtarëve në një familje

| Nr i anëtarëve | Nr i familjeve |
|----------------|----------------|
| 1              | 153            |
| 2              | 225            |
| 3              | 335            |
| 4              | 564            |
| 5              | 346            |
| 6              | 133            |
| 7              | 75             |
| 8              | 49             |
| <b>Shuma</b>   | 1.880          |

Gjeni:

- Mesataren aritmetike;
- Mesataren harmonike;
- Mesataren gjeometrike;

### Zgjidhje

- Shprehja analitike e mesatares aritmetike në një shpërndarje frekuencash është:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i}$$

ku  $x_i, i = 1, 2, \dots, k$ , janë variabellet e vëzhguara dhe  $n_i, i = 1, 2, \dots, k$ , frekuencat korresponduese. Atëherë kemi:

| $x_i$         | $n_i$ | $x_i n_i$ |
|---------------|-------|-----------|
| 1             | 153   | 153       |
| 2             | 225   | 450       |
| 3             | 335   | 1.005     |
| 4             | 564   | 2.256     |
| 5             | 346   | 1.730     |
| 6             | 133   | 798       |
| 7             | 75    | 525       |
| 8             | 49    | 392       |
| <b>Totale</b> | 1.880 | 7.309     |

Duke përdorur formulën për të dhëna në tabelë, mesatarja aritmetike është:

$$\bar{x} = \frac{7.309}{1.880} = 3,88$$

Mesatarja harmonika në një shpërndarje frekuencash është:

$$H = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \frac{n_i}{x_i}}$$

Për të përdorur formulën, duhet të ndjekim skemën e llogaritjeve si në tabelë:

| $x_i$        | $n_i$ | $n_i/x_i$ |
|--------------|-------|-----------|
| 1            | 153   | 153,00    |
| 2            | 225   | 112,50    |
| 3            | 335   | 111,67    |
| 4            | 564   | 141,00    |
| 5            | 346   | 69,20     |
| 6            | 133   | 22,17     |
| 7            | 75    | 10,71     |
| 8            | 49    | 6,13      |
| <b>Shuma</b> | 1.880 | 626,37    |

Mesatarja harmonike është :

$$H = \frac{1.880}{626,37} = 3$$

b. Mesatrja gjeometrike në një shpërndarje frekuencash është:

$$G = \sqrt[n_1]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}} = \sqrt[n_1]{\prod_{i=1}^k x_i^{n_i}}$$

ose:

$$G = \exp \left\{ \frac{1}{\sum_{i=1}^k n_i} \sum_{i=1}^k n_i \log x_i \right\}$$

Për të përdorur formulën, duhet të ndjekim skemën e llogaritjeve si në tabelë:

| $x_i$         | $n_i$ | $\log x_i$ | $n_i \log x_i$ |
|---------------|-------|------------|----------------|
| 1             | 153   | 0,00       | 0,00           |
| 2             | 225   | 0,30       | 67,73          |
| 3             | 335   | 0,48       | 159,84         |
| 4             | 564   | 0,60       | 339,56         |
| 5             | 346   | 0,70       | 241,84         |
| 6             | 133   | 0,78       | 103,49         |
| 7             | 75    | 0,85       | 63,38          |
| 8             | 49    | 0,90       | 44,25          |
| <b>Totale</b> | 1.880 |            | 1.020,10       |

Mesatarja gjeometrike në një shpërndarje frekuencash është:

$$G = \exp \left\{ \frac{1}{1.880} \cdot 1.020,10 \right\} = \exp \{0,543\} = 3,49$$

c) Gjeni devijacionin standart:

| $x_i$        | $f_i$        | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $(x_i - \bar{x})^2 * f_i$ |
|--------------|--------------|-----------------|---------------------|---------------------------|
| 1            | 153          | -3              | 9                   | 1377                      |
| 2            | 225          | -2              | 4                   | 1100                      |
| 3            | 335          | -1              | 1                   | 335                       |
| 4            | 564          | 0               | 0                   | 0                         |
| 5            | 346          | 1               | 1                   | 346                       |
| 6            | 133          | 2               | 4                   | 532                       |
| 7            | 75           | 3               | 9                   | 675                       |
| 8            | 49           | 4               | 16                  | 784                       |
| <b>Shuma</b> | <b>1.880</b> |                 |                     | <b>5149</b>               |

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2 * f}{n}} = \sqrt{\frac{5149}{1880}} = 1,6$$

### Ushtrime

1. Në tabelën më poshtë janë paraqitur peshat në kg në 12 bebe.

Tabela. Peshat e femijeve ne kg

| Pesha (kg) | Numri i rasteve |
|------------|-----------------|
| 3          | 2               |
| 4          | 4               |
| 6          | 2               |
| 7          | 2               |

Gjeni modën.

Gjeni medianën dhe vlerën e saj.

Gjeni mesataren aritmetike.

Llogarisni devijacionin standard për këto të dhëna.

2. Tabela më poshtë jep shpërndarjen e frekuencave të niveleve të zinkut (në mg/dl) në meshkuj nga moshë 15 – 17 vjeç.

Tabela Paraqitja e niveleve të Zn në meshkujt e moshës 15-17 vjeç

| Nivelet e Zn | Numri i rasteve |
|--------------|-----------------|
| 50-59        | 6               |

|         |     |
|---------|-----|
| 60-69   | 35  |
| 70-79   | 110 |
| 80-89   | 116 |
| 90-99   | 91  |
| 100-109 | 63  |
| 110-119 | 30  |
| 120-129 | 5   |
| 130-139 | 2   |
| 140-149 | 2   |
| 150-159 | 2   |

Gjeni modën.

Gjeni medianën dhe vlerën e saj.

Gjeni mesataren aritmetike.

Llogarisni devijacionin standard për këto të dhëna.

#### 4.4 KOEFICIËNTI I VARIACIONIT

Koeficienti i variacionit përkufizohet si raporti i devijacionit standart me mesataren aritmetike i shumëzuar me 100%. (Sokal&Rohlf, 1981, p 59).

Koeficienti i variacionit përdoret për të krahasuar vlerën relative të variacionit.

Tab. Koeficienti i variacionit për variable me shkallë të ndryshme matjeje

| Variabli      | Mesatarja | SD | CV     |
|---------------|-----------|----|--------|
| Pesha (kg)    | 75        | 10 | 13,3 % |
| Gjatësia (cm) | 170       | 20 | 11,7 % |

Daëson\_Saunders&Trapp (1994) e cilësojnë koeficientin e variacionit si një tregues shumë të përdorshëm në testimet laboratorike dhe në kontrollin cilësor të procedurave laboratorike.

$$CV = \frac{SD}{\bar{x}} 100\% \quad \text{ose} \quad CV = \frac{\sigma}{\mu} 100\%$$

#### 4.5 SHPERNDARJA SIPAS TRUNGUT DHE GJETHES

Lidhet me shpërndarjen e variableve në një diagram, duke shfrytëzuar mesataret e pozicionit (shpërndarja bazuar në percentile).

Ky lloj diagrami ka dy komponentë kryesore:

- **trungun**, që paraqet kolonën e dhjetësheve
- **gjethen**, që paraqet kolonën e njesheve.

Ta sqarojmë këtë me një shembull:

Është dhënë vargu i vlerave si më poshte. Bëni shpërndarjen sipas trungut dhe gjethes.

2, 4, 4, 10, 18, 19, 29, 30, 31, 44, 44, 49, 51

Zhvillim

Fillimisht bëjmë zbërthimin e shifrës në dhjetëshe dhe njëshe (fig 5). Për këtë veprojmë në këtë mënyrë. Në anën e majtë të linjës vertikale vendosim dhjetëshet (nga 0 tek nënta), kurse në anën e djathtë, vendosim njëshet. Kështu numri 2 ka 0 dhjetëshe dhe 2 njëshe, pra numrin dy, do ta vendosim në rradhën e parë. Numri 4 po ashtu do të vendoset në rradhën e parë. Edhe pse kemi dy numra 4, ato do t'i paraqesim të dy në grafik për të treguar prezencën e tyre. Mos harroni se çdo numër prej tyre ka pozicionin e vet në varg.

Numri 10 ka një dhjetëshe dhe 0 njëshe. Për këtë ai vendoset në rradhën e dytë, duke shënuar 1 në anën e majtë të linjës vertikale dhe 0 në anën e djathtë të saj. Kështu do të veprojmë për të gjitha shifrat e tjera.

(1) 5|1  
(4) 4| 4,4,9  
(6) 3| 0,1  
(7) 2| 9  
(6) 1| 0,8,9  
(3) 0| 2,4,4

(fig.5)

Duke qenë se normalisht kemi shumë vlera në varg, të cilat është shpeshherë e lodhshme, ose dhe e pamundur t'i numërosh, përdorim masën kumulative (në grafik janë vlerat e shënuara brenda kllapave, në anën e majtë të grafikut). Për këtë veprojmë kështu: zgjedhim me hamendje një nivel të mesit në grafik, p.sh. midis 20 dhe 30 dhe heqim një linjë imagjinare horizontale. Mbi dhe nën këtë linjë, mbledhim me rradhë shifrat duke u ngjitur nga një "kat" në tjetrin. P.sh: në rradhën e parë kemi tre shifra (gërma); shënojmë tre në kllapa. Në rradhën e dytë kemi dhe tre shifra (gërma) të tjera, që ia shtojmë tre të parave; pra  $3+3=6$ . Këtë e vendosim në kllapa në rradhën (katin) e dytë. Në të tretin kemi  $6+1=7$ . Meqenëse arritëm në nivelin e mesit, që zgjedhëm në fillim, këtu ndalojmë dhe rinisim numërimin për gjysmën tjetër, duke u nisur nga lart-poshtë dhe kemi:

1 element në rradhën e pesëdhjetësheve,  
 $1+3=4$  elementë në rradhën e dyzetësheve,  
 $4+2=6$  elementë në rradhën e tridhjetësheve.

Duke mbledhur tani vetëm dy masat  $6+7=13$ , ne gjejmë numrin e elementëve në varg, pra  $n = 13$ .

Pas kësaj do të gjejmë pozicionin  $75^\circ$ , trecerekun e vargut, që quhet ndryshe kuartili i sipërm dhe që jepet me formulën:

$$Q_s = \frac{3}{4}(n+1)$$

Kështu, do të kemi :

$$Q_s = \frac{3}{4}(13+1) = 10,5$$

Quartili i sipërm ndodhet në pozicionin 10,5 dhe vlera e tij llogaritet si mesatarja aritmetike e vlerës në pozicionin dhjetë dhe asaj në pozicionin e njëmbëdhjetë:

$$Vlera\ e\ Q_s = \frac{(Vlera\ e\ pozicionit\ dhjete + vlera\ e\ pozicionit\ njembdhjete)}{2}$$

$$Vlera\ e\ Q_s = \frac{(44 + 44)}{2} = 44$$

Pozicioni  $25^\circ$  në varg, cereku i vargut, ose kuartili i poshtëm;

$$Q_p = \frac{1}{4}(n+1)$$

Mediana (pozicioni 50%)

$$Me = \frac{1}{2}(n+1)$$

Rangu interkuartil =  $Q_s - Q_p$       Rangu kompleksiv =  $51 - 2 = 49$

Pasi gjetëm këto, ndërtojmë Box-plott -in që është bashkësia e pesë pikave: mediana,  $Q_s$ ,  $Q_p$ , ekstremiteti i sipërm dhe ekstremiteti i poshtëm (fig 3).



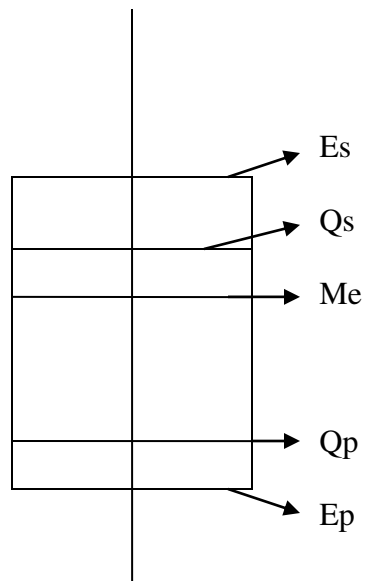


fig 3

### Ushtrime.

1. Është dhënë vargu i vlerave si më poshtë: 0,19, 23, 7, 54, 0, 23, 32, 9. Gjeni të gjitha mesataret që dini, SD dhe CV.
2. Për vargun e vlerave ndërtoni Box-plott-in. Vargu i vlerave është: 11,9,33,24,55,42,7,24,50.
3. Për të dhënat në tabelën më poshtë, gjeni mesataren aritmetike, moden, medianën dhe ndërtoni grafikun e duhur.

| Peshat (kg) | Numri i rasteve |
|-------------|-----------------|
| 21-30       | 25              |
| 31- 40      | 20              |
| 41-50       | 40              |
| 51-70       | 15              |

Tabela 33. Paraqitja e peshave dhe frekuencave

4. Në një studim të realizuar mbi prognozën e fëmijëve që kanë kaluar meningit bacterial, u analizuan ataket e bëra tyre në intervale kohore të matura në muaj, nga ku u morrën të dhënat si më poshtë:

0, 10, 0, 25, 4, 12, 24, 0, 50, 12, 31, 24, 36, 42, 55, 96

Gjeni modën, medianën, mesataren aritmetike, devijacionin standard, rangun interkuartil dhe ndërtoni box-plott-in.

## Statistika deskriptive

Prof. Dr. Elizana Petrela

---

---

---

---

---

---

---



## Numrat, të Dhënat dhe Statistika Deskriptive

---

---

---

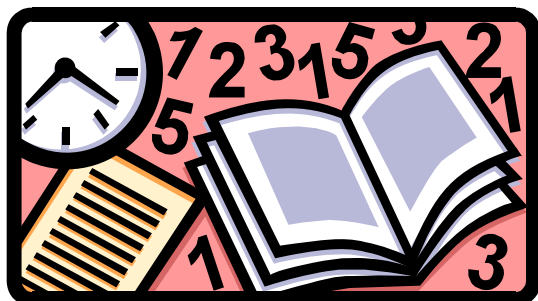
---

---

---

---

## Të Dhënat dhe Numrat



---

---

---

---

---

---

---

Përdorimi i duhur i procedurave statistikore kërkon të kuptuarit e marrëdhënies midis të dhënave dhe numrave.

Karakteristikat që maten quhen variablat. Vlerat e variablave janë karakteristikat e të dhënave që dëshironi të hulumtoni.

Vlerat të cilat u caktohen variablave quhen observime.

---

---

---

---

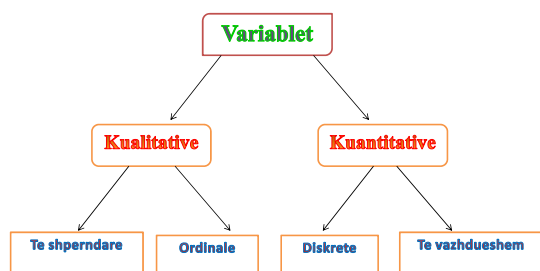
---

---

---

---

#### Tipet e variableve



9/11/2018

5

---

---

---

---

---

---

---

---

#### Variablet

|                                                               |                                                                  |
|---------------------------------------------------------------|------------------------------------------------------------------|
| <b>Nominal</b><br>-diagnoza<br>-grup gjaku (O, A, B, AB)      | <b>Ordinal</b><br>-shkallët e sëmundjes<br>-nivelet e ushqyerjes |
| <b>Dikotomik</b><br>Gjinia (m/f)<br>Mbijetesa (gjallë/vdekur) | <b>Vazhdueshëm</b><br>-niveli i kolesterol<br>-TA<br>-moshat     |
| <b>Discrete</b><br>-numri i shtretërve<br>-numri i QSH        |                                                                  |

E. Petrela

---

---

---

---

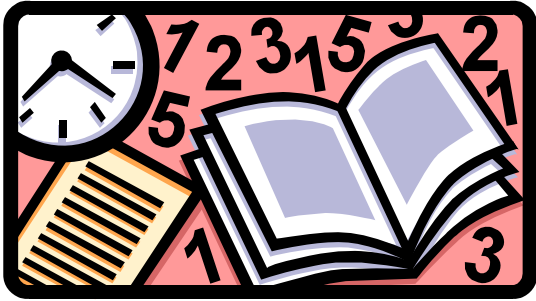
---

---

---

---

## Statistika Deskriptive



---

---

---

---

---

---

---

## Përmbledhja

Përmbledhja i referohet reduktimit të vëzhgimeve të papërpunuara në një format të përdorshëm dhe të interpretueshëm.

Të dhënat mund të përmblihen në një nivel bivariate duke përdorur teknikat e paraqitura më poshtë.

Përmbledhjet numerike të të dhënave mund të marrin formën e tabelave të frekuencave të vlerave të cilat ka një ndryshore, ose

një ndryshore mund të përmblihet duke përdorur masat e tendencës qendrore, ose së treti,

një variabël mund të përmblihet me vlera që tregojnë sasinë e ndryshueshmërisë së një grupi i të dhënave.

---

---

---

---

---

---

---

## Statistika deskriptive

- Organizimin, përmbledhjen dhe përshkrimin e të dhënave
- Shembuj: përqindja, frekuencat, madhësitë mesatare, tabelat dhe grafikët
- Nuk ka qëllim përgjithësimin e rezultateve

---

---

---

---

---

---

---

## Mesataret e pozicionit

### Moda :

Në një seri të dhënash, modë quhet varianti me frekuencën më të lartë.

### Mediana

Mediana i korrespondon variablit që ze pozicionin e mesit në një varg vlerash, kur vlerat janë të renditura nga vlera më e madhe tek më e vogla, ose anasjelltas.

$$Me = \frac{n+1}{2}$$

9/11/2018

E. Petrela

---

---

---

---

---

---

---

---

## Mesatarja aritmetike

raporti mes shumës së variableve dhe numrit total të varianteve (obervimeve) të marra (te kryera).

$$\bar{x} = \frac{\sum (x \cdot fr)}{n}$$

ku:

x -variabli ;

n -numri i të gjithë variableve =  $\sum fr_i$ .

9/11/2018

E. Petrela

---

---

---

---

---

---

---

---

## Veçoritë e mesatares dhe medianës

### Mesatarja aritmetike

- është e vetme në një varg vlerash;
- është e thjeshtë në llogaritje;
- Ndikohet nga vlerat ekstreme në varg

### Mediana

- është e vetme në një varg vlerash;
- Është e thjeshtë në llogaritje;
- Nuk ndikohet nga vlerat ekstreme në varg

9/11/2018

E. Petrela

---

---

---

---

---

---

---

---

## Treguesit e variacioneve

- Variaconi shprehet përmes treguesve të cilët paraqesin nivelin e shmangieeve të të dhënave të tiparit ,nga madhësia mesatare e dukurisë së analizuar.
- Treguesit e variacioneve përdoren për të zbuluar zhvillimin dhe ligjshmëritë e shpërndarjes së dukurisë që analizohet.
- Ata pasqyrojnë masën dhe përmes tyre, mund të paraqitet ndikimi i faktoreve të ndryshëm në dukurinë që analizohet
- Sa më të vegjël treguesit e variacionit, aq më e drejt ëdhe reale është vlera e mesatares aritmetike, që gjenerohet nga këto të dhëna.

---

---

---

---

---

---

---

## Treguesit e variacionit

### ABSOLUT

- 1. Gjerësia e variacionit,
- 2. Shmangia mesatare absolute,
- 3. Devijimi standard,
- 4. Varianca,

### RELATIV

- 1.Koefiçienti i variacionit,

---

---

---

---

---

---

---

## Gjerësia e variacionit

- Gjerësia e variacionit llogaritet si ndryshimi absolut mes vlerës më të madhe dhe vlerës më të vogël.

$$GJv = X_{max} \pm X_{min}$$

---

---

---

---

---

---

---

## Shmangia mesatare absolute

- ShMA paraqet madhësinë mesatare të diferencave absolute në mes të mesatares aritmetike dhe vlerave individuale të vargut të të dhënave.

$$\text{Sh M A} = \frac{\sum (\bar{x} - x)}{n}$$

---

---

---

---

---

---

---

## Varianca

- Varianca paraqet madhësinë mesatare të shmangieve të të dhënave në katror nga mesatarja aritmetike dhe ka vlera nga 0 deri +1.

$$\sigma^2 = \frac{\sum_{i=1}^n (\bar{x} - x)^2}{n}$$

---

---

---

---

---

---

---

## Koeficienti i variacionit

- CV është tregues relativ, që shpreh madhësinë relative të raportit në mes SD dhe mesatares

$$CV = \frac{\sigma}{\bar{x}}$$

- Sa më i vogël CV, aq më homogjene janë të dhënat e sasisë së dukurisë, pra, ndikimi i faktorëve të jashtëm është shumë i vogël.

---

---

---

---

---

---

---

## Devijacioni standart

• Kur  $n \leq 30$   $SD = \sqrt{\frac{\sum(x-x)^2}{n-1}}$

dhe kur  $n > 30$   $SD = \sqrt{\frac{\sum(x-x)^2}{n}}$

Në një grup të dhënash të observuara, termi  $\bar{x} \pm SD$ , paraqet një devijacion standart nën dhe mbi mesatare.

9/11/2018

E. Petrella

---

---

---

---

---

---

---

---

## Paraqitja grafike e modës



9/11/2018

E. Petrella

---

---

---

---

---

---

---

---

## Një mënyrë e thjeshtë dhe efikase për të përmbledhur të dhënat është me tabela:

- tabela të thjeshta (një variabël)
  - Univariate
- Tabela të përbëra (dy variable)
  - Bivariate (rxc)
- Tabela me më shumë se 2 variabla
  - Mantel-Haenszel
  - Tabelat Pivot

E. Petrella

---

---

---

---

---

---

---

---



## Frekuencat

- Frekuenca Absolute (fr)
- Frekuenca Relative (FR)
- Përqindja (%)
- Frekuenca Cumulative (Fc)
- Frekuenca relative ( $F_{c_R}$ )

---

---

---

---

---

---

---

## Tabela e frekuencave

Supozojmë se pyetëm 18 të parët që hynë në parking, për vitin akademik në të cilin ata janë?

| Person # | Viti akademik             | Person # | Viti akademik             |
|----------|---------------------------|----------|---------------------------|
| 1        | student i vitit të parë   | 10       | student i vitit të fundit |
| 2        | student i vitit të parë   | 11       | student i vitit të fundit |
| 3        | student i vitit të parë   | 12       | stazhier                  |
| 4        | stazhier                  | 13       | student i vitit të parë   |
| 5        | stazhier                  | 14       | stazhier                  |
| 6        | stazhier                  | 15       | stazhier                  |
| 7        | student i vitit të parë   | 16       | student i vitit të parë   |
| 8        | student i vitit të fundit | 17       | stazhier                  |
| 9        | stazhier                  | 18       | Junior                    |

---

---

---

---

---

---

---

Krijoni një tabelë të thjeshtë.

---

---

---

---

---

---

---

### Tabela e thjeshtë

| Grada                     | fr |
|---------------------------|----|
| student i vitit të parë   | 6  |
| stazhier                  | 9  |
| student i vitit të fundit | 3  |
| Total                     | 18 |

---

---

---

---

---

---

---

---

### Frekuenca Relative

• FR

= Frekuenca për këtë kategori/Shumën e të gjithë frekuencave

$$= fr / \Sigma f$$

---

---

---

---

---

---

---

---

### Frekuenca Relative

| Grade                     | Fr | FR          |
|---------------------------|----|-------------|
| student i vitit të parë   | 6  | 6/18 = 0.33 |
| stazhier                  | 9  | 9/18 = 0.5  |
| student i vitit të fundit | 3  | 3/18 = 0.17 |
| Total                     | 18 | 1.0         |

---

---

---

---

---

---

---

---

## Përqindja (%)

| Grade                     | Fr | FR            | %                 |
|---------------------------|----|---------------|-------------------|
| student i vitit të parë   | 6  | $6/18 = 0.33$ | $0.33 * 100 = 33$ |
| stazhier                  | 9  | $9/18 = 0.5$  | 50                |
| student i vitit të fundit | 3  | $3/18 = 0.17$ | 17                |
| Total                     | 18 | 1.0           | 100               |

---

---

---

---

---

---

---

---

## Frekuenca Cumulative

| Grade                     | Fr | Fc |
|---------------------------|----|----|
| student i vitit të parë   | 6  | 6  |
| stazhier                  | 9  | 15 |
| student i vitit të fundit | 3  | 18 |
| Total                     | 18 |    |

---

---

---

---

---

---

---

---

## Tabela e frekuencave

| Raca      | Fr  | FR | Fc |
|-----------|-----|----|----|
| Bardhë    | 376 |    |    |
| Ngjyrë    | 59  |    |    |
| Egjiptian | 24  |    |    |
| Të tjerë  | 41  |    |    |
| Total     | 500 |    |    |

- Plotësoni tabelën

---

---

---

---

---

---

---

---

## Tabela e frekuencave

| Raca      | Fr  | FR    | Fc    |
|-----------|-----|-------|-------|
| Bardhë    | 376 | 0.752 | 0.752 |
| Ngjyrë    | 59  | 0.118 | 0.870 |
| Egjiptian | 24  | 0.048 | 0.918 |
| Të tjerë  | 41  | 0.082 | 1.000 |
| Total     | 500 | 1.000 |       |

**Pyetje** – Sa është përqindja e popullatës me ngjyrë?

**Përgjigje** – 11.8% (% =  $FR \cdot 100 = 0.118 \cdot 100$ )

---

---

---

---

---

---

---

---

## Tabelë e përbërë

| Race or Gender | Meshkuj | Femra | Total |
|----------------|---------|-------|-------|
| Bardhë         | 177     | 199   | 376   |
| Ngjyrë         | 24      | 35    | 59    |
| Egjiptian      | 11      | 13    | 24    |
| Të tjerë       | 15      | 26    | 41    |
| <b>Total</b>   | 227     | 273   | 500   |

**Cila është frekuenca absolute dhe frekuenca relative e femrave të bardha?**

Frekuenca absolute është 199 dhe frekuenca relative është  $199/500 = 0.398$ , 39.8% e mostrës totale janë femra të bardha.

---

---

---

---

---

---

---

---

## TË DHËNA TË GRUPUARA

---

---

---

---

---

---

---

---

## Të dhëna të grupuara

| Variabli-paga mujore | Numri i punonjesve (Fr) |
|----------------------|-------------------------|
| 121-140              | 31                      |
| 141-170              | 39                      |
| klasa III 171-200    | 15 frekuenca per kl III |
| 201-300              | 9                       |
| 301-500              | 6                       |
| Shuma                | 100                     |

Vlera me e vogel e  
klases se trete

Vlera me e madhe e  
klases se trete

---

---

---

---

---

---

---

---

| Klasat Limit | Klasa kufij                         | Gjerës e klasës Δi | Klasa midpoint m |
|--------------|-------------------------------------|--------------------|------------------|
| 121-140      | 120.5 deri tek me te vogel se 140.5 | 20                 | 130.5            |
| 141-170      | 140.5 deri tek me te vogla se 170.5 | 30                 | 155.5            |
| 171-200      | 170.5 deri tek me te vogla se 200.5 | 30                 | 185.5            |
| 201-300      | 200.5 deri tek me te vogla se 300.5 | 100                | 250.5            |
| 301-500      | 300.5 deri tek me te vogla se 500.5 | 200                | 400.5            |

---

---

---

---

---

---

---

---

## Të dhëna të grupuara

**Pikën mesatare** (midpoint) te limitit te siperm per klasën e parë dhe atij të poshtëm, për klasën e dytë

$$M = \frac{140 + 141}{2} = 140.5$$

Gjerësia e klasës = kufiri i sipërm – kufiri i poshtëm

**Gjerësia** per kl. e pare:  $\Delta i = 140.5 - 120.5 = 20$

Klasa midpoint ( per kl I):  $m = \frac{121+140}{2} = 130.5$

---

---

---

---

---

---

---

---

## Mesatarja aritmetike dhe Mediana

$$\bar{x} = \frac{\sum mf}{N}$$

$$Me = x_i + \Delta_i * \frac{0.5 - F_{c_B}}{Fr}$$

---

---

---

---

---

---

---

## Devijacioni standart

• Kur  $n \leq 30$   $SD = \sqrt{\frac{\sum (m - \bar{x})^2 fr}{n-1}}$

dhe kur  $n > 30$   $SD = \sqrt{\frac{\sum (m - \bar{x})^2 fr}{n}}$

9/11/2018

E. Petrela

---

---

---

---

---

---

---

## Statistika deskriptive

**Grafiku është paraqitja e të dhënave me anë të pikave, vijave, figurave gjeometrike, hartave te gatshme apo dhe simboleve.**

Grafiku e bën më të lexueshëm fenomenin.  
Çdo grafik ka të domosdoshëm tri elementë:

- titullin;
- shkallën dhe;
- legjendën.

9/11/2018

E. Petrela

---

---

---

---

---

---

---

### Për të paraqitur shpërndarjet e variablave sasiore

- Trungu dhe gjethet,
- Histogramët,
- Poligonët e frekuencave,
- Box-plott,
- Bar-diagramë,
- Grafiku vijor,
- Scatter-diagramë
- Grafiku me pika.

---

---

---

---

---

---

---

### SHPËRNDARJA SIPAS TRUNGUT DHE GJETHES

---

---

---

---

---

---

---

### Shpërndarja sipas trungut dhe gjethes

Më i përshtatshëm për numër të vogël observimesh

Ky lloj diagrami ka dy komponentë kryesore:

- **trungun**, që paraqet kolonën e dhjetësheve dhe
- **gjethen**, që paraqet kolonën e njësheve.

---

---

---

---

---

---

---

## Qëllimi

- të qartësojë formën e shpërndarjes;
- i përdorshëm për krahasimin e shpërndarjeve.

---

---

---

---

---

---

---

- është dhënë vargu i vlerave si më poshtë:

2, 4, 4, 10, 18, 19, 29, 30, 31, 44, 44, 49, 51

Bëni shpërndarjen sipas trungut dhe gjethes.

---

---

---

---

---

---

---

## Shembull

- (1) 5 | 1
- (4) 4 | 4,4,9
- (6) 3 | 0,1
- (7) 2 | 9
- (6) 1 | 0,8,9
- (3) 0 | 2,4,4


$$N=7+6=13$$

---

---

---

---

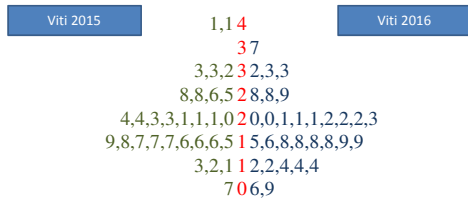
---

---

---



## Back-to-back




---

---

---

---

---

---

---

---

## Të dhëna të vazhdueshme

- Vijor;
- Histogramë;
- Scatter diagramë;
- Box-plot

9/11/2018

E. Petrela

---

---

---

---

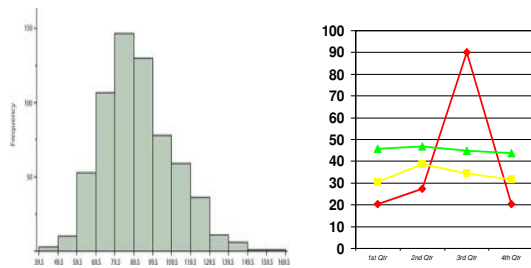
---

---

---

---

## Paraqitja grafike për të dhëna të vazhdueshme



9/11/2018

E. Petrela

---

---

---

---

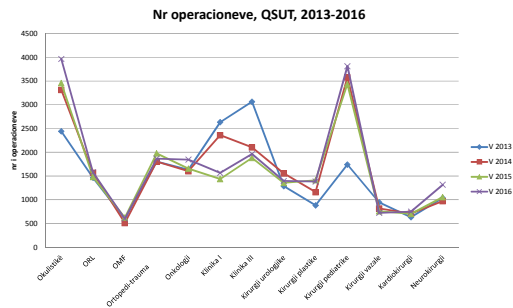
---

---

---

---

!!!



---

---

---

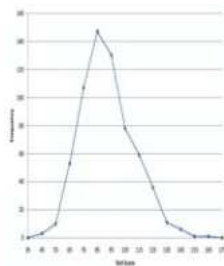
---

---

---

## Poligoni i frekuencave

- Poligoni i frekuencave quhet vija qe formohet nga bashkimi i pikave te mesit te kolonave te njepasnjeshme te histogramit



---

---

---

---

---

---

## Histogrammi

Histogrami mund të marrë disa forma:

- Simetrik
- I pjerrët (majtas ose djathtas)
- Uniform
- Multimodal

---

---

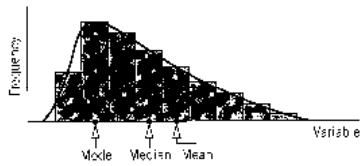
---

---

---

---

## Lidhja mes mesatares medianes dhe modes



9/11/2018

E. Petreli

---

---

---

---

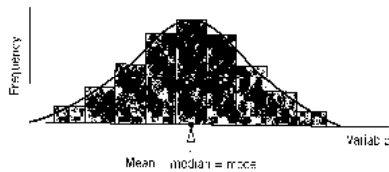
---

---

---

---

## Shpërndarje normale



9/11/2018

E. Petreli

---

---

---

---

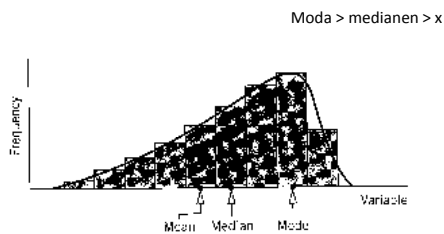
---

---

---

---

## Pjerrësi majtas



9/11/2018

E. Petreli

---

---

---

---

---

---

---

---

BOX-PLOTT

Koha e shkrimit tek femrat (sek)

14171819202129

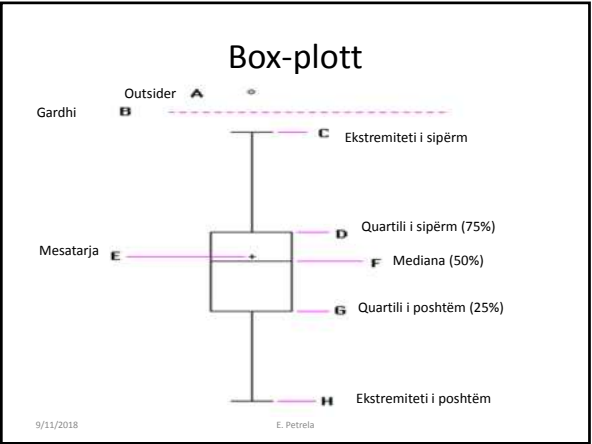
151718192022

161718192023

161718202024

171818202124

| Emri                                    | Formula                                                                      | Vlera |
|-----------------------------------------|------------------------------------------------------------------------------|-------|
| Quartili i sipërm (Qs)                  | Percentili 75                                                                | 20    |
| Quartili i poshtëm (Qp)                 | Percentili 25                                                                | 17    |
| Rangu interkuartil (RI)                 | $Qs - Qp$                                                                    | 3     |
| Hapi (H)                                | $1.5 \times QI$                                                              | 4.5   |
| Kufiri i sipërm i gardhit të brendshëm  | $Qs + 1 H$                                                                   | 24.5  |
| Kufiri i poshtëm i gardhit të brendshëm | $Qp - 1 H$                                                                   | 12.5  |
| Kufiri i sipërm i gardhit të jashtëm    | $Qs + 2 H$                                                                   | 29    |
| Kufiri i poshtëm i gardhit të jashtëm   | $Qp - 2 H$                                                                   | 8     |
| Ekstremiteti i sipërm                   | Vlera më e madhe poshtë kufirit të sipërm të gardhit të brendshëm            | 24    |
| Ekstremiteti i poshtëm                  | Vlera më e vogël mbi vlerën më të ulët të gardhit të brendshëm               | 14    |
| Vlerat jashtë (outsider)                | Një vlerë përtej gardhit të brendshëm, por jo më e madhe se gardhi i jashtëm | 29    |
| Vlera e largët                          | Një vlerë përtej gardhit të jashtëm                                          | s'ka  |



---

---

---

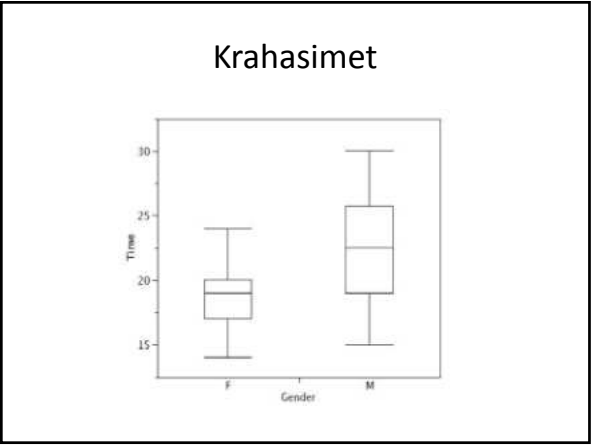
---

---

---

---

---



---

---

---

---

---

---

---

---

• është dhënë vargu i vlerave si më poshtë:

2, 4, 4, 10, 18, 19, 29, 30, 31, 44, 44, 49, 51

Bëni shpërndarjen sipas trungut dhe gjethes.

---

---

---

---

---

---

---

---

### Quartili i sipërm

- kuartilin e 75°, (kuartili i sipërm)  $Q_s$
- 

$$Q_s = \frac{3}{4} (n+1)$$

---

---

---

---

---

---

---

### Kuartili i poshtëm

Kuartilin e 25° ose kuartilin e poshtëm ( $Q_p$ )

$$Q_p = \frac{1}{4} (n+1)$$

---

---

---

---

---

---

---

### Mediana

- Mediana (50% vlerave në varg)

$$Me = \frac{1}{2} (n+1)$$

---

---

---

---

---

---

---

## Rangu interkuartil

- Rangu interkuartil=  $Q_s - Q_p$ .
- Rangu kompleksiv =  $51 - 2 = 49$ .

---

---

---

---

---

---

---

## Paraqitjet grafike për të dhëna diskrete

- Bar –diagrami;
- Diagrami me sipërfaqe;
- Hartogrami dhe hartëdiagrami.

9/11/2018

E. Petrela

---

---

---

---

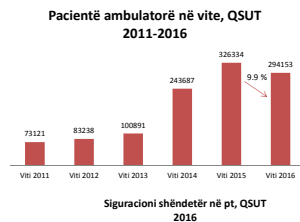
---

---

---

## PARAQITJA GRAFIKE E TË DHËNAVE CILËSORE

a) bar-diagram dhe



b) diagrami me sipërfaqe (pie-chart).



9/11/2018

E. Petrela

---

---

---

---

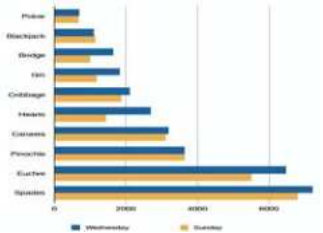
---

---

---

## Krahasimi mes shpërndarjeve

- Bar-diagrami është i shkëlqyer për të ilustruar dallimet midis dy shpërndarjeve.




---

---

---

---

---

---

---

---

## Disa gabime grafike për t'u shmangur

Mos i shtoni grafikëve veçori që nuk ndihmojnë për të përcjellë informacionin e tyre.

Psh.: grafikët 3-D nuk janë aq efektive sa 2-D.

---

---

---

---

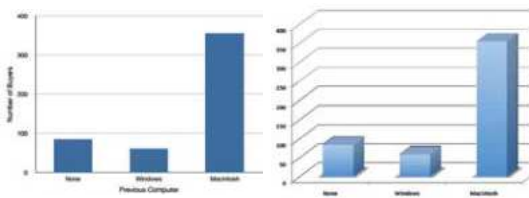
---

---

---

---

## 2D vs 3D




---

---

---

---

---

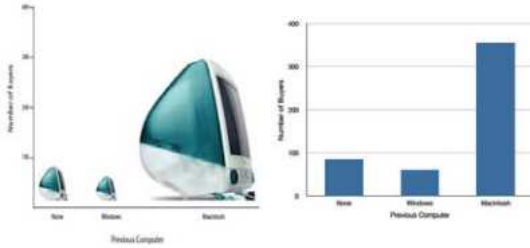
---

---

---



## Mos përdorni figura (lie factors)



---

---

---

---

---

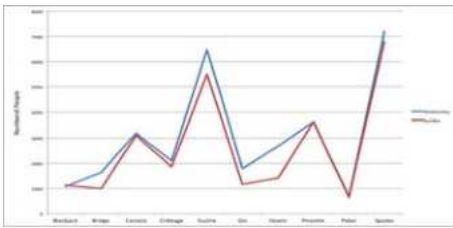
---

---

---

## Kujdes!

- është një gabim serioz të përdoret një grafik vijor, kur boshti X përmban ndryshore thjesht cilësore.



---

---

---

---

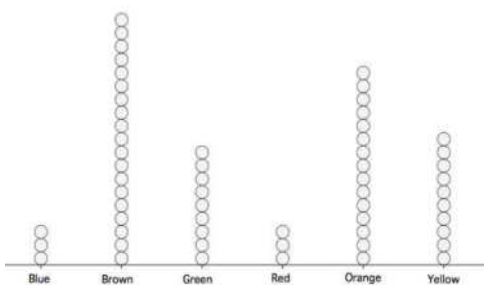
---

---

---

---

## Grafiku me pika



---

---

---

---

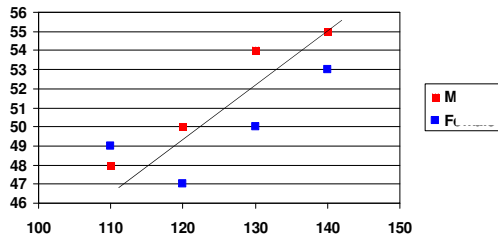
---

---

---

---

## Scatter-diagram



11/09/2018

E. Petrella,

---

---

---

---

---

---

---

---

## Hartogrammi



9/11/2018

E. Petrella

---

---

---

---

---

---

---

---

## PROBABILITETI

Daniel e përkufizon kështu probabilitetin: Nëse një proces përsëritet “N” herë, në një numër të madh rastesh dhe rezultatet e një ngjarje me karakteristikë A përsëriten “m” herë, atëherë frekuenca relative e ndodhjes së A-së, duhet të jetë afërsisht e barabartë me probabilitetin e A-së. Nëse ne përdorim simbolin “p” për të treguar probabilitetin, atëherë probabiliteti i E-së është:

$$P(A) = \frac{m}{N}$$

Para se të flasim për probabilitetin, është e nevojshme të japim disa përkufizime dhe të sqarojmë disa simbole:

**Prova** është eksperimenti që zhvillohet në kushtet e rastësisë, rezultati përfundimtar i të cilës, nuk mund të dihet asnjëherë para realizimit të provës.

**Ngjarje** janë të gjitha rezultatet e mundshme të provës. Ngjarjet janë të fundme (të numërueshme), ose të pafundme (të panumërueshme).

Mbledhja e të gjitha ngjarjeve të një eksperimenti quhet hapësirë e provës, dhe shënohet me **S**. Psh: në hedhjen e monedhës, hapësira e provës është:

$$S = \{\text{stemë, ana tjetër e stemës}\}$$

Çdo element i hapësirës së provës quhet pikë e provës.

Hapësira e provës për një eksperiment mund të ilustrohet përmes një diagrami Venn ose përmes diagramit pemë “tree diagram”.

Venn diagrami është një pikturë (që mund të jetë një figurë gjeometrike si p.sh.: një katërkëndor, një sipërfaqe, ose rreth), e cila pasqyron të gjitha ngjarjet e mundshme të një eksperimenti. Në figurën 6 paraqitet hedhja e zarit me anë të diagrami Venn.

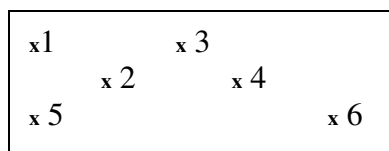


Figura 6. Diagrami Venn

Diagramë pemë paraqet eksperimentin me një pikë (pika E në **figurën 7**), nga ku dalin degë. Secila “degë” e pemës prezanton një ngjarje të mundshme të provës. Pra, në një diagramë pemë kemi aq degë, sa kemi dhe rezultate të mundshme të provës. Kështu në hedhjen e zarit, nga origjina E dalin gjashtë degë që i korrespondojnë ngjarjeve 1, 2, 3, 4, 5 dhe 6 të hedhjes së zarit (fig. 7).

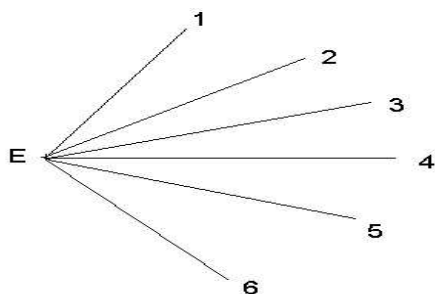


Figura 7. Diagrami pemë

Diagrami Venn dhe diagrami pemë, na ndihmojmë ne të kuptojmë konceptin e probabilitetit, duke bërë një paraqitje vizuale te dukurisë në studim.

### Ushtrim 1.

Vizatoni diagramën Venn dhe diagramën pemë në eksperimentin e hedhjes së monedhës.

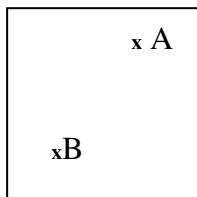
#### Zgjidhje:

Në këtë eksperiment kemi dy ngjarje të mundshme: stemë (A), ana tjetër e stemës (B). Pra, hapësira e provës është:

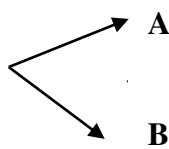
$$S = \{A, B\}$$

Për të vizatuar diagramin Venn, ne vizatojmë një katërkëndor, brenda të cilit do të vendosim dy pika që nuk janë gjë tjetër, veçse dy ngjarjet e mundshme A dhe B. Katërkëndori quhet S, sepse paraqet hapësirën e provës (fig.8a).

8a



S



8b

Për të vizatuar diagramin pemë, vizatojmë dy degë, që nisen nga e njëjta pikë dhe që pasqyrojnë dy ngjarjet e mundshme të eksperimentit. Në fund të degëve shënohen ngjarjet finale të provës, A dhe B (fig.8b).

Formalizimi i ndodhjes se dy ngjarjeve:

$A \cup B \rightarrow$  nënkupton ngjarjen që ndodh kur ndodh ngjarja A ose ngjarja B.

$A \cap B \rightarrow$  nënkupton ngjarjen që ndodh vetëm kur ndodhin të dy ngjarjet A dhe B njëkohësisht.

$A^c$  ose  $(A^c) \rightarrow$  ngjarja që ndodh kur nuk ndodh ngjarja A, ose e thënë ndryshe, ngjarja e kundërt e ngjarjes A.

$\{ \}$  → bashkësia e ngjarjeve të sigurta.

$\emptyset$  → bashkësia boshe. Kur  $A \cap B = \emptyset$ , atëherë ngjarjet janë të papajtueshme.

## LLOGARITJA E PROBABILITETIT

**Probabiliteti**, i cili na jep mundësinë e ndodhjes së një ngjarje, shënohet me  $P$ .

**Probabiliteti** është një masë numerike e mundësisë që ka një ngjarje specifike për të ndodhur.

### 5.2 Vetite elementare të probabilitetit

1. Probabiliteti i një ngjarjeje është gjithmonë një numër pozitiv, që merr vlerat nga 0 në 1.

$$0 \leq P(A) \leq 1$$

Për një ngjarje të pamundur  $M$   $P(M) = 0$

Për një ngjarje të sigurt  $C$   $P(C) = 1$

2. Shuma e probabiliteteve të të gjitha ngjarjeve të mundshme të një prove (eksperimenti) është e barabartë me një.

Kjo është vetia e mjaftueshmërisë, e cila mbështet faktin që studjuesi gjatë realizimit të një eksperimenti, duhet të lejojë të ndodhin të gjitha ngjarjet e mundshme të provës, probabiliteti total i të cilave është i barabartë me një.

$$\sum P(A_i) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$$

P.sh.: në eksperimentin e hedhjes së monedhës:

$$P(\text{A-stemë}) + P(\text{B-ana e kundërt e stemës}) = 1$$

### TRE PËRNGJASIME KONCEPTUALE MBI PROBABILITETIN.

Tre përngjasime konceptuale mbi probabilitetin janë a) probabiliteti klasik; b) koncepti i frekuencës relative në probabilitet dhe c) koncepti i probabilitetit subjektiv.

#### 1. PROBABILITETI KLASIK

Ngjarjet, të cilat kanë të njëjtin probabilitet ndodhjeje quhen ngjarje barazishtë të ngjashme.

Lidhur me rregullin e probabilitetit klasik, probabiliteti i ndodhjes së një ngjarje të thjeshtë është i barabartë me një, pjesëtuar me numrin e të gjitha ngjarjeve të mundshme të provës. Kjo është e qartë, sepse shuma e probabiliteteve për të gjitha ngjarjet në një eksperiment është 1 dhe të gjitha ngjarjet finale janë barazishtë të ngjashme. Në të kundërt, probabiliteti i një ngjarje të përbërë është i barabartë me numrin e ndodhjes së ngjarjes në fjalë, pjesëtuar me numrin e të gjitha ngjarjeve të mundshme të provës.

Ngjarje e thjeshtë quhet ngjarja që ka një dhe vetëm një rezultat final në një eksperiment; zakonisht shënohet me  $Ei$ .

Ngjarje e përbërë quhet ngjarja që ka më shumë se një rezultat final në një eksperiment; zakonisht shënohet me  $Ai$ .

Nisur nga kjo rregulli i probabilitetit klasik paraqitet në të dy situatat më poshtë:

$P(Ei) = 1 / \text{numrin total të të gjitha ngjarjeve të mundshme të provës.}$

$P(Ai) = \text{numrin e ndodhjes së ngjarjes } A / \text{numrin total të të gjitha ngjarjeve të mundshme të provës.}$

Shembull:

- a) Sa është probabiliteti për të rënë stemë, gjatë eksperimentit të hedhjes së monedhës:

$$P(\text{steme}) = \frac{1}{2}$$

- b) Sa është probabiliteti që gjatë eksperimentit të hedhjes së kubit, të na bjerë një numër çift?

$$A = \{2, 4, 6\}$$

Kështu:

$$P(A) = \frac{3}{6} = \frac{1}{2} = 0,5$$

## 2. KONCEPTI I FREKUENCËS RELATIVE NË PROBABILITET

Supozojmë se duam të llogarisim probabilitetet më poshtë:

1. Sa është probabiliteti që autobusi tjetër, që do të vijë në stacion të jetë i verdhë?
2. Sa është probabiliteti që fëmija tjetër që do të lindë në maternitet të jetë vajzë?
3. Sa është probabiliteti që një familje e zgjedhur rastësisht të ketë dy makina në përdorim?

Në këto raste nuk mund të përdorim rregullin e probabilitetit klasik, pasi ngjarje të ndryshme në eksperimentet më lart nuk janë barazisht të ngjashme. P.sh.: autobusi që do të vijë në stacion mund të jetë ose jo i verdhë. Të dy ngjarjet “I verdhë” dhe “jo i verdhë” nuk janë barazisht të ngjashme, ndaj, për të llogaritur probabilitetin e tyre ne duhet ta përsërisim eksperimentin disa herë (në një numër të madh herësh) me qëllim që të merret një rezultat i përafërt me vlerën e vërtetë.. Kjo metodë e përafërimit të probabiliteteve të një ngjarje është quajtur koncepti i frekuencës relative të probabilitetit. Meqë frekuencat relative përcaktohen duke realizuar eksperimentin, probabiliteti i llogaritur duke përdorur frekuencat relative mund të ndryshojë sa herë që në përsërisim eksperimentin. Sidoqoftë, një gjë është e sigurt, variacioni i përqindjes së autobusave të verdhë do të jetë i vogël, n.q.se madhësia e mostrës është e madhe. Nisur nga sa thamë më lart, mund të pohojmë se:

*Në një eksperiment që përsëritet  $n$  herë dhe ku ngjarja  $A$  haset  $f$  herë, atëherë duke j'u referuar konceptit të frekuencës relative të probabilitetit, kemi:*

$$P(A) = \frac{fr}{n}$$

**Shembull:**

Në një galanteri dhjetë në 50 çanta të zgjedhura rastësisht janë të verdha. Sa është probabiliteti që çanta tjetër, që do të prodhohet, të jetë e verdhë, duke shtuar që prodhimi i çantave të verdha bëhet në mënyrë të rastësishme?

$n = 50, f = 10$

$P(\text{çantë e verdhë}) = 10 / 50 = 0.2$

Duhet të kemi parasysh që frekuencat relative nuk janë probabilitete, por probabilitete të përafërta. Por, n.q.se ne do ta përsërisim provën shumë herë, probabilitet e përafërta do t'i afrohen shumë probabilitetit aktual.

Pikërisht ky ligj quhet *ligji i numrave të mëdhenj* që thotë:

*Në qoftë se një eksperiment përsëritet disa herë, probabiliteti i një ngjarje, i përfshirë nga frekuencat relative, i afrohet shumë probabilitet aktual ose teorik të ndodhjes së saj.*

### 3. PROBABILITETI SUBJEKTIV

1. Sa është probabiliteti që Tirona të fitojë kupën e Kampioneve?

Këto dhe të tjera eksperimente nuk kanë ngjarje barazishtë të ngjashme, njëkohësisht nuk mund të llogariten me frekuencën relative. Në këtë rast nëse “Tirona do ta fitojë kupën e kampioneve” bazohet jo vetëm në gjykimin, eksperiencën, apo besimin që ka skuadra e Tironës për të fituar, por edhe nga përgatitja e saj, nga gjendja e skudrës kundërshtare, etj. Në këtë rast kemi të bëjmë me probabilitet subjektiv.

*Probabilitet subjektiv i një ngjarje, është probabiliteti i bazuar në gjykimin tonë subjektiv, në eksperiencën, informacionin dhe në besim.*

Probabiliteti subjektiv zgjidhet në mënyrë arbitrare. Ai zakonisht influencohet nga biaset (gabimet e mundshme).

### NGJARJET PËRJASHTUESE

*Ngjarje përjashtuese quhen ngjarjet të cilat nuk mund të ndodhin njëkohësisht.*

Në një eksperiment, ngjarjet finale janë përjashtuese, sepse një dhe vetëm një prej tyre mund të ndodhë gjatë kryerjes së një eksperimenti. P.sh.: marrim rastin e hedhjes së dy monedhëve. Ngjarjet e mundshme janë:

{AA, AB, BA, BB}.

Të gjitha këto ngjarje janë përjashtuese, sepse vetëm njëra prej tyre mund të ndodhë kur ne hedhim dy monedha.

**Shembull:**

Mendoni hedhjen e kubit. Shenojmë me:

A ngjarjen për të rënë numër tek = { 1, 3, 5 }

B ngjarjen për të rënë numër çift = { 2, 4, 6 }

C ngjarjen për të rënë një numër më i vogël se 5 = { 1, 2, 3, 4, 5 }.

A janë ngjarjet A dhe B përjashtuese? A janë ngjarjet A dhe C përjashtuese?

**NGJARJET E PAVARURA**

Ngjarjet quhen të pavarura atëherë kur ndodhia e njëres ngjarje nuk ndikon në probabilitetin e ndodhjes së ngjarjes tjetër. A dhe B quhen të pavarura n.q.se :

$$P(A/B) = P(A) \quad \text{ose} \quad P(B/A) = P(B)$$

Ngjarjet quhen të varura atëherë kur ndodhia e njëres ngjarje ndikon në probabilitetin e ndodhjes së ngjarjes tjetër. A dhe B quhen të varura n.q.se :

$$P(A/B) \neq P(A) \quad \text{ose} \quad P(B/A) \neq P(B)$$

**Shembull:**

Në tabelën më poshtë, shihen rezultatet e votimit për të zgjedhur shefin e shërbimit në një klinikë të spitalit në QSUT. Sa është probabiliteti i ngjarjes për të qenë “femër (F)” dhe “pro (A)”?

Tab. Paraqitja e rezultateve të votimit sipas gjinisë

|             | Pro (A) | Kundër (B) | Totali |
|-------------|---------|------------|--------|
| Femra (F)   | 4       | 45         | 49     |
| Meshkuj (M) | 15      | 36         | 51     |
| Totali      | 19      | 81         | 100    |

$$P(A) = 19/100 = 0.19 \quad P(B) = 81/100 = 0.81$$

Probabiliteti me kusht është probabiliteti që një ngjarje mund të ndodhë ndërkohë, që ngjarja tjetër ka ndodhur. Kështu probabiliteti i ngjarjes A ndaj B paraqitet si  $P(A/B)$ .

Në rastin tonë probabiliteti për të qenë femër dhe pro është:

$$P(F/A) = 4/19 = 0.2105$$

Që ngjarjet F dhe A të jenë të pavarura duhet që:

$$P(F/A) = P(F)$$



Probabiliteti për të qenë femër është:

$$P(F) = 40/100 = 0.4$$

Siç shihet :

$$P(F/A) \neq P(F)$$

Ndaj ngjarjet F dhe A janë të varura.

### KUJDES!

*Këtu gjykojmë se janë me vend dy sugjerime:*

1. Dy ngjarje janë ose përjashtuese ose të pavarura. Përjashtim nga ky rregull kemi vetëm në rastin kur një nga të dy ngjarjet ka probabilitet zero.
  - a) ngjarjet përjashtuese janë gjithmonë të pavarura.
  - b) ngjarjet e pavarura asnjëherë nuk janë përjashtuese.
2. Ngjarjet e varura mundet ose jo të jenë përjashtuese.

### NGJARJET KOMPLEMENTARE

Dy ngjarje të pavarura të cilat përbëjnë të gjitha rezultatet e mundshme të provës në një eksperiment, quhen ngjarje komplementare. Dy ngjarje komplementare janë gjithmonë përjashtuese. Komplementi i ngjarjes A, shënohet me  $\bar{A}$  dhe lexohet “A bar” ose “A komplement”. Meqënëse dy ngjarje komplementare përfaqësojnë të gjitha ngjarjet e mundshme në një eksperiment, shuma e tyre është 1 (fig 5).

Kështu:

$$P(A) + P(\bar{A}) = 1 \Rightarrow P(A) = 1 - P(\bar{A}) \text{ dhe } P(\bar{A}) = 1 - P(A)$$

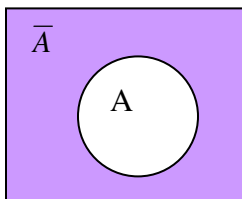


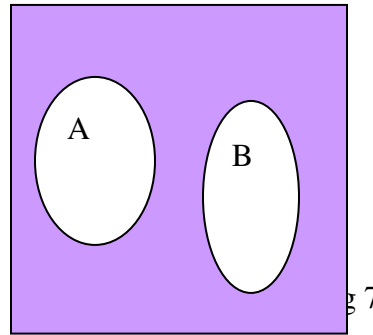
Fig 9 Praraqitja grafike e ngjarjeve komplementare

**Prerja e ngjarjeve** jepet si rezultat i përbashkët i të dyja ngjarjeve që mund të ndodhin njëkohsisht. Probabiliteti i tyre jepet si probabilitet “lidhës” dhe simbolikisht paraqitet (fig.10)



**Fig 10**  $P(A \cap B)$

Probabiliteti lidhës për dy ngjarje të papajtushme është zero.  
(fig 11)  
 $P(A \cap B) = 0$



**Fig.11**  $P(A \cap B) = 0$

### **Llogaritja e probabilitetit të ndodhjes së një ngjarjeje.**

Probabiliteti i ndodhjes së një ngjarjeje A quhet raporti mes numrit të ndodhjes së ngjarjes A (n) dhe numrit të të gjitha ngjarjeve të mundshme të provës (N).

$$P(A) = \frac{n}{N}$$

### **Probabiliteti me kusht:**

Probabiliteti i ngjarjes A me kusht ngjarjen B, është i barabartë me probabilitetin e ngjarjes A dhe  $P(A \cap B)$ , pjesëtuar me probabilitetin e ngjarjes B, kur ky i fundit është i ndryshëm nga zero.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{ku } P(B) \neq 0$$

### **Rregulli i shtesës në probabilitet:**

Për dy ngjarje të pajtueshme A dhe B që mund të ndodhin njëkohësisht probabiliteti i ndodhjes së ngjarjes A ose B ( $A \cup B$ ) jepet si probabiliteti i ngjarjes A plus probabilitetin e ngjarjes B minus probabilitetin që të dy ngjarjet të ndodhin njëkohësisht.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

### **Rregulli i shumezimit në probabilitet**

Rregulli i shumëzimit në probabilitet, rrjedh nga rregulli i probabilitetit me kusht. Për dy ngjarje A dhe B që mund të ndodhin njëkohësisht, probabiliteti i ndodhjes së tyre njëkohësisht është i barabartë me probabilitetin e ngjarjes A me kusht B, shumëzuar me probabilitetin e ngjarjes B.

$$P(A \cap B) = P(A/B) * P(B)$$

Shembull: Në tabelën më poshtë shihen rezultatet e votimit në spital.

Tab. Paraqitja e rezultateve të votimit sipas gjinisë

|           | Pro (A) | Kundër (B) | Totali |
|-----------|---------|------------|--------|
| Femra (F) | 4       | 45         | 49     |

|             |    |    |     |
|-------------|----|----|-----|
| Meshkuj (M) | 15 | 36 | 51  |
| Totali      | 19 | 81 | 100 |

$$P(A) = 19/100 = 0.19 \quad P(B) = 81/100 = 0.81$$

Po t'i referohemi shembullit më lart, gjeni

1. Sa është probabiliteti i ngjarjes për të qenë “femër (F)” ?
2. Sa është probabiliteti që personi i zgjedhur të jetë mashkull me kusht që të ketë votuar kundër?
3. Sa është probabiliteti që personi i zgjedhur të jetë femër ose të ketë votuar pro?
4. Sa është probabiliteti që personi i zgjedhur të jetë mashkull dhe të ketë votuar pro?

### Zgjidhje:

1. Sa është probabiliteti i ngjarjes për të qenë “femër (F)” ?

Dimë që probabiliteti i një ngjarje jepet nga raporti

$$P(A) = \frac{n}{N}$$

Kështu, probabiliteti për të qenë femër, llogaritet nga raporti i numrit të femrave me numrin e popullatës totale.

$$P(F) = \frac{49}{100} = 0.49$$

2. Sa është probabiliteti që personi i zgjedhur të jetë mashkull me kusht që të ketë votuar kundër?

Në këtë rast kërkohet probabiliteti me kusht, i cili llogaritet me formulën:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Duke bërë zëvendësimet e nevojshme në formulën e probabilitetit me kusht, ajo merr këtë formë:

$$P(M/B) = \frac{P(M \cap B)}{P(B)} = \frac{\frac{36}{100}}{\frac{81}{100}} = \frac{36}{81} = 0.44$$

3. Sa është probabiliteti që personi i zgjedhur të jetë femër ose të ketë votuar pro?  
Në këtë rast kemi të bëjmë me rregullin e shtesës në probabilitet. Rregulli I shtesës në probabilitet jepet me formulën:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Duke e përshtatur me të dhënat tona, formula më lart do të paraqitet:

$$P(F \cup A) = P(F) + P(A) - P(F \cap A) = \frac{49}{100} + \frac{19}{100} - \frac{4}{100} = \frac{64}{100} = 0.64$$

4. Sa është probabiliteti që personi i zgjedhur të jetë mashkull dhe të ketë votuar pro?

Kjo kërkesë përkon me rregullin e shumëzimit në probabilitet, i cili është rrjedhojë e probabilitetit me kusht dhe llogaritet me formulën:

$$P(A \cap B) = P(A / B) * P(B)$$

duke bërë zëvendësimet e vlerave të ushtrimit tonë, gjejmë se probabiliteti që personi i zgjedhur të jetë mashkull dhe të ketë votuar pro, është:

$$P(M \cap A) = P(M / A) * P(A) = \frac{15}{51} * \frac{19}{100} = 0,29 * 0,19 = 0,055$$

## TEOREMA E BAYES

Testet diagnostike

Sensibiliteti dhe specificiteti

Kurbat Roc

Llogaritja e prevalences

Risku relative dhe risku absolute

Për të sqaruar më mirë teoremën e Bayes i referohemi shembullit të mëposhtëm, ku një popullatë, bazuar në gjendjen e punësimit, ndahet në tre kategori: të punësuar, të papunë dhe pensioniste, në të cilat janë përcaktuar, respektivisht, dëmtimet e dëgjimit.

Tabela Paraqitja e dëmtimeve të dëgjimit sipas pozicionit të punës

| Pozicioni   | Nr i individeve | Demtime te degjimit |
|-------------|-----------------|---------------------|
| Te punesuar | 98917           | 552                 |
| Te papune   | 7462            | 27                  |
| Pensioniste | 56778           | 368                 |
| Shuma       | 163157          | 947                 |

Shënojmë me  $A_1$  ngjarjen për të qenë i punësuar, me  $A_2$  ngjarjen për të qenë i papunë dhe me  $A_3$  ngjarjen për të qenë në pension. Duke supozuar që të dhënat janë të mjaftueshme për të llogaritur probabilitetin (pra popullata është përfaqësuese), atëherë kemi:

$$P(A_1) = \frac{98917}{163157} = 0,6063$$

$$P(A_2) = \frac{7462}{163157} = 0,0457$$

$$P(A_3) = \frac{56778}{163157} = 0,3480$$

Nëse me  $B$  shënojmë ngjarjen që një person, i përfshirë në studim të jetë i punësuar, i papunë ose në pension atëherë mund të shkruajmë:

$$B = A_1 \cup A_2 \cup A_3$$

Meqënëse të tre ngjarjet janë përjashtuese, mund të aplikohet parimi i shumës në probabilitet, sipas të cilit mund të shkruajmë:

$$\begin{aligned}
 B &= A_1 \cup A_2 \cup A_3 = \\
 &= P(A_1) + P(A_2) + P(A_3) = \\
 &= 0,6063 + 0,0457 + 3480 = 1.000
 \end{aligned}$$

Kur shuma e probabiliteteve për të gjitha ngjarjet përjashtuese është një, atëherë thuhet se janë paraqitur të gjitha ngjarjet e mundshme, pra që nuk ka rezultate të tjera të mundshme. Cdo person i përfshirë në këtë studim bën pjesë vetëm në njërin prej tre grupeve.

Nëse shënojmë me D ngjarjen që një person të ketë dëmtime të dëgjimit, atëherë probabiliteti që një person i zgjedhur rastësisht nga kjo popullatë të jetë me dëmtim të dëgjimit është:

$$P(D) = \frac{947}{163157} = 0,0058$$

Duke analizuar secilën kategori të punësimit më vete, kemi:

$P(D/A_1)$  = probabiliteti që personi të ketë dëmtime të dëgjimit me kusht që të jetë i në punë)

$$P(D/A_1) = \frac{552}{98917} = 0,0056$$

$P(D/A_2)$  = probabiliteti që personi të ketë dëmtime të dëgjimit me kusht që të jetë i pa punë)

$$P(D/A_2) = \frac{27}{7462} = 0,0036$$

$P(D/A_3)$  = (probabiliteti që personi të ketë dëmtime të dëgjimit me kusht që të jetë në pension)

$$P(D/A_3) = \frac{368}{56778} = 0,0065$$

Duke qenë se D tregon ngjarjen që një person të ketë dëmtime të dëgjimit, kjo mund të shprehet si bashkimi i tre ngjarjeve përjashtuese  $A_1 \cap D$  (i punësuar dhe me defekte të dëgjimit),  $A_2 \cap D$  (i papunë dhe me defekte të dëgjimit)  $A_3 \cap D$  (pension dhe me defekte të dëgjimit) dhe këtë mund ta shprehim:

$$D = (A_1 \cap D) \cup (A_2 \cap D) \cup (A_3 \cap D)$$

Personat me dëmtime të dëgjimit mund të bëjnë pjesë vetëm në njërin prej tre grupeve dhe meqë ngjarjet janë përjashtuese, aplikojmë rregullin e shumës në probabilitet:

$$P(D) = P(A_1 \cap D) \cup P(A_2 \cap D) \cup P(A_3 \cap D)$$

$$P(D) = P(A_1 \cap D) + P(A_2 \cap D) + P(A_3 \cap D)$$

Duke zbatuar rregullin e shumezimit per ngjarje te papajtueshme, për secilën nga kategoritë më vete, mund të shkruajmë:

$$\begin{aligned}
 P(D) &= P(A_1 \cap D) + P(A_2 \cap D) + P(A_3 \cap D) = \\
 &= P(A_1) * P(D / A_1) + P(A_2) * P(D / A_2) + P(A_3) * P(D / A_3) = \\
 &= (0,6063)(0,0056) + (0,0457)(0,0036) + (0,0348)(0,0065) = \\
 &= 0,0058
 \end{aligned}$$

Kjo është vlera që llogaritem që në fillim të cilën e quajtmë probabiliteti për të patur dëmtime të dëgjimit, vlerë të cilën e gjetëm si:

$$P(D) = \frac{947}{163157} = 0,0058$$

Shprehja më e plotë:

$$P(D) = P(A_1) * P(D / A_1) + P(A_2) * P(D / A_2) + P(A_3) * P(D / A_3) \quad (1)$$

mund të përdoret në rastet kur ne nuk mund ta gjejmë direkt vlerën e probabilitetit.

Supozojme tani se duam te gjejmë probabilitetin  $P(A_1/D)$ , pra probabilitetin që një person i zgjedhur rastësisht nga kjo popullatë të jetë i punësuar dhe të ketë dëmtime të dëgjimit. Bazuar në rregullin e shumezimit në probabilitet kemi:

$$P(A_1 \cap D) = P(D) * P(A_1 / D) \quad (a)$$

nga ku kemi:

$$P(A_1 / D) = \frac{P(A_1 \cap D)}{P(D)} \quad (b)$$

Duke zëvendësuar (a) tek barazimi (b) kemi:

$$P(A_1 / D) = \frac{P(A_1) * P(A_1 / D)}{P(D)} \quad (c)$$

Duke e zëvendësuar (c) në atë që thamë më sipër tek ekuacioni (1)

$$P(D) = P(A_1) * P(D / A_1) + P(A_2) * P(D / A_2) + P(A_3) * P(D / A_3) \quad (1)$$

kemi:

$$P(A_1 / D) = \frac{P(A_1) * P(A_1 / D)}{P(A_1) * P(D / A_1) + P(A_2) * P(D / A_2) + P(A_3) * P(D / A_3)}$$

Kjo është shprehja e teoremës së Bayes.

Duke bërë zëvendësimet e vlerave kemi:

$$P(A_1 / D) = \frac{0,6063 \times 0,0056}{(0,6063 \times 0,0056) + (0,0457 \times 0,0036) + (0,3480 \times 0,0065)} = 0,583$$

Probabiliteti që një person të jetë i punësuar, me kusht që të ketë dëmtime të dëgjimit është 0,583. Në këtë shembull të vecantë ky probabilitet mund të llogaritet dhe drejtpërdrejt nga tabela; ne 947 personat me dëmtime të dëgjimit, 552 prej tyre janë me dëmtime të dëgjimit, ndaj kemi:

$$P(A_1 / D) = \frac{552}{947} = 0,583$$

Në përfundim mund të themi se nëse ngjarjet  $A_1, A_2, A_3, \dots, A_n$  janë ngjarje përjashtuese dhe janë të gjitha ngjarjet e mundshme, atëherë:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1,$$

Atëhere teorema e Bayes kornfirmon që:

$$P(A_i / B) = \frac{P(A_i) * P(A_i / B)}{P(A_1) * P(B / A_1) + P(A_2) * P(B / A_2) + P(A_3) * P(B / A_3) + \dots + P(A_i) * P(B / A_i)}$$

për cdo "i", që merr vlerat  $1 \leq i \leq n$ .



Teorema e Bayes është shumë e rëndësishme, sepse bën të mundur ri-llogaritjen e probabilitetit, bazuar në një informacion të ri.

### TESTET diagnostike

Teorema e Bayes shpesh përdoret për vlerësimin e testeve diagnostike ose në screening. Screening-u është aplikimi i një testi në persona që nuk kanë shfaqur akoma asnjë shenjë të sëmundjes, me qëllim llogaritjen e probabilitetit që këta persona të preken nga sëmundja në fjalë. Ata që rezultojnë pozitiv ndaj një testi, kanë probabilitet mjaft të madh për të patur sëmundjen dhe këta duhet t'i nënshtrohen procedurave specifike diagnostikuese. Screening-u përdoret nga punonjësit e shëndetësisë në rastet kur identifikimi paraprak i sëmundjes mund të sjellë si pasojë një prognozë më të mirë të sëmundjes, për individin apo për popullatën në tërësi. Teorema e Bayes përdoret pikërisht për vlerësimin e lidhjeve të dyshimta.

### SENSIBILITETI dhe SPECIFICITETI

Tabela

| Rezultatet e testit | Sëmundje |       | Shuma         |
|---------------------|----------|-------|---------------|
|                     | Po       | Jo    |               |
| Test pozitiv        | a        | b     | a + b         |
| Test negativ        | c        | d     | c + d         |
| Shuma               | a + c    | b + d | a + b + c + d |

**Sensibiliteti** i një testi është probabiliteti që një test të rezultojë pozitiv në persona të cilët realisht janë të sëmurë.

$$P(\text{test pozitiv} / \text{te sëmurë}) = \frac{a}{a + c}$$

**Specificiteti** i një testi është probabiliteti që rezultati i një testi të jetë negativ, për ata persona që vërtet janë të shëndoshe.

$$P(\text{test negativ} / \text{te shëndoshe}) = \frac{d}{b + d}$$

### Risku Relativ dhe Odds Ratio (raporti i gjasave)

Risku relativ (RR) përdoret në rastet kur duam të krahasojmë probabilitetet për të ndodhur sëmundja në dy situata të ndryshme ose në dy grupe të ndryshme.

RR është probabiliteti që një individ, i cili i perket grupit që ekspozohet ndaj një faktori risk, të zhvillojë sëmundjen, krahasuar me një individ të grupit që nuk ekspozohet ndaj një faktori risk, por që zhvillon sëmundjen në fjalë. Me fjalë të tjera, RR paraqet probabilitetin për t'u

sëmure në grupin e të ekspozuarve pjesëtuar me probabilitetin që kanë të paekspozuarit, për t'u sëmure nga e njëjta sëmundje.

$$RR = \frac{P(\text{te semure} / \text{te ekspozuar})}{P(\text{te semure} / \text{te paekspozuar})}$$

$RR = 1$  tregon se probabiliteti për t'u sëmure, si në grupin e të ekspozuarve dhe në atë të të paekspozuarve ndaj faktorit/faktorëve risk, është i njëjtë, pra s'ka lidhje mes faktorit risk dhe sëmundjes.

$RR > 1$  tregon se në të ekspozuarit ka një probabilitet më të madh për t'u sëmure, krahasuar me të paekspozuarit.

$RR < 1$  tregon që ekziston një risk më i vogël për t'u sëmure në personat e ekspozuar.

Një masë tjetër që përdoret për të krahasuar probabilitetet është Odds Ratio (OD).

Nëse një ngjarje ndodh me probabilitet "p", Odds ratio në favor të ngjarjes është  $p / (1-p)$ .

Kështu, nëse  $p = 1/2$ , OD është  $1/2 / 1/2 = 1$ . Në këtë rast ngjarja mund të verifikohet ose jo me të njëjtin probabilitet.

Nëse  $p = 2/3$ ,  $OD = 2/3 / 1/3 = 2$  në 1, pra probabiliteti që ngjarja të verifikohet është dy herë më i madh, sesa ngjarja të mos verifikohet.

Odds ratio përkufizohet si odds në favor të sëmundjes në individë të ekspozuar, ndaj odds të sëmundjes në individë të paekspozuar.

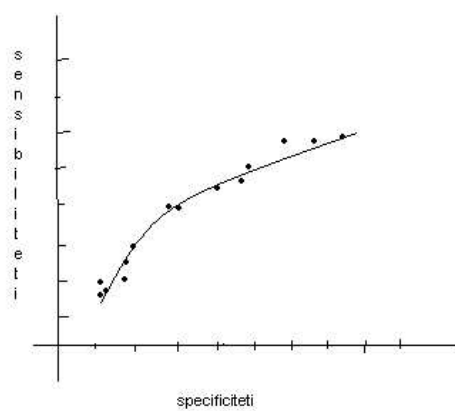
$$OD = \frac{P(\text{te semure} / \text{te ekspozuar}) / [1 - P(\text{te semure} / \text{te ekspozuar})]}{P(\text{te semure} / \text{te paekspozuar}) / [1 - P(\text{te semure} / \text{te paekspozuar})]}$$

RR dhe OD janë dy masa që tentojnë të shpjegojnë të njëjtin fenomen; edhe pse RR duket me intuitiv, OD ka më shumë prioritet statistikore.

## Kurbat ROC

Lidhja mes sensibilitetit dhe specifikitetit ilustron përmes kurbës ROC. Kurba ROC është **një grafik linear**, i cili tregon probabilitetin e një rezultati vërtet pozitiv në funksion të probabilitetit të një rezultati fals pozitiv, për një seri pikash cut-off.

Një test sensibil është i preferueshëm atëherë kur mungesa e zbulimit të një sëmundje, ka pasoja të rrezikshme dhe një test specifik është i rëndësishëm, atëherë kur rezultati fals pozitiv është i dëmshëm.



Graf. Kurba Roc

statistika

Elizana PETRELA (ZAIMI)

---

---

---

---

---

---

---

---

Probabiliteti



---

---

---

---

---

---

---

---

Ceshtjet qe do te diskutojme:

- Perkufizimi i probabilitetit
- Terminologji dhe simbolike
- Vetite elementare te probabilitetit
- Rregulli i shteses (mbledhjes) ne probabilitet
- Rregulli i probabilitetit me kusht
- Rregulli i shumezimit ne probabilitet
- Testimi i hipotezes
- Vlera sinjifikante e probabilitetit

---

---

---

---

---

---

---

---



Para se te nisim kete leksion duhet te keni parasysh qe probabiliteti eshte nje nga konceptet baze dhe nder me te veshtirat.

Por kini besim, ketu do te merrni nocionet baze, me qellim qe njohurite e marra t'i perdorni per te interpretuar testet statistikore te duhura...

*Mos harroni ,Roma nuk u ndertua ne nje dite!*



---

---

---

---

---

---

---

“A pinch of probability is worth a pound of perhaps”

James Thurber

---

---

---

---

---

---

---

Secili nga ne ka nje ide intuitive se c'eshte probabiliteti. Para se te flasim per probabilitetin duhet te sqarojme disa koncepte dhe simbole.



---

---

---

---

---

---

---

Probabilitet është mundësia e ndodhjes së një dukurie (ngjarje)



$$P(A) = \frac{m}{N}$$

---

---

---

---

---

---

---

---

## Perkufizim

- Probabiliteti është mundësia e ndodhjes së një ngjarje në tërësinë e ngjarjeve të mundshme të provës.
- Shembull:
  - Në q. se ne hedhim një zar, sa është probabiliteti për të rënë numri 2?
  - Përgjigje:  $1/6$



---

---

---

---

---

---

---

---

## Prova dhe ngjarja

- **Prova** është një eksperiment që zhvillohet në kushtet e rastësisë, pra, rezultati i të cilit nuk mund të parashikohet para realizimit
- **Ngjarja** quhen të gjitha rezultatet e mundshme të provës

---

---

---

---

---

---

---

---

## Simbolika

- $A \cup B \rightarrow A$  ose  $B$
- $A \cap B \rightarrow A$  dhe  $B$
- $\bar{A}$  ( $A^c$ )  $\rightarrow$  ngjarja e kundërt e  $A$      $1 - (A)$
- $\Omega$  (omega)  $\rightarrow$  ngjarje e sigurtë (bashkësia)
- $\Phi$  (fi)  $\rightarrow$  ngjarje e pamundur
- N.q.s.  $A \cap B = \Phi \rightarrow A$  dhe  $B$  janë te papajtueshme
- $\omega$  (rro)  $\rightarrow$  një ngjarje e veçantë

---

---

---

---

---

---

---

## Te dhena te pergjithshme per Probabilitetin

**P ose Pr** (probabiliteti)

$\{ \quad \}$  tregon bashkesine e rezultateve te mundshme

Ex: Probabiliteti qe ne hedhje e zarit te bien numrat 1,3,5 eshte  $\Pr\{A\} = \Pr\{1,3,5\}$

---

---

---

---

---

---

---

## Perkufizime

- ne.q.se dy ngjarje  $A$  dhe  $B$  jane ngjarje qe nuk mund te ndodhin ne te njejten kohe ato quhen *perjashtuese*;
- ne.q.se ngjarjet  $A$  dhe  $B$  mund te ndodhin bashke, por ndodhia e njerës nuk ndikon ne ndodhine e ngjarjes tjeter, *ngjarjet quhen te pavarura*;
- ne.q.se ngjarjet  $A$  dhe  $B$  mund te ndodhin bashke, por ndodhia e njerës ndikon ne ndodhine e ngjarjes tjeter, *ngjarjet quhen te varura*.

---

---

---

---

---

---

---

## Probabiliteti



Ngjarja me interes  $A$ :  $\Pr\{A\}$  është proporcioni i  $A$  në tërë hapësirën e provës

Box = hapësirë e provës = pra tërë mundësia e ngjarjeve

Hapësira me të zeze – quhet komplement i  $A$ , të gjitha ngjarjet e tjera quhen komplement i  $A = A^c$   $\Pr\{A^c\} = 1 - \Pr\{A\}$

---

---

---

---

---

---

---

---

## Vetite e probabilitetit

- Për çdo ngjarje  $A$ , probabiliteti i ndodhjes së saj është një numër jo negativ  
 $0 \leq \Pr(A) \leq 1$

- Vetia e mjaftueshmërisë:  
Shuma e probabiliteteve të të gjitha ngjarjeve të papajtueshme të provës është e barabartë me 1,

$$\Pr(A) + \Pr(B) + \Pr(C) + \dots + \Pr(n) = 1$$

- Vetia e komplementaritetit:  
 $\Pr(A^c) = 1 - \Pr(A)$

---

---

---

---

---

---

---

---

## vazhdim

- Probabiliteti i  $A$  ose  $B$  jepet si  $\Pr(A \cup B)$ ;
- Probabiliteti  $A$  dhe  $B$  kur dy ngjarje ndodhin bashkë është  $\Pr(A \cap B)$ ; tregon prerjen e bashkësive “intersection”

---

---

---

---

---

---

---

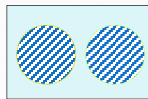
---



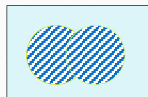
## Ilustrimi i “bashkimi” dhe “prerja”



$A \cap B$ ;



$A \cup B = \Pr(A) + \Pr(B)$ ,  
A dhe B janë të pavarura



$A \cup B = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ ,  
A dhe B janë të varura

---

---

---

---

---

---

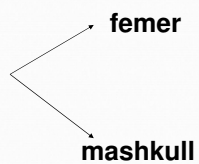
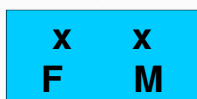
---

---

## Paraqitja grafike

• Veinn

Peme




---

---

---

---

---

---

---

---

## Rregullat ne probabilitet

- Rregulli i shteses ne probabilitet  
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Probabiliteti me kusht  
 $P(A / B) = P(A \cap B) / P(B)$ ;
- Rregulli i shumezimit  
 $P(A \cap B) = P(A / B) * P(B)$ ;

---

---

---

---

---

---

---

---

**X dhe Y jane perjashtuese?**

**po**

1.  $P(X \text{ dhe } Y) = 0$  sepse ngjarjet (X dhe Y) s'mund te ndodhin bashke,  $(X \text{ dhe } Y = X \cap Y)$
2.  $P(X \text{ ose } Y) = P(X) + P(Y)$ ; regulli i shteses

**jo**

1.  $P(X \text{ dhe } Y) > 0$
2.  $P(X \text{ ose } Y) = P(X) + P(Y) - P(X \cap Y)$

X dhe Y jane te pavarura

**po**

$P(X \text{ dhe } Y) = P(X)P(Y)$ , regulli i shumezimit

**jo**

1.  $P(X|Y) = P(X \cap Y)/P(Y)$ ; probabiliteti me kusht

---

---

---

---

---

---

---

---

## Ngjarjet e pavarura

- Dy ngjarje quhen te pavarura, nese plotesohet kushti:

$$P(A \cap B) = P(A)P(B)$$


---

---

---

---

---

---

---

---

## Tre përngjasime konceptuale mbi probabilitetin.

- Probabiliteti Klasik
- Frekuences relative ne probabilitet
- Probabiliteti Subjektiv

---

---

---

---

---

---

---

---

## Teorema e limitit qendror

|               |    |    |     |     |     |      |      |      |      |
|---------------|----|----|-----|-----|-----|------|------|------|------|
| Nr. i hedhjes | 10 | 50 | 100 | 200 | 300 | 400  | 600  | 800  | 1000 |
| Nr. Stemës    | 6  | 23 | 46  | 98  | 147 | 199  | 287  | 383  | 499  |
| % Stemës      | 60 | 46 | 46  | 49  | 49  | 49.7 | 49.5 | 49.1 | 49.9 |

- “ Në një numër shumë të madh dukurish të rastit rezultati i tyre mesatar **pushon** së qeni i rastit; ai mund të parashikohet me një shkallë të lartë saktësie”

---

---

---

---

---

---

---

---



## Probabiliteti

Ne.q.se hedhim tri monedha, sa eshte probabiliteti per te marre dy koka dhe nje steme?

u 1: percaktoni hapesiren e proves (te gjitha ngjarjet e mundshme).

u 2: probabiliteti qe te marrim dy koka dhe nje steme

Pergjigja ne slaidin tjeter

---

---

---

---

---

---

---

---

## Pergjigja

- : {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}
- hapesira e ngjarjes
  - {HHT, HTH, THH}
- Probabiliteti eshte  $\frac{3}{8}$

---

---

---

---

---

---

---

---

Ju kaluat me sukses kete  
prove!!!

---

---

---

---

---

---

---

---

## Pyetje

**P.1:** nese dy ngjarje jane perjashtuese sa eshte  
probabiliteti i prerjes se tyre = \_\_\_\_\_.

---

---

---

---

---

---

---

---

**P.1:** nese dy ngjarje jane perjashtuese sa eshte probabiliteti i  
prerjes se tyre = \_\_\_\_\_.

Zgjidhje: probabiliteti prerjes se tyre eshte **zero**.  
Ne kete rast s;ka mbivendosje.

---

---

---

---

---

---

---

---

**P.2:** Nese  $\Pr(A)$  eshte 0.79, atehere  
Komplementi I saj eshte  $\Pr(A^c) = \underline{\hspace{2cm}}$ .

---

---

---

---

---

---

---

**P.2:** Nese  $\Pr(A)$  eshte 0.79, atehere Komplementi I saj eshte  
 $\Pr(A^c) = \underline{\hspace{2cm}}$ .

**Zgjidhje:** Dime qe komplementi I nje ngjarje  
eshte  $1 - \Pr(A)$ . Keshtu,

$$\Pr(A^c) = 1 - 0.79 = 0.21$$

---

---

---

---

---

---

---

**P.3:** Jane Y dhe Z ngjarje te pavarura?  
 $\Pr(Z) = 0.05$ ;  $\Pr(Y) = 0.14$ ;  $\Pr(Y \text{ dhe } Z) = 0.007$

---

---

---

---

---

---

---

Zgjidhje: Jo – nese Y dhe Z jane te pavarura, atehere probabiliteti qe Y dhe Z te ndodhin rastesisht bashke duhet te jete i barabarte me prodhimin e probabiliteteve individuale te tyre.  $\Pr(Z) * \Pr(Y) = 0.05 * 0.14 = 0.070$ . meqenese ky prodhim I ndodhjes se bashku te te dyja ngjarjeve se bashku eshte I # nga probabiliteti I dhene (.007), atehere ngjarjet Y dhe Z nuk jane te pavarura.

---

---

---

---

---

---

---

---

**P. 4:** A, B, dhe C jane ngjarje te pavarura me vleren e probabiliteteve respektivisht 0.4, 0.28, 0.0091. Gjeni probabilitin qe te tria keto ngjarje te ndodhin rastesisht se bashku ( $A \cap B \cap C$ ).

---

---

---

---

---

---

---

---

Zgjidhje: Meqenese ngjarjet jane te pavarura, ne perdorim rregullin e shumezimit ne probabilitet, qe ngjarjet te ndodhin rastesisht se bashku

$$0.4 * 0.28 * 0.0091 = 0.0010192$$

---

---

---

---

---

---

---

---

**P.5:** A dhe B jane ngjarje perjashtuese.  $\Pr(A) = .4$  dhe  $\Pr(B) = .01$ . Sa eshte probabiliteti of  $A \cup B$  ( A ose B)?

---

---

---

---

---

---

---

---

Zgjidhje: Ne do te perdorim **regullin e shiteses ne probabilitet**.

$$\Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Meqe A dhe B jane perjashtuese,  $(A \cap B)=0$ .

Keshtu, **pergjigja eshte  $0.4 + 0.01 + 0 = 0.41$ .**

---

---

---

---

---

---

---

---

**Testimi I hipotezes**

---

---

---

---

---

---

---

---

- $H_0$  - hipoteza e zeros – (null hypothesis)

- $H_A$  - hipoteza alternative

---

---

---

---

---

---

---

## $H_0$

- Hipoteza e mungeses se ndryshimit  
(midis dy metodave te ndryshme qe perdoren nuk ka ndryshim te rendesishem)

---

---

---

---

---

---

---

## $H_A$

- Hipoteza e pranise se ndryshimit  
(midis dy metodave te ndryshme qe perdoren, ka nje ndryshim te rendesishem, qe nuk shkaktohet nga rasti, por nga faktore teresisht te matshem)

---

---

---

---

---

---

---



### Gabimet e hipotezes

|          | H0 e vertete | H0 fallse |
|----------|--------------|-----------|
| Pranohet | +            | $\beta$   |
| Terhiqet | $\alpha$     | +         |

---

---

---

---

---

---

---

### Gabimet e hipotezes

- Gabimi falls pozitiv,  $\alpha$ , gabimi i tipit pare, ndodh atehere kur ne rrezojme nje hipoteze te zeros si fallso, nderkohe qe ajo eshte e vertete.
- Gabimi falls negativ,  $\beta$ , gabimi i tipit te dyte, ndodh atehere kur ne pranojme nje hipoteze te zeros si te vertete, nderkohe qe ajo eshte fallso.

---

---

---

---

---

---

---

Ne rast se:

- $p \leq 5\%$

**Probabiliteti  
eshte  
sinjifikant**

---

---

---

---

---

---

---



---

---

---

---

---

---

---

## TEOREMA E BAYES

Prof. Dr. Elizana PETRELA Zaimi

---

---

---

---

---

---

---

## TEOREMA E BAYES

| POZICIONI             | NR I<br>INDIVIDEVE | DEMTIME TE<br>DEGJIMIT |
|-----------------------|--------------------|------------------------|
| Te punesuar ( $A_1$ ) | 98917              | 552                    |
| Te papune ( $A_2$ )   | 7462               | 27                     |
| Pensioniste ( $A_3$ ) | 56778              | 368                    |
| Shuma                 | 163157             | 947                    |

---

---

---

---

---

---

---

$$P(A_1 \text{ Te punesuar}) = \frac{98917}{163157} = 0,6063$$
$$P(A_2 \text{ Te papune}) = \frac{7462}{163157} = 0,0457$$
$$P(A_3 \text{ Pensioniste}) = \frac{56778}{163157} = 0,3480$$

---

---

---

---

---

---

---

## NGJARJET B

- B - ngjarjen qe nje person i perfshire ne studim te jete i punesuar, i papune ose ne pension ateherë:

$$B = A1 \cup A2 \cup A3$$

- Meqe te tre ngjarjet jane te perjashtuese, mund te aplikohet parimi i shumes ne probabilitet, sipas se ciles kemi:

$$\begin{aligned} P(B) &= P(A1) \cup P(A2) \cup P(A3) = \\ &= P(A1) + P(A2) + P(A3) = \\ &= 0,6063 + 0,0457 + 0,348 = 1.0000 \end{aligned}$$

## SHENOJME ME D NGJARJEN QE NJE PERSON TE KETE DEMTIME TE DEGJIMIT

Probabiliteti qe nje person i zgjedhur rastesisht nga kjo popullate te jete me demtim te degjimit eshte:

$$P(D) = \frac{947}{163157} = 0,0058$$

Duke analizuar secilen kategori te punesimit me vete,

| POZICION I                    | NR I INDIVIDIVE VE | DEMTIME TE DEGJIMIT D |
|-------------------------------|--------------------|-----------------------|
| Te punesuar (A <sub>1</sub> ) | 98917              | 552                   |
| Te papune (A <sub>2</sub> )   | 7462               | 27                    |
| Pensioniste (A <sub>3</sub> ) | 56778              | 368                   |
| Shuma                         | 163157             | 947                   |

$$\begin{aligned} P(D/A1) &= \\ &= (\text{probabiliteti qe nje person te kete demtime te degjimit me kusht qe te jete i punesuar}) = \\ &= 552 / 98917 = 0,0056 \end{aligned}$$

$$\begin{aligned} P(D/A2) &= \\ &= (\text{probabiliteti qe nje person te kete demtime te degjimit me kusht qe te jete i pa pune}) = \\ &= 27 / 7462 = 0,0036 \end{aligned}$$

$$\begin{aligned} P(D/A3) &= \\ &= (\text{probabiliteti qe nje person te kete demtime te degjimit me kusht qe te jete ne pension}) = \\ &= 368 / 56778 = 0,0065. \end{aligned}$$

Duke qene se D tregon ngjarjen qe nje person te kete demtime te degjimit, kjo mund te shprehet si **bashkimi i tre ngjarjeve perjashtuese  $A1 \cap D$**  (i punesuar dhe me defekte te degjimit),  **$A2 \cap D$**  (i papune dhe me defekte te degjimit)  **$A3 \cap D$**  (pension dhe me defekt te degjimit) dhe kete mund ta shprehim:

$$D = (A1 \cap D) \cup (A2 \cap D) \cup (A3 \cap D)$$

---

---

---

---

---

---

---

---

- Personat me demtime te degjimit mund te bejne pjese **vetem ne njerin prej tre grupeve** dhe meqe **ngjarjet jane perjashtuese**, aplikojme **rregullin e shumes ne probabilitet**:

$$P(D) = P(A1 \cap D) \cup P(A2 \cap D) \cup P(A3 \cap D) \\ = P(A1 \cap D) + P(A2 \cap D) + P(A3 \cap D) =$$

---

---

---

---

---

---

---

---

Duke zbatuar rregullin e shumezimit per ngjarje te pappajtueshme per secilen nga kategorite me vete, mund te shkruajme:

$$P(D) = P(A1 \cap D) \cup P(A2 \cap D) \cup P(A3 \cap D) = \\ = P(A1 \cap D) + P(A2 \cap D) + P(A3 \cap D) = \\ = P(A1) * P(D / A1) + P(A2) * P(D / A2) + \\ P(A3) * P(D / A3) = \\ = (0,6063) * (0,0056) + (0,0457) * (0,0036) + \\ + (0,03480) * (0,0065) = \\ = 0,0058$$

---

---

---

---

---

---

---

---

## SHPREHJA ME E PLOTE:

$$P(D) = P(A1) * P(D / A1) + P(A2) * P(D / A2) + P(A3) * P(D / A3) \quad (1)$$

mund te perdoret ne rastet kur ne nuk mund ta gjejme direkt vleren e probabilitetit.

---

---

---

---

---

---

---

Supozojme tani se duam te gjemje probabilitetin  $P(A1/D)$ , pra probabilitetin qe nje person i zgjedhur rastesisht nga kjo popullate, **te jete i punesuar dhe te kete demtime te degjimit.**

Bazuar ne rregullin e shumezimit ne probabilitet kemi:

$$P(A1 \cap D) = P(D) * P(A1 / D) \quad (a)$$

nga ku kemi:

$$P(A1 / D) = \frac{P(A1 \cap D)}{P(D)} \quad (b)$$

Duke zevendesuar (a) tek barazimi (b) kemi

---

---

---

---

---

---

---

$$P(A1 / D) = \frac{P(A1) * P(A1 / D)}{P(D)} \quad (c)$$

Duke e zevendesuar (c) ne ate qe thame me siper tek ekuacioni (1)

$$P(D) = P(A1) * P(D / A1) + P(A2) * P(D / A2) + P(A3) * P(D / A3) \quad (1)$$

---

---

---

---

---

---

---

## SHPREHJA E TEOREMES SE BAYES

$$P(A_1 / D) = \frac{P(A_1) * P(A_1 / D)}{P(A_1) * P(D / A_1) + P(A_2) * P(D / A_2) + P(A_3) * P(D / A_3)}$$

---

---

---

---

---

---

---

- Duke bere zevendesimet e vlerave kemi:

$$0,6063 \times 0,0056$$

$$P(A_1 / D) = \frac{(0,6063 \times 0,0056) + (0,0457 \times 0,0036) + (0,3480 \times 0,0065)}{0,583}$$

- Probabiliteti qe nje person te jete i punesuar me kusht qe te kete demtime te degjimit eshte 0,583.

---

---

---

---

---

---

---

- Ne perfundim mund te themi se nese ngjarjet  $A_1, A_2, A_3, \dots, A_n$  jane ngjarje perjashtuese dhe jane te gjitha ngjarje te mundshme, atehere:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1,$$

Atehere teorema e Bayes kornfirmon qe:

$$P(A_i / B) = \frac{P(A_i) * P(A_i / B)}{P(A_1) * P(B / A_1) + P(A_2) * P(B / A_2) + P(A_3) * P(B / A_3)}$$

Per cdo "i", qe merr vlerat  $1 \leq n \leq n$ .

Teorema e Bayes eshte shume e rendesishme, sepse ben te mundur **ri-llogaritjen e probabilitetit, bazuar ne nje informacion te ri.**

---

---

---

---

---

---

---

## RISKU RELATIV DHE ODDS RATIO (RAPORTI I GJASAVE)

- Risku relativ (RR) perdoret ne rastet kur duam te krahasojme probabilitetet per te ndodhur semundja ne dy situata te ndryshme ose ne dy grupe te ndryshme.
- RR eshte probabiliteti qe nje individ, i cili i perket grupit qe ekspozohet ndaj nje faktori risk, te zhvilloje semundjen krahasuar me nje nje individ te grupit qe nuk ekspozohet ndaj nje faktori risk, por qe zhvillon semundjen ne fjale.
- Me fjale te tjera RR paraqet **probabilitetin per t'u semure ne grupin e te ekspozuarve, pjesetuar me probabilitetin e te paekspozuarve per t'u semure nga e njejta semundje.**

---

---

---

---

---

---

---

---

$$RR = \frac{P(\text{te semure} / \text{te ekspozuar})}{P(\text{te semure} / \text{te paekspozuar})}$$

---

---

---

---

---

---

---

---

## Interpretimi i vlerave te RR

- $RR = 1$  tregon se probabiliteti per t'u semure si ne grupin e te ekspozuarve dhe ne ate te te paekspozuarve ndaj faktorit/faktoreve risk eshte i njejte, pra **nuk ka lidhje** mes faktorit risk dhe semundjes.
- $RR > 1$  tregon se ne te ekspozuarit ka nje **probabilitet me te madh per t'u semure**, krahasuar me te paekspozuarit.
- $Rr < 1$  tregon **qe ekziston nje risk me i vogel per t'u semure** ne personat e ekspozuar.

---

---

---

---

---

---

---

---



## OR

- Nje mase tjeter qe perdoret per te krahasuar probabilitetet eshte Odds Ratio (OD).
- OD perkufizohet si odds ne favor te semundjes ne individe te ekspozuar, ndaj odds te semundjes ne individe te paekspozuar.

---

---

---

---

---

---

---

---

- Nese nje ngjarje ndodh me probabilitet "p", ods ne favor te ngjarjes eshte  $p / (1-p)$ . Keshtu nese  $p = \frac{1}{2}$ , OD eshte  $\frac{1}{2} / \frac{1}{2} = 1$ . Ne kete rast ngjarja mund te verifikohet ose jo me te njejtin probabilitet.
- Nese  $p = \frac{2}{3}$ , **OD =  $\frac{2}{3} / \frac{1}{3} = 2$**  ne 1, pra probabiliteti qe ngjarja te verifikohet eshte dy here me i madh sesa ngjarja te mos verifikohet.

---

---

---

---

---

---

---

---

## Si llogaritet OR

$$OD = \frac{P(\text{te semure} / \text{te ekspozuar}) / [1 - P(\text{te semure} / \text{te ekspozuar})]}{P(\text{te semure} / \text{te paekspozuar}) / [1 - P(\text{te semure} / \text{te paekspozuar})]}$$

RR dhe OD jane dy masa qe tentojne te shpjegojne te njejtin fenomen; edhe pse RR duket me intuitiv, OD ka me shume prioritet statistikore.

---

---

---

---

---

---

---

---

## **SHPËRNDARJET PROBABILITARE DHE VARIABLET E RASTËSISHËM**

Karakteristika që analizohet në një studim quhet variabël. Variable mund të jenë gjinia, raca, mosha, koha në kurbën epidemike që korrespondon me shfaqjen e sëmundjes në pacientë, etj. Termi “variabël”, merr këtë kuptim pasi vlerat e të njëjtës karakteristike ndryshojnë nga një subjekt në tjetrin. Ky variacion mund të lidhet me veçoritë biologjike të gjithsecilit, me gabimet e bëra në matjen e kësaj karakteristike tek çdo subjekt, etj. Variabël i rastësishëm quhet variabli i cili studjohet në persona të cilët janë zgjedhur rastësisht.

Ashtu sikurse vlerat e një karakteristike si psh: niveli i sheqerit në gjak, mund të përmbliken në një shpërndarje frekuencash, edhe variablet random mund të përmbliken në një shpërndarje frekuencash, që quhet shpërndarje probabilitare.

Në statistikë ka disa shpërndarje probabilitare teorike, por ne do të analizojmë vetëm tri prej tyre, të cilat janë më të përdorshme në mjekësi. Shpërndarjet që do të analizohen janë shpërndarja normale (Gauss), shpërndarja binomiale dhe shpërndarja Poisson.

Shpërndarja normale (e Gauss-it) është një shpërndarje probabilitare e vazhdueshme, pra përdoret për variable të vazhdueshme. Shpërndarja binomiale dhe ajo e Poisson përdoren për variable diskrete, pra për variable diskrete që marrin vlera të plota si psh 0, 1, 2, 3, ..., n.

### **SHPËRNDARJA BINOMIALE**

*Shpërndarja binomiale është një shpërndarje probabilitare që përdoret për të dhëna diskrete.*

Supozojme se një eksperiment ka vetëm dy mundësi ndodhjeje / ngjarje (pra është dikotomike ose binare), rezultatet e të cilës janë dy (po/jo; pozitiv/negativ; shëndoshë/sëmure etj) dhe kanë secila probabilitet që verifikohet si ‘p’ për njërin nga ngjarjet dhe ‘1-p’ për ngjarjen tjetër. Kështu pra nëse shënojmë me A njërin nga dy mundësitë e ndodhjes, probabiliteti i A-së është p, ose  $P(A) = p$ . Ky probabilitet është i pandryshueshëm sa herë që ndodh kjo ngjarje. Probabiliteti i ngjarjes B është 1-p, ose  $P(B) = 1 - p$ , sepse ngjarja B ndodh atëherë kur nuk ndodh ngjarja A. Nëse një eksperiment e përsërisim “n” herë, sa është probabiliteti që ngjarja A të ndodhë saktësisht X herë, duke konsideruar që ngjarjet janë të pavarura? E thënë ndryshe, në cilin proporcion të n-së ndodh ngjarja A?

Parimet bazë të shpërndarjes binomiale i kanë fillimet që në shekullin e XVII, nga matematikani zviceran Jakob Bernulli, me kontributin e tij në teorinë e probabilitetit. Për nder të tij, çdo eksperiment që përfshin një probabilitet binomial quhet eksperimenti Bernulli dhe çdo pjesë e këtij eksperimenti quhet procesi Bernulli.

Variabël rastësor, i pritshëm quhet variabël binomial .

Variabël rastësor quhet variabli i cili, në një shpërndarje probabilitare, mund të marrë vlera të ndryshme. Psh: numri i meshkujve në një familje, apo dhe numri i femrave.

Në shëmbullin tonë le të quajmë përkatësisht A dhe B gjasat për të qënë mashkull dhe femër. Quajmë “n” numrin e rasteve (në shëmbullin tonë numrin e femijëve). Secilin rast mund ta quajmë si një nga numrat 0, 1, 2, 3...n-r, n .

Quajmë gjithashtu; .

$$p = r / n$$

frekuencën relative të ngjarjes A (të qënurit mashkull) .

dhe  $q = n-r / n = 1 - p$  frekuencën relative të ngjarjes B (të qënurit femër).

$(n / r) \rightarrow$  quhet koeficient binomial.

Formula e shpërndarjes binomiale është:

$$f(r) = \frac{n!}{r!(n-r)!} p^r q^{n-r}$$

Ku:

p - është probabiliteti për të ndodhur njëra nga ngjarjet (psh: për të qënë mashkull);

q - është probabiliteti për të ndodhur ngjarja tjetër ( psh.per te qene femer);

r - është numri i rasteve që na interesojnë ne.

### **Shembull**

Një çift ka tre fëmijë. Sa është probabiliteti që dy nga këta të jenë meshkuj?

Pra kemi:

$n=3$ ,  $r=2$ .

$p=1/2$ , probabiliteti që një fëmijë të jetë djalë.

$q = 1-p = 1/2$ , është probabiliteti për të ndodhur ngjarja tjetër, pra të qënurit vajzë.

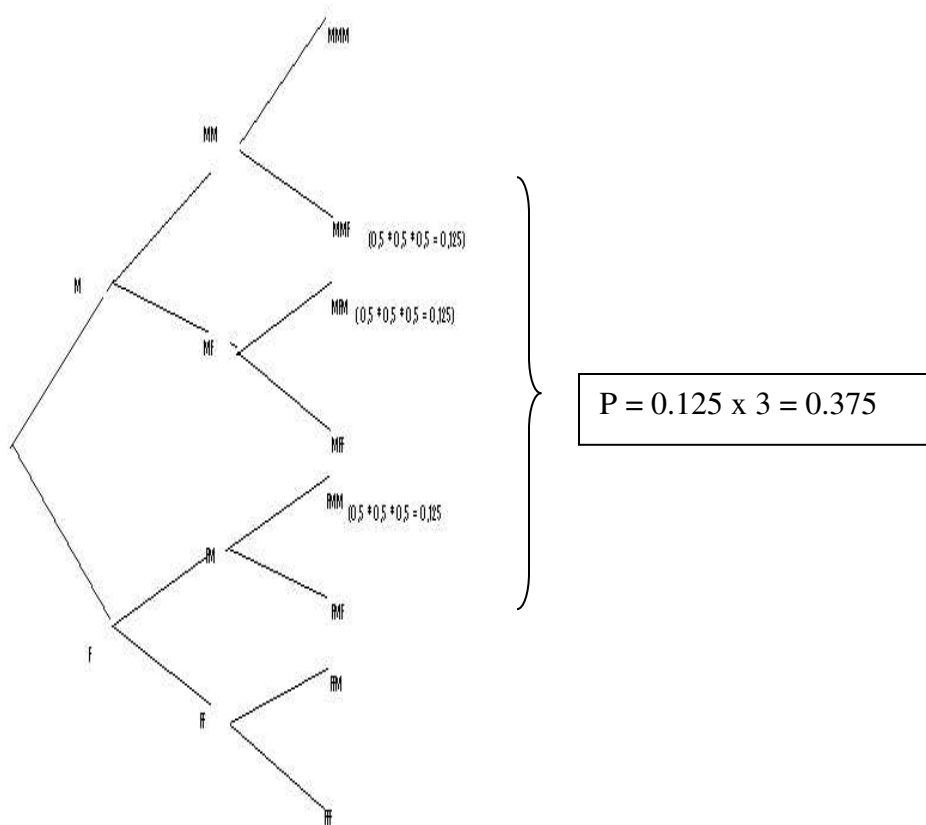
Atëhere, duke aplikuar formulën e shpërndarjes binomiale mund të shkruajmë:

$$f(r) = \frac{3!}{2!(3-2)!} p^2 q^{3-2} \quad \text{ku } p = q = 1/2 = 0,5$$

$$f(r) = \frac{3!}{2!1!} 0,5^2 \cdot 0,5 = \frac{3 \cdot 2 \cdot 1}{2} \cdot 0,125 = 0,375$$

Pra, probabliliteti që dy nga tre fëmijët të jenë meshkuj, është 0,375.

Kjo paraqitet qartë dhe thjesht me anë të grafikut pemë .



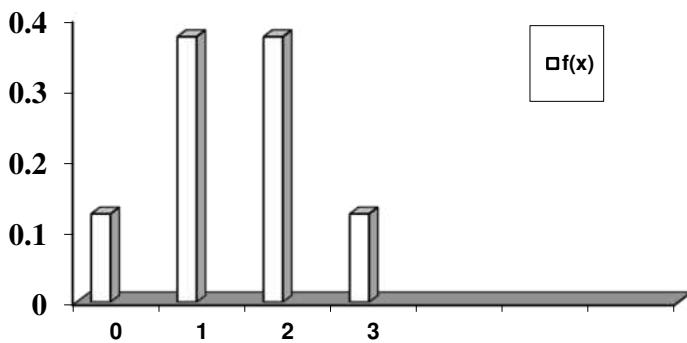
Grafiku 9. Diagami Pemë

Në mënyrë të përmbledhur, probabilitetet për shpërndarjen binomiale me  $n=3$  dhe  $p=0,5$  jepen në tabelën më poshtë dhe grafikun më poshtë:

Tabela . Probabilitetet për shpërndarjen binomiale me  $n=3$  dhe  $p=0,5$

| Numri i fëmijëve meshkuj | $\frac{n!}{r!(n-r)!}$ | $p^r$ | $q^{(n-r)}$ | $f(r)$ |
|--------------------------|-----------------------|-------|-------------|--------|
| 0                        | 1                     | 1     | 0,125       | 0,125  |
| 1                        | 3                     | 0,5   | 0,25        | 0,375  |
| 2                        | 3                     | 0,25  | 0,5         | 0,375  |
| 3                        | 1                     | 0,125 | 1           | 0,125  |

Graf. Shpërndarja binomiale për  $n=3$  dhe  $p = 0,5$



Mesatarja e shpërndarjes binomiale llogaritet si

$$np \text{ dhe në rastin tonë është } 3 \cdot 0,5 = 1,5.$$

Numri “1,5”, përfaqëson numrin mesatar të fëmijëve meshkuj që mund të ketë çifti. Devijacioni standart llogaritet me formulën:

$$\sqrt{n \cdot p(1-p)}$$

Në shëmbullin tonë, devijacioni standart është:

$$\sqrt{n \cdot p(1-p)} = \sqrt{3 \cdot 0,5(1-0,5)} = 0,86$$

Për të përcaktuar shpërndarjen binomiale janë të domosdoshëm dy parametra kryesorë

n dhe p

të cilët quhen parametra të shpërndarjes binomiale.

Kjo lloj shpërndarje lejon llogaritjen e probabiliteteve dhe në rastet kur kërkohet, psh (duke iu referuar problemës më lart), sa është probabiliteti që më pak se dy fëmijë të jenë meshkuj?

Në këtë rast:

$$f(r) = f_{(r=0)} + f_{(r=1)}$$

Figura më poshtë tregon grafikun e shpërndarjes binomiale dhe format e ndryshme të shpërndarjes për kombinime të ndryshme të vlerave të p-së dhe n-së. Shihet se duke i dhënë vlera të ndryshme n-së, shpërndarja ndryshon formë. Pikërisht duke iu referuar një vlere të p-së, shihet se asimetria bie me rritjen e n-së. Grafiku është simetrik për  $p = 1/2$  dhe asimetrik për p më të vogël se  $1/2$ , ose më të madhe se  $1/2$ .

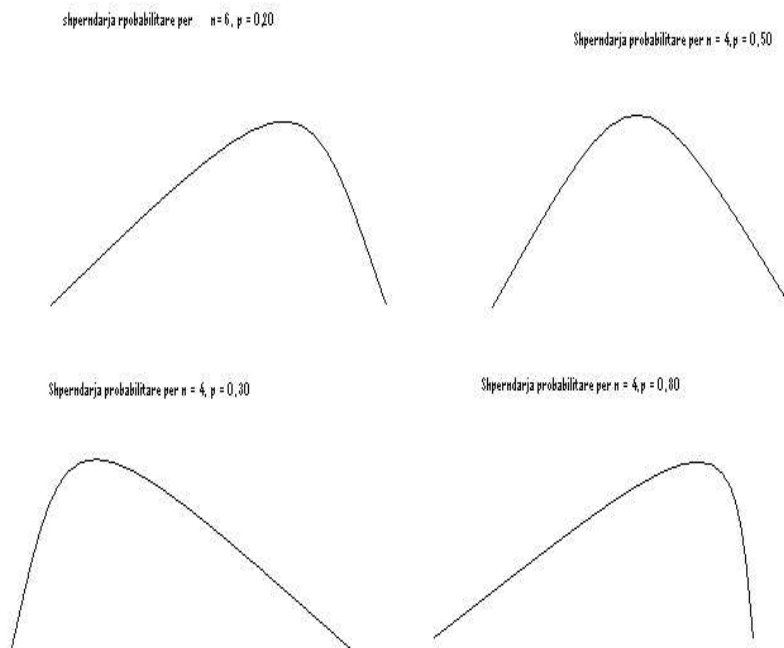


Fig.12 Grafiku i shpërndarjes binomiale dhe format e ndryshme të shpërndarjes për kombinime të ndryshme të vlerave të p-së dhe n-së

### Karakteristikat e shpërndarjes binomiale

1. rezultati i një eksperimenti mund të konsiderohet si rezultat i n eksperimenteve më të vegjël të quajtura ngjarje, ku secila prej tyre ka vetëm dy mundësi ndodhjeje (gjallë-vdekur; shëndoshë-sëmurë; vajzë-djalë);
2. ndodhitë e ngjarjeve janë të pavarura;
3. probabiliteti i ndodhjes së një ngjarje nga njëra provë në tjetrën është i pandryshueshëm, pra është homogjen.

## SHPËRNDARJA POISSON

(Simeon D. Poisson, matematikan francez ,1780-1840)

Shpërndarja Poisson është një shpërndarje probabilitare për variable diskretë dhe që përdoret për ngjarje që ndodhin rrallë në intervalin e kohës “t”. Ndodhitë e ngjarjeve janë random dhe nuk ndjekin ndonjë rregull, pra janë të paparashikueshme. Pavarësia e ngjarjeve do të thotë që ndodhia e një ngjarje nuk ndikohet nga ndodhia e ngjarjes paraardhëse. Ndodhia e ngjarjeve vlerësohet vetëm e lidhur me intervalin e kohës (psh një herë në 15 vjet).

Karakteristika të shpërndarjes Poisson janë:

- Ndodhitë e ngjarjeve janë të pavarura;
- Probabiliteti i ndodhjes së një ngjarje është proporcional me gjatësinë e intervalit;
- Në një porcion shumë të vogël të kohës, probabiliteti i ndodhjes së më shumë se një ngjarje është pothuajse zero

Modeli probabilistik i shpërndarjes Poisson është:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Supozojmë se observojmë në periudha kohë të përsëritura, me zgjatje “t”, një proces të Poissonit me përqindje mesatare  $\lambda$ . Numri “x” i ngjarjeve të pritshme i ndodhur në kohën t, do të variojë nga një interval në tjetrin.(është një variabël rastësor).

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{ku } e = 2,718.$$

Siç shihet shpërndarja e Poisson-it shquhet tërësisht nga një parametër dhe ky është  $\lambda$ . Pra të gjitha karakteristikat e kësaj shpërndarjeje janë funksione të mesatares  $\lambda$ . Me rritjen e  $\lambda$ , asimetria zvogëlohet.

### Ushtrime

Një kompani sigurimesh kontrollon mesatarisht tri herë në muaj nga institucionet kompetente. Duke përdorur shpërndarjen Poisson, sa është probabiliteti që kjo ndërmarrje të kontrollohet:

1. vetëm dy herë në vit,
2. deri në një herë në vit.

Shënojmë me  $\lambda$  numrin mesatar të kontrolleve që ushtrohen në një vit në kompani dhe me x shënojmë numrin e kontrolleve që ne dëshirojmë që të ushtrohet në këtë kompani. Kështu do të kemi:

a) Probabiliteti që kontrolli të ushtrohet vetëm dy herë në vit në kompaninë e sigurimeve:

$$\lambda = 3$$

$$\text{dhe } x = 2$$

duke përdorur formulën e shpërndarjes Poisson gjejmë:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{3^2 \cdot e^{-3}}{2!} = \frac{9 \cdot 0,4978707}{2} = 0,2240$$

b) Probabiliteti që kontrolli të ushtrohet deri në një herë në vit në kompaninë e sigurimeve:

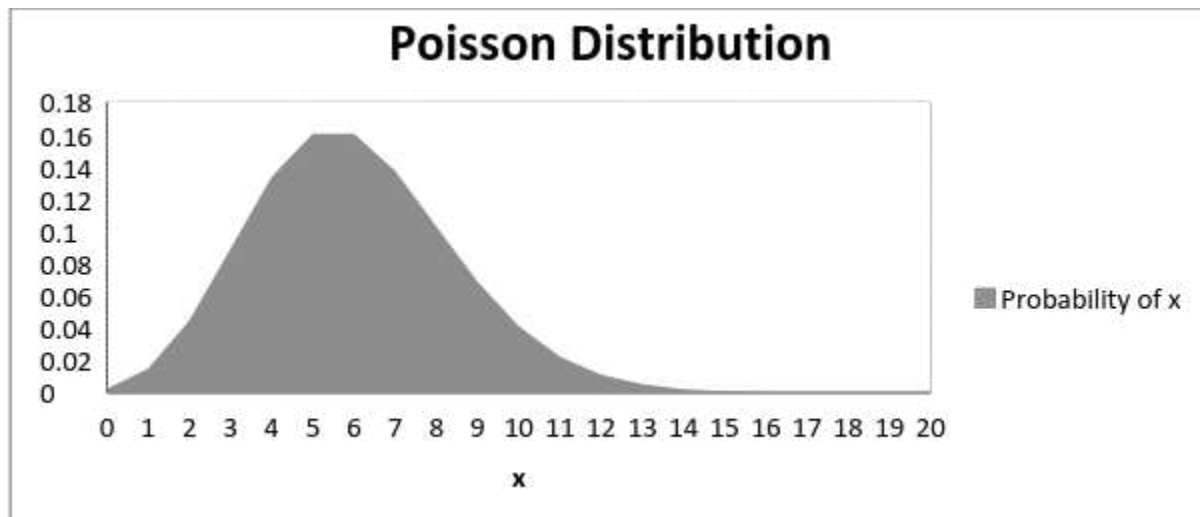
Kur themi deri në një herë kuptojmë që flitet për rastin që kontrolli mund të mos ushtrohet asnjëherë ( pra  $x = 0$ ) të ushtrohet vetëm një herë ( $x = 1$ ). Në këtë rast mund të shkruajmë:

$P(\text{deri në një herë në vit}) =$

$$P(x=0) + P(x=1) =$$

$$\frac{3^0 \cdot e^{-3}}{0!} + \frac{3^1 \cdot e^{-3}}{1!} = \frac{1 \cdot 0,4978707}{1} + \frac{3 \cdot 0,4978707}{1} = 0,0498 + 0,1494 = 0,1992$$

Vini re: në shpërndarjen Poisson duhet që intervali kohor për  $\lambda$  dhe  $x$  duhet të jetë i njëjtë. Kur intervali nuk është i njëjtë, atëherë mesatarja  $\lambda$  riformulohet në mënyrë të tillë që të ketë interval të njëjtë me  $x$ -in.





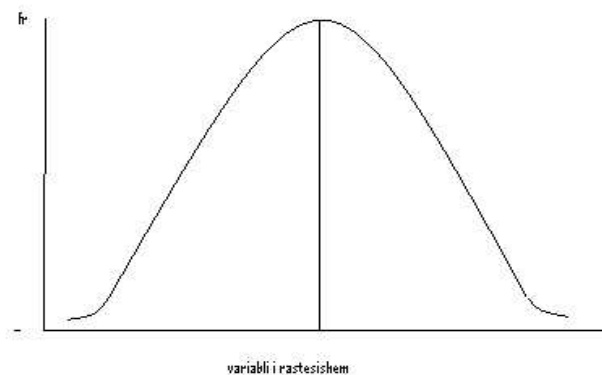
## **SHPËRNDARJA NORMALE e Gauss-it** (*Carl Friedrich Gauss, shek. XIX*)

Quhet ndryshe kurba normale ose shpërndarja probabilitare normale. Colton thekson se: “**Normal**” ndoshta nuk është tamam termi i përshtatshëm për të quajtur këtë lloj shpërndarje, për arsye se kuptimi “normal “ në statistikë është i ndryshëm nga kuptimi që ka kjo fjalë në mjekësi, megjithatë është pranuar që kjo lloj shpërndarje të quhet normale. Shpërndarja normale është një model matematikor që jepet me formulën:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Ku  $\pi = 3,14$ ;  $e = 2,71828$ ,  $x$  është variabli i rastësishëm që mund të marrë çdo lloj vlere  $(-\infty; +\infty)$ .

Densiteti i funksionit  $f(x)$  varet vetëm nga vlerat  $\mu$  dhe  $\sigma$  për çdo  $x$ , sepse këta parametra janë të vetmit komponentë të funksionit që mund të ndryshojnë. Grafiku që gjenerohet nga ky funksion (formula më lart) jep grafikun më poshtë (graf 4), ku në boshtin  $x$  (abshisa) vendosen vlerat e variablit të rastësishëm dhe në boshtin  $y$  (ordinatat) vendosen frekuencat relative të variablit të shprehur si  $f(x)$ .



Graf. 11 Kurba e Gauss-it

Kurbat normale mund të kenë të njëjtën mesatare, por devijacione standarte të ndryshme, mesatare të ndryshme, por devijacione standarte të njëjta dhe vlera mesatare të ndryshme dhe devijacione standarte të ndryshme, të cila paraqiten në figura 13, përkatësisht 13a, 13b dhe 13c.

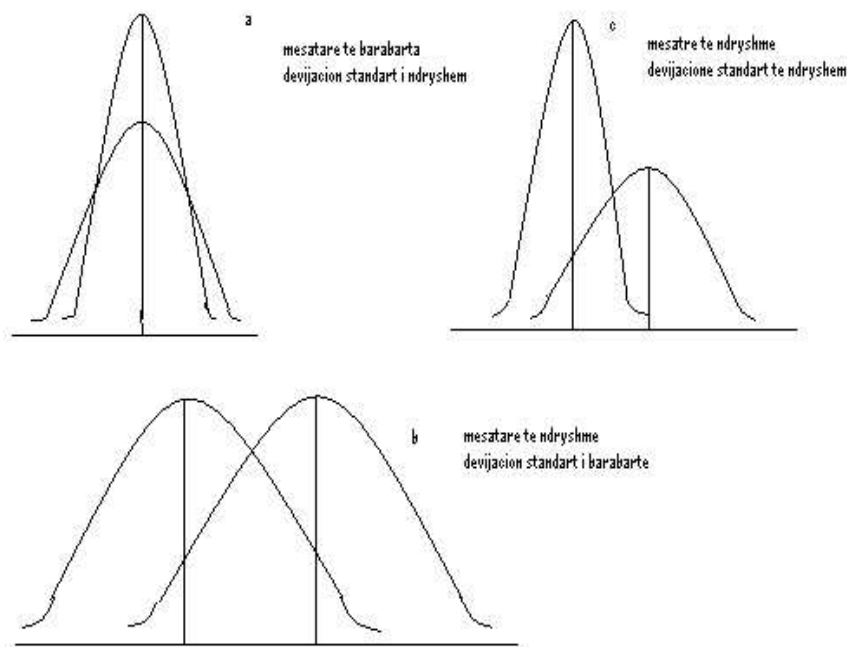


Figura 13 Kurbat normale sipas vlerave te mesatares aritmetike dhe devijacionit standart

Shpërndarja e Gaussit është shpërndarja më e përdorshme në biostatistikë dhe ka këto karakteristika:

- është një shpërndarje probabilitare e vazhdueshme, që paraqet një numër infinit observimesh të matura në mënyrë të vazhdueshme;
- është simetrike ndaj mesatares;
- është me rang infinit, çka do të thotë që buzët e grafikut i afrohen pambarimisht boshtit x, pa e prekur asnjëherë atë;
- është një modale dhe meqënëse shpërndarja është simetrike, atëhere moda, medianja dhe mesatarja përputhen në të njëjtën pikë;
- grafiku i saj është në formën e një këmbane;
- sipërfaqja nën kurbë deri në linjën e abshisës është 1 njësi katrore. Duke kujtuar që shpërndarja normale e Gauss-it është një **shpërndarje probabilitare, simetrike, një modale** dhe duke u mbështetur në një nga vetitë elementare të probabilitetit, që thotë se **“shuma e probabiliteteve për të gjitha ngjarjet e mundëshme të provës është e barabartë me 1”**, atëherë 50% e sipërfaqes ndodhet në të majtë të vijës pingule (me probabilitet 50%) dhe 50% e sipërfaqes ndodhet në të djathtë të po kësaj linje ( $p = 50\%$ ).

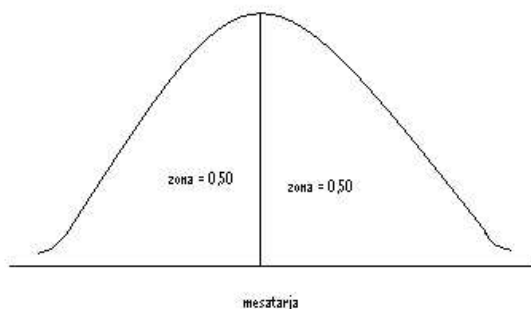


Figura 14 Shpërndarje probabilitare, simetrike, një modale

Në shpërndarjen normale të Gauss-it kanë vend tri barazimet e mëposhtme:

- $\bar{x} \pm SD = 68.3\%$  e të gjithë sipërfaqes;
- $\bar{x} \pm 2SD = 95.5\%$  e të gjithë sipërfaqes;
- $\bar{x} \pm 3SD = 99.7\%$  e të gjithë sipërfaqes.

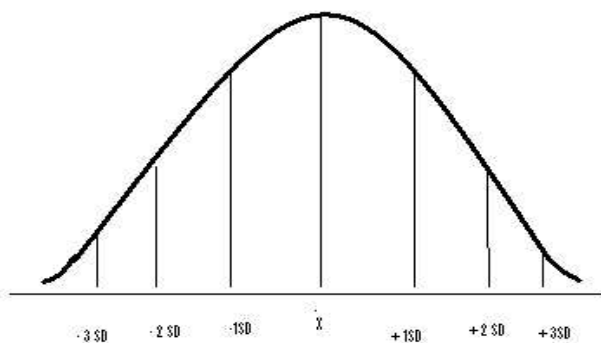


Fig. 15 Paraqitja grafike e tri barazimeve të Shpërndarjes Normale

Grafiku i shpërndarjes normale standarte ka një vlerë mesatare  $\mu = 0$  dhe  $\sigma = 1$ . Në jetën e përditëshme, ku kemi situata krejt të ndryshme, larg asaj ideale,  $\mu \neq 0$  dhe  $\sigma \neq 1$ . Në këto raste përdorim atë që quhet transformimi z, ose standarti normal i devijuar, që shprehet si:

$$z = \frac{X - \mu}{\sigma}$$

ku  $x$  është mesatarja e mostrës;

$\mu$  mesatarja e popullatës;

devijacioni standart i mostrës.

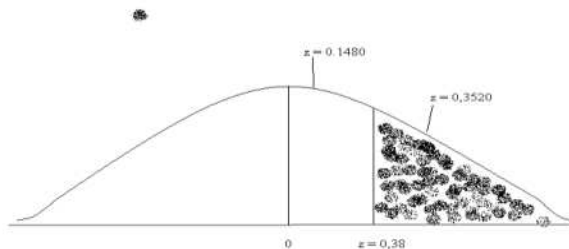
Ushtrim: Supozojmë se në një popullatë me 10 000 individë niveli mesatar i sheqerit në gjak është 85mg/dl dhe devijacioni standart është 13 mg/dl. Gjeni ç'proporcion i popullatës sonë ka një nivel mesatar të glicemisë  $x \geq 90$  mg/dl?

Standarti normal i devijuar është:

$$z = \frac{X - \mu}{\sigma} = \frac{90 - 85}{13} = \frac{5}{13} = 0,38$$

Niveli i glicemisë 90 mg/dl është afërsisht 0,38 devijacione standart kundrejt vlerës mesatare 85 mg/dl. Hyjmë në tabelën Z. Gjejmë  $z = 0,38$ . Ngjitemi vertikalisht lart në kolonën “zona në një anë”. Kjo kollonë përmban zonën që i përket vlerës  $z$  prej 0,38 dhe shihet se afërsisht 35% e individëve të popullatës e ka vlerën e sheqerit në gjak 90 mg/dl. Pra nëse rastësisht do të zgjidhnim një individ nga kjo popullatë, probabiliteti që ky person të ketë nivelin e sheqerit në gjak 90 mg/dl, është 0,35. (Fig. 16)

Figura 16 Paraqitja për  $z = 0,38$



2. Cili porcion i popullatës ka vlera të glicemisë mes 105 – 114 mg/dl?

Fillimisht llogarisim vlerën  $z$  për  $x = 105$  mg/dl:

$$z = \frac{X - \mu}{\sigma} = \frac{105 - 85}{13} = \frac{20}{13} = 1,54$$

Gjejmë zonën korresponduese në një anë (vlere e probabilitetit në një anë) për

$$z = 1,54 \rightarrow 0,0618$$

Logarisim vlerën z për  $x = 114$  mg/dl:

$$z = \frac{X - \mu}{\sigma} = \frac{114 - 85}{13} = \frac{29}{13} = 2,23$$

Gjejmë zonën korresponduese në një anë (vlere e probabilitetit në një anë) për

$$z = 2,23 \rightarrow 0,0129$$

Zona mes dy vlerave të “z” është:

$$0,0618 - 0,0129 = 0,0489 \quad (\text{shih fig. 17})$$

Duke shumëzuar këtë vlerë (porcionin e popullatës) me numrin e përgjithshëm të individëve të popullatës (10 000) marrim numrin e individëve që kanë vlerat e glicemisë mes 105 – 114 mg/dl. Kështu kemi:

$$0,0489 \times 10000 = 489 \text{ individë}$$

Pra 489 individë do të kenë vlerat e glicemisë mes 105 – 114 mg/dl.

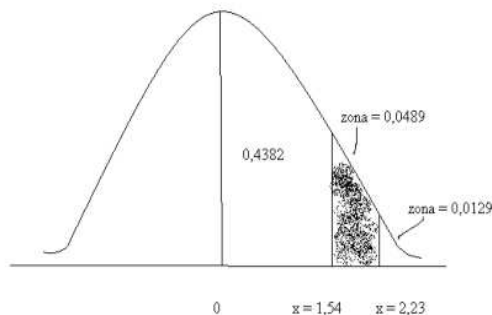


Fig. 17 Zona mes dy vlerave të “z” për glicemisë mes 105 – 114 mg/dl.

**Shembull 2:**

Supozojmë se një popullatë me 100 banorë i nënshtrohet shpërndarjes normale. Peshë mesatare është 70 kg dhe SD 10 kg.

Beni te gjitha konkluzionet e mundshme per kete popullatë.

**Zgjidhje:**

Derisa popullata i nënshtrohet shpërndarjes normale, kanë vend tre barazimet e Gauss-it.

$$\bar{x} \pm SD = 68.3 \% \text{ të popullatës.}$$

Popullata përbëhet nga 100 banorë, pra  $n = 100$ .  $\bar{x} = 70$  kg dhe  $SD = 10$  kg.

$$70 \pm 10 = \begin{matrix} 60 \text{ kg} \\ 80 \text{ kg} \end{matrix} \} \text{ këtu përfshihen } 68.3\% \cdot 100 = 68.3 \text{ persona.}$$

$$70 \pm (2 \times 10) = \begin{matrix} 50 \text{ kg} \\ 90 \text{ kg} \end{matrix} \} \text{ këtu përfshihen } 95.5\% \cdot 100 = 95.5 \text{ persona.}$$

$$70 \pm (3 \times 10) = \begin{matrix} 40 \text{ kg} \\ 100 \text{ kg} \end{matrix} \} \text{ këtu përfshihen } 99.7\% \cdot 100 = 99.7 \text{ persona.}$$

**Shembull 2:** Është dhënë popullata e mëposhtme. Gjeni a i nënshtrohet ajo shpërndarjes normale?

| <i>Pesha(kg)</i> | 2 | 3 | 4 | 5 | 6 |
|------------------|---|---|---|---|---|
| <b>Frekuenca</b> | 2 | 4 | 2 | 3 | 1 |

Tab Paraqitja e peshës sipas frekuencës

**Zhvillim:** Meqënëse në mjekësi mjafton që rezultati të vërtetohet në 95% të rasteve që ai të konsiderohet si i vërtetë, atëherë mjafton të vërtetohet vetëm barazimi i dytë i Gauss-it që të themi që popullata i nënshtrohet shpërndarjes normale. Në rast të kundërt, popullata nuk i nënshtrohet shpërndarjes normale.

Barazimi i dytë i Gauss-it është:

$$\bar{x} \pm 2SD = 95.5\% \text{ të popullsisë.}$$

$$n = \sum f = 12 ; \quad \bar{x} = \frac{\sum xfr}{n} = 4 \text{ kg} ; \quad SD = \frac{\sqrt{\sum (x - \bar{x})^2 fr}}{\sqrt{n-1}} = 1,2 \text{ kg}$$

Atëherë kemi :

$$4 + (2 \times 1.2) = \begin{array}{l} \nearrow 6.4 \text{ kg} \\ \searrow 1.6 \text{ kg} \end{array}$$

Pikërisht në këtë segment [1.6-6.4 ] përfshihet e gjithë popullata jonë, pra 100% e individëve. Si përfundim themi se, popullata në studim i nënshtrohet shpërndarjes normale.

## Shpërndarja binomiale, përqasja e saj me shpërndarjen normale

Dr. Elizana (Petrela) ZAIMI

---

---

---

---

---

---

---

## Shpërndarja binomiale

- Kjo lloj shpërndarje përdoret për të dhëna diskrete.
- Supozojmë se ndodhja e një ngjarjeje, rezultatet e të cilës janë dy (A dhe B), kanë secila probabilitetet që cilësohen si “p” dhe “1 – p”.
- Variabli rastësor, i pritshëm quhet variabli binomial
- Variabel rastësor quhet variabli i cili, në një shpërndarje probabilitare, mund të marrë vlera të ndryshme. P.sh.: numri i meshkujve, apo numri i femrave në një familje.

---

---

---

---

---

---

---

## Shembull

- Supozojmë se një çift ka tre fëmijë. Sa është probabiliteti që, dy nga këta fëmijë të jenë meshkuj?

???

---

---

---

---

---

---

---



?

- “n” numri i rasteve (ne shembullin tone numri i femijeve).
- $p = r / n$  frekuencen relative te ngjarjes A (te qenit mashkull) .
- $q = n-r / n = 1 - p$  frekuencen relative te ngjarjes B (te qenit femer).
- $n / r \rightarrow$  quhet koeficient binomial.

---

---

---

---

---

---

---

### Formula e shperndarjes binomiale

$$f(r) = \frac{n!}{r!(n-r)!} p^{(r)} q^{(n-r)}$$

---

---

---

---

---

---

---

Pra kemi:  
 $n = 3, r = 2$ .

$p = 1/2$  pra probabiliteti qe nje femije te jete djale.  
 $q = 1/2 = 1 - p$  eshte probabiliteti per te ndodhur ngjarja tjeter, pra vajze.

---

---

---

---

---

---

---

$$f(r) = \frac{3!}{2!(3-2)!} p^2 * q^{(3-2)}, \text{ ku } p = q = \frac{1}{2} = 0,5$$

$$f(r) = \frac{3!}{2! 1!} 0,5^2 * 0,5^1 = \frac{3*2*1}{2*1} 0,125 = 0,375$$

---

---

---

---

---

---

---

---

- Pra, probabiliteti qe dy nga tre fwmijet te jene meshkuj, eshte 0,375.

---

---

---

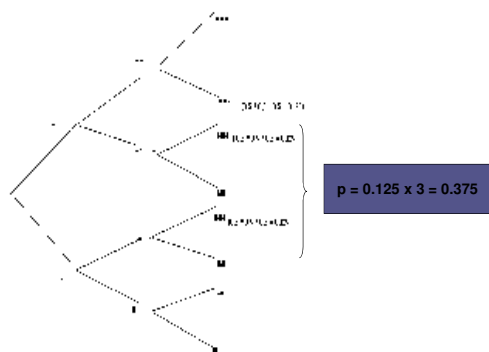
---

---

---

---

---




---

---

---

---

---

---

---

---

## Grafiku i Sh. binomiale




---

---

---

---

---

---

---

---

Paraqitja grafike e SHB, (per kombinime te ndryshme te vlerave te  $p$ -se dhe  $n$ -se)

- Duke i dhene vlera te ndryshme  $n$ -se shperndarja ndryshon forme.
- Pikerisht, duke j'u referuar nje vlere te  $p$ -se, shihet se asimetria bie me rritjen e  $n$ -se.
- Grafiku eshte simetrik per  $p = 1/2$  dhe asimetrik per  $p$  me te vogel se  $1/2$ , ose me e madhe se  $1/2$ .

---

---

---

---

---

---

---

---

Fig. 10.1.1. Sh. binomiale per  $n=10$  dhe  $p=0.5$

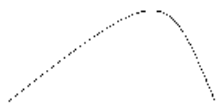


Fig. 10.1.2. Sh. binomiale per  $n=10$  dhe  $p=0.3$

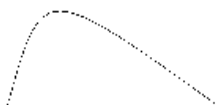
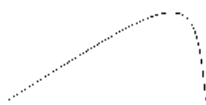


Fig. 10.1.3. Sh. binomiale per  $n=10$  dhe  $p=0.7$



Fig. 10.1.4. Sh. binomiale per  $n=10$  dhe  $p=0.1$




---

---

---

---

---

---

---

---

## Karakteristikat e shperndarjes binomiale

1. Rezultati i nje eksperimenti mund te konsiderohet si rezultat i n eksperimenteve me te vegjel te quajtura ngjarje,ku secila prej tyre ka vetem dy mundesi ndodhjeje (gjalle -vdekur; shendoshe- semure; vajze- djale);
2. Ndodhite e ngjarjeve jane te pavarura;
3. Probabiliteti i ndodhjes se nje ngjarje, nga njera prove ne tjetren eshte i pandryshueshem, pra eshte homogjen.

---

---

---

---

---

---

---

# SHPËRNDARJA POISSON

## Shpërndarja Poisson (1)

- Shpërndarja Poisson është një shpërndarje probabilitare për variabël diskrete dhe që përdoret për ngjarje që ndodhin rrallë në intervalin e kohës "t".

## Shpërndarja Poisson (2)

- Ndodhjet e ngjarjeve janë random dhe nuk ndjekin ndonjë rregull, pra janë të paparashikueshme.

### Shpërndarja Poisson (3)

- Pavarësia e ngjarjeve do të thotë që ndodhia e një ngjarjeje nuk ndikohet nga ndodhia e ngjarjes para-ardhëse.
- Ndodhia e ngjarjeve vlerësohet vetëm e lidhur me intervalin e kohës (p.sh. një herë në 15 vjet).

---

---

---

---

---

---

---

### Karakteristikat e shpërndarjes Poisson

- Ndodhjet e ngjarjeve janë të pavarura;
- Probabiliteti i ndodhjes së një ngjarjeje është proporcional me gjatësinë e intervalit;
- Në një porcion shumë të vogël të kohës, probabiliteti i ndodhjes së më shumë se një ngjarjeje është pothuajse zero

---

---

---

---

---

---

---

### Formula:

$$f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \quad \lambda = \mu \cdot t$$

ku  $e = 2,718$ .

---

---

---

---

---

---

---

## Shpërndarja standarte normale “z”

Prof. Dr. Elizana PETRELA

Dime se...

- Shpërndarja normale perdoret kur  $x = 0$  dhe  $SD = 1$
- Kur  $x \neq 0$  dhe  $SD \neq 1$ , atehere perdorim

*transformimin*  $z$

## Transformimi $z$

- *Transformimi*  $z$  shpreh devijimin prej mesatares ne njesite e devijacionit standart.

$$Z = \frac{x - \mu}{\sigma}$$

Quhet  $z$  score, vlera kritike

## Ushtrim

- TA mes =  $\mu$  = 120 mmHg
- SD =  $\sigma$  = 10 mm Hg
- Duke ditur qe popullata eshte e shperndare normalisht, gjeni cila zone ne kurbe ka TA > 130 mmHg?
- Cila zone ne kurbe ka TA > 140 mmHg?
- Cila zone ne kurbe ka TA mes 100-140 mmHg?

---

---

---

---

---

---

---

## Pyetje?

- Cila zone ne kurbe ka TA > 150 mmHg?
- Cila zone ne kurbe ka TA < 90 mmHg dhe mbi 150 mmHg?
- Cila eshte vlere e TA qe e ndan kurben nen 95% dhe mbi 5%?
- Cila eshte vlere e TA qe e ndan kurben nen 97,5% dhe mbi 2,5%?

---

---

---

---

---

---

---

Gjeni cila zone ne kurbe ka TA > 130 mmHg?

$$Z = \frac{x - \mu}{\sigma} = \frac{130 - 120}{10} = 1$$

$Z=1$    $p = 0,159$  (15,9%)

---

---

---

---

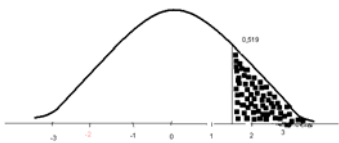
---

---

---



Gjeni cila zone ne kurbe ka TA > 130 mmHg?




---

---

---

---

---

---

---

---

Cila zone ne kurbe ka TA > 140 mmHg?

$$Z = \frac{x - \mu}{\sigma} = \frac{140 - 120}{10} = 2$$

$$Z = 2 \quad p = 0,023 \quad (2,3\%)$$

---

---

---

---

---

---

---

---

Cila zone ne kurbe ka TA > 140 mmHg?




---

---

---

---

---

---

---

---

Cila zone ne kurbe ka TA mes 100-140 mmHg?

$$Z = \frac{x - \mu}{\sigma} = \frac{140-120}{10} = 2$$

$$Z = 2 \quad p = 0,023 \quad (2,3\%)$$

$$Z = \frac{x - \mu}{\sigma} = \frac{100-120}{10} = (-2)$$

$$Z = (-2) \quad p = 0,023 \quad (2,3\%)$$

$$1 - 0,46 = 0,954$$

---

---

---

---

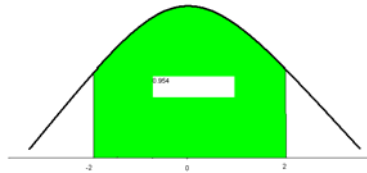
---

---

---

---

Cila zone ne kurbe ka TA mes 100-140 mmHg?




---

---

---

---

---

---

---

---

Cila zone ne kurbe ka TA > 150 mmHg?

$$Z = \frac{x - \mu}{\sigma} = \frac{150-120}{10} = 3$$

$$Z = 3 \quad p = 0,0015 \quad (1\%)$$

---

---

---

---

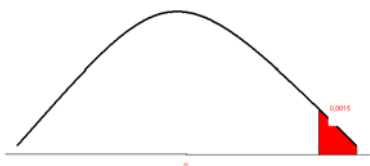
---

---

---

---

Cila zone ne kurbe ka TA > 150 mmHg?




---

---

---

---

---

---

---

---

Cila zone ne kurbe ka TA < 90 mmHg dhe mbi 150 mmHg?

$$Z = \frac{x - \mu}{\sigma} = \frac{90 - 120}{10} = (-3)$$

$$Z = (-3) \quad p = 0,0015 \quad (1\%)$$

$$Z = \frac{x - \mu}{\sigma} = \frac{150 - 120}{10} = 3$$

$$Z = 3 \quad p = 0,0015 \quad (1\%)$$

---

---

---

---

---

---

---

---

Cila zone ne kurbe ka TA < 90 mmHg dhe mbi 150 mmHg?




---

---

---

---

---

---

---

---

Cila është vlera e TA që e ndan kurben nën 95% dhe mbi 5%?

$$Z = \frac{x - \mu}{\sigma}$$

per 0,05%,  $z = 1,645$

$$1,645 = \frac{x - 120}{10} \rightarrow x = 136,45 \text{ mm Hg}$$

---

---

---

---

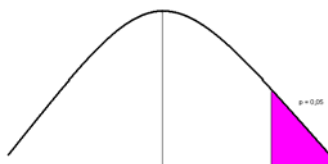
---

---

---

---

Cila është vlera e TA që e ndan kurben nën 95% dhe mbi 5%?



95% e individëve të popullatës kanë TA 136,45 mmHg

---

---

---

---

---

---

---

---

Cila është vlera e TA, që e ndan kurbën nën 97,5% dhe mbi 2,5%?

$$Z = \frac{x - \mu}{\sigma}$$

për 0,025%  $z = 1,96$

$$1,96 = \frac{x - 120}{10} \rightarrow 1,96 * 10 = x - 120$$

$$X = 19,6 + 120 \quad x = 139,6 \text{ mm Hg}$$

---

---

---

---

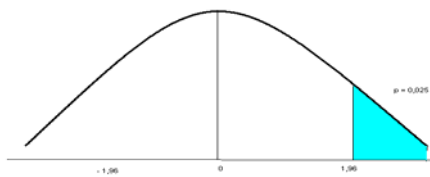
---

---

---

---

Cila eshte vlera e TA qe e ndan kurben  
nën 97,5% dhe mbi 2,5%



97,5% e individeve te popullates kane TA 139,6 mmHg

---

---

---

---

---

---

---

---

## KAPITULLI "7" KONTROLLI I HIPOTEZAVE STATISTIKORE

( 1 POPULLIMI )  
METODA: PARAMETRIKE)

Ne kapitujt 7 dhe 8 pame se si mund te perdoret nje zgjedhje per ndertimin e vleresimeve pikesore dhe intervale per parametrat e nje popullimi. Ne kete kapitull do te vazhdojme me gjykimet statistikore, duke treguar se si mund te perdoret kontrolli i hipotezave per percaktimin ne se nje pohim rreth nje vlere per parametrin e nje popullimi duhet ose jo te hidhet poshte.

Kontrolli i hipotezave fillon me ngritjen e supozimit per vlerat e mundshme te nje parametri te popullimit. Ky supozim quhet hipoteza zero ose hipoteza baze dhe shenohet me  $H_0$ . Verifikimi i vertetesise te supozimit te bere s'mund te behet pa nje baze krahasimi. Per kete qellim krahas  $H_0$  ngrihet edhe nje tjetër supozim ne thelb i kundert me supozimin baze. Ky supozim do te quhet hipoteza alternative dhe do te shenohet me  $H_a$ . Procedura e kontrollit te hipotezave perfshin perdorimin e te dhenave nga nje zgjedhje per kontrollin e dy pohimeve konkuruese qe tregohen me anen e  $H_0$  dhe  $H_a$ .

Qellimi i ketij kapitulli eshte te tregojme se si mund te behet kontrolli i hipotezave per mesataren dhe perpjesen e nje popullimi.

### 9.1 Hipotezat zero dhe alternative

Me siper pame se gjate kontrollit te hipotezave shtrohen dy pohime konkuruese, ose dy hipoteza per parametrin e nje popullimi, njeri prej te cileve konsiderohet si hipoteza baze ose zero, kurse tjetri hipoteza alternative. Ne disa aplikime nuk eshte e qarte se sa hipoteza zero dhe sa alternative duhet te formulohen. Ne menyre qe te merret nje informacion i sakte, hipotezat duhet te formulohen me kujdes. Ne kete kapitull do te paraqiten tre nga llojet me kryesore te hipotezave statistikore.

Le te shenojme me  $\mu$  vleren numerike qe shqyrtohet ne hipotezat zero dhe alternative. Formulimi i hipotezes zero dhe asaj alternative do te kete nje nga tre format e meposhtme.

$$H_0: \mu \geq \mu_0$$

$$H_0: \mu \leq \mu_0$$

$$H_0: \mu = \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu > \mu_0$$

$$H_a: \mu \neq \mu_0$$

Shpesh ne praktike percaktimi i  $H_0$  dhe  $H_a$  nuk eshte aq i qarte. Ne kete rast zgjedhja e formave me te pershtatshme per  $H_0$  dhe  $H_a$  varet nga gjykimi i studiuesit. Megjithate sic tregojne format e mesiperme, pjesa e barazimit te shprehjeve te tyre gjithnje shfaqet ne hipotezen zero. Ne zgjedhjen e formave me te pershtatshme per  $H_0$  dhe  $H_a$ , duhet te kihet parasysh se hipoteza alternative eshte ajo qe kerkohet te vendoset nga studimi i zgjedhjes.

### 9.2 Gabimet e llojit te pare dhe te dyte

Hipotezat zero dhe alternative jane pohime qe konkurojne rreth vlere se vertete te parametrin. Duke qene se hipoteza baze dhe ajo alternative jane kundërshtuese atehere është e qarte se vetem njera nga keto hipoteza eshte e vertete. Teorikisht procedura e kontrollit te hipotezave duhet te coje ne pranimin e hipotezes zero, ne qofte se  $H_0$  eshte e vertete dhe ne hedhjen poshte te saj, kur është e vertete  $H_a$ . Per fat te keq keto rezultate nuk jane gjithnje te mundshme. Meqenese kontrolli i hipotezave bazohet ne informacionin e zgjedhjes, atehere ne proceduren e kontrollit te hipotezave duhet te pranojme mundesine e gabimeve. Ne tabelen 9.1 tregohen dy llojet e gabimeve qe mund te behen gjate kontrollit te hipotezave. Megjithese gabimet ne kontrollin e hipotezave nuk mund te eliminohen, eshte e mundur te llogaritim probabilitetin e ndodhjes se tyre. Probabiliteti i berjes te gabimeve te llojit te pare zakonisht shenohet me  $\alpha$ , nderkohe qe probabiliteti i berjes te gabimeve te llojit te dyte shenohet me  $\beta$ .



Tabela 9.1 Perfundimet e gabuara dhe korrekte ne kontrollin e hipotezave

| Perfundimi      | Gjendja e natyres   |                       |
|-----------------|---------------------|-----------------------|
|                 | Ho eshte e vertete  | Ha eshte e vertete    |
| Ho pranohet     | Perfundimi i drejte | Gabimi i llojit te II |
| Ho nuk pranohet | Gabimi i llojit I   | Perfundim i drejte    |

Ne praktike, personi qe kryen kontrollin e hipotezes specifikon probabilitetin maksimal te lejueshem te berjes se gabimeve te llojit te pare, i ashtuquajtur nivel i rendesise per kriterin. Vlerat e nivelit te rendesise qe perdoren me shume ne praktike jane 0,05 dhe 0,01. Duke iu referuar rreshtit te dyte te tabelës 9. 1, veme re se konkluzioni per te hedhur poshte hipotezen Ho, ose mund te permbaje gabime te llojit te pare, ose mund te jete i drejte. Keshtu ne qofte se gabimi i llojit te pare mund te kontrollohet nepermjet zgjedhjes te nje vlere te vogel te nivelit te rendesise, ne kemi nje shkalte te larte besimi qe konkluzioni per hedhjen poshte te hipotezes zero eshte i drejte. Ne raste te tilla ne kemi nje mbeshtetje statistikore per te arritur ne konkluzionin se Ho eshte dhe gabuar dhe Ha eshte e drejte. Si rrjedhoje ndermarrja e nje veprimi ne perputhje me hipotezen alternative eshte e drejte.

Meqenese ne shume zbatime te kontrollit te hipotezave probabiliteti i berjes te gabimeve te llojit te pare eshte nen kontroll, gabimet e llojit te dyte jo gjithnje jane te kontrolluara. Keshtu kurdohere qe vendosim te pranojme Ho nuk mund te percaktojme se sa te sigurte mund te jemi me vendimin per te pranuar Ho. Per shkak te pasigurise qe na shoqeron ne berjen e gabimeve te llojit te dyte, shpesh statisticienet rekomandojne te perdorin pohimin " Ho nuk hidhet poshte " ne vend te pohimit " Ho pranohet". Duke pranuar pohimin Ho nuk hidhet poshte, ne fakt pranohet edhe gjykimi dhe veprimi. Ne te vertete duke mos pranuar drejtperdrejte Ho, statisticieni menjanon berjen e gabimeve te llojit te dyte. Kurdohere qe probabiliteti i gabimeve te llojit te dyte nuk mund te percaktohet dhe te kontrollohet, nuk duhet te konkludojme per pranimin e Ho. Ne kete rast vetem dy perfundime jane te mundshme: Ho te mos hidhet poshte, ose Ho te hidhet poshte.

Pikerisht sepse ne praktike zakonisht gabimet e llojit te dyte nuk kontrollohen, kjo nuk do te thote se ato nuk mund te kontrollohen. Ne paragrafet me poshte do te tregojme se si mund te percaktohen dhe te kontrollohen gabimet e llojit te dyte. Ne qofte se ky kontroll eshte bere sic duhet, atehere do te ishte e pershtatshme qe te ndermerrej nje veprim i bazuar ne pranimin e Ho.

### 9.3 Kontrolli i hipotezave te njeanshme per mesataren e popullimit, rasti i zgjedhjes se madhe

Kontrolli i hipotezave gjen perdorim te gjere per kontrollin e pretendimeve te prodhuesve per produktin e tyre. P.sh. nje shenje ne nje kuti te madhe kafeje pohon se kutia permban 500 gram kafe. Le ta zeme se duam te kontrollojme kete pretendim me anen e kontrollit te hipotezave statistikore.

Etapa e pare ka te beje me ngritjen e hipotezave zero dhe alternative. Per kete do te fillojme me supozimin se pretendimi i prodhuesit eshte i drejte. Veme re se ne qofte se popullimi i kutive me kafe ka nje mesatare prej 500 g ose me shume, atehere pretendimi i prodhuesit per produktin e tij eshte korrekt. Megjithate ne qofte se popullimi i kutive me kafe ka nje peshe mesatare me te vogel se 500g, pretendimi i prodhuesit nuk eshte i drejte.

Po te shenojme me  $\mu$  peshen mesatare te popullimit te kutive me kafe, hipotezat zero dhe alternative do te jene si me poshte.

$$H_0: \mu \geq 500$$

$$H_a: \mu < 500$$

Ne qofte se te dhenat e zgjedhjes tregojne se Ho nuk mund te hidhet poshte, atehere kjo do te thote se faktet statistikore jane ne mbeshtetje te pretendimit te prodhuesit per peshen e kutive te kafes. Megjithate, ne qofte se te dhenat e zgjedhjes tregojne se Ho mund te hidhet poshte,



konkludojme se hipoteza alternative  $H_a$  është e vertete. Në këtë rast do të ishte më vend të pretendimi se kutitë janë të mbushura nën peshën e tyre.

Le ta zëmë se janë zgjedhur në mënyrë të rastit 36 kuti me kafe. Në qoftë se pesha mesatare e këtyre kutive është më e vogël se 500 g, rezultatet e zgjedhjes do të çojnë në një dyshim për hipotezën  $H_0$  se  $\mu \geq 500$ . Por shtrohet problemi, sa më pak se 500 duhet të jetë mesatarja e zgjedhjes  $\bar{x}$  përpara se të rrezikonim të benim një gabim të llojit të parë dhe të akuzojmë padrejtesisht firmën për vendosjen e shenjës për peshën mbi kutitë me kafe?

Për t'iu përgjigjur kësaj pyetje le të supozojmë se hipoteza  $H_0: \mu \geq 500$  është e vertete. Nga studimi i shpërndarjes së zgjedhjes në kapitullin 7 dimë se kurdoherë që vëllimi i zgjedhjes është i madh ( $n \geq 30$ ), shpërndarja e zgjedhjes për mesataren e zgjedhjes  $\bar{x}$  mund të përafrohet me shpërndarjen probabilitare normale. Figura 9.1 tregon shpërndarjen e zgjedhjes për  $\bar{x}$  kur hipoteza zero është e vertete,  $\mu \geq 500$ .

Figura 9.1 Shpërndarja e zgjedhjes për mesataren kur hipoteza zero është e vertete

$\mu$

Vlera e  $z = (\bar{x} - 500) / \sigma_{\bar{x}}$ , jep numrin e devijimeve standarte që mesatarja e zgjedhjes  $\bar{x}$  është larg mesatares së popullimit  $\mu$ . Për rastin e kontrollit të hipotezës për mesataren e zgjedhjes, madhësia  $z$  do të përdoret si statistikë kontrolli për të përcaktuar se sa larg mesatarja e zgjedhjes duhet të jetë nga mesatarja e popullimit  $\mu = 500$ , në mënyrë që të justifikohet hedhja poshtë e hipotezës  $H_0$ . Vëmë në dukje se për  $z = -1$  mesatarja e zgjedhjes ndodhet 1 devijim standart poshtë mesatares së popullimit  $\mu = 500$  g, vlera  $z = -2$  do të thotë se mesatarja e zgjedhjes ndodhet 2 devijime standarte nën mesataren e popullimit, kurse  $z = -3$  tregon se mesatarja e zgjedhjes bie 3 devijime standarte më poshtë se mesatarja e popullimit e kështu me radhë. Në qoftë se hipoteza zero është e vertete, marrja e një vlere të  $z < -3$  ka shumë pak të ngjarë që të ndodhë. Pyetja që shtrohet është: sa e vogël duhet të jetë statistika e kontrollit  $z$  përpara se të kemi një evidence të mjaftueshme për të hedhur poshtë hipotezën zero?

Nga tabela e shpërndarjes normale mund të gjendet probabiliteti që të vrojtohet një vlerë e  $\bar{x}$  me shumë se 1,645 devijime standarte poshtë mesatares së popullimit  $\mu = 500$  ose cka njelloj  $P\{|Z| < 1.645\}$ , probabilitet i cili rezulton i barabartë me 0,95. Kështu në qoftë se do të hedhim poshtë hipotezën zero, sa herë që vlera e  $z = (\bar{x} - 500) / \sigma_{\bar{x}}$  është më e vogël se -1,645, probabiliteti i berjes së gabimeve të llojit të parë do të jetë 0,05. Në qoftë se 0,05 e konsiderojmë si një nivel të pranueshëm të probabilitetit të gabimeve të llojit të parë, atëherë hipoteza zero do të hidhet poshtë sa herë që statistika e kontrollit të tregojë se mesatarja e zgjedhjes të jetë me shumë se 1,645 poshtë mesatares së popullimit  $\mu = 500$ . Kështu hipoteza zero do të hidhet poshtë në qoftë se  $z < -1.645$ .

Metodologjia e kontrollit të hipotezave kërkon që të specifikojmë probabilitetin maksimal të lejueshëm të berjes të gabimeve të llojit të parë. Siç kemi theksuar në paragrafin e mëparshëm, ky probabilitet maksimal quhet nivel i rëndësishëm të kriterit. Ai shënohet me  $\alpha$  dhe paraqet probabilitetin e hedhjes poshtë të  $H_0$  në kushtet kur ajo është e vertete, që në literaturën



statistikore njihet me emrin gabim i llojit te pare. Percaktimi i nivelit te rendesise eshte detyre e studjuesit ose e marresit te vendimit. Ne qofte se kostoja e berjes te gabimeve te llojit te I eshte e madhe, atehere per nivelin e rendesise do te zgjidhej nje vlere e vogel. Ne qofte se kostoja nuk eshte teper e madhe, nivelit te rendesise mund t'i jepej nje vlere me e madhe.

Le ta zeme se ne rastin e shembullit te peshes te kutive me kafe bejme pohimin e meposhtem. Ne qofte se firma ploteson standartin per peshen ekzakt ( $\mu = 500$ ), atehere do te donim qe ne 99% te rasteve ndaj firmes te mos ndermerrim asnje veprim. Meqenese nuk do te donim te akuzonim padrejtesisht firmen per mosmbushjen sa duhet te kutijave, kete gabim do te donim ta benim shume rralle, pra vetem ne 1% te rasteve.

Nga pohimi i mesiperme rrjedh se probabiliteti i berjes te gabimeve te llojit te pare eshte 0,01. Nga tabela e shperndarjes normale mund te shohim se vlere  $z_{kr}$ , e tille qe  $P\{|Z| < z_{kr}\} = 0.01$  është 2,33. Si rrjedhim, per hedhjen poshte te hipotezes zero mund te vendosim rregullin e meposhtem: hipoteza zero hidhet poshte ne qofte se vlere e vrojtuar e kriterit  $z_v = \frac{\bar{x} - \mu}{\sigma_x} < -2.33$ . Zona  $]-\infty; z_{kr}[$  e vlerave te kriterit quhet zone kritike, kurse komplimenti i saj zone e lejuar e vlerave te kriterit.

Ne qofte se vlere e  $\bar{x}$  eshte e tille qe statistika e kontrollit  $z$  bie ne zonen kritike, atehere hipoteza  $H_0$  hidhet poshte dhe pranohet hipoteza  $H_a$ . Nga ana tjeter ne qofte se vlere e  $\bar{x}$  eshte e tille qe statistika  $z$  nuk bie ne zonen e hedhjes poshte, hipoteza zero nuk mund te hidhet poshte. Ne keto raste themi se kriteri i kontrollit te hipotezave eshte kriter i njeanshem.

Le ta zeme se merret nje zgjedhje prej 36 kuti, ku pesha mesatare e tyre ka rezultuar e barabarte me  $\bar{x} = 480$  g dhe nga nje studim i meparshem kemi gjetur nje vleresim per devijimin standart te popullimit  $\sigma = 45$ . Per  $\sigma_{\bar{x}} = \sigma / \sqrt{n}$  vlere e statistikes se kontrollit rezulton si me poshte.

$$z = \frac{\bar{x} - 500}{\sigma / \sqrt{n}} = \frac{480 - 500}{45 / \sqrt{36}} = -2.67$$

Duke qene se  $z_v = -2.67$  ben pjese ne zonen kritike jemi te justifikuar qe per nivelin e rendesise  $\alpha = 0.01$  te konkludojme se  $\mu < 500$ . Me kete rezultat nga ana statistiore kemi te drejte te themi qe firma nuk eshte korrekte ne mbushjen e kutive sipas standartit.

Le ta zeme tani se 36 kutite e zgjedhura na kane dhene nje peshe mesatare  $\bar{x} = 485$  g. Ne kete rast, me qene se vlere e statistikes se kontrollit do te ishte  $z_v = -1$ , vlere e cila bie ne zonen e lejuar te vlerave te kriterit, konkludojme se hipoteza zero nuk mund te hidhet poshte. Asnje gjykim tjeter nuk mund te behet dhe asnje veprim nuk mund te ndermerret ndaj firmes prodhuese.

Vlere e  $z$  qe vendos kufirin e zones te hedhjes poshte te hipotezes zero quhet vlere kritike. Per percaktimin e vleres kritike ne bejme supozimin se hipoteza zero eshte e vertete. Por per firmen e kafese hipoteza zero eshte e vertete ne rastin kur  $\mu \geq 500$ , kurse ne kemi marre ne konsiderate vetem rastin kur  $\mu = 500$ . Shtrojme pyetjen, c'mund te thuhet ne rastin kur  $\mu > 500$ ? Ne qofte se  $\mu > 500$ , probabiliteti i berjes te gabimeve te llojit te I do te jete me i vogel se ne rastin kur  $\mu = 500$ ; d.m.th. ne kete rast madje eshte me i vogel se ai per vleren e statistikes se kontrollit qe bie ne zonen kritike. Meqenese objektivi i procedures se kontrollit te hipotezes eshte te kufizojme probabilitetin maksimal te berjes te gabimeve te llojit te I, vlere kritike per kriterin vendoset per supozimin se  $\mu = 500$ .

### Permbledhje: Kontrolli i njeanshem per mesataren e popullimit

Le te pergjithsojme proceduren e kontrollit te hipotezave per kontrollin e njeanshem te mesatares se popullimit. Per kete kufizohemi me rastin e zgjedhjes se madhe ( $n \geq 30$ ) ku teorema qendrore limite na lejon te supozojme per nje shperndarje normale te zgjedhjes per  $\bar{x}$ . Ne rastin e



zgjedhjes se madhe, kur  $\sigma$  eshte e panjohur ate e zevendesojme me devijimin standart te zgjedhjes. Forma e pergjithshme e kontrollit te hipotezes per mesataren e nje popullimi ne kete rast eshte si me poshte.

$$H_0: \mu \geq \mu_0 \quad H_a: \mu < \mu_0 \quad \text{ose} \quad H_0: \mu \leq \mu_0 \quad H_a: \mu > \mu_0$$

Statistika e kontrollit do te mbetet po ajo:  $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ . Ne qofte se  $\sigma$  eshte e panjohur ne vend te

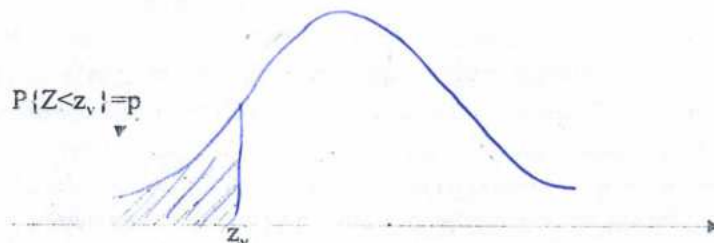
saj marrim devijimin standart te zgjedhjes s. Rregulli i hedhjes poshte te hipotezes per nivelin e rendesise  $\alpha$  eshte:  $H_0$  rrezohet kur  $z_v < -z_{kr}$  per rastin e pare dhe  $z_v > z_{kr}$  per rastin e dyte

### Perdorimi i vlerave te p

Le te shikojme nje menyre tjeter qe mund te perdoret ne kontrollin e hipotezave, e cila bazohet ne te ashtuquajturen vlere e p. Ketu do te tregojme se midis vlerave te p dhe vlerave te z ekziston nje lidhje, d.m.th. ne rastin e nje zgjedhje vlere e p mund te llogaritet nga vlere e z. Vlere e p mund te perdoret gjithashtu ne marrjen e vendimit per hedhjen poshte ose jo te hipotezes zero.

Me supozimin se hipoteza zero eshte e vertete, vlere e p shpreh probabilitetin e marrjes te nje rezultati zgjedhje qe eshte te pakten po aq e pamundur te ndodhe sa c'eshte vlere e vrojtuar. Keshtu ne rastin e firmes se kafese, zona e hedhjes poshte te hipotezes zero eshte e njeanshme e majte, prandaj vlere e p shpreh probabilitetin qe nje mesatare e zgjedhjes te jete me e vogel ose e barabarte me ate qe eshte vrojtuar.

Figura 9.2 Zona e hedhjes poshte te hipotezes per mesataren per kriterin e njeanshem



Le te llogaritim vleren e p qe i perket mesatares se zgjedhjes  $\bar{x} = 480$  gr per rastin e shembullit te firmes se kafese. Vlere e p ne kete rast eshte probabiliteti qe te marrim nje vlere te mesatares se zgjedhjes me te vogel ose te barabarte me vleren e vrojtuar  $\bar{x} = 480$ , ne rastin kur vlere hipotetike e mesatares se popullimit eshte 500. Duke u bazuar ne njevleshmerine e ngarjeve

$\{\bar{x} < 480\}$ ;  $\{\frac{\bar{x} - 500}{\sigma_{\bar{x}}} < \frac{480 - 500}{\sigma_{\bar{x}}}\}$  dhe  $\{Z < -2.67\}$ , si dhe duke shfrytezuar tabelen e shprndarjes

normale gjejme  $p = P\{\bar{x} < 480\} = 0,0038$ . Kjo eshte pikerisht vlere e p, e cila tregon nje vlere shume te vogel te probabilitetit qe te merret nje mesatare e zgjedhjes me e vogel se  $\bar{x} = 480$ , nga popullimi me mesatare te barabarte me  $\mu = 500$ .

Vlere e p mund te perdoret per marrjen e vendimit ne kontrollin e nje hipoteze, duke vene ne dukje se ne qofte se vlere e p eshte me e vogel se niveli i rendesise  $\alpha$ , atehere vlere e statistikes se kontrollit duhet te jete ne zonen e hedhjes poshte te hipotezes. Ne menyre analoge, ne qofte se vlere e p eshte me e madhe ose e barabarte me  $\alpha$ , vlere e statistikes se kontrollit nuk eshte ne zonen e hedhjes poshte te hipotezes. Per shembullin e firmes se kafese fakti se vlere e p = 0,0038



është me e vogël se niveli i rendesise  $\alpha = 0,01$ , tregon se hipoteza zero duhet të hidhet poshtë. Për një nivel rendesie të dhënë  $\alpha$  dhe për çdo kontroll hipoteze, vendimi për të hedhur poshtë ose jo një hipotezë  $H_0$  në termat e vlerës së p është:

*Hipoteza  $H_0$  hidhet poshtë në qoftë se vlera e  $p < \alpha$ .*

Për kontrollin e një hipoteze për një nivel rendesie të dhënë  $\alpha$ , vlera e  $p$  dhe statistika koresponduese e saj  $z$  çojnë gjithnjë në të njëjtin perfundim. Kur zona e hedhjes poshtë të hipotezës është e njëanshme dhe në anën e majtë të shpërndarjes së zgjedhjes, vlera e  $p$  është sipërfaqja nën vijën në të majtë të vlerës së statistikës së kontrollit. Kur zona e hedhjes poshtë të hipotezës është në anën e sipërme të shpërndarjes së zgjedhjes, vlera e  $p$  është sipërfaqja nën vijën në të djathtë të vlerës së statistikës së kontrollit. Kështu një vlerë e vogël e  $p$  tregon një rezultat të zgjedhjes që nuk është i zakonshëm me supozimin se  $H_0$  është e vertetë. D.m.th. vlerat e vogla të  $p$ , çojnë në hedhjen poshtë të  $H_0$ . Nga ana tjetër vlerat relativisht të mëdha të  $p$ , tregojnë se hipoteza zero nuk mund të hidhet poshtë.

#### 9.4 Kontrolli i hipotezave të dyanshme për mesataren e popullimit, rasti i zgjedhjes së madhe

Kontrolli i hipotezave të dyanshme ndryshon nga rregulli i kontrollit të njëanshem të tyre për shkak të zones së dyanshme të hedhjes poshtë të hipotezës, e cila shtrihet në të dy anët e shpërndarjes së zgjedhjes. Le të tregojmë me një shembull se si dhe përse ndërtohet kontrolli i hipotezave të dyanshme.

Supozojmë se një linjë prodhimi shërben për mbushjen e kutive me një peshe standarte të barabartë me 500 g. Mbushja me shumë ose me pak e kutijave përben një problem serioz për firmën dhe në qoftë se kjo ndodh, prodhimi i linjës duhet të ndërpritet. Nga të dhënat e mëparshme ka rezultuar se devijimi standart është i barabartë me  $\sigma = 25$  g. Le të zëmë se inspektori i cilesisë ben një zgjedhje prej 30 kutish në çdo dy ore dhe gjatë kësaj kohe merr vendim të ndalojë ose jo linjën e prodhimit për të nderhyrë në rregullimin e saj. Për këtë kërkojmë të përcaktojmë për një nivel rendesie  $\alpha = 0,05$  të dhënë që ne pare, se cili është rregulli i hedhjes poshtë të  $H_0$  në procedurën e kontrollit të hipotezave. Kështu p.sh. mund të shtrohen keto probleme: c'fare veprimi rekomandojmë në qoftë se mesatarja e zgjedhjes ka rezultuar e barabartë me 550 g, po në qoftë se mesatarja e zgjedhjes ka rezultuar e barabartë me 460 g?. Cilat janë vlerat e  $p$  në keto raste?

Si pjesë e programit të kontrollit të cilesisë, në mënyrë periodike kryhet një zgjedhje për një numër të caktuar të kutiave nga linja e prodhimit duke matur peshën e tyre. Duke mos patur arsye të dyshojmë se procesi i prodhimit funksionon normalisht, ngrejme keto hipoteza zero dhe alternative.

$$H_0: \mu = 500$$

$$H_a: \mu \neq 500$$

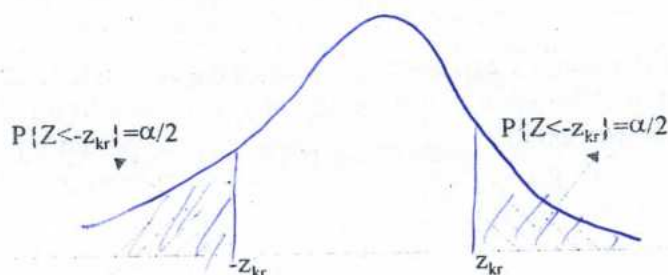
Kriteri i kontrollit do të merret po ai që pame në paragrafin e mëparshëm:  $Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ . Hapi i

dytë në kontrollin e hipotezës të ngritur është përcaktimi i zones së hedhjes poshtë të hipotezës ose të zones kritike. Kështu, kërkojmë të hedhim poshtë pretendimin që  $\mu = 500$ , në rastin kur vlera e  $z$  tregon se mesatarja e zgjedhjes është në mënyrë të rendesishme me e vogël, ose me e madhe se 500. Kështu hipoteza zero do të hidhet poshtë, në qoftë se vlerat e statistikës së kontrollit janë larg mesatares së zgjedhjes, të cilat mund të ndodhen si në anën e majtë ashtu edhe në anën e djathtë të saj. Për këtë arsye kriteri i kontrollit të hipotezave quhet kriteri i dyanshem. Duke patur parasysh procedurën e kontrollit të hipotezave të paraqitur në paragrafin e mëparshëm në fillim specifikojmë nivelin e rendesisë si probabiliteti maksimal i lejueshëm të gabimeve të llojit të parë. Le të zëmë se kemi zgjedhur nivelin e rendesisë  $\alpha = 0,05$ . Kjo do të thotë se probabiliteti që të konkludojmë se pesha mesatare e kutijave nuk është 500 gr. kur në fakt ajo është e tillë është e barabartë me 0,05.



Ne figuren 9.3 tregohet shpërndarja e zgjedhjes së  $z$  me zonën e dyaneshme të hedhjes poshte të hipotezës brenda kritereve  $\alpha = 0,05$ . Për kriterin e kontrollit të dyaneshëm zona e hedhjes poshte të hipotezës përcaktohet në mënyrë të tillë që sipërfaqja ose probabiliteti në të dy anët e shpërndarjes të jetë e barabartë me  $\alpha/2$ . Vlera e  $z$  që jep një sipërfaqe të barabartë me 0.05 në të dy anët e shpërndarjes mund të gjendet në tabelën e shpërndarjes probabilitare standarte:  $z_{kr}=1.96$ . Për arsye simetrie skaji tjetër i zonës së lejuar do të merret  $-z_{kr}$ .

Fig 9.3 Zona e hedhjes poshte të hipotezës për kriterin e dyaneshëm



Le ta zëmë se kemi marrë një zgjedhje prej 30 kutish ku pesha mesatare e tyre ka rezultuar 550 gr. dhe devijimi standart i zgjedhjes  $s = 25$  gr. Duke përdorur vlerën e  $\mu$  nga hipoteza zero dhe devijimin standart të zgjedhjes si vlerësim për devijimin standart të popullimit vlera e statistikës së kontrollit është

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{550 - 500}{25 / \sqrt{36}} = 0,33$$

Në përputhje me rregullin e hedhjes poshte të hipotezës,  $H_0$  nuk mund të hidhet poshte. Rezultatet e zgjedhjes tregojnë se manaxheri i kontrollit të cilësise nuk ka arsye të dyshojë në supozimin se linja e prodhimit për mbushjen e kutijave është në përputhje me standartin e caktuar.

#### Permbledhje: Kriteri i dyaneshëm i kontrollit për mesataren e popullimit

Le të shënojmë me  $\mu$  vlerën mesatare sic pretenduam në hipotezën zero. Forma e përgjithshme e kriterit të dyaneshëm për mesataren e popullimit në rastin e zgjedhjes së madhe ( $n \geq 30$ ) është si më poshte.

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

$$\text{Statistika e kontrollit: } z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Në qoftë se  $\sigma$  është e panjohur në vend të saj marrim vlerësimin pikësor  $s$ . Rregulli i hedhjes poshte të hipotezës për nivelin e rëndësise  $\alpha$  është:  $H_0$  hidhet poshte në qoftë se  $z < -z_{kr}$  ose kur  $z > z_{kr}$ .

#### Vlerat e p për kriterin e dyaneshëm

Më supozimin se hipoteza zero është e vertetë, vlera e  $p$  shpreh probabilitetin e marrjes të një rezultati të zgjedhjes, që është të paktën po aq e pamundur sa ç'është ai i vërtetuar. Më supozimin se hipoteza  $H_0$  është e vertetë, një vlerë e vogël e  $p$  tregon se rezultati i zgjedhjes është praktikisht



i pamundur. Keshtu si ne rastin e kontrollit te njeanshem te hipotezave, nje vlere e vogel e p con ne hedhjen poshte te hipotezes  $H_0$ .

Le te llogaritim vleren e p per shembullin e firmes prodhuese te mbushjes te kutive me kafe.  $P\{|\bar{x}| > 550\} = P\left\{\left|\frac{\bar{x} - 500}{\sigma_{\bar{x}}}\right| > \frac{550 - 500}{\sigma_{\bar{x}}}\right\} = P\{|Z| > 0.33\} = P\{Z < -0.33\} + P\{Z > 0.33\}$ . Per arsye simetrie dy probabilitetet e fundit jane te barabarte njeri me tjetrin. Perfundimisht,  $P\{|\bar{x}| > 550\} = 2 * 0.3707 = 0.7414$ . Me qene se  $0.7414 > 0.05$  shohim se hipoteza zero nuk mund te hidhet poshte.

*Lidhja midis vleresimit intervalor dhe kontrollit te hipotezave.* Ne kapitujt 8 dhe 9 kemi pare procedurat statistikore qe mund te perdoren per te bere gjykime rreth vleres se nje mesatareje. Ne kapitullin 8 u perqendruam ne vleresimin intervalor, kurse ne kapitullin 9 ne kontrollin e hipotezave. Ne rastin e vleresimit intervalor mesatarja e popullimit  $\mu$  ishte e panjohur. Porsa ishte bere zgjedhja dhe ishte llogaritur mesatarja e zgjedhjes  $\bar{x}$ , rreth vleres se  $\bar{x}$  ndertuam nje interval qe kishte nje shans te mire per perfshirjen e vleres se mesatares se popullimit  $\mu$ . Keshtu ne proceduren e vleresimit intervalor percaktohej intervali i besimit per mesataren e popullimit me koeficient besimi  $1 - \alpha$ . Ne rastin e zgjedhjes se madhe vleresimi intervalor i mesatares se popullimit u paraqit ne formen e meposhtme.

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Gjate kontrollit te hipotezave ne fillim kerkohej te benim nje supozim rreth nje vlere te parametrut te pergjithshem. Ne rastin e mesatares se popullimit, kontrolli i dyanshem i hipotezave ka formen:

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

ku  $\mu_0$  eshte nje vlere hipotetike per mesataren e popullimit. Duke perdorur rregullin e hedhjes poshte te hipotezes se dhene me anen e 9.3, shohim se zona ku hipoteza nuk hidhet poshte perfshin vlerat e mesatares se zgjedhjes qe bien brenda - z dhe + z devijime standarte nga mesatarja. Keshtu zona e mos hedhjes poshte te hipotezes  $H_0$  per mesataren e zgjedhjes, ne rastin e zgjedhjes se madhe dhe per nivelin e rendesise  $\alpha$ , ka formen e meposhtme.

$$\mu \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Nje veshtrim i afert i ketyre formulave do te na jape nje pamje te brendshme per lidhjen midis vleresimit intervalor dhe kontrollit te hipotezave. Ne vecanti theksojme se te dyja procedural kerkojne llogaritjen e vlerave te  $z_{\alpha/2}$  dhe  $\sigma / \sqrt{n}$ . Duke e perqendruar vemendjen tek  $\alpha$ , shohim se nje koeficient besimi  $(1 - \alpha)$  per vleresimin intervalor, i pergjigjet nje niveli rendesie  $\alpha$  per kontrollin e hipotezave. Per shembull intervali i besimit 95%- esin per vleresimin intervalor i korespondon nje niveli rendesie 0,05 per kontrollin e hipotezave. Perverc kesaj duke pare shprehjet e formulave te mesiperme shohim se meqenese  $z_{\alpha/2} (\sigma / \sqrt{n})$  eshte e njejta vlere qe shtohet dhe hiqet per te dy shprehjet rrjedh se, ne qofte se  $\bar{x}$  bie ne zonen e lejuar, vlere hipotetike  $\mu_0$  do te bjere ne intervalin e besimit dhe anasjelltas. Keto arsyetime cojne ne perdorimin e rezultateve te intervalit te besimit per kontrollin e hipotezave si me poshte.

1. Marrim nje zgjedhje te rastit nga popullimi ne studim dhe duke perdorur mesataren e zgjedhjes  $\bar{x}$  ndertojme intervalin e besimit per mesataren e popullimit  $\mu$ :  $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ .



2. Ne qofte se intervali i besimit permban vleren hipotetike  $\mu_0$ , hipoteza  $H_0$  nuk hidhet poshte, ne rast te kundert  $H_0$  hidhet poshte.

Le te kthehemi tek shembulli i firmes se prodhimit te kafese per te demonstruar perdorimin e intervalit te besimit per kontrollin e hipotezave. Hipotezat qe shtrihen ne kete rast jane te dyanshme.

$$H_0: \mu = 500$$

$$H_a: \mu \neq 500$$

Per kontrollin e kesaj hipoteze me nje nivel rendesie  $\alpha = 0,05$ , marrim nje zgjedhje prej 36 kutish nga peshimi i te cilave ka rezultuar pesha mesatare 515gr dhe devijimi standart i barabarte me 60 gr. Ne baze te ketyre rezultateve per  $z = 1,96$ , intervali 95% i besimit per mesataren e popullimit merr formen:

$$\bar{x} \pm z_{0,025} \frac{\sigma}{\sqrt{n}}$$

$$515 \pm 1,96 \frac{60}{\sqrt{36}} = 515 \pm 19,6, \text{ ose } [495,4; 534,6]$$

Ky rezultat e ben manaxherin e kontrollit te cilesise te konkludoje me koeficient besimi 95%, se pesha mesatare e kutive do te jete nga 495,4 deri ne 534,6 g. Meqenese vlera hipotetike e mesatares se popullimit qe eshte  $\mu = 500$  g bie ne intervalin e besimit, arrijme ne perfundimin se hipoteza  $H_0: \mu = 500$  nuk mund te hidhet poshte.

Veme re se analiza qe beme dhe shembulli qe paraqitem i perkasin rastit te kriterit te dyanshem te kontrollit te hipotezave. Megjthate nje lidhje e njejte per kontrollin e hipotezave ekziston edhe per parametrat e tjere te popullimit. Pervec kesaj kjo lidhje mund te zgjerohet edhe per rastin e kriterit te njeanshem te kontrollit te njeanshem te hipotezave.

### 9.5 Kontrolli i hipotezave per mesataren e zgjedhjes, rasti i zgjedhjes se vogel

Le ta zeme, tani se velli i zgjedhjes eshte i vogel, d.m.th.  $n < 30$ . Ne qofte se popullimi ka shperndarje normale dhe devijimi standart i popullimit njihet, atehere statistika qe perdoret per zgjedhjen e madhe mund te perdoret si statistike kontrolli te hipotezave edhe ne rastin e zgjedhjes se vogel. Ne rastin kur devijimi standart i popullimit  $\sigma$  nuk njihet, atehere si vleresim per te merret devijimi standart i zgjedhjes  $s$ . Ne qofte se eshte e arsyeshme te supozojme se popullimi ka shperndarje normale, per te bere gjykime rreth mesatares se popullimit mund te perdoret shperndarja  $t$ . Statistika e kontrollit te hipotezave rreth, mesatares se popullimit ne kete rast ka formen e meposhtme.

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Provohet se ne kushtet kur  $H_0$  eshte e vertete, kjo statistike ka shperndarje studenti,  $t$  me  $n - 1$  shkalle lirie. Procedura e zgjedhjeve te vogla per kontrollin e hipotezave per mesataren e popullimit, eshte shume e ngjashme me procedurat per zgjedhjen e madhe te paraqitura me lart.

*Nje kriter i dyanshem.* Si shembull te zbatimit te kriterit te dyanshem per kontrollin e hipotezave per mesataren e popullimit ne rastin e zgjedhjes se vogel marrim problemin e meposhtem. Le ta zeme se nje proces prodhimi eshte destinuar per mbushjen e shisheve me peshe 400 g. Ne qofte se shishet nuk mbushen sic tregohet ne kushtet e standartit, atehere konsumatoret nuk do te marrin sasine e produktit qe tregohet ne etiketen e shishes. Nga ana tjeter ne qofte se shishet mbushen me shume se sa parashikohet sipas standartit, atehere firma prodhuese do te dale me humbje. Keshtu mbushja nden peshen standarte, ose mbi peshen standarte ne te dy rastet nuk eshte gje e mire. Per



kete arsye personeli i kontrollit te cilesise ne menyre periodike kryen zgjedhje te rastit prej 8 shishesh dhe kontrollon hipotezen  $H_0: \mu = 400$  kundrejt hipotezes alternative  $H_a: \mu \neq 400$ .

Ne qofte se hipoteza zero hidhet poshte, manaxheri i prodhimit do te kerkoje nderprerjen e procesit te prodhimit per te nderhyre ne rregullimin e linjes per mbushjen e shisheve sipas standartit, d.m.th. me peshe mesatare 400 g. Le ta zeme se nga zgjedhja kane rezultuar keto vlera per peshat: 410, 420, 390, 405, 425, 415, 420, 395. Shtrohet pyetja cfare veprimi duhet te ndermarre manaxheri per nivelin e rendesise  $\alpha = 0,05$ ? Per kete ne fillim le te supozojme se popullimi i peshave eshte i shperndare normalisht.

Ne fillim llogaritim mesataren e zgjedhjes dhe devijimin standart te saj:  $\bar{x} = 410$  dhe  $s = 12,5$ .

Per kriterin e dyanshem dhe nivelin e rendesise 0,05, vlerat e t qe percaktojne zonen e hedhjes poshte te hipotezes i kerkojme ne formen  $-t_{0,025}$  dhe  $t_{0,025}$ . Ne tabelen e shperndarjes t, per numrin e shkalleve te lirise  $n - 1 = 8 - 1 = 7$ , gjejmë se  $-t_{0,025} = -2,365$  dhe  $t_{0,025} = 2,365$ . Keshtu, rregulli i hedhjes poshte te hipotezes është:  $H_0$  hidhet poshte nqs.  $t_v < -2,365$  ose per  $t_v > 2,365$ . Duke qene se ne rastin tone  $t_v = 2,264$  nuk bie ne zonen e hedhjes poshte te hipotezes, hipoteza zero  $\mu = 400$  nuk mund te hidhet poshte. Prej kendej rrjedh se nuk ka nje evidence te mjaftueshme per nderhyrjen ne procesin e prodhimit.

### 9.6 Kontrolli i hipotezave per perpjesen e popullimit

Ne qofte se me simbolin p quajme perpjesen e popullimit dhe me  $p_0$  quajme nje vlere hipotetike te vecante, atehere format e kontrollit te hipotezave per perpjesen e popullimit jane si me poshte.

$$\begin{array}{lll} H_0: p \geq p_0 & H_0: p \leq p_0 & H_0: p = p_0 \\ H_a: p < p_0 & H_a: p > p_0 & H_a: p \neq p_0 \end{array}$$

Dy format e para jane kritere te njeanshme, kurse forma e trete eshte kriter i dyanshem. Kontrolli i hipotezave per perpjesen e popullimit bazohet ne diferencen midis perpjeses se zgjedhjes  $\bar{p}$  dhe vleres hipotetike te saj  $p_0$ . Metodot qe perdoren per ndertimin e kritereve te kontrollit ne rastin e perpjeses jane shume te ngjashme me procedurat e perdorura per kontrollin e hipotezave per mesataren e popullimit. I vetmi ndryshim eshte qe per ndertimin e statistikes se kontrollit perdorim perpjesen e zgjedhjes  $\bar{p}$  dhe devijimin standart te saj  $\sigma_p$ .

### 9.7 Kontrolli i hipotezave dhe marrja e vendimit

Ne paragrafin 9.1 dalluam tre lloje te situatave per kontrollin e hipotezave. Procedura e kontrollit te hipotezave e paraqitur deri ketu ka nje zbatim te kufizuar ne situaten e marrjes se vendimit, sepse nuk eshte e pershtatshme te pranohet hipoteza  $H_0$  dhe te ndermerret nje veprim i bazuar ne perfundimin se  $H_0$  eshte e vertete. Arsyeja per mos ndermarrjen e vendimit kur rezultatet e kontrollit nuk kundershtojne hipotezen  $H_0$ , eshte se vendimi per te pranuar hipotezen  $H_0$  ekspozon marresin e vendimit te beje nje gabim te llojit te II-te, d.m.th. te pranoje  $H_0$  kur ajo eshte e gabuar. Per proceduren e kontrollit te hipotezes te pershkruar ne paragrafin e meparshem, probabiliteti i berjes te gabimit te llojit te I kontrollohej me dhenien e nivelit te rendesise per kriterin. Megjithate probabiliteti i berjes te gabimeve te llojit te II-te nuk vihej nen kontroll.

Ne situatat e marrjes e vendimit rekomandohet qe procedura e kontrollit te hipotezave te zgjerohet, duke marre ne konsiderate edhe probabilitetin e berjes te gabimeve te llojit te II-te. Meqenese ne kushtet ku hipoteza zero nuk hidhet poshte duhet marre nje vendim dhe nje veprim qe pason kete vendim, njohja e probabilitetit te berjes te gabimeve te llojit te II-te ka te beje me percaktimin e besimit qe mund te kemi me perfundimin per te pranuar se hipoteza  $H_0$  eshte e vertete. Ne paragrafet qe pasojne do te llogaritim probabilitetin e berjes te gabimeve te llojit te dyte dhe do te shikojme se si mund te rregullohet vellimi i zgjedhjes per te kontrolluar gabimet e llojit te II-te



### 9.8 Llogaritja e probabilitetit të gabimeve të llojit të II-te

Ne kete paragraf do te tregojme si te llogaritim probabilitetin e berjes te gabimeve te llojit të II ne rastin e kontrollit te hipotezave per mesataren e popullimit. Le te-illustrime proceduren e kontrollit te hipotezes per rastin e shembullit te pranimit te nje malli te paraqitur me poshte. Hipotezat zero dhe alternative per afatsherbimin mesatar te kontigjentit te baterive mund te paraqiten si me poshte.

$$H_0: \mu \geq 120$$

$$H_a: \mu < 120 \quad (\mu_0 = 120 \text{ ore jetegjatesi mestare, } n = 36)$$

Ne qofte se hipoteza  $H_0$  rrezohet, kjo do te thote se duhet te merret vendimi per kthimin e mallit tek furnizuesi, sepse afatsherbimi mesatar eshte me i vogel se 120 ore. Ne qofte se hipoteza  $H_0$  nuk hidhet poshte, do te merret vendimi per te pranuar kontigjentin e baterive.

Supozojme se per kontrollin e hipotezes zgjidhet niveli i rendesise  $\alpha = 0,05$ . Statistika e kontrollit eshte:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 120}{\sigma \sqrt{n}}$$

Per  $\alpha=0,05$ ,  $z_{\alpha} = -1,645$ .  $H_0$  hidhet poshte ne qofte se  $z_v < -1,645$ .

Le ta zeme se merret nje zgjedhje prej 36 baterish dhe se nga nje vrojtim i meparshem ka rezultuar se devijimi standart per popullimin eshte  $\sigma = 12$  ore. Rregulli i meparshem i hedhjes poshte te hipotezes behet si me poshte.

$$z = \frac{\bar{x} - 120}{12 / \sqrt{36}} < -1,645$$

Duke zgjidhur ekuacionin e mesiperme kundrejt  $\bar{x}$  gjejmë se hipoteza  $H_0$  do te hidhet poshte kur

$$\bar{x} < 120 - 1,645 (12 \sqrt{36}) = 116,71$$

Hedhja poshte e hipotezes  $H_0$  kur  $\bar{x} < 116,71$ , do te thote se do te marrim vendimin per te pranuar mallin kurdohere qe  $\bar{x} \geq 116,71$ . Ne baze te ketij informacioni mund te llogaritim probabilitetin qe lidhet me berjen e gabimeve te llojit te dyte. Ne fillim kujtojmë se gabimet e llojit te II-te behen kurdohere qe afatsherbimi mesatar i vertete i baterive eshte me i vogel se 120 ore, nderkohe qe marrim vendimin qe te pranojmë hipotezen  $H_0: \mu \geq 120$ . Keshtu per llogaritjen e probabilitetit te berjes te gabimeve te llojit te II, duhet te zgjedhim nje vlere te  $\mu$  me te vogel se 120 ore. Per shembull le ta zeme se malli i ardhur konsiderohet jo i cilesise se mire, ne rastin kur afatsherbimi fillesatar i baterive eshte me i vogel se  $\mu = 112$  ore. Le te shohim sa do te jete probabiliteti i pranimit te hipotezes  $H_0: \mu \geq 120$  dhe si rrjedhoje probabiliteti i berjes te gabimeve te llojit te dyte, ne kushtet kur  $\mu = 112$  eshte e vertete. Ky eshte probabiliteti qe mesatarja e zgjedhjes te jete  $\bar{x} \geq 116,71$ , nderkohe qe  $\mu = 112$ .

Duke perdorur shperndarjen normale standarte shohim se  $z_v = 2,36$ , kurse vlere e p per kete vlere te vrojtuar te Z do te jete 0,0091. Ky eshte probabiliteti i pranimit te  $H_0$  kur  $\mu = 112$ . Me fjale te tjera ky eshte probabiliteti i berjes te gabimeve te llojit te dyte kur  $\mu=112$ . Duke shenuar probabilitetin e berjes te gabimeve te llojit te II me  $\beta$ , shohim se kur  $\mu=112$ ,  $\beta=0,0091$ . Per kendej rrjedh se ne qofte se mesatarja e popullimit eshte 112 ore, probabiliteti i berjes te gabimeve te llojit te dyte eshte i barabarte 0,0091.

Keto llogaritje mund te behen edhe per vlere te tjera te  $\mu$  me vogla se 120. Duke vepruar ne kete menyre shohim se probabiliteti i berjes te gabimeve te llojit te II eshte i ndryshem per vlere te ndryshme te  $\mu$  me te vogla se 120. Per shembull supozojme se marrim rastin kur dergesa e baterive ka afatsherbim mesatar  $\mu = 115$  ore. Meqenese hipoteza  $H_0$  pranohet kurdohere qe

$$\bar{x} \geq 116,71, \text{ vlere e } z \text{ per } \mu = 115 \text{ rezulton } z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{116,71 - 115}{12 \sqrt{36}} \approx 0,86.$$



Tabela 9.2 Probabiliteti i berjes te gabimeve te llojit te II per pranimin e mallit

| $\mu$ | $z$   | $\beta$ | $1 - \beta$ |
|-------|-------|---------|-------------|
| 112   | 2,36  | 0,0091  | 0,9909      |
| 114   | 1,36  | 0,0869  | 0,9131      |
| 115   | 0,86  | 0,1949  | 0,8051      |
| 116   | 0,00  | 0,5000  | 0,5000      |
| 117   | -0,15 | 0,5596  | 0,4404      |
| 118   | -0,65 | 0,7422  | 0,2578      |

Nga tabela e shpërndarjes probabilitare normale standarte gjejmë se sipërfaqja në anën e sipërme të shpërndarjes probabilitare normale standarte për  $z = 0,86$  është  $0,5000 - 0,3051 = 0,1949$ . Kështu probabiliteti i berjes te gabimeve te llojit te II-te, kur mesatarja e vertete është  $\mu = 115$ , është  $\beta = 0,1949$ .

Në tabelën 9.2 janë paraqitur probabilitetet e berjes te gabimeve te llojit te II për vlera të  $\mu$  që janë më të vogla se 120. Nga tabela shihet se kur  $\mu$  rritet deri në 120, probabiliteti i berjes te gabimeve te llojit te dyte gjithashtu rritet deri në 0,95. Kjo është ajo që pritej. Më të vertete në qoftë se mesatarja e vertete e popullimit  $\mu$  është e afërt me vlerën e mesatares së popullimit  $\mu = 120$  në hipotezën zero, atëherë ekziston një probabilitet i madh që të pranojmë hipotezën  $H_0$  dhe të bëjmë një gabim të llojit te dyte. Megjithatë kur mesatarja e vertete e popullimit  $\mu$  është tepër larg nga vlera e hipotezës  $H_0$  për mesataren e popullimit  $\mu = 120$ , probabiliteti që të pranojmë hipotezën  $H_0$  dhe të bëjmë një gabim të llojit te dyte është i vogël.

Probabiliteti që të hedhim poshtë hipotezën  $H_0$  kur ajo është e gabuar quhet fuqi e kriterit. Për çdo vlerë të vecantë të  $\mu$  fuqia e kriterit është  $1 - \beta$ . D.m.th. probabiliteti i hedhjes më të drejtë poshtë të hipotezës  $H_0$  është i barabartë me 1 minus probabilitetin e berjes te gabimeve te llojit te II-te. Në tabelën 9.2 gjithashtu tregohen vlerat e fuqisë së kriterit. Në bazë të këtyre vlerave mund të ndërtohet grafiku për lidhjen e fuqisë së kriterit me vlerat e mesatares së popullimit  $\mu$ . Një grafik i tillë quhet vija e fuqisë së kriterit. Vëmë re se vija e fuqisë së kriterit është shprehur në varësi të vlerave të  $\mu$  për të cilat hipoteza  $H_0$  është e gabuar. Lartësia e vijës së fuqisë në çdo vlerë të  $\mu$ , jep probabilitetin e hedhjes poshtë më të drejtë të hipotezës  $H_0$  në rastin kur ajo është e gabuar.

Në përfundim po japim një procedurë që mund të përdoret për llogaritjen e probabilitetit të gabimeve te llojit te II, në rastin e kontrollit të hipotezës për mesataren e popullimit.

1. Formulohet hipotezat zero dhe alternative.
2. Përcaktojmë nivelin e rëndësisë  $\alpha$  që përcakton rregullin e hedhjes poshtë të hipotezës zero të bazuar në statistikën e zgjedhjes.
3. Duke përdorur rregullin e vendosur për rrezimin e hipotezës, zgjidhim ekuacionin për gjetjen e vlerës të mesatares së zgjedhjes, që përcakton zonën e hedhjes poshtë të hipotezës për kriterin e kontrollit.
4. Në bazë të rezultateve të etapës së tretë përcaktojmë vlerat e mesatares së zgjedhjes që çojnë në pranimin e  $H_0$  të cilat karakterizojnë zonën e hedhjes poshtë të hipotezës për kriterin.

Le ta zëmë se kemi marrë një zgjedhje prej 30 kutish, ku pesha mesatare e tyre ka rezultuar 550 g dhe devijimi standart i zgjedhjes  $s = 25$ -g. Duke përdorur vlerën e  $\mu$  nga hipoteza zero dhe devijimin standart të zgjedhjes si vlerësim për devijimin standart të popullimit, vlera e statistikës së kontrollit është.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{550 - 500}{25\sqrt{36}} = 0.33$$

Në përputhje me rregullin e hedhjes poshtë të hipotezës,  $H_0$  nuk mund të hidhet poshtë. Rezultatet e zgjedhjes tregojnë se manaxheri i kontrollit të cilësise nuk ka arsye të dyshojë në supozimin se linja e prodhimit për mbushjen e kutijave është në përputhje me standartin e caktuar.

## **TEMA 7:**

### **TESTIMI I HIPOTEZAVE (RASTI I NJË POPULLIMI): TESTET ‘t’**

#### **Ushtrime**

**Ushtrim 1:** Në qendrën spitalore XXX është marrë një zgjedhje prej 49 pacienteve me qëllim që të përcaktohet nëse mosha mesatare e pacientëve është mbi 41 vjeç. Mosha mesatare e pacientëve në zgjedhje ishte 43 vjeç me një devijim standart prej 4.67 vjet

1. Për nivelin e rëndësisë  $\alpha = 0.01$  përcaktoni nëse mosha mesatare e pacientëve është mbi 41 vjeç.
2. Për nivelin e rëndësisë  $\alpha = 0.01$  përcaktoni nëse mosha mesatare e pacientëve është mbi 41 vjeç kur dihet se numri i pacientëve në këtë qendër spitalore për vitin e kaluar ishte 2000.

**Ushtrimi 2:** Shefi i kuzhinës së një qendre spitalore pretendon që pesha mesatare e një pjate supe është të paktën 300 gram. Shpërndarja e masës së supës është normale, me devijim standart 50 gram. Janë zgjedhur rastësisht 16 pjata me peshë mesatare zgjedhjeje 293 gram. Testoni me një nivel rëndësie 5%, deklaratën e shefit të kuzhinës së kësaj qendre spitalore.

**Ushtrimi 3: Janë** zgjedhur 200 pacientë të shtruar në një qendër spitalore, 87 nga këta ishin femra. Testoni hipotezen zero që gjysma e të gjithë pacientëve janë femra ( $\alpha = 0.05$ ).

**Ushtrimi 4:** Një anëtar bashkie deklaron që në lagjet e varfra, së paku 20% e meshkujve të moshës 18 - 65 vjeç janë të papunë. Janë zgjedhur rastësisht 120 vetë nga ky popullim, ku 20 janë të papunë. Testoni deklaratën e anëtarit të Bashkisë ( $\alpha = 0.05$ ).

**Ushtrimi 5:** Nga një vrojtim i bërë rastësisht në 50 pasagjerë të një linje ajrore u vu re se pesha mesatare e bagazheve të çdo udhetari ishte 18.5 kg ndërsa varianca e saj ishte  $4 \text{ kg}^2$  duke ditur se pesha e lejuar është 20 kg për çdo pasagjer,  $\alpha = 0.05$  testoni nëse janë respektuar rregullat e udhëtimit në lidhje me peshën mesatare ose jo. (sup. Që pesha ka shpërndarje normale)

**Ushtrimi 6:** Një administrator i një qendre spitalore po teston vërtetësinë e deklaratës së një oferte në lidhje me tenderin e hapur për furnizimin e qendrës me karburant. Kompania ofertuese e naftës deklaron se çmimi mesatar i benzinës së saj është më i ulët se 1.25 dollarë për litër. Me qëllim që të kontrollohet kjo deklaratë, zgjidhen rastësisht 49 pika furnizimi karburanti dhe përcaktohet se çmimi mesatar i benzinës në pikat e karburantit të zgjedhjes është 1.2 dollarë. Për më tepër ne supozojmë se devijimi standart është 0.14 dollarë.

1. Përdorni procedurën e testimit të standartizuar dhe testoni deklaratën.
2. Cila është vlera p së shoqëruar me rezultatet e zgjedhjes së mësipërme, dhe në ç'konkluzion mund të arrihet duke u bazuar në vlerën p?

**3.** Sa është probabiliteti që të bëhet një gabim i llojit të dytë nëse çmimi i vërtetë për litër është 1.19 dollarë?



## KAPITULLI 8

### GJYKIMET STATISTIKORE PER MESATARET DHE PERPJESAT E DY POPULLIMEVE

Ne kapitujt 6 dhe 7 paraqitem metodat statistikore per marrjen e vleresimeve intervalore, si edhe kriteret e kontrollit te hipotezave per mesataren dhe perpjesen e nje popullimi. Ne kete kapitull do te shqyrtojme rastet kur kemi te bejme me dy popullime. Keshtu p.sh. mund te duam te krahasojme popullimin e meshkujve me popullimin e femrave, popullimin e mallrave nga furnizuesi A me popullimin e mallrave nga furnizuesi B etj. Vlesimin intervalor dhe kontrollin e hipotezave per ndryshimin e mesatareve dhe te perpjesave te dy popullimeve, meqenese ato nuk njihen, do ta bejme ne menyre te terthorte ne baze te te dhenave te nje zgjedhje te rastit.

#### 10.1 Vleresimi i ndryshimit te mesatareve te dy popullimeve, rasti i zgjedhjeve te pavarura

Ne shume situata ne praktike kemi te bejme me dy popullime te ndryshme ku krahasimi i mesatareve ka nje rendesi te vecante. Ne kapitullin 7 dhe 8 pame se si mund te perdoret mesatarja e nje zgjedhje te rastit per vleresimin e mesatares te popullimit. Ne rastin e dy popullimeve mund te kryejme dy zgjedhje te rastit te pavarura, njera nga popullimi i pare dhe tjetra nga popullimi i dyte. Shenojme me:  $\mu_1$  mesataren e popullimit te pare;  $\mu_2$  mesataren e popullimit te dyte;  $\bar{x}_1$  = mesataren e nje zgjedhje te rastit nga popullimi i pare dhe me  $\bar{x}_2$  = mesataren e nje zgjedhje te rastit nga popullimi i dyte.

Ndryshimi midis mesatareve te dy popullimeve eshte  $\mu_2 - \mu_1$ , kurse vleresimi pikesor i ndryshimit te tyre eshte:

$$\bar{x}_2 - \bar{x}_1 \quad 10.1$$

Keshtu, si vleresim per ndryshimin e mesatareve te dy popullimeve marrim diferencen e mesatareve te dy zgjedhjeve te rastit te pavarura.

#### Shperndarja e zgjedhjes per diferencen e mesatareve te dy zgjedhjeve $\bar{x}_2 - \bar{x}_1$

Diferenca e mesatareve te dy zgjedhjeve te rastit te pavarura, qe sherben si vleresim pikesor i differences te mesatareve te dy popullimeve, ka shpemdardjen e saj te zgjedhjes. Vetite e shpemdardjes te differences se mesatareve te zgjedhjeve  $\bar{x}_2 - \bar{x}_1$  jane si me poshte:

$$\text{Vlera e pritur: } E(\bar{x}_2 - \bar{x}_1) = \mu_2 - \mu_1 \quad 10.2$$

$$\text{Devijimi standart: } \sigma_{\bar{x}_2 - \bar{x}_1} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad 10.3$$

Per vleresimin intervalor te differences te mesatareve te dy popullimeve do te perdorim shperndarjen e zgjedhjes per  $\bar{x}_2 - \bar{x}_1$ . Per kete dallojme dy raste; rasti kur te dy vellimet e zgjedhjeve jane te medha ( $n_1 \geq 30$  dhe  $n_2 \geq 30$ ) dhe rasti tjetër kur njeri nga vellimet e zgjedhjeve, ose qe te dy jane te vogla ( $n_1 < 30$  dhe / ose  $n_2 < 30$ ). Le te shikojme ne fillim rastin kur zgjedhja eshte e madhe.

**Rasti i zgjedhjes se madhe.** Le ta zeme se nje firme ka dy dyqane shitjeje, njeri eshte ne qendren e qytetit dhe tjetri ne nje qender tregtare jashte qytetit. Supozojme se gjate nje periudhe kohe manaxheri ka vene re se produktet qe shiten mire ne njerin dyqan nuk shiten sic duhet ne dyqanin tjetër. Nje nga spjegimet e mundshme per diferencen e shitjeve ne te dy dyqanet jane ndryshimet qe ekzistojne ne vete klientet e ketyre dyqaneve. Keshtu klientet mund te ndryshojne nga moshja, arsimi, te ardhurat etj. Per kete eshte organizuar nje vrojtim me zgjedhje duke marre te dhena per keto karakteristika dhe te tjera si keto. Megjithate ne do te kufizohemi vetem me



moshen e klienteve, per te pare ne se ekzistojne ndryshime midis moshes mesatare te klienteve qe blejne ne dy dyqane te ndryshme. Te dhenat e grumbulluara per dy zgjedhjet e rastit te pavarura, per klientet qe kane ardhur perkatesisht ne secilin dyqan, kane dhene rezultatet e meposhtme.

| Dyqani         | Numri i klienteve ne zgjedhje | Mosha mesatare ne zgjedhje (ne vjet) | Devijimi standart ne zgjedhje (ne vjet) |
|----------------|-------------------------------|--------------------------------------|-----------------------------------------|
| Brenda qytetit | 36                            | 40                                   | 9                                       |
| Jashte qytetit | 49                            | 35                                   | 10                                      |

Duke perdorur 10.1 vleresimi pikesor i differences midis moshave mesatare te te dy popullimeve eshte  $\bar{x}_2 - \bar{x}_1 = 40 - 35 = 5$  vjet. Keshtu mund te besojme se klientet e dyqanit brenda qytetit kane moshen mesatare 5 vjet me te madhe se klientet e dyqanit jashte qytetit. Megjithate dime se diferenca prej 5 vjetesh eshte vetem nje peraftrim per diferencon midis moshave mesatare per dy popullimet qe krahasohen. Si rrjedhoje duhet te llogaritim edhe intervalin e besimit per diferencon midis mesatareve te dy popullimeve.

Ne rastin e zgjedhjes se madhe dime se shperndarja e zgjedhjes per diferencon e mesatareve mund te perafrohet me nje shpemdaje probabilitare normale. Mbi kete baze mund te bejme vleresimin intervalor per diferencon midis mesatareve te dy popullimeve. Ky vleresim paraqitet si me poshte:

$$(\bar{x}_2 - \bar{x}_1) \pm z_{\alpha/2} \sigma_{\bar{x}_2 - \bar{x}_1} \quad 10.4$$

ku  $1-\alpha$  eshte koeficienti i besimit.

Le te perdorim kete formule per ndertimin e intervalit te besimit per te dhenat e shembullit te klienteve te firmes. Ne kushtet kur variancat e dy popullimeve  $\sigma_1$  dhe  $\sigma_2$  nuk njihen, atehere 10.3 nuk mund te perdoret per llogaritjen e  $\sigma_{\bar{x}_2 - \bar{x}_1}$ . Megjithate ne kete rast si vleresim per devijimet standarte te popullimeve mund te perdorim devijimet standarte te zgjedhjeve. Ky vleresim paraqitet si me poshte.

$$s_{\bar{x}_2 - \bar{x}_1} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad 10.5$$

Kur vellimet e zgjedhjeve jane te medha, vleresimi i gjetur me ane formule 10.5 sherben si nje vleresim i mire per devijimin standart te differences te mesatareve te dy zgjedhjeve. Ne rastin e shembullit tone ky vleresim eshte:

$$s_{\bar{x}_2 - \bar{x}_1} = \sqrt{\frac{9^2}{36} + \frac{10^2}{49}} = \sqrt{4.29} = 2.07$$

Ne baze te vleresimit te gjetur per  $\sigma_{\bar{x}_2 - \bar{x}_1}$  dhe per  $z_{\alpha/2} = z_{0.025} = 1.96$  marrim intervalin 95%-sh  $5 \pm (1.96)(2.07)$  ose  $5 \pm 4.06$ . Keshtu per nivelin e besimit 95% shohim se vleresimi intervalor per diferencon midis moshave mesatare te klienteve te dy dyqaneve eshte nga 0.94 deri ne 9.06.

*Rasti i zgjedhjes se vogel.* Shikojme tani proceduren e vleresimit intervalor per diferencon midis mesatareve te dy popullimeve kur te pakten njeri nga vellimet e zgjedhjeve eshte me i vogel se 30. Ky quhet rasti i zgjedhjes se vogel.

Ne kapitullin 8, per rastin e zgjedhjes se vogel, paraqitem nje procedure per vleresimin intervalor te mesatares per nje popullim te vetem. Kjo procedure mbeshetetej ne supozimin se



popullimi kishte shpërndarje probabilitare normale. Vlerësimi për mesataren e popullimit mbështetet në vlerësimin  $s$  për  $\sigma$  dhe në përdorimin e shpërndarjes student,  $t$ .

Për ndërtimin e vlerësimit intervalor për diferencën e mesatareve të dy popullimeve në kushtet e zgjedhjes së vogël, është e nevojshme që për të dy popullimet dhe zgjedhjet e marra nga ato të behen dy supozime:

1. Të dy popullimet kanë shpërndarje normale.

2. Variancat e dy popullimeve janë të barabarta  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ .

Me supozimet e bera më lart shpërndarja e zgjedhjes të  $\bar{x}_2 - \bar{x}_1$  pavarësisht nga vëllimet e zgjedhjeve është normale dhe vlera e pritur e  $\bar{x}_2 - \bar{x}_1$  është  $\mu_1 - \mu_2$ . Në bazë të supozimit për variancat e barabarta, 10.3 mund të paraqitet në formën e mëposhtme:

$$\sigma_{\bar{x}_2 - \bar{x}_1} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad 10.6$$

Në qoftë se varianca e popullimit  $\sigma^2$  është e njohur, atëherë për vlerësimin intervalor të diferencës midis mesatareve mund të përdoret 10.4. Megjithatë në shumicën e rasteve varianca është e panjohur, kështu që për vlerësimin e saj mund të përdoren variancat e zgjedhjeve  $s_1^2$  dhe  $s_2^2$ . Mëqenese 10.6 bazohet në supozimin se  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , nuk ka nevojë që të behen vlerësime të vecanta për variancat e popullimeve. Prandaj të dhënat e zgjedhjeve mund të kombinohen për të marrë një vlerësim të përgjithshëm për variancën e popullimit. Vlerësimi  $s^2$  për variancën e popullimit  $\sigma^2$  mund të gjendet si mesatare e ponderuar e variancave të dy zgjedhjeve të vecanta. Ky vlerësim i përbashkët ka formën e mëposhtme:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad 10.7$$

Me këtë vlerësim për variancën e popullimit dhe duke përdorur 10.6, vlerësimi për variancën e diferencës të mesatareve ka formën:

$$s_{\bar{x}_2 - \bar{x}_1} = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad 10.8$$

Në këtë rast për ndërtimin e vlerësimit intervalor të diferencës të mesatareve të dy popullimeve mund të përdorim kriterin  $t$ . Mëqenese për popullimin e parë kemi  $n_1 - 1$  shkallë lirie dhe për popullimin e dytë  $n_2 - 1$ , atëherë numri i shkallëve të lirise që lidhet me kriterin  $t$  është i barabartë me  $n_1 + n_2 - 2$ . Intervali i besimit në këtë rast është si më poshtë:

$$\bar{x}_2 - \bar{x}_1 \pm t_{\alpha/2} s_{\bar{x}_2 - \bar{x}_1} \quad 10.9$$

ku  $t_{\alpha/2}$  bazohet në shpërndarjen  $t$  me  $n_1 + n_2 - 2$  shkallë lirie dhe  $1 - \alpha$  është koeficienti i besimit.

Le të ilustrojmë procedurën për gjetjen e intervalit të besimit për diferencën e mesatareve të dy popullimeve në kushtet e zgjedhjes së vogël, në rastin e shembullit për të parë ndryshimin e bilanceve mesatare për dy filiale të një banke. Vëllimet e zgjedhjeve, mesataret dhe devijimet standarte të tyre paraqiten në tabelën e mëposhtme:

| Filialet e<br>bankes | Numri i llogarive<br>rrjedhese | Bilanci mesatar për<br>zgjedhjet | Devijimet standarte të<br>zgjedhjeve |
|----------------------|--------------------------------|----------------------------------|--------------------------------------|
| Filiali 1            | 12                             | 1000                             | 150                                  |
| Filiali 2            | 10                             | 920                              | 120                                  |



Ne baze te ketyre te dhenave duam te ndertojme intervalin 90%-sh te besimit per diferencen midis balanceve mesatare te llogarive rrjedhese te tyre. Duke perdorur 10.7 vleresimi pikesor i variances se popullimit eshte.

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(12 - 1)150^2 + (10 - 1)120^2}{12 + 10 - 2}$$

Prej kendej mund te gjejme vleresimin per devijimin standart te differences te dy mesatareve:

$$s_{x_2 - x_1} = \sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{18.855 \left( \frac{1}{20} + \frac{1}{10} \right)} = 58.79$$

Ne kete rast shpemdarda e t per proceduren e vleresimit intervalor ka  $12 - 10 - 2 = 20$  shkalle lirie. Per nivelin e rendesise  $\alpha = 0.10$  gjejme  $t_{\alpha/2} = t_{0.05} = 1.725$ . Vleresimi intervalor per diferencen e mesatareve do te jete:

$$\bar{x}_2 - \bar{x}_1 \pm t_{\alpha/2} s_{x_2 - x_1} = (1000 - 920) \pm 1.725 * 58.79 = 80 \pm 101.41$$

Keshtu mund te themi se per nivelin e besimit 90%, vleresimi intervalor i differences te balanceve mesatare te llogarive rrjedhese te dy filialeve te nje banke eshte intervali:

$[-21.41; 181.41]$ . Fakti se ky interval perfshin nje vlere negative tregon se diferenca aktuale midis dy mesatareve mund te jete edhe negative.

## 10.2 Kontrolli i hipotezes per diferencen e mesatareve te dy popullimeve ne rastin e zgjedhjeve te pavarura.

Ne kete paragraf do te paraqesim procedurat qe mund te perdoren per kontrollin e hipotezave per diferencen e mesatareve te dy popullimeve.

*Rasti i zgjedhjes se madhe.* Me qellim qe te vleresojme ndryshimin per cilesine e kualifikimit ne dy qendra trainimi, u be nje testim per individet qe ishin trainuar ne keto qendra. Rezultatet e ekzaminimit perbejne nje faktor te rendesishem per vleresimin e ndryshimeve cilesore qe ekzistojne midis dy qendrave.

Shenojme me:  $\mu_1$ -rezultatin mesatar per individet e kualifikuar ne qendren A dhe me  $\mu_2$ -rezultatin mesatar per individet e kualifikuar ne qendren B. Do te supozojme se nuk ekziston ndryshim ne cilesine e kualifikimit midis dy qendrave. Keshtu ne termat e rezultatit mesatar te kontrollit, kjo do te thote se hipoteza zero qe duhet te kontrollohet eshte  $\mu_2 - \mu_1 = 0$ . Ne qofte se rezultatet e zgjedhjes cojne ne hedhjen poshte te hipotezes zero, atehere konkludojme se ekziston nje ndryshim cilesor midis dy qendrave te trainimit. Hipoteza zero mund te shkruhet si me poshte:

$$H_0: \mu_2 - \mu_1 = 0$$

$$H_a: \mu_2 - \mu_1 \neq 0$$

Duke ndjekur proceduren e kontrollit te hipotezave si ne kapitullin 9, bejme supozimin se hipoteza zero eshte e vertete. Ne keto kushte dhe kur zgjedhjet jane te medha, diferenca e mesatareve te zgjedhjeve  $\bar{x}_2 - \bar{x}_1$ , ka shperndarje normale.

Meqenese shpemdarda eshte normale, atehere per te kontrolluar hipotezen mund te perdoret kriteri:

$$z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad 10.10$$

Kurdohere qe  $n_1 > 30$  dhe  $n_2 > 30$  ne vend te  $\sigma_1^2$  dhe  $\sigma_2^2$  do te perdorim vleresimet perkatese te tyre  $s_1^2$  dhe  $s_2^2$ . Ne kushtet kur hipoteza  $H_0$  eshte e vertete  $Z \sim N(0; 1)$ .



Vlera e  $z$  që jepet me anën e 10.10 mund të interpretohet si numri i devijimeve standarte që  $\bar{x}_2 - \bar{x}_1$ , është larg  $\mu_2 - \mu_1$ , në kushtet kur ka vend hipoteza zero. Zona kritike përcaktohet mbi bazën e vlerës së  $\alpha$ . Vete vlera kritike gjendet në tabelën e shpërndarjes normale, kështu psh. për  $\alpha=0.05$ ,  $z_{kr}=1.96$ , kurse zona kritike do të jetë  $]-\infty; -1.96[ \cup ]1.96; \infty[$ .

Le ta zëmë se rezultatet e ekzaminimit të individëve të zgjedhur nga të dy qendrat janë paraqitur në tabelën e mëposhtme:

|                     | Qendra e kualifikimit A | Qendra e kualifikimit B |
|---------------------|-------------------------|-------------------------|
| Vellimi i zgjedhjes | 30                      | 40                      |
| Mesataret           | 82                      | 578                     |
| Devijimet           | 8                       | 10                      |

Duke përdorur në vend të  $\sigma_1^2$  dhe  $\sigma_2^2$  vlerësimet përkatëse të tyre  $s_1^2$  dhe  $s_2^2$  Vlera e vrojtuar e kriterit do të jetë:  $z_v = \frac{(82.5 - 78) - 0}{\sqrt{\frac{8^2}{30} + \frac{10^2}{40}}} = 2.09$ . Për këtë vlerë të  $z_v$  përfundimi është se

hipoteza zero duhet hedhur poshtë. Kështu rezultatet e zgjedhjes tregojnë se cilësia e kualifikimit për të dy qendrat që krahasohen është e ndryshme. Për  $z_v=2.09$  në bazë të tabelës të shpërndarjes normale standarte mund të gjejmë vlerën e probabilitetit  $p$ , vlerë e cila rezulton:

$$p=2*(0.5-0.4817)=0.0366.$$

Për këtë procedurë të kontrollit të hipotezës qëllimi është të përcaktojmë nëse ekziston një diferencë midis mesatareve të dy popullimeve. Mëqenëse nuk kemi ndonjë informacion nëse një mesatare mund të jetë më e madhe ose më e vogël se tjetra, hipoteza mbi barazinë e mesatareve  $H_0: \mu_2 - \mu_1 = 0$  dhe hipoteza alternative se ato ndryshojnë nga njëra tjetra  $H_a: \mu_2 - \mu_1 \neq 0$  janë që të dyja të përshtatshme. Në rastin e kontrollit të hipotezave të tjera për diferencën e mesatareve të dy popullimeve, mund të ndodhë që të kontrollojmë nëse një mesatare është më e madhe ose më e vogël se mesatarja tjetër. Në këto raste statistika e kontrollit është e njëanshme. Të dy format e kriterit të njëanshem të kontrollit të diferencës të mesatareve të dy popullimeve paraqiten si më poshtë:

$$H_0: \mu_2 - \mu_1 \leq 0$$

$$H_a: \mu_2 - \mu_1 > 0$$

$$H_0: \mu_2 - \mu_1 \geq 0$$

$$H_a: \mu_2 - \mu_1 < 0$$

Keto hipoteza kontrollohen me anën e statistikes  $z$  e cila jepet nga 10.10. Zona e hedhjes poshtë të hipotezës përcaktohet në të njëjtën mënyrë si zona e njëanshme që është paraqitur në kapitullin 9.

*Rasti i zgjedhjes së vogël.* Le të shikojmë tani kontrollin e hipotezës për diferencën e mesatareve të dy popullimeve në rastin e zgjedhjes së vogël, d.m.th. nerastin kur  $n_1 < 30$  dhe/ose  $n_2 < 30$ . Kjo procedurë bazohet në shpërndarjen  $t$  me  $n_1 + n_2 - 2$  shkallë lirie. Për këtë sic pame në paragrafin 10.1 bazohemi në supozimet për shpërndarjen normale të dy popullimeve dhe në barazinë e variancave të tyre.

Le ta zëmë se kerkohen të krahasohen dy mënyra të prodhimit nga pikpamja e kohës së nevojshme për përgatitjen e produktit me fjalë të tjera rendimentet e tyre. Për këtë behet një zgjedhje e rastit prej 24 punetoresh, ku 12 prej tyre të zgjedhur në mënyrë të rastit do të punojnë sipas mënyrës ekzistuese, dhe 12 të tjerët do të punojnë sipas mënyrës së re që eksperimentohet.

Në këtë studim ka dy popullime: popullimi i punetoreve që punojnë sipas mënyrës ekzistuese dhe popullimi i punetoreve që punojnë sipas mënyrës së re. Nga 24 punetoret e zgjedhur gjithësej, 12 prej tyre vihen të punojnë sipas mënyrës së parë dhe 12 të tjerë sipas mënyrës së dytë. Pas eksperimentimit për secilin grup janë llogaritur koha mesatare e nevojshme për prodhimin e njesisë së produktit dhe variancat përkatëse të të dy zgjedhjeve. Keto rezultate paraqiten në tabelën e mëposhtme:



| Treguesit | Menyra<br>$n_1=12$ | ekzistuese<br>prodhimit, | Menyra e re e prodhimit, $n_2=12$ |
|-----------|--------------------|--------------------------|-----------------------------------|
| Mesatarja | 325 minuta         |                          | 288 minuta                        |
| Varianca  | 40 minuta          |                          | 44 minuta                         |

Shenojme me  $\mu_1$  kohen mesatare per prodhimin e njesise se produktit sipas menyres se pare te prodhimit dhe me  $\mu_2$  kohen mesatare per prodhimin e njesise se produktit sipas menyres se dyte te prodhimit. Meqenese pritet qe menyra e re e prodhimit te jape rendiment me te larte se menyra ekzistuese, ateher ne presim qe mesatarja e popullimit te dyte te jete me madhe se mesatarja e popullimit te pare, dm.th.  $\mu_1 - \mu_2 \leq 0$ . Prandaj hipotezat qe kerkohen te kontrollohen jane si me poshte:

$$H_0: \mu_2 - \mu_1 \leq 0$$

$$H_a: \mu_2 - \mu_1 > 0$$

Me tentativen e supozimit se hipoteza zero eshte e vertete, ne pranojme se menyra e re e prodhimit qe propozohet eshte me efektive se e para. Prandaj ne kete aspekt perpiqemi qe me te dhenat e zgjedhjes te hedhim poshte hipotezen zero, d.m.th. qe menyra e dyte e prodhimit nuk jep nje rendiment me te larte se e para.

Me supozimin se variancat e popullimeve jane te barabarta, vleresimii i variances se pergjithshme mund te behet me anen e 10.7:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(12 - 1)40^2 + (12 - 1)44^2}{12 + 12 - 2} = 1768$$

Statitika e kontrollit per rastin e zgjedhjes se vogel eshte:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}, \text{ kurse vlera e vrojtuar e saj } t_v = \frac{(325 - 288) - 0}{\sqrt{1768 \left( \frac{1}{12} + \frac{1}{12} \right)}} = 2.16$$

Per nivelelin e rendesise  $\alpha=0,05$  dhe numrin e shkalleve te lirise  $12+12-2=22$ , ne tabelen e shperndarjes t gjejme vleren kritike te djathte  $t_{kr}=1,717$ . Meqenese vlera faktike e kriterit t eshte me e madhe se vlera kritike e tij  $2,16 > 1,717$ , nga kjo rrjedh se hipoteza zero hidhet poshte dhe pranohet qe menyra e dyte e prodhimit eshte me efektive se e para.

### 10.3 Gjykimet statistikore per mesataret dy popullimeve ne rastin e zgjedhjeve te varura.

Le ta zeme se nje ndermarje kerkon te eksperimentoje nje menyre te re prodhimi me qellim qe te rrise prodhimin e saj. Per kete ajo kerkon te identifikojte se per cilen nga metodat koha per njesi prodhimi eshte me e vogel. Shenojme me  $\mu_1$  kohen mesatare per prodhimin e njesise se produktit sipas metodes se pare dhe me  $\mu_2$  kohen mesatare per njesi sipas metodes se dyte. Keshtu hipoteza zero merr formen  $H_0: \mu_1 - \mu_2 = 0$ . Ne qofte se hipoteza zero hidhet poshte, arrijme ne perfundimin se te dy metodat ndryshojne lidhur me kohen mesatare te nevojshme per njesi prodhimi. Hipoteza zero dhe ajo alternative mund te paraqiten ne formen e meposhtme:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

Per percaktimin e procedures se zgjedhjes qe do te perdoret per grumbullimin e te dhenave lidhur me kohet per njesi prodhimi per te dy metodat, si edhe per kontrollin e hipotezave te mesiperme shqyrtojme dy alternativa. Njera bazohet ne zgjedhjet e pavarura dhe tjetra bazohet ne zgjedhjet e varura ose te ciftuara. Keto modele te zgjedhjeve mund te pershkruhen si me poshte:

1. Modeli i zgjedhjeve te pavarura. Ne kete rast behet nje zgjedhje e rastit per nje numur te caktuar punonjesis qe punojne me metodën e pare te prodhimit dhe nje zgjedhje tjetër e rastit per



nje numur te caktuar punetoresh qe punojne me metoden e dyte. Kontrolli per diferencen e mesatareve bazohet ne proceduren e paraqitur ne paragrafin 10.2.

2. Modeli i zgjedhjeve te varura. Kryhet nje zgjedhje e rastit per nje numur te caktuar punetoresh te cilet ne fillim punojne sipas metodes se pare dhe pastaj sipas metodes se dyte. Radha sipas te ciles punetoret punojne sipas cdo metode eshte e rastit dmth disa mund te punojne ne fillim me metoden e pare dhe disa te tjere me metoden e dyte, por ajo qe ka rendesi ne kete rast eshte se cdo punetor punon ne te dy menytrat. Keshtu per cdo punetor merren cifte te dhenash, ku nje vlere i korespondon metodes se pare te prodhimit, kurse vlera tjeter i perket metodes se dyte.

Ne rastin e zgjedhjeve te varura ose te ciftuara, meqenese te dy metodat e prodhimit kontrollohen ne kushte te njejta (d.m.th. jane te njejtet punetore qe punojne me te dy metodat), ky model con ne gabim me te vogel te zgjedhjes se sa rasti i zgjedhjeve te pavarura. Shkaku kryesor per kete eshte se ne rastin e zgjedhjeve te ciftuara, variacioni midis punetoreve nuk sherben me si burim i gabimit te zgjedhjes.

Le te ilustrojme analizen e modelit te zgjedhjeve te ciftuara per shembullin e kontrollit te diferences midis dy metodave te prodhimit. Per kete behet nje zgjedhje e rastit qe permban 6 punetore. Veme re se cdo punetor jep nje cift vlerash, nje per cdo metode prodhimi. Gjithashtu veme re se shtylla e fundit permban diferencat e koheve per cdo punetor ne zgjedhje.

Sipas modelit te zgjedhjeve te ciftuara merret ne konsiderate vetem shtylla e diferencave per te dy metodat. Si rezultat marrim vlerat e tyre per 6 te dhena (0.6; -0.2; 0.5; 0.3; 0.0 dhe 0.6) dhe vetem keto do te perdoren me tej per analizen e diferences te mesatareve te dy popullimeve, qe u korespondojne dy metodave te prodhimit qe eksperimentohen.

Tabela 10. Koha per njesi produkti per dy metoda prodhimi ne rastin e zgjedhjeve te ciftuara

| Puntori | Koha per njesi sipas<br>metodes A | Koha per njesi sipas<br>metodes B | Diferenca midis koheve |
|---------|-----------------------------------|-----------------------------------|------------------------|
| 1       | 6                                 | 5.4                               | 0.6                    |
| 2       | 5                                 | 5.2                               | -0.2                   |
| 3       | 7                                 | 6.5                               | 0.5                    |
| 4       | 6.2                               | 5.9                               | 0.3                    |
| 5       | 6                                 | 6                                 | 0                      |
| 6       | 6.4                               | 5.8                               | 0.6                    |

Shenojme me  $\mu_d$  mesataren e vlerave te diferencave per popullimin e puntoreve qe punojne me te dy metodat. Me kete shenim hipotezat zero dhe alternative kane formen e meposhtme:

$$H_0: \mu_d = 0$$

$$H_a: \mu_d \neq 0$$

Ne qofte se hipoteza zero hidhet poshte konkludojme se ekziston nje ndryshim per kohen per njesi prodhimi per te dy metodat.

Indeksi d perdoret per te treguar se zgjedhja e ciftuar jep te dhenat per vlerat e diferencave. Mesatarja e zgjedhjes dhe devijimi standart per vlerat e diferencave te vrojtuar te tabelës 10.13 jane si me poshte:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{1.8}{6} = 0.30$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{0.56}{5}} = 0.335$$

Ne kapitullin 9 kemi pare se, ne qofte se popullimi eshte i shperndare normalisht, atehere per kontrollin e hipotezes per mesataren e popullimit mund te perdoret shperndarja t me n - 1 shkalle lirie. Me te dhenat per diferencat statistika e kontrollit merr formen:



$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

10.12

Per nivelin e rendesise  $\alpha=0,05$  dhe  $n-1=5$  shkalle lirie vlera kritike e kriterit  $t$  eshte  $t_{0,025}=2,571$ . Keshtu rregulli i hedhjes poshte te hipotezes zero eshte: Ho hidhet poshte ne qofte se  $t < -2,571$  ose ne qofte se  $t > 2,571$ . Llogaritim tani vleren e vrojtuar te kriterit tone, ne kushtet kur hipoteza zero eshte e vertete:

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{0.3 - 0}{0.335\sqrt{6}} = 2.19$$

Meqenese vlera faktike e kriterit  $t$  eshte me e vogel se vlera kritike e tij, te dhenat e zgjedhjes nuk perbejne nje baze te mjaftueshme per hedhjen poshte te hipotezes zero.

Ne kete rast vleresimi intervalor per diferencen e mesatareve te dy popullimeve mund te ndertohtet njelloj sic beme ne rastin e nje mesatareje ne kapitullin 8. Keshtu vleresimi intervalor i differences te mesatareve te dy popullimeve eshte si me poshte:

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = 0.3 \pm 2.571 \frac{0.335}{\sqrt{6}} = 0.3 \pm 0.35$$

Si perfundim intervali 95%-sh i besimit per diferencen midis koheve mesatare per njesi prodhimi per te dy metodat e prodhimit eshte -0,5 minuta deri ne 0,65 minuta.

#### 10.4 Gjykimet statistikore per ndryshimin e perpjesëve te dy popullimeve

Nje firme prodhuese me dy dege eshte e interesuar per produktin e saj me cilesi te larte. Pas marrjes se masave ne procesin e prodhimit firma do te shikoje ne se ekziston ndonje ndryshim ne cilesine e produktit te prodhuar ne keto dy dege. Keshtu problemi qe shtrohet eshte: te kontrollohet nga ana statistikore ne se ekziston ndryshim midis perpjesëve te dy popullimeve.

Bejme shenimet e meposhtme:

$p_1$  = perpjesa e produktit me cilesi jo te larte per popullimin 1 (dega A)

$p_2$  = perpjesa e produktit me cilesi jo te larte per popullimin 2 (dega B)

$\bar{p}_1$  = perpjesa ne zgjedhje per nje zgjedhje te rastit nga popullimi 1

$\bar{p}_2$  = perpjesa ne zgjedhje per nje zgjedhje te rastit nga popullimi 2

Keshtu diferenca midis perpjesëve te dy popullimeve jepet nga  $p_1 - p_2$ , kurse vleresimi i kesaj difference me te dhenat e zgjedhjes eshte  $\bar{p}_1 - \bar{p}_2$ . Shohim se vleresimi pikesor i differences midis pjesëve te dy popullimeve eshte sa diferenca midis perpjesëve te dy zgjedhjeve te rastit te pavarura.

**Shperndarja e zgjedhjes te  $\bar{p}_1 - \bar{p}_2$ .** Per krahasimin e differences te perpjesëve te dy popullimeve do ti referohemi vleresimit pikesor te saj, e cila eshte sa diferenca e perpjesëve te dy zgjedhjeve te rastit te pavarura nga keto popullime  $\bar{p}_1 - \bar{p}_2$ . Sic kemi pare ne mjaft raste, shperndarja e zgjedhjes te nje vleresimi pikesor eshte faktori kryesor per ndertimin e vleresimit intervalor dhe per kontrollin e hipotezave per parametrat qe na interesojne. Vetite e shperndarjes se zgjedhjes te  $\bar{p}_1 - \bar{p}_2$  jane si me poshte:

$$E(\bar{p}_1 - \bar{p}_2) = p_1 - p_2$$

10.13



$$\sigma \bar{p}_1 - \bar{p}_2 = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad 10.14$$

$n_1$ =vëllimi i zgjedhjes nga teresia 1

$n_2$ =vëllimi i zgjedhjes nga teresia 2

Forma e shpërndarjes së  $\bar{p}_1 - \bar{p}_2$  në qoftë se vëllimet e zgjedhjeve janë të mëdha (d.m.th. kur  $n_1 p_1$ ;  $n_1(1-p_1)$ ;  $n_2 p_2$  dhe  $n_2(1-p_2)$  janë më të mëdha ose të barabarta se 5), mund të përafrohet me shpërndarjen probabilitare normale. Shpërndarja e zgjedhjes së  $\bar{p}_1 - \bar{p}_2$  tregohet në figurën 10.7.

**Vlerësimi i intervalit të  $\bar{p}_1 - \bar{p}_2$ .** Le ta zëmë se është kryer nga një zgjedhje e rastit për secilin popullim, d.m.th. nga prodhimi i cdo dege është zgjedhur në mënyrë të rastit një numër i caktuar njesish produkti.

|                                         | Dega prodhuese A | Dega prodhuese B |
|-----------------------------------------|------------------|------------------|
| Vëllimi i zgjedhjes                     | 250              | 300              |
| Numri i njesive të cilësive jo të lartë | 35               | 27               |

Përpjesët e produktit të cilësive jo të lartë për të dy degët janë përkatesisht  $\bar{p}_1 = 35/250 = 0.14$  dhe  $\bar{p}_2 = 27/300 = 0.09$ . Vlerësimi pikësor i diferencës midis përpjesëve të produktit të cilësive jo të lartë të dy popullimeve është  $\bar{p}_1 - \bar{p}_2 = 0.14 - 0.09 = 0.05$ . Kështu kjo të çon të besosh se dega A prodhon 5% të produktit me cilësi më të ulët se dega B. Megjithatë ashtu si me të gjithë vlerësimet pikësore, ndryshimi prej 0.05 është njëra nga shumë vlerat e mundshme të diferencës midis përpjesëve të dy teresive. Vlerësimi i intervalit të diferencës midis përpjesëve të dy popullimeve në kushtet e zgjedhjeve të mëdha është si më poshtë:

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sigma \bar{p}_1 - \bar{p}_2 \quad 10.15$$

ku  $1-\alpha$  është koeficienti i besimit.

Le të përdorim këtë procedurë për ndërtimin e intervalit të besimit të diferencës midis përpjesëve të produktit të cilësive jo të lartë të dy degeve të prodhimit për shembullin që kemi marrë më lart. Meqenëse  $p_1$  dhe  $p_2$  janë të panjohura llogaritja e devijimit standart të diferencës të përpjesëve të dy popullimeve nuk mund të bëhet me anën e formulës 10.14. Megjithatë në qoftë se në vend të përpjesëve të dy popullimeve përdorim vlerësimet pikësore të tyre, mund të marrim një vlerësim për devijimin standart të diferencës të këtyre përpjesëve si më poshtë:

$$s \bar{p}_1 - \bar{p}_2 = \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \quad 10.16$$

Për të dhënat e shembullit tonë ky vlerësim është i barabartë me:

$$s \bar{p}_1 - \bar{p}_2 = \sqrt{\frac{0.14(1-0.14)}{250} + \frac{0.09(1-0.09)}{300}} = 0.0275$$

Për koeficientin e besimit 90% gjejmë  $z_{\alpha/2} = z_{0.05} = 1.645$  dhe prej kësaj intervalin e besimit për diferencën e përpjesëve të dy popullimeve  $(0.14 - 0.09) \pm 1.645(0.0275) = 0.05 \pm 0.045$ . Kështu intervali i besimit 95% për diferencën e përpjesëve të produktit me cilësi jo të lartë për të dy degët prodhuese është nga 0.005 deri në 0.095; ose nga 0.5% deri në 9.5%.



Per kontrollin e hipotezes lidhur me diferencen e perpjeseve te dy popullimeve, i referohemi edhe nje here shembullit qe morem ne fillim per te pare ne se ekziston ndryshim per perpjeset e produktit me cilesi jo te larte te prodhuara ne dy deget e nje firme prodhuese. Hipotezat qe kerkojme te kontrollojme jane si me poshte:

$$H_0: p_1 - p_2 = 0$$

$$H_a: p_1 - p_2 \neq 0$$

Ne figuren 10.8 tregohet shpemdarrja e zgjedhjes per diferencen  $p_1 - p_2$  e bazuar ne supozimin se midis perpjeseve te dy popullimeve nuk ekzistojne ndryshime, d.m.th.  $p_1 - p_2 = 0$ . Me shpemdarrjen e zgjedhjes afersisht normale statistika  $t$  per diferencen midis perpjeseve te dy popullimeve mund te shkruhet si me poshte:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sigma_{p_1 - p_2}} \quad 10.17$$

Per nivelin e rendesise  $\alpha = 0.10$  dhe  $z_{\alpha/2} = z_{0.05} = 1.645$ , rregulli i hedhjes poshte te hipotezes eshte:

Hipoteza  $H_0$  hidhet poshte ne qofte se  $z < -1.645$ , ose kur  $z > 1.645$ .

Llogaritja e  $z$  me anen e formule 10.17 kerkon paraprakisht njohjen e vleres per gabimin standart te diferences midis perpjeseve te dy popullimeve,  $\sigma_{\bar{p}_1 - \bar{p}_2}$ . Meqenese ky gabim standart eshte i panjohur, ai duhet te vleresohet nga te dhenat e zgjedhjes. Per analogji me procedures e vleresimit intervalor ne mund te donim te perdornim  $\bar{p}_1$  dhe  $\bar{p}_2$ , ne fomulen 10.16, megjithate per kontrollin e kesaj hipoteze, formulen 10.16 do ta transformojme ne nje forme tjeter. Keshtu ne baze te vleresimeve te perpjeseve te popullimeve do te llogaritim variancen e perbashket te popullimit ne kushtet kur hipoteza zero eshte e vertete. Vleresimi per perpjesen e popullimit eshte si me poshte:

$$\bar{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} \quad 10.18$$

Me perdorimin e vleresimit te perbashket ne vend te vleresimeve te vecanta 10.16 merr formen:

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad 10.19$$

Ne mund te perdorim tani 10.18 dhe 10.19 per kryerjen e llogaritjeve ne rastin e shembullit tone si me poshte:

$$\bar{p} = \frac{250(0.14) + 300(0.09)}{550} = \frac{62}{550} = 0.113$$

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{0.113(1 - 0.113)\left(\frac{1}{250} + \frac{1}{300}\right)} = 0.027$$

Ne baze te 10.17 vlera e statistikes se kontrollit eshte:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{s_{p_1 - p_2}} = \frac{(0.14 - 0.09) - 0}{0.0271} = 1.85$$

Meqenese  $1.85 > 1.645$ , arrijme ne perfundimin se per nivelin e rendesise  $\alpha = 0.10$  hipoteza zero hidhet poshte. Te dhenat e zgjedhjes tregojne se ekziston nje ndryshim midis dy degeve te prodhimit persa i perket perpjeses te produktit me cilesi jo te larte te prodhuara nga ato dege.

Sikunder pame gjate kontrollit te hipotezave per diferencen e mesatareve te dy popullimeve edhe per rastin e diferences te dy perpjeseve mund te ndertohen kriteret e njeanshem. Zonat kritike te njeanshme ndertohen ne menyre analoge me procedurat per perpjesen e nje popullimi.

## **TEMA 9:**

### **TESTIMI I HIPOTEZAVE (RASTI I 'K' POPULLIMEVE): ANOVA (NJË DHE DY FAKTORIALE) , ANALIZA E MATJEVE TË PËRSËRITURA**

#### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje të:

- Njohin situatat në të cilat të përdorin Analizën e Variancës .
- Të performojnë një ANOVA një-faktoriale dhe të interpretojnë rezultatet .
- Të performojnë një ANOVA dy-faktoriale dhe të interpretojnë rezultatet .
- Të aplikojnë në SPSS një teknikë ANOVA duke shpjeguar pse, si dhe të interpretojnë rezultatet .
- Njohin situatat në të cilat të përdorin Analizën e Matjeve të Përsëritura .
- Të performojnë një AMP një-faktoriale dhe të interpretojnë rezultatet .
- Të aplikojnë në SPSS një teknikë AMP duke shpjeguar pse, si dhe të interpretojnë rezultatet .

#### **LITERATURA**

- “ Applied Statistics”, Warner A.R , 2008
- Statistika”- Grup Autoresh, Shtëpia botuese Pegi, (2016)
- Statistical for Business nad Economics”, Newbold P, Carlson W, Thorne B, 2011
- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011
- “Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014



## KAPITULLI 9/1 ANALIZA E VARIANCES DHE MODELIMI I EKSPERIMENTIT

### Hyrje ne analizen e variaces

Analiza e variaces është vazhdim i punes per kontrollin e hipotezave, veçse kete rradhe synohet te kontrollohet barazia ose jo e pritjeve matematike te disa popullimeve.

Le ta zeme se duam te dime se çfare njohurish kane punonjesit e reparteve te nje fabrike rreth drejtimit te pergjithshem te cilesise. Per kete supozojme se eshte kryer nje zgjedhje e rastit duke marre nga 6 punonjes per çdo repart, ku rezultatet e vrojtimit per 18 punonjesit i kemi paraqitur ne tabelen 13.1. Pervec kesaj ne table per çdo grup jane llogaritur mesataret, variancat dhe devijimet standarte. Drejtimi i ndermarrjes kerkon te kontrolloje hipotezen qe rezultatet mesatare te testimit per te tre repartet nuk kane ndryshime nga njera tjetra. Secili nga tre repartet mund te konsiderohet si nje popullim.

Tabela 13.1 Rezultatet e vrojtimit per 18 punon es

|           |    |    |    |    |    |    | Mesatarja | Varianca | Devijimi standart |
|-----------|----|----|----|----|----|----|-----------|----------|-------------------|
| Reparti 1 | 85 | 74 | 82 | 76 | 71 | 85 | 79        | 34       | 5.83              |
| Reparti 2 | 71 | 75 | 73 | 74 | 69 | 82 | 74        | 20       | 4.47              |
| Reparti 3 | 59 | 64 | 62 | 69 | 75 | 67 | 66        | 32       | 5.66              |

Shenojme me  $\mu_i$  rezultatin mesatar per popullimin e "i". Me qene se ne raste te tilla vlerat reale te  $\mu_i$  nuk njihen, qellimi i analizes eshte te kontrollohet hipoteza e meposhtme:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_a$ : Jo te gjitha mesataret e popullimeve jane te barabarta

Zbatimi i analizes se variaces per kontrollin e hipotezes se mesiperme mbeshetet ne supozimet e meposhtme:

1. Variabli qe shqyrtohet ne seicilin nga popullimet ka shperndarje normale. Nga kjo rrjedh se per rastin e shembullit qe kemi marre rezultati i vrojtimit ne çdo repart duhet te kete shperndarje normale.
2. Varianca e variablit ne shqyrtim qe shenohet me  $\sigma^2$  eshte e njejte per seicilin popullim. Nga kjo rrjedh se varianca e rezultateve te vrojtimit per te dhenat e shembullit tone duhet te jete e njejte per te gjitha repartet.
3. Vrojtimit duhet te jene te pavarura. Nga kjo rrjedh se rezultati i vrojtimit te nje punonjesi duhet te jete i pavarur nga rezultati i vrojtimit te çdo punonjesi tjetër.

### Thelbi i analizes se variaces

Ne kushtet kur supozimet e mesiperme plotesohen, ne qofte se hipoteza zero per barazine e mesatareve eshte e vertete, çdo vrojtim me zgjedhje duhet te kete ardhur nga e njejta shpemdardje probabilitare normale me mesatare  $\mu$  dhe variance te barabarte me  $\sigma^2$ . Ne nje rast te tille  $\bar{x}$  do te kete shperndarje normale me te njejten mesatare dhe variance te barabarte me  $\sigma^2/n$ .

Keshtu, ne qofte se mesataret per te tre popullimet jane te barabarata, atehere duhej te prisnim qe edhe mesataret e zgjedhjeve te jene te aferta me njera tjetren. Ne fakt, sa me pak qe mesataret e zgjedhjeve te ndryshojne midis tyre aq me i bazuar do te jete konkluzioni qe mesataret e popullimeve jane te barabarta. E kunderta, sa me shume qe mesataret e zgjedhjeve te ndryshojne nga njera tjetra, aq me shume mbeshtetje do te kete konkluzioni qe mesataret e popullimeve ndryshojne midis tyre. Me fjale te tjera ne qofte se variacioni midis mesatareve te zgjedhjeve eshte i vogel kjo perben nje mbeshtetje per hipotezen  $H_0$ , ndersa kur variacioni midis mesatareve te zgjedhjeve eshte i madh ajo eshte mbeshtetje per hipotezen alternative  $H_a$ .

Ne qofte se hipoteza zero eshte e vertete, d.m.th.  $\mu_1 = \mu_2 = \mu_3$ , atehere variacioni ndermjet mesatareve te zgjedhjeve mund te perdoret per marrjen e nje vleresimi per variancen e popullimit  $\sigma^2$ .

Ne qofte se hipoteza zero eshte e vertete, atehere mund te mendojme se te tre mesataret ne tabelen 13.1,  $\bar{x}_1 = 79$ ,  $\bar{x}_2 = 74$  dhe  $\bar{x}_3 = 66$ , jane marre nga shpemdardja e zgjedhjes normale. Ne kete rast mesatarja dhe varianca e te tre mesatareve mund te perdoren perkatesisht si vleresime per mesataren dhe variancen e shpemdardjes se zgjedhjes. Ne rastin e shembullit tone vleresimi me i mire i mesatares se zgjedhjes  $\bar{x}$  eshte mesatarja e te tre mesatareve te



zgjedhjes, d.m.th.  $(79+74+66)/3=73$ . Kete vleresim do ta quajme mesatarja e pergjitheshme e zgjedhjes.

Vleresimi per variancen e shperndarjes se zgjedhjes merret sa varianca e mesatare te zgjedhjes.

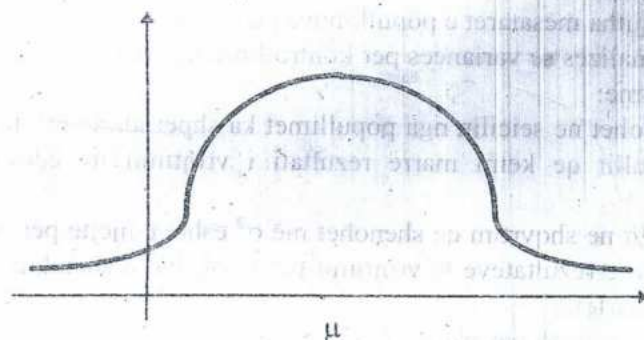
$$S^2 = [(79-73)^2 + (74-73)^2 + (66-73)^2] / (3-1) = 86/2 = 43$$

$$S^2 = \frac{\sum (\bar{x}_i - \bar{x})^2}{K-1}$$

Me qene se  $\sigma_x^2 = \sigma^2 / n$ , per  $\sigma^2$  do te kemi  $\sigma^2 = n\sigma_x^2$ . Meqenese vleresimi per  $\sigma_x^2$  eshte  $s_x^2 = 43$  vleresimi per  $\sigma^2$  do te jete i barabarte me  $6 \cdot 43 = 258$ .

Vleresimi i variaces midis zgjedhjeve bazohet ne supozimin mbi vertetesine e hipotezes zero, d.m.th. se cdo zgjedhje vjen nga i njejti popullim dhe ka vetem nje shperndarje te zgjedhjes per mesataren e zgjedhjes. Per te ilustruar se cfare ndodh ne qofte se hipoteza zero nuk eshte e vertete, le ta zeme se mesataret e popullimeve ndryshojne nga njera tjetra. Me qene se ne kete rast te tre zgjedhet jane nga popullime me shperndarje normale me mesatare te ndryshme, atehere shperndarjet e zgjedheve do te jene te ndryshme ose me sakte te zhvendosura nga njera tjetra. Per kete situatë mesataret e zgjedheve nuk do te jeve aq te aferta si ne rastin kur hipoteza zero eshte e vertete. Kjo do te beje qe edhe vleresimi midis zgjedhjeve per variancen e pergjithshme te jete me i madh. Pra ne pergjithesi kur mesataret e popullimeve nuk jane te barabarta, vleresimi midis zgjedhjeve perben nje mbivleresim per variancen e popullimit  $\sigma^2$ .

Figura 13.1 Shperndarja e zgjedhjes per mesataren kur hipoteza zero eshte e vertete.



Variacioni brenda cdo zgjedhje gjithashtu mund te kete efekt ne perfundimet qe arrihen me analizen e variaces. Ne qofte se nga cdo popullim eshte kryer nje zgjedhje krejt e rastit, atehere seicila nga variancat e zgjedhjes perben nje vleresim te pazhvendosur per variancen e pergjitheshme  $\sigma^2$ . Keshtu te gjitha vleresimet e vecanta per variancen mund te kombinohen ne nje te tere per te marre nje vleresim te vetem te perbashket per variancen e popullimit. Vleresimi i gjetur ne kete menyre per  $\sigma^2$ , d.m.th. si mesatare e variancave te zgjedhjeve te vecanta, quhet vleresimi brenda zgjedhjeve i saj. Me qene se cdo variance e zgjedhjes perben nje vleresim per variancen e pergjithshme te bazuar vetem ne variacionin brenda zgjedhjes, vleresimi i variaces brenda zgjedhjeve nuk ndikohet nga fakti ne se mesataret e popullimeve jane ose jo te barabarta. Ne rastin kur vellimet e zgjedhjeve jane te barabarta, vleresimi i variaces brenda zgjedhjeve mund te llogaritet si mesatare e thjeshte e variancave te zgjedhjeve te vecanta. Per shembullin tone ky vleresim eshte  $\sigma^2 = (34+20+32)/3 = 28,67$ .

Shohim se vleresimi i variaces midis zgjedhjeve (258) eshte mjaft me i madh se vleresimi i variaces brenda zgjedhjeve (28,67). Raporti i variaces midis zgjedhjeve me variancen brenda zgjedhjeve rezulton i barabarte me  $258 / 28,67 = 9$ . Mirepo varianca midis zgjedhjeve eshte vleresim i mire per variancen e pergjithshme vetem ne qofte se hipoteza zero eshte e vertete.

Ne qofte se hipoteza zero eshte e vertete te dy vleresimet do te jene te aferta dhe raporti i tyre do te jete afer njeshit. Ne qofte se hipoteza zero eshte e gabuar, atehere vleresimi i variaces midis zgjedhjeve do te tentoje te jete me i madh se vleresimi i variaces brenda zgjedhjeve dhe si rrjedhoje raporti i variaces midis zgjedhjeve me variancen brenda



zgjedhjeve do te jete ne menyre te konsiderueshme me i madh se 1. Ne paragrafin tjeter do te shohim se sa i madh duhet te jete ky raport qe te hidhet poshte hipoteza zero.

Duke permbledhur, llogjika qe qendron pas analizes se variaces bazohet ne ndertimin e dy vleresimeve te pavarura te variaces se popullimit. Njeri nga vleresimet bazohet ne variacionin e vete mesatare te zgjedhjeve, kurse vleresimi tjeter ne variacionin e te dhenave per brenda cdo zgjedhje. Duke krahasuar keto vleresime te variaces se pergjithshme, ose me sakt duke krahasuar raportin e tyre me nje koeficient normativ, do te jemi ne gjendje te percaktojme ne se mesataret e zgjedhjeve jane ose jo te barabarta. Me qene se metodologjia e perdorur ka te beje me krahasimin e dy variancave, kjo metode mban emrin analiza e variaces dhe do te shenohet shkurt ANOVA<sup>1</sup>.

### 13.2 Analiza e variaces: Kontrolli i barazise te k mesatareve.

Ne pergjithesi analiza e variaces mund te perdoret per kontrollin e barazise te k mesatareve. Forma e pergjithshme e hipotezes qe kontrollohet eshte:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$H_a$ : Jo te gjitha mesataret e popullimeve jane te barabarta

Le ta zeme se nga seicili prej k popullimeve kemi marre nga nje zgjedhje te rastit me velleim  $n_j$ , ku  $j = 1, 2, \dots, k$ . Shenojme me  $x_{ij}$  - vrojtimin e i-te ne zgjedhjen e j-te,  $n_j$  - numrin e vrojtimeve ne zgjedhjen e j -te  $\bar{x}_j$  - mesataren e zgjedhjes se j -te,  $s_j^2$  - variancen ne zgjedhjen e j -te dhe  $s_j$  - devijimin standart ne zgjedhjen e j -te.

Formulat per mesataren dhe variancen ne zgjedhjen e j -te jane si me poshte:

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \quad s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1} \quad \text{kurse mesataren e pergjithshme te popullimit}$$

$$\bar{\bar{x}} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}}{n_T} \quad \text{ku } n_T = n_1 + n_2 + \dots + n_k$$

Veme ne dukje se mesatarja e popullimit mund te paraqitet edhe si mesatare e ponderuar e mesatare te zgjedhjeve:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^k \bar{x}_j n_j}{n_T}$$

Keshtu ne rastin e shembullit tone mesatarja e pergjithshme do te jete:  $(6*79 + 6*74 + 6*66)/18 = 73$ . Ne qofte se hipoteza zero eshte e vertete, atehere mesatarja e pergjithshme e zgjedhjes 73 eshte vleresimi me i mire per mesataren e popullimit  $\mu$ .

#### Vleresimi i variaces midis zgjedhjeve

Ne paragrafin e meparshem dhame kuptimin e vleresimit te variaces midis zgjedhjeve. Ky vleresim quhet katrori mesatar midis zgjedhjeve dhe shenohet me KMM. Formula per llogaritjen e KMM eshte si me poshte:

$$KMM = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2}{k - 1} = SKM / k - 1$$

SKM eshte shuma e katrore te midis zgjedhjeve. Emeruesi i raportit  $(k-1)$  paraqet shkallet e lirise te SKM. Ne qofte se hipoteza  $H_0$  eshte e vertete, KMM perben nje vleresim te pazhvendosur per  $\sigma^2$ . Megjithate ne qofte se mesataret e k popullimeve nuk jane te barabarta, KMM nuk eshte vleresim i pazhvendosur per  $\sigma^2$ , por nje vleresim me i madh ose nje mbivleresim per te. Per te dhenat e paraqitura ne tabelen 13.1 SKM=516 dhe KMM=258.

<sup>1</sup> Shkuritmi i togfjalshit Analysis of Variance



### Vleresimi i variaces brenda zgjedhjeve

Vleresimi tjetër për  $\sigma^2$  bazohet në variacionin e të dhënave të zgjedhjes brenda çdo zgjedhje. Vleresimi i gjetur në këtë mënyrë quhet katrori mesatar brenda zgjedhjeve dhe shenohet me KMB. Formula për llogaritjen e KMB është si më poshtë:

$$KMB = \frac{\sum_{j=1}^k (n_j - 1)s_j^2}{n_T - k} = \frac{SKB}{n_T - k}$$

Veme re se KMB bazohet në variacionin brenda çdo zgjedhje dhe si rrjedhojë nuk influencohet nga fakti nëse hipoteza zero është ose jo e vertetë. Kështu KMB jep gjithnjë një vleresim të pazhvendosur për variancën e përgjithshme. Për të dhënat e shembullit të marrim  $SKB=430$  dhe  $KMB=28,67$ .

### Krahasimi i vleresimeve të variaces; kriteri F

Le të supozojmë për momentin se hipoteza zero është e vertetë. Në këtë rast KMM dhe KMB janë dy vleresime të pavarura dhe të pazhvendosura për variancën e popullimit. Në kapitullin 11 pame se për popullimin normal, shpërndarja e raportit të dy vleresimeve të pavarura të variaces ka shpërndarje F. Kështu në qoftë se hipoteza zero është e vertetë dhe supozimet për ANOVA-n plotësohen, atëherë shpërndarja e zgjedhjes për KMM/KMB është shpërndarja F, me numrin e shkallëve të lirise për numruesin të barabartë me  $k-1$  dhe numrin e shkallëve të lirise për emruesin  $n_T - k$ .

Në qoftë se mesataret e popullimeve nuk janë të barabarta, raporti KMM/KMB do të tentojë të jetë më i madh se 1, sepse KMM është një mbivleresim për variancën e popullimit. Kështu hipoteza  $H_0$  do të hidhet poshtë, në qoftë se vlera e këtij raporti është tepër e madhe. Vlera e raportit KMM/KMB që shërben si kriter për hedhjen poshtë të hipotezës zero varet nga niveli i rendesise  $\alpha$ . Mbi bazën e vlerës të tij dhe të numrit të shkallëve të lirise mund të gjejmë vlerën kritike të raportit F. Në figurën 13.1 tregohet shpërndarja e raportit KMM/KMB dhe zona e hedhjes poshtë të hipotezës për nivelin e rendesise  $\alpha$ , ku  $F_k$  tregon vlerën kritike të kriterit.

Le të zbatojmë këto që thamë deri këtu për shembullin që kemi marrë në shqyrtim. Për këtë supozojmë se manageri përgjegjës për marrjen e vendimit do të pranonë një gabim të llojit të parë të barabartë me  $\alpha=0,05$ . Nga tabela e shpërndarjes së Fisherit bazuar në shkallët e lirise për numruesin dhe emruesin gjejmë vlerën kritike të  $F_{kritike}=3,29$ . Kështu rregulli i hedhjes poshtë të hipotezës në rastin e shembullit të marrë është: hipoteza zero hidhet poshtë në qoftë se  $KMM/KMB > 3,29$ . Vlera e vrojtuar e kriterit është  $F_v = KMM/KMB = 258/28,67 = 9$ , pra më e madhe se vlera kritike e kriterit  $F=3,29$ . Kjo do të thotë se hipoteza bazë mbi barazimin e mesatareve të tre reparteve nuk mund të pranohet.

### Tabela e analizës së variaces, ANOVA

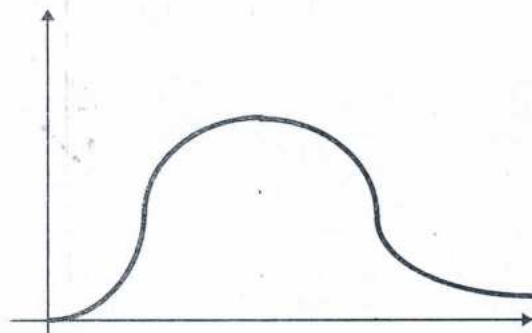
Rezultatet e llogaritjeve mund të paraqiten në një tabelë të posaçme që quhet tabela e analizës së variaces, ose tabela ANOVA. Tabela 13.2 tregon analizën e variaces për të dhënat e shembullit të marrë. Shuma e katrore që i përket burimit të përgjithshëm të variacionit quhet shuma e katroreve totale, SKT. Veme re se rezultatet e shembullit që kemi marrë tregojnë se plotësohet barazimi  $SKT=SKM + SKB$  dhe numri i shkallëve të lirise për SKT është i barabartë me shumën e numrit të shkallëve të lirise për SKM me numrin e shkallëve të lirise për SKB.

Tabela e analizës së variaces

| Burimi i variacionit | Shuma e katroreve | Shkallët e lirise | Katrori mesatar  | F |
|----------------------|-------------------|-------------------|------------------|---|
| Midis zgjedhjeve     | 516 <i>SKM</i>    | 2 <i>k-1</i>      | 258 <i>KMM</i>   | 9 |
| Brenda zgjedhjeve    | 430 <i>SKB</i>    | 15 <i>n_T-k</i>   | 28,67 <i>KMB</i> |   |
| Totali               | 946 <i>SKT</i>    | 17 <i>n_T-1</i>   | <i>KMM/KMB</i>   |   |



Figura 13.4 Shperndarja e zgjedhjes per raportin KMMIKMB. Vlera kritike per hedhien poshte te hipotezes zero



$$SKT = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

Theksojme se SKT e pjestuar me numrin e shkalleve te lirise  $n_T - 1$  nuk eshte gje tjeter vecse varianca e pergjitheshme e zgjedhjes, ne qofte se bashkesia e te dhenave prej 18 vrojtimesh do te trajtohej si nje popullim i vetem. Per te gjitha te dhenat e zgjedhjes, formula per llogaritjen e

shumes se katroreve totale SKT eshte  $\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$ . Per me teper, rezultatet e mesiperme

jane te verteta ne rastin me te pergjithshem. Me fjale te tjera SKT mund te ndahet ne dy shuma katroresh; ne shumen e katroreve midis zgjedhjeve dhe ne shumen e katroreve brenda zgjedhjeve. Veme ne dukje gjithashtu se edhe numri i shkalleve te lirise qe i pergjigjet shumes se katroreve totale  $(n_T - 1)$  shperbehet ne dy pjese, shkallet e lirise qe i pergjigjen shumes se katroreve midis grupeve  $(k - 1)$  dhe shkallet e lirise qe i korespondon shumes se katroreve brenda grupeve  $(n_T - k)$ . Keshtu, analiza e variaces mund te shikohet si nje proces i ndarjes se shumes se katroreve totale dhe i numrit te shkalleve te lirise ne pjese perkatesisht sipas burimeve te variacionit: midis dhe brenda zgjedhjeve. Duke pjestuar shumet me numrat perkates te shkalleve te lirise, marrim vleresimet per variancen dhe pastaj duke bere raportin e tyre marrim vleren e kriterit F qe perdoret per kontrollin e hipotezes per barazine e mesatare te popullimit.

#### Procedurat e krahasimeve te shumefishta

Duhet te kemi parasysh se kur perdorim analizen e variaces per kontrollin e barazise te k mesatareve, hedhja poshte e hipotezes zero na lejon te konkludojme vetem se mesataret e popullimeve nuk jane te gjitha te barabarta. Ne disa raste nga ky perfundim mund te jemi te kenaqur, por ne raste te tjera do te donim te shkonim nje hap me tej, per te percaktuar se cilat mesatare ndryshojne nga njera tjetra. Qellimi i ketij paragrafi eshte te japim disa metoda qe mund te perdoren per te bere krahasimet midis cifteve te mesatare te popullimeve.

#### Ndryshimi me i vogel i rendesishem (NVR) i Fisherit

Procedura e ndryshimit me te vogel te rendesishem te Fisherit (NVR) eshte nje nga metodat qe perdoren me shpesh per krahasimet e cifteve te mesatare te popullimeve. Ne rastin e shembullit tone ka tre popullime, hipotezat per te cilat paraqiten si me poshte.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_a$ : Jo te gjitha mesataret jane te barabarta

Duke perdorur analizen e variaces hipoteza zero u hodh poshte dhe u arrit ne perfundimin se rezultatet mesatare te testimit nuk jane te njejta per cdo repart. Ne kete rast mund te shtrohet pyetja: cilat mesatare do te konsiderohen se ndryshojne midis tyre ato te popullimeve 1 dhe 2, dhe 3 apo 1 dhe 3 apo te tre ciftet jane te ndryshem?

P.sh. le te shikojme ne se ekziston nje ndryshim i rendesishem midis mesatareve 1 dhe 2. Ne gjuhen e statistikes kjo do te thote te shtrosh per kontroll hipotezat e meposhtme:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

485677 484 me



Ne kapitullin 10 paraqitem nje procedure standarte per kontrollin e hipotezes qe mesataret e dy popullimeve jane te barabarta. Procedura NVR bazohet ne statistiken  $t$  te paraqitur me pare per rastin e dy popullimeve, me nje modifikim te vogel per menyren e vleresimit te variances se popullimit. Statistika e kontrollit per proceduren NVR eshte si me poshte:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{KMB(\frac{1}{n_1} + \frac{1}{n_2})}} \quad \checkmark$$

$$t_{2/2} \rightarrow KMB$$

Hipoteza  $H_0$  hidhet poshte ne qofte se  $t < -t_{\alpha/2}$ , ose  $t > t_{\alpha/2}$ . Ne rastin e shembullit tone kemi gjetur vleren e  $KMB = 28,67$ . Ky eshte vleresimi per variancen e pergjithshme i bazuar ne 15 shkalle lirie. Per nivelin e rendesise te barabarte me 0,05 dhe per numrin e shkalleve te lirise 15, ne tabelen e shpemdardjes  $t$  gjejme vleren kritike te kriterit  $t = 2,131$ . Keshtu, hipoteza  $H_0$  do te hidhet poshte ne qofte se vlera faktike e kriterit  $t$  e kalon ne madhesi absolute vleren kritike te tij. Ne rastin e shembullit tone vlera faktike e kriterit  $t$  eshte 1,62. Me qene se vlera e kriterit  $t = 1,62$  bie ne zonen e lejuar, ateher hipoteza zero nuk hidhet poshte dhe si rrjedhoje nuk mund te konkludojme se rezultatet e testimit per repartin 1 dhe repartin 2 te fabrikes ne fjale ndryshojne nga njera tjetra.

Disa zbatues te kesaj analize e gjejne me te pershtateshme te percaktojne se sa e madhe duhet te jete diferenca e mesatareve te zgjedhjeve per te hedhur poshte hipotezen  $H_0$ . Keshtu ne qofte se:

$$\bar{x}_1 - \bar{x}_2 > t_{\alpha/2} \sqrt{KMB(\frac{1}{n_1} + \frac{1}{n_2})} \quad \text{ose} \quad \bar{x}_1 - \bar{x}_2 < -t_{\alpha/2} \sqrt{KMB(\frac{1}{n_1} + \frac{1}{n_2})}$$

hipoteza zero hidhet poshte. Keshtu ne qofte se percaktojme ndryshimin me te vogel te rendesishem si:

$$NVR > t_{\alpha/2} \sqrt{KMB(\frac{1}{n_1} + \frac{1}{n_2})}$$

$$NVR = t_{\alpha/2} \sqrt{KMB(\frac{1}{n_1} + \frac{1}{n_2})}$$

ateher hipoteza  $H_0$  do te hidhet poshte ne qofte se  $\bar{x}_1 - \bar{x}_2 > NVR$  ose  $\bar{x}_1 - \bar{x}_2 < -NVR$

Ne qofte se marrim parasysh vetem madhesine absolute te ndryshimit, ateher rregullin e hedhjes poshte te hipotezes per ndryshimin me te vogel te rendesishem do ta shkruanim  $|\bar{x}_1 - \bar{x}_2| > NVR$ . Si ilustrim per kete po te llogaritim vleren e NVR per shembullin qe kemi marre me lart do te kemi  $NVR = 6,59$ .

Veme ne dukje se kur vellimet e zgjedhjeve jane te gjitha te barabarta, mjafton te llogaritim vetem nje vlere te NVR. Ne keto raste mund te krahasojme thjesht madhesine e differences midis cdo dy mesatareve me vleren e NVR. Per rastin e shembullit tone diferenca e mesatareve te zgjedhjeve per popullimin 1 dhe 3 eshte  $79 - 66 = 13$ . Me qene se kjo difference eshte me e madhe se 6,59, ateher ne baze te ketij kriteri hipoteza zero hidhet poshte dhe mesataret e popullimeve konsiderohen te ndryshme. Po ashtu me qene se diferenca midis mesatareve te zgjedhjeve per popullimet 2 dhe 3 eshte  $74 - 66 = 8 > 6,59$ , perseri hipoteza mbi barazine e mesatareve hidhet poshte.

Procedura e krahasimit me anen e NVR mund te perdoret gjithashtu per ndertimin e intervalit te besimit per diferencon midis dy mesatareve te zgjedhjeve. P.sh. nje vleresim per intervalin e besimit per diferencon e mesatareve te zgjedhjeve per repartet 1 dhe 2 jepet me anen e shprehjes se meposhtme:

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{KMB(\frac{1}{n_1} + \frac{1}{n_2})} \quad \checkmark$$

Duke patur parasysh percaktimin e NVR intervali i mesiperm mund te shkruhet  $\bar{x}_1 - \bar{x}_2 \pm NVR$ .

Ne qofte se intervali i besimit perfshin vleren 0, hipoteza qe mesataret e dy popullimeve jane te barabarta nuk mund te hidhet poshte. Megjithate, ne qofte se intervali i besimit nuk e







praktikues te tij u kthehen metodave te tjera qe lejojne gabime me te vogla te llojit te pare per eksperimentet.

### Korrigjimi Bonferroni

Ne nje problem ku zbatohet procedura NVR e Fisherit, koeficienti i gabimeve te llojit te pare te eksperimentit varet nga koeficienti i gabimeve te llojit te pare te krahasimit  $\alpha$  dhe nga numri i çifteve qe krahasohen. Keshtu ne vend te specifikimit te koeficientit te gabimeve te krahasimit, le te marrim te dhene vleren e  $\alpha_c$  qe deshirojme dhe pastaj te gjejmë vleren perkatese te  $\alpha$ .

Me lart pame probabilitetin e berjes te pakten te nje gabimi te llojit te pare ne rastin e problemeve qe permbajne k krahasime te çifteve te mundshme. Ky probabilitet eshte i kufizuar nga siper:  $-(1-\alpha) C_n^k < C_n^k \alpha$ , per çdo  $\alpha$  nga intervali  $]0; 1[$ . Prej kendej rrjedh se ne qofte se duam qe probabiliteti maksimal i berjes te gabimeve te llojit te pare per te gjithë eksperimentin te jete  $\alpha_c$ , mjafton te marrim koeficientin e gabimeve te llojit te pare te krahasimeve te barabarte me  $\alpha_c / C_n^k$ .

Per shembull le te supozojme se duam te perdorim proceduren NVR per kontrollin e tre krahasimeve te çifteve per shembullin qe kemi marre ne shqyrtim. Pervet kesaj le ta zeme se marrim koeficientin e gabimeve te llojit te pare per eksperimentin te barabarte me 0,05. Ne qofte se marrim koeficientin e gabimeve te krahasimit te barabarte me  $0,05/3=0,017$ , do te sigurohemi qe koeficienti i gabimeve te eksperimentit do te jete me i vogel ose i barabarte me 0,05. Keshtu kur perdorim proceduren NVR me korrigjimin e Bonferronit vlera e  $t$  eshte  $t_{0,017/2} = t_{0,0085}$ . Me qene se nuk ka tabela per percaktimin e vleres se sakte te  $t$  qe i korespondon ketij niveli te rendesise, shohim se nje vlere e perafert e saj eshte  $t_{0,01} = 2,602$ . Kujtojmë se vlera e NVR per shembullin tone eshte 6,59. Ne qofte se ne vend te vleres 2,131 perdorim vleren e  $t=2,602$ , atehere gjejmë nje vlere tjeter te NVR, qe do ta quajme ndryshimi i rendesishem sipas Bonferronit, NRB:

$$NRB = 2,602 \sqrt{28,67 \left( \frac{1}{6} + \frac{1}{6} \right)} = 8,04$$

*defekt  $\alpha_c \Rightarrow \alpha = \frac{\alpha_c}{C_n^k}$*   
 $NRB = t_{\frac{\alpha_c}{2}} \sqrt{C_n^k \left( \frac{1}{n} + \frac{1}{n} \right)}$   
 *$C_n^k = \frac{n(n-1)}{2}$  ku  $n$  m i zgjedhjes.*

Keshtu per te hedhur poshte hipotezen qe dy mesataret jane te barabarta me anen e NRB duhet te kemi nje ndryshim me te madh midis mesatareve se sa kur perdorim kriterin NVR.

Keshtu p.sh. per shembullin qe kemi marre diferenca e mesatareve midis repartit 1 dhe 2 eshte  $79-74=5$ . Meqenese kjo deference nuk kalon  $NRB=8,04$  hipoteza zero mbi barazine e dy mesatareve nuk mund te hidhet poshte. Persa i perket ndryshimit te mesatareve per repartet 1 dhe 3, ajo eshte  $79-66=13$  dhe e kalon vleren e  $NRB=8,04$ . Prej kendej rrjedh se hipoteza mbi barazine e tyre hidhet poshte dhe ndryshimi i mesatareve konsiderohet i rendesishem. Se fundi ndryshimi midis mesatareve te reparteve 2 dhe 3 eshte  $74-66=8$ , i cili eshte me i vogel se NRB dhe si rriedhoje ndryshimi midis tyre konsiderohet i parendesishem.

Ne qofte se problemi do te permbante 5 popullime dhe si rrjedhoje do te kishim 10 çifte qe krahasohen, korrigjimi do te rezultonte ne nje koeficient gabimi krahasimi te llojit te pare te barabarte me  $0,05/10=0,005$ . Sic dime nga kontrolli i hipotezave ne kapitullin 9, ne rastin e nje zgjedhje me vellim te kufizuar, çdo zvoglim i probabilitetit te gabimeve te llojit te pare do te conte ne rritjen e probabilitetit te gabimeve te llojit te dyte, gje qe do ti korespondonte situates kur hipoteza mbi barazine e mesatareve pranohet si e vertete kur ne fakt ato ndryshojne nga njera tjeter. Pikerisht kjo eshte aresyeja qe ne praktike nuk perdoret shpesh kriteri individual me koeficient te vogel te gabimeve te llojit te pare te krahasimit, meqenese ai rrezikon ne rritjen e gabimeve te llojit te dyte. Per te menjanuar kete rrezik mund te perdoret nje procedure tjeter, te cilen do ta shikojme me poshte.

### Procedura Tukey

Kjo procedure lejon eksperimentuesin te beje kontrollin e te gjithë krahasimeve te mundshme dhe te ruaje te pandryshuar gjate gjithë kohes nje koeficient gabimi te llojit te pare per eksperimentin te barabarte me 0,05. Kriteri ndertohet mbi bazen e gjerjesise se variacionit. Ne qofte se quajme me  $\bar{x}_{\max}$  mesataren me te madhe te zgjedhjes dhe me  $\bar{x}_{\min}$  mesataren me te vogel te zgjedhjes, atehere ne rastet kur vellimi i çdo zgjedhje eshte i njejte  $n$  dhe variancat e popullimeve jane te barabarta, shperndarja e statistikes:



$$q = \frac{\bar{x}_{\max} - \bar{x}_{\min}}{\sqrt{\frac{KMB}{n}}}$$

ndjek shperndarjen e gjerësisë së variacionit të studentit. Vlerat kritike të kësaj shpërndarje paraqiten në tabelën 11 të shtojcës B njëkohësisht për nivelet e rëndësise 0,05 dhe 0,01. Le të shikojmë zbatimin e këtij kriteri për rastin e shembullit të mëparshëm.

Kështu ne kemi 3 zgjedhje secila me 6 njësi dhe kërkojmë të zbatojmë kriterin e krahasimit të të gjithë çiftëve të mundshme të mesatareve të zgjedhjeve, duke patur gjithnjë një koeficient të njëjte gabimi për eksperimentin  $\alpha_e = 0,05$ . Vlerat që jepen në tabelën 13.6 u korespondojnë numrit të terësive nga ku janë bërë zgjedhjet (shtyllat e tabelës) dhe numrit të shkallëve të lirise (rreshtat) që lidhen me vlerësimin me të mirë të variancës së popullimit KMB. Për shembullin tone zgjedhim shtyllën 3 meqenëse numri i popullimeve është i barabartë me  $k=3$ . Nga tabela për numurin e shkallëve të lirise 15, rezulton se vlera kritike e kriterit është  $q = 3,67$ .

Tani ne baze të vlerave të  $q=3,67$ ,  $n=6$  dhe  $KMB=28,67$  mund të përdorim 13,22 për të përcaktuar se sa e madhe duhet të jetë diferenca midis dy mesatareve për të hedhur poshtë hipotezën që ato janë të barabarta. Kete vlerë do ta quajmë ndryshimi i rëndësishëm i Tukey-it, NRT.

ku  $\bar{x}_{\max} - \bar{x}_{\min} > NRT \Rightarrow H_0 \downarrow$

$$NRT = q \sqrt{\frac{KMB}{n}}$$

ku  $q = q_{k, n-k, \frac{\alpha}{2}}$  3 shkallë li

Për shembullin tone  $NRT=8,02$ . Në qoftë se diferenca e dy mesatareve në vlerë absolute është më e madhe se 8,02, atëherë kjo është e mjaftueshme për të hedhur poshtë hipotezën mbi barazinë e mesatareve të popullimeve dhe të konkludojmë se ato janë të ndryshme midis tyre. Procedura Tukey është e ngjashme me korigjimin Bonferoni, sepse duhet vërtetuar një difference më e madhe midis mesatareve të zgjedhjeve për të konkluduar nëse mesataret e popullimeve përkatëse ndryshojnë nga njëra tjetra.

Shohim se ndryshimi midis repartit 1 dhe 2 është  $79-74=5$ , kështu që sipas kësaj procedure nuk mund të hidhet poshtë hipoteza se mesataret e tyre janë të barabarta. Në rastin kur krahasojmë mesataret për repartet 1 dhe 3 diferenca midis tyre është e barabartë me  $79-66=13 > 8,02$  dhe si rrjedhojë hipoteza mbi barazinë e mesatareve të këtyre popullimeve hidhet poshtë. Në fundi për repartin 2 dhe 3 arrijmë në përfundimin se ndryshimi i mesatareve është i përendësishëm, sepse  $74-66=8 < 8,02$ , d.m.th. nuk ka baze që hipoteza zero mbi barazinë e tyre të hidhet poshtë.

Me anën e NRT mund të ndërtojmë gjithashtu intervalin e besimit për diferençën e dy mesatareve. Intervale të gjetura quhen intervale të njehershme të besimit për ndryshimet e çiftëve të mesatareve që krahasohen. Shprehja e përgjithshme e këtij intervali ka formën

$$\bar{x}_i - \bar{x}_j \pm q \sqrt{\frac{KMB}{n}}$$

Për rastin tone vlera e NRT pavarësisht nga popullimet që krahasohen është e barabartë me 8,02. Kështu bashkësia e intervaleve të njehershme të besimit është: për repartin 1 dhe 2  $(79-74) \pm 8,02$ , d.m.th. nga -3,02 në 13,02; për repartin 1 dhe 3  $(79-66) \pm 8,02$ , d.m.th. nga 4,98 në 21,02 dhe për repartin 2 dhe 3  $(74-66) \pm 8,02$ , d.m.th. nga -0,02 në 16,02.

Meqenëse për repartet 1 dhe 2 si edhe për repartet 2 dhe 3, vlera 0 përfshihet në intervalet përkatëse, nuk ka baze të mjaftueshme që hipoteza mbi barazinë e çiftëve të mesatareve të hidhet poshtë. Megjithatë meqenëse për repartet 1 dhe 3 intervali nuk përfshin vlerën zero, ekziston një baze statistikore për të konkluduar se mesataret e popullimeve përkatëse ndryshojnë midis tyre.

#### Disa rekomandime për kryerjen e krahasimeve të shumfishta

Ketu janë paraqitur tre procedure për kryerjen e krahasimeve të shumfishta: kriteri NVR i Fisherit, korigjimi Bonferoni dhe procedura Tukey. Kur përdorem kriterin NVR pame se ekzistonte një ndryshim i rëndësishëm midis mesatareve të popullimeve 3 dhe 4. Në qoftë se përdorim procedurën NVR, korigjimin Bonferoni, ose kur përdorim procedurën Tukey, arrijmë në përfundimin se ekziston ndryshim i rëndësishëm vetëm midis mesatareve të popullimeve 1 dhe 3. Kështu vlerësimi i ndryshimit të çiftëve të mesatareve që krahasohen varet nga metoda e përdorur.

$$1 - \alpha_{\text{total}} = \alpha_e$$

$$1 - \alpha_{\text{total}} = 1 - \alpha_e$$

$$1 - \alpha_{\text{total}} = \alpha_e$$

$$1 - \alpha_{\text{total}} = \alpha_e - 1$$

$$1 - \alpha_{\text{total}} = 1 - \alpha_e$$

$$\alpha_e = 1 - \alpha_{\text{total}}$$

$$\alpha_e = \frac{\alpha_{\text{total}}}{k}$$



Per zgjedhjen e metodave te vleresimit ka mendime te ndryshme se cila prej tyre duhet konsideruar me e mira. E verteta eshte se asnjera nga metodat nuk mund te konsiderohet me e mira per cdo tip problemi. Megjithate ne qofte se numri i krahasimeve te cifteve eshte i vogel rekomandohet te perdoret korigjimi Bonferroni. Ne qofte se numri i krahasimeve te cifteve eshte i madh, me e pershtatshme eshte procedure Tukey. P.sh. le ta zeme se duam te kryejme krahasimin e dy ose tre cifteve te mesatareve, ku seicila prej tyre perben nje vleresim paraprak te te dhenave. Ne kete rast rekomandohet korigjimi Bonferoni. Ne qofte se qellimi eshte kontrolli i gjithë cifteve te mundshme, atehere rekomandohet te perdoret procedura Tukey.

Studimet kane treguar se ne situatat kur rezultatet e analizës shikohen vetem si nje menyre per te sugjeruar se cfare duhet te shqyrtohet ne studimet ne te ardhmen, atehere preferohet te perdoret procedure NVR e Fisherit. Keshtu, kjo procedure na lejon te bazojme hedhjen poshte te hipotezes qe mesataret e dy popullimeve jane te barabarta per vlera me te vogla te diferencave te vrojtuar te mesatareve te zgjedhjeve, se sa do te kerkohej qe te ishin keto diferenca ne rastin e perdorimit te procedurave te tjera.

### **Hyrje ne modelimin e eksperimentit**

Te dhenat statistikore te grumbulluara mund te jene rezultat i vrojtimit ose i eksperimentit. Ne nje studim eksperimental ne filim percaktohen variablat qe na interesojne, dmth, variablat per te cilet kryhet studimi. Pastaj percaktohen nje ose me shume faktore te cilet gjate studimit mund te ndryshohen ose te vihen nen kontroll per te pare se si ata ndikojne ne variablat ne shqyrtim ne kushtet e tjera te pandryshueshme. Ne studimet jo eksperimentale, ose me vrojtim, nga ana e studiuesit nuk behen perpjekje per te kontrolluar faktoret qe ndikojne ne variablat ne shqyrtim.

Shembulli me i perhapur i studimeve me vrojtim eshte vrojtimi me zgjedhje. Nje shembull te tille pame ne analizen e variances per studimin e drejtimit te pergjithshem te cilesise. Keshtu per te pare se sa punonjesit e nje ndermarje dine rreth drejtimit te pergjithshem te cilesise, u krye nje zgjedhje tipike e rastit ku ne cdo repart u zgjedhen 6 punonjes ndaj te cileve u zbatua nje kriter testimi. Me tej rezultatet e vrojtimit per punonjesit e zgjedhur u perpunuan me anen e analizës se variances, per te pare ne se mesataret e popullimeve ishin ose jo te barabarta.

Si shembull te nje studimi statistikor eksperimental marrim rastin e nje ndermarje qe kerkon te prodhoje nje sistem te ri filtrimi per ujin e pishem. Komponentet per sistemin e ri te filtrimit do te blihen nga furnizues te ndryshem, kurse firma ne fjale do te beje montimin e tyre. Nje grup inxhinieresh ka pergjegjesine per percaktimin e metodës me te mire te montimit te sistemit te ri te filtrimit. Le ta zeme se pas analizës te nje numri te madh variantesh, firma eshte kufizuar ne tre menyra te montimit te sistemit te filtrimit, metoda A, metoda B dhe metoda C.

Ne pergjithesi nje faktor eshte nje variabel qe eksperimentuesi ka zgjedhur per vrojtim, kurse nje trajtim i korespondon nivelit te faktorit ne shqyrtim. Seicila nga metodat ndryshon nga te tjerat nga vargu i etapave te perdorura per montimin e produktit. Drejtimi i firmës kerkon te percaktojë se cila nga metodat çon ne numrin me te madh te sistemeve te filtrimit te prodhuara ne nje jave, d.m.th. qe siguron nje rendiment me te madh.

Ne eksperimentin e mesiperm ne rolin e faktorit qe analizohet eshte metoda e montimit. Meqenese jane tre metoda montimi, ose tre nivele qe i korespondojne ketij faktori, themi se ka tre trajtime qe lidhen me eksperimentin, te cilat u korespondojne metodave te montimit. Shembulli qe paraqitem eshte nje model eksperimenti me nje faktor te vetem, i cili eshte cilesor, metoda e montimit. Eksperimente te tjere mund te permbajne me shume faktore, disa nga te cilet mund te jene cilesore dhe disa te tjere sasiore.

Tre metodat e montimit ose trajtimet ne eksperimentin e mesiperm percaktojne tre popullime per te cilat interesohemi. Nje popullim u korespondon punetoreve qe perdorin metoden e montimit A, tjetri u korespondon atyre qe perdorin metoden B dhe i treti punetoreve qe perdorin metoden C. Veme re se variabli i rastit per cdo popullim eshte numri i sistemeve te filtrimit te montuara ne nje jave dhe qellimi primar statistikor per eksperimentin eshte te percaktojme ne se numri i njesive qe prodhohen ne jave eshte i njejte per te tre popullimet. Ne terminologjine e modelit eksperimental, variabli i rastit qe na intereson quhet variabel i varur, ose variabli pergjegjes.

Le ta zeme se nga teresia e punetoreve qe merren me procesin e montimit merret nje zgjedhje e rastit prej 3 punetoresh. Ne terminologjine e modelit te eksperimentit, 3 punetoret e



zgjedhur ne menyre te rastit quhen njesi eksperimentale. Modeli eksperimental qe do te perdorim per problemin e montimit quhet nje model plotesisht i rastit. Ky lloj modeli kerkon qe seicila nga metodat e montimit ose trajtimet tu jene shoqeruar ne menyre te rastit njesive te eksperimentit, ose punetoreve. P. sh. metoda A mund te jete caktuar per punetorin e dyte, metoda B per punetorin e pare dhe metoda C per punetorin e trete. Kuptimi i rastesise, ose i randomizimit sic eshte bere zakon te thuhet, eshte nje parim i rendesishem per te gjitha modelet eksperimentale.

Veme re se rezultati i eksperimentit eshte i vetem dhe pikerisht sa numri i njesive te montuara per cdo metode montimi ose trajtimi. Me fjale te tjera nje njesi e zgjedhjes i korespondon nje trajtimi, ose nje menyre montimi. Keshtu per te marre te dhena te tjera per cdo metode montimi, do te duhet te perserisim procesin eksperimental. P. sh. le ta zeme se ne vend te 3 punetoreve kemi zgjedhur ne menyre te rastit 15 punetore dhe pastaj cdo metode montimi ua shoqerojme cdo grupi prej 5 punetoresh te zgjedhur ne menyre te rastit nga 15 punetoret qe kemi zgjedhur gjithesej. Meqenese cdo menyre montimi u shoqerohet 5 punetoreve, themi se merren 5 perseritje. Ne figuren 13.3 paraqitet modeli eksperimental plotesisht i rastit per shembullin e eksperimentit per metodat e montimit.

#### Grumbullimi i te dhenave

Mbas percaktimit te modelit eksperimental, duhet te mbledhim te dhenat dhe ti analizojme ato. Ne rastin e shembullit tone ne fillim duhen instruktuar punetoret si te kryejne procesin e montimit ne perputhje me metodën e zgjedhur dhe pastaj te filloje te kryhet montimi i sistemit te ri te filtrimit. Le ta zeme se nga cdo puntor ne perputhje me metodën e caktuar eshte kryer montimi i sistemit te filtrimit dhe te dhenat mbi numurin e njesive te montuara nga seicili prej tyre gjate nje jave jane paraqitur ne tabelen 13.4. Numri mesatar i njesive te montuara per te tre metodat paraqitet si me poshte: 62 per metodën A, 66 per metodën B dhe 52 per metodën C.

Figura 13.3 Modeli plotesisht i rastit per eksperimentimin e metodës se montimit

Puntoret ne repartet A, B, C

Zgjedhja e 15 puntoreve  
per eksperimentin

Shperndarja e puntoreve sipas  
metodave te montimit

| Metoda A | Metoda B | Metoda C |
|----------|----------|----------|
| $n_1=5$  | $n_2=5$  | $n_3=5$  |

Tabela 13.4 Numri i njesive te prodhuara ne jave nga seicili prej puntoreve te zgjedhur

|          |    |    |    |    |    |              |               |                        |
|----------|----|----|----|----|----|--------------|---------------|------------------------|
| Metoda A | 68 | 54 | 55 | 66 | 67 | Mesatarja=62 | Varianca=27,5 | Devijimi standart=5,24 |
| Metoda B | 58 | 69 | 71 | 64 | 68 | Mesatarja=66 | Varianca=27,5 | Devijimi standart=5,15 |
| Metoda C | 48 | 57 | 59 | 47 | 49 | Mesatarja=52 | Varianca=27,5 | Devijimi standart=5,57 |

Nga te dhenat e mesiperme shihet se metoda e dyte ka nje rendiment me te larte se metodat e tjera. Megjithate rendimenti i punes te puntoreve eshte i ndryshem, si ne rastin kur aplikohen metoda te ndryshme montimi, ashtu dhe ne rastin kur aplikohet e njejta metode. Keshtu me qene se zgjedhja e puntoreve eshte e rastit dhe numri i tyre eshte i vogel, mund te ndodhe qe ndryshimi i rendimenteve per puntoret individuale te jete ne kufijte e luhatjeve te rastit dhe si rrjedhoje edhe ndryshimi i mesatareve per seicilen nga metodat e montimit te jete i



rastit. Keshtu problemi qe shtrahet eshte te percaktojme ne se ky ndryshim ne mesataret e zgjedhjeve te vrojtura eshte aq i madh sa te arrijme ne perfundimin se vete mesataret e popullimeve qe u korespondojne metodave te montimit jane te ndryshme. Per te paraqitur kete problem ne termat statistikore bejne shenimet e meposhtme:

$\mu_1$  = numri mesatar i njesive te prodhuara ne jave me metoden A

$\mu_2$  = numri mesatar i njesive te prodhuara ne jave me metoden B

$\mu_3$  = numri mesatar i njesive te prodhuara ne jave me metoden C

Hipotezat statistikore qe shtrahen per kontroll jane:

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_a$ : Jo te gjitha mesataret jane te barabarta

Keshtu ne rastin e modeleve plotesisht te rastit, problemi me te cilin perballohem eshte analog me ate qe pame ne rastin e analizes se variaces, d.m.th. kerkohet te kontrollohet hipoteza qe mesataret e dy ose me shume popullimeve jene te barabarta. Ne paragrafin tjeter do te tregojme se si mund te perdoret analiza e variaces ne situatat e ngjashme me ate qe paraqitem me lart.

2

### Modelet plotesisht te rastit

Kur analizojme te dhenat ne nje model plotesisht te rastit, hipotezat qe duam te kontrollojme jane krejtesisht te njejta me formen e hipotezave qe paraqitem ne paragrafin 13.2.

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$

$H_a$ : Jo te gjitha mesataret e teresive jane te barabarta

ku  $\mu_j$  = mesatarja e teresise j

Kontrolli i ketyre hipotezave bazohet ne supozimet e bera ne paragrafin 13.1 per analizen e variaces, prandaj edhe ne rastin e modelit eksperimental plotesisht te rastit mund te perdorim analizen e variaces. Kujtojmë se analiza e variaces kerkon llogaritjen e dy vleresimeve te pavarura per variancen e popullimit  $\sigma^2$ .

**Vleresimi i variaces midis trajtimeve.** Ne rastin e modelit eksperimental vleresimi midis trajtimeve quhet katrori mesatar per trajtimet dhe shenohet KMTR. Kjo eshte e njejte me ate qe quajtem ne paragrafin 13.2 katrori mesatar midis zgjedhjeve KMM. Formula per llogaritjen e katrorit mesatar per trajtimet eshte si me poshte:

$$KMTR = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2}{k - 1}$$

Numruesi i formule 13.25 quhet shuma e katroreve midis trajtimeve dhe shenohet me SKTR. Emruesi i saj k-1 shpreh numrin e shkalleve te lirise qe lidhet me SKTR. Duke e zbatuar kete formule per te dhenat e shembullit tone te paraqitura ne tabelen 13.4, marrim SKTR=520 dhe KMTR=260.

**Vleresimi i variaces brenda trajtimeve.** Vleresimi i dyte per  $\sigma^2$  bazohet ne variacionin e te dhenave brenda trajtimeve. Ne rastin e analizes se variaces kete vleresim e quajtem vleresimi brenda zgjedhjeve per variancen e popullimit. Ky vleresim quhet katrori mesatar per gabimin dhe shenohet me KMG. Ai eshte i njejte me katrorin mesatar te zgjedhjeve, i cili gjithashtu quhet katrori mesatar brenda trajtimeve. Formula per llogaritjen e KMG eshte si me poshte:

$$KMG = \frac{\sum_{j=1}^k (n_j - 1) s_j^2}{n_T - k}$$

Numruesi i formule se mesiperme quhet shuma e katroreve brenda trajtimeve, ose shuma e katroreve per gabimin dhe shenohet me SKG, ndersa emruesi i shprehjes tregon numrin e shkalleve te lirise qe lidhet me vleresimin e variaces brenda trajtimeve. Per te dhenat e shembullit te tabelës 13.4 marrim SKG=340 dhe KMG=340/(15-3)=28,33.

**Krahasimi i vleresimeve te variaces: Kriteri F.** Ne qofte se hipoteza zero eshte e vertete dhe supozimet per ANOVA-n plotesohen, raporti KMTR/KMG ka shpemdardje F me numur te shkalleve te lirise per numruesin te barabarte me k -1 dhe per emruesin te barabarte



me  $n_T-k$ . Kujtojmë gjithashtu se në qoftë se mesataret e  $k$  popullimeve nuk janë të barabarta, vlera e raportit  $KMTR/KMG$  do të tentojë të jetë e madhe, sepse  $KMTR$  përben një mbivlerësim për  $\sigma^2$ . Kështu hipoteza zero do të hidhet poshtë në qoftë se gjejmë një vlerë tepër të madhe të këtij raporti, e cila nuk mund të ketë ardhur nga një zgjedhje e rastit me shpërndarje  $F$  me numur të shkallëve të lirise për numruesin  $k-1$  dhe për emruesin  $n_T-k$ .

Për problemin tonë vlera e raportit është  $KMTR/KMG=260/28,33=9,18$ . Vlera kritike e  $F$  për numrin e shkallëve të lirise 2 për numruesin dhe 12 për emruesin dhe për nivelin e rendesisë  $\alpha=0,05$  është e barabartë me 3,89. Meqëse vlera e vërtetuar e  $F$  është më e madhe se vlera kritike e saj, hipoteza zero hidhet poshtë dhe arrihet në përfundimin se jo të gjitha mesataret e popullimeve janë të barabarta.

**Tabela ANOVA për modelet plotësisht të rastit.** Duke përdorur terminologjinë që dhame për modelet eksperimentale plotësisht të rastit, mund të shkruajmë rezultatim për shpërberjen e shumës totale të katroreve  $SKT$  si shumë të katrorevebrenda trajtimeve me ato të gabimeve:  $SKT=SKTR+SKG$ . Vëmë re se një lidhje e tillë është e vërtetë edhe për numrin e shkallëve të lirise, dmth numuri i shkallëve të lirise për  $SKT$  është i barabartë me shumën e numurit të shkallëve të lirise për  $SKTR$  dhe numrin e shkallëve të lirise për  $SKG$ . Forma e përgjithshme për tabelën ANOVA për modelin plotësisht të rastit tregohet në tabelën 13.5. Në tabelën 13.6 jepet tabela ANOVA për shembullin e metodave të montimit.

Tabela 13.5. Tabela ANOVA për modelin plotësisht të rastit

| Burimi i variacionit | Shuma e katroreve | Shkallet e lirise | Katrori mesatar   | F          |
|----------------------|-------------------|-------------------|-------------------|------------|
| Trajtimet            | SKTR              | $k-1$             | $KMTR=SKTR/(k-1)$ | $KMTR/KMG$ |
| Gabimi               | SKG               | $n_T-k$           | $KMG=SKG/(n_T-k)$ |            |
| Totali               | SKT               | $n_T-1$           |                   |            |

Tabela 13.6 Tabela ANOVA për problemin e metodave të montimit

| Burimi i variacionit | Shuma e katroreve | Shkallet e lirise | Katrori mesatar | F    |
|----------------------|-------------------|-------------------|-----------------|------|
| Trajtimet            | 520               | 2                 | 260             | 9.18 |
| Gabimi               | 340               | 12                | 28.33           |      |
| Totali               | 860               | 14                |                 |      |

**Krahasimet e çifteve të mesatare.** Për kontrollin e krahasimeve të të gjithë çifteve të mundshme sic dihet mund të përdorim procedurën Tukey. Për nivelin e rendesisë 5% për të dhënat e tabelës 13.8, mund të gjejmë vlerën kritike të shpërndarjes të gjërësise së variacionit të Studentit për  $k=3$  popullime dhe 12 shkallë lirie. Kjo vlerë është  $q=3,77$ . Duke zëvendësuar në formulë vlerën perkatese për  $KMG$  gjejmë  $NRT=8,97$ . Kështu në qoftë se madhësia e differences midis çdo dy zgjedhjeve e kalon 8,97 atëherë hipoteza zero hidhet poshtë duke arritur në përfundimin se mesataret e popullimeve nuk janë të barabarta. Për shembullin tonë rezultatet e krahasimit të çifteve janë si më poshtë:

| $k-1+n-k$ | Diferencat midis mesatareve të zgjedhjeve | Vleresimi i diferencave |
|-----------|-------------------------------------------|-------------------------|
| 2-1=1     | Metoda A - Metoda B = $62 - 66 = -4$      | E përendesishme         |
| 2-2=0     | Metoda A - Metoda C = $62 - 52 = 10$      | E rendesishme           |
| 2-3=0     | Metoda B - Metoda C = $66 - 52 = 14$      | E rendesishme           |

Kështu diferenca midis mesatareve të popullimeve për metodat A dhe C dhe për metodat B dhe C paraqiten të rendesishme, ndërsa për metodat A dhe B ndryshimi midis mesatareve paraqitet i përendesishëm. Si përfundim metodat A dhe B janë më të preferuara se metoda C. Megjithatë një kontroll tjetër duhet bërë për metodat A dhe B, të cilat në studimin aktual nuk vlerësohet se janë të ndryshme nga njera tjetra.

### 3 Modelet e blloqeve të rastit

Deri këtu kemi parë modelet e blloqeve plotësisht të rastit. Kujtojmë se për kontrollin e differences ndërmjet mesatareve të trajtimeve kemi llogaritur raportin  $F=KMTR/KMG$ . Gjate krahasimeve të mesatareve mund të ndodhë që për shkak të faktoreve të papërfshirë në modelin eksperimental, termi i gabimit  $KMG$  të jetë më i madh se ç'pritet dhe si rrjedhojë vlera e



ketij variacioni nga termi i KMG. Keshu studimi nuk u modelua per kontrollin e ndryshimeve te individeve nga pikpamja e stresit ne pune.

Tabela 13.9. Tabela ANOVA per metodat e organizimit te punes

| Burimi i variacionit | Shuma e katrore | Shkallet e lirise | Katrori mesatar | F            |
|----------------------|-----------------|-------------------|-----------------|--------------|
| Trajtimet            | 21              | 2                 | 10.5            | 10.5/1.9=5.5 |
| Bllloqet             | 30              | 5                 | 6               |              |
| Gabimi               | 19              | 10                | 1.9             |              |
| Totali               | 70              | 17                |                 |              |

Disa analiste llogaritin edhe kriterin  $F = KMG/KMG$ , duke e perdorur kete statistike edhe per kontrollin e rendesise te ndryshimit midis bllloqeve. Pastaj rezultatin e gjetur e perdorin per te pare ne se ky model eksperimenti eshte i pershtatshem per modelimin e eksperimenteve per te perfshire edhe kete faktor ne te ardhmen. Megjithate shumica mendojne se ne qofte se ndryshimi individual per stresin ne pune perben nje faktor per studim, atehere per kete duhet te perdoret nje model tjeter eksperimental dhe pikerisht ne rastin tone modeli eksperimental dy faktorial.

### 3 Eksperimentet me disa faktore

Modelet eksperimentale qe kemi pare deri tani na ndihmojne per nxjerrjen e konkluzioneve statistikore per nje faktor. Megjithate ne disa eksperimente lind nevoja te nxirren konkluzione per me shume se nje variabel ose faktor. Modelimi faktorial dhe procedure ANOVA jane vecanerisht te nevojshme te perdoren ne rastin kur kerkohet te nxirren per dy ose me shume faktore ne te njejten kohe. Perdorimi i termit faktorial lidhet me faktin se kushtet eksperimentale perfshijne te gjitha kombinimet e mundshme te faktoreve te perfshire ne eksperiment. P.sh. ne qofte se faktori A ka  $a$  nivele dhe faktori B  $b$  te tille, atehere eksperimenti do te perfshije  $a*b$  kombinime te trajtimeve. Ky model eksperimenti mund te zgjerohet edhe per me shume se dy faktore.

Si ilustrim per eksperimentin dy faktorial marrim nje shembull qe ka te beje me provimin e konkursit te matematikes per pranimet ne universitet. Per kete kerkohet te kontrollohet ne se ndikon ne rezultatin e provimit te konkursit te matematikes, vendbanimi qytet-fshat i studentit dhe konsultimet qe studentet bejne me mesues perpara se ata te hyjne ne provimin e matematikes. Studentet jane ndare ne 3 grupe sipas vendbanimit: qytete te medha, qytete te tjera dhe fshat. Persa i perket faktorit te dyte, konsultimit, ose pergatitjes paraprake per provimin e matematikes, studentet jane ndare gjithashtu ne 3 grupe, studente qe nuk kane bere konsultim me mesues, studente qe kane bere deri ne nje jave konsultim dhe studente qe kane bere me shume se dy jave konsultim me mesues.

Tabela 13.10. Rezultatet e konkursit ne eksperimentin dy faktorial

| Vendbanimi i studentit | Koha e konsultimit me mesues |             |              |
|------------------------|------------------------------|-------------|--------------|
|                        | Mbi 1 jave                   | Deri 1 jave | Pa konsultim |
| Qytete te medha        | 60; 58                       | 56; 60      | 48; 41       |
| Qytete te tjera        | 56; 62                       | 46; 54      | 42; 48       |
| Fshat                  | 54; 46                       | 50; 58      | 48; 40       |

Keshu modeli faktorial per kete eksperiment me 3 trajtime per faktorin A (vendbanimi qytet-fshat) dhe 3 trajtime per faktorin B (programi pergatitor per provim) paraqitet me  $3*3=9$  kombinime te trajtimeve. Le ta zeme se per seicilin nga keto 9 kombinime kemi bere nga nje zgjedhje te rastit prej 2 studentesh, pra gjithesej ne eksperiment jane perfshire 18 studente, ku piket e fituar nga seicili prej tyre jane paraqitur ne tabelen 13.10.

Duke zbatuar analizen e variances per te dhenat e tabeles 13.10 jemi ne gjendje tu pergjigjemi pyetjeve: a ndikon vendbanimi qytet-fshat ne rezultatin e provimit te matematikes (Efekti i faktorit A); a ndikon pergatitja paraprake me mesues ne provimin e matematikes (Efekti i faktorit B); a ka ndikim kombinimi i niveleve te faktoreve ne rezultatin e provimit te matematikes (Efekti i bashkeveprimit te faktoreve)?



Ne qofte ndikimi i faktorit A eshte i rendesishem, atehere duhet te merren masa per mbarevajtjen e shkollave ne keto zona; ndersa neqofte se ndikimi i faktorit B eshte i rendesishem, atehere mund te organizohen konsultime te rregullta prane shkollave per tu ardhur ne ndihme studenteve.

**Procedura ANOVA.** Eksperimenti dy faktorial mund te ndertohet ose sipas modelit te blloqeve plotesisht te rastit, ose sipas modelit te blloqeve te rastit, ne kuptimin qe ne te gjitha rastet shuma e katroreve totale dhe numri i pergjithshem i shkalleve te lirise te shperbehet sipas burimeve te tyre perkatese. Formula per ndarjen e shumes se katroreve ne eksperimentin dy faktorial eshte:  $SKT = SKA + SKB + SKAB + SKG$ . Ndarja e shumes se katroreve dhe e numrit perkates te shkalleve te lirise eshte permbledhur ne tabelen 13.11. Per kete jane perdorur simbolet: a- numuri i niveleve te faktorit A; b-numuri i niveleve te faktorit B; r-numuri i perseritjeve;  $n_r$ -numuri total i vrojtimeve ne eksperiment.

Per llogaritjen e statistikes F qe nevojitet per kontrollin e rendesise te faktorit A, te faktorit B dhe te bashkeveprimit te tyre duhet te llogaritim KMA, KMB, dhe KMAB. Per llogaritjen e katroreve mesatare duhet te llogaritim me pare shumet e katroreve perkatese, SKA, SKB, SKAB dhe SKG, por krahas tyre edhe SKT. Pervec shenimeve te bera me lart per a, b, r dhe  $n_r$ , bejme edhe shenimet e meposhtme.

$x_{ijk}$ -vrojtimi qe i perket perseritjes k per trajtimin i te faktorit A dhe trajtimin j per faktorin B.

$\bar{x}_i$  -mesatarja e zgjedhjes per vrojtimet ne trajtimin i ( faktori A)

$\bar{x}_j$  -mesatarja e zgjedhjes per vrojtimet ne trajtimin j ( faktori B)

$\bar{x}_{ij}$  -mesatarja e zgjedhjes per vrojtimet qe u korespondojne kombinimit te trajtimit i te faktorit A dhe trajtimit j te faktorit B.

$\bar{\bar{x}}$  mesatarja e pergjithshme e zgjedhjes per te gjitha vrojtimet.

Per thjeshtimin e llogaritjeve sugjerohet procedura e meposhtme:

Etapa 1. Llogaritim shumen e katroreve totale:  $SKT = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (x_{ijk} - \bar{\bar{x}})^2 = 824,5$

Etapa 2. Llogaritim shumen e katroreve per faktorin A:  $SKA = br \sum_{i=1}^a (\bar{x}_i - \bar{\bar{x}})^2 = 61$

Etapa 3. Llogaritim shumen e katroreve per faktorin B:  $SKB = ar \sum_{j=1}^b (\bar{x}_j - \bar{\bar{x}})^2 = 453$

Etapa 4. Llogaritim shumen e katroreve per bashkeveprimin AB:

$$SKAB = r \sum_{i=1}^a \sum_{j=1}^b (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2 = 112$$

Etapa 5. Llogaritim shumen e katroreve per gabimin SKG:  $SKG = SKT - SKA - SKB - SKAB = 198,5$

Shumat e katroreve i pjetojme me numrat perkates te shkalleve te lirise, sic tregohet ne tabelen 13.12, per te marre vlerat e katroreve mesatare qe jane te nevojshem per kontrollin e efektit te faktoreve dhe te bashkeveprimit te tyre. Per kontrollin e ndryshimit te mesatareve perdorim kriterin F, qe ne qdo rast paraqitet si raport i katroreve mesatare perkates. Vlera e kritike e kriterit ne rastin e faktorit A gjendet sipas nivelit te rendesise  $\alpha = 0,05$  dhe numrit te shkalleve te lirise per numruesin dhe emruesin, perkatesisht te barabarte me 2 dhe 9,  $F_{0,05} = 4,26$ . Vlera faktike e kriterit F rezulton 1,38, e cila eshte me vogel se vlera kritike e tij e gjetur me lart. Prandaj hipoteza zero nuk mund te hidhet poshte dhe si rrjedhoje nuk vihet re ndonje ndryshim i madh ne rezultatin e provimit te matematikes ne varesi nga vendbanimi qytet-fshat i studenteve. Persa i perket faktorit te dyte, parapregatitjes per provim, vlera faktike e kriterit  $F = 10,27$  e kalon vleren kritike te tij 4,26, prandaj arrijme ne perfundimin se efekti i tij ne rezultatin e provimit eshte i rendesishem.

Se fundi per efektin e bashkeveprimit te ketyre faktoreve vlera faktike e kriterit F eshte 1,27 dhe vlera kritike e tij 3,63. Si rrjedhoje nuk vihet re nje ndikim i rendesishem i bashkeveprimit te faktoreve ne rezultatin e provimit.



Tabela 13.12. Tabela ANOVA per eksperimentin dy faktorial me r perseritje

| Burimi i variacionit | Shuma e katroreve | Shkallet e lirise | Katrori mesatar | F        |
|----------------------|-------------------|-------------------|-----------------|----------|
| Faktori A            | SKA               | a-1               | KMA/(a-1)       | KMA/KMG  |
| Faktori B            | SKB               | b-1               | KMB/(b-1)       | KMB/KMG  |
| Bashkeveprimi        | SKAB              | (a-1)(b-1)        | KMAB/(a-1)(b-1) | KMAB/KMG |
| Gabimi               | SKG               | ab(r-1)           | KMG/ab(r-1)     |          |
| Totali               | SKT               | n <sub>T</sub> -1 |                 |          |

Tabela 13.13 Te dhenat permbledhese per eksperimentin dy faktorial

| Faktori A<br>Lokalizimi i shkolles | Faktori B<br>Koha e konsultimit me mesues |                |                |                          |
|------------------------------------|-------------------------------------------|----------------|----------------|--------------------------|
|                                    | Mbi 1 jave                                | Deri ne 1 jave | Pa konsultim   | Mesataret per faktorin A |
| Qytetet e medha<br>Mesataret       | 60; 58<br>59                              | 56; 60<br>58   | 48; 41<br>44,5 | 53,83                    |
| Qytetet e vogla<br>Mesataret       | 56; 62<br>59                              | 46; 54<br>50   | 42; 48<br>45   | 51,33                    |
| Fshat<br>Mesataret                 | 54; 46<br>50                              | 50; 58<br>54   | 48; 40<br>44   | 49,33                    |
| Mesataret per faktorin B           | 56                                        | 54             | 44,5           | 51,5                     |

Tabela 13.14 Tabela ANOVA per matjen e ndikimit te faktoreve

| Burimi i variacionit | Shuma e katroreve | Shkallet e lirise | Katrori mesatar | F     |
|----------------------|-------------------|-------------------|-----------------|-------|
| Faktori A            | 61                | 2                 | 30.5            | 1.38  |
| Faktori B            | 453               | 2                 | 226.5           | 10.27 |
| Bashkeveprimi        | 112               | 4                 | 28              | 1.27  |
| Gabimi               | 198.5             | 9                 | 22.06           |       |
| Totali               | 824.5             | 17                |                 |       |

Keshtu, parapregatitja per provim paraqitet nje faktor i rendesishem per provimin e matematikes. Nga tabela 13.15 shohim se mesataret e rezultateve ne provim sipas kohes se konsultimit paraqiten si me poshte: per studentet qe kane bere konsultime me mesues me shume se nje jave eshte 56, per ata qe kane bere konsultime deri ne nje jave eshte 54 dhe per studentet qe nuk jane konsultuar me mesues numri i pikeve te fituara ne provimin e konkursit eshte 44,5. Per kete analiza duhet te vazhdoje me krahasimin e çifteve te mesatareve. Nga krahasimi i tyre rrjedh se nuk ka ndikim ne rezultatin ne provim per studentet qe kane bere konsultime me mesues me shume se nje jave dhe te studenteve qe jane konsultuar deri ne nje jave. Megjithate puna e studenteve qe nuk jane konsultuar me mesues ne krahasim me studentet e tjere qe kane bere konsultime te rregullta duket se eshte e pamjaftueshme lidhur me rezultatin e arritur ne provim. Prandaj ky problem mund kihet parasysh nga shkollat e mesme per te organizuar kurse te rregullta konsultimesh gjate veres, per tu ardhur ne ndihme nxenesve te tyre per provimin e konkurimit ne matematike per tu pranuar ne universitet.

$$2 \cdot 61 - 1 = (n_T) - 1 \Rightarrow$$

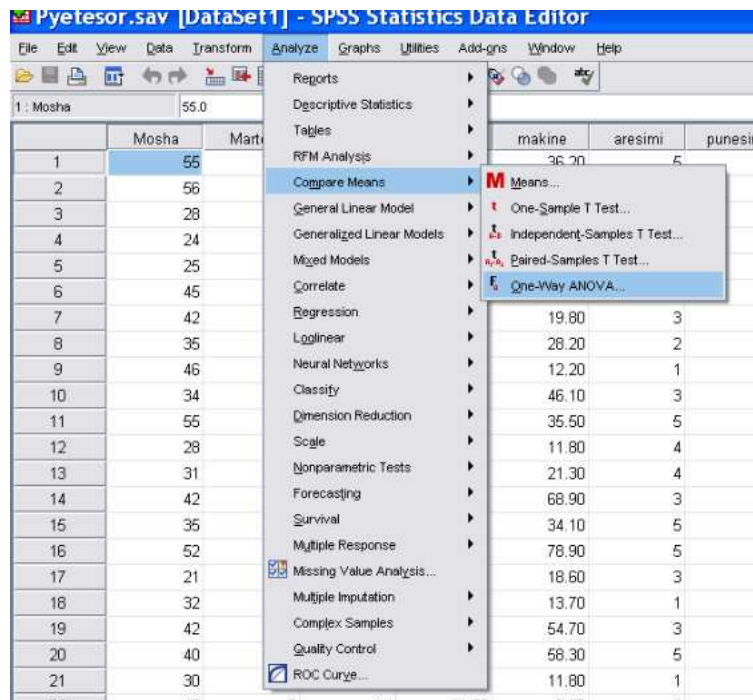
$$\Rightarrow$$

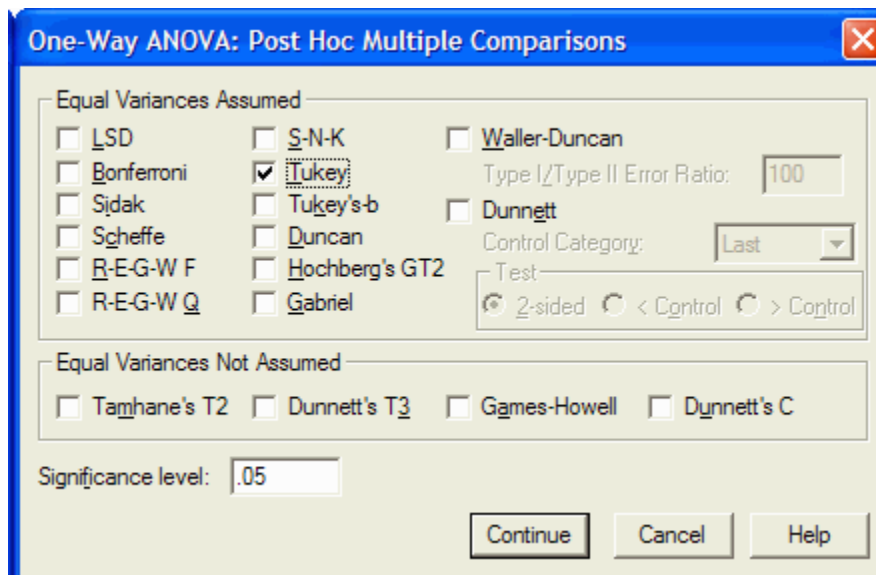
## ANOVA NE SPSS

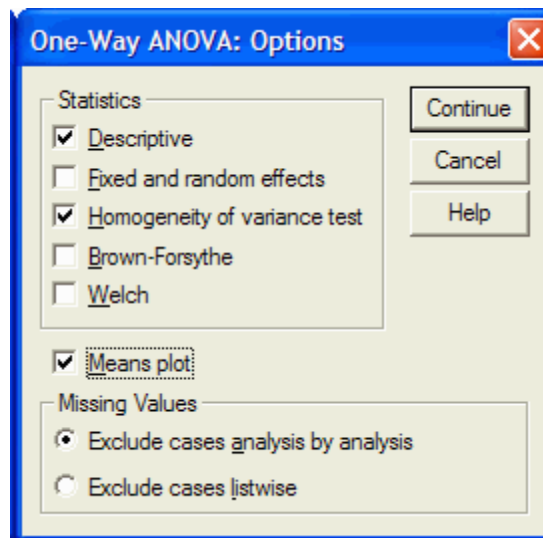
Do të testojmë nëse faktori arësimim ndikon në të ardhurat mesatare mujore të të intervistuarve ?  
E thënë ndryshe:

H<sub>0</sub>:  $\mu_{\text{shk.mesme}} = \mu_{\text{kolegj}} = \mu_{\text{universitet}} = \mu_{\text{pasuniversitet}}$

H<sub>1</sub>: të paktën një nga  $\mu$  të e ndryshme







#### Test of Homogeneity of Variances

Te ardhura te familjes ne mije Dollare

| Levene Statistic | df1 | df2  | Sig. |
|------------------|-----|------|------|
| 3.547            | 4   | 1995 | .007 |

$Sig = 0.007$ , tregon që variancat nuk janë të barabarta. Në kushte reale do të përdornim metodat joparametrike. Për efekt ushtrimi sup. Se plotësohet kushti për barazi variancash, për të vazhduar me ANOVA

#### ANOVA

Të ardhura të familjes në mijë Dollarë

|                | Sum of Squares | df   | Mean Square | F     | Sig. |
|----------------|----------------|------|-------------|-------|------|
| Between Groups | 90516.267      | 4    | 22629.067   | 3.227 | .012 |
| Within Groups  | 1.399E7        | 1995 | 7011.832    |       |      |
| Total          | 1.408E7        | 1999 |             |       |      |

Sig. =0.012, do të thotë se Ho nuk qëndron pra ka ndryshim statistikisht të rëndësishëm midis mesatareve të të ardhurave në varësi të nivelit të arësimimit.

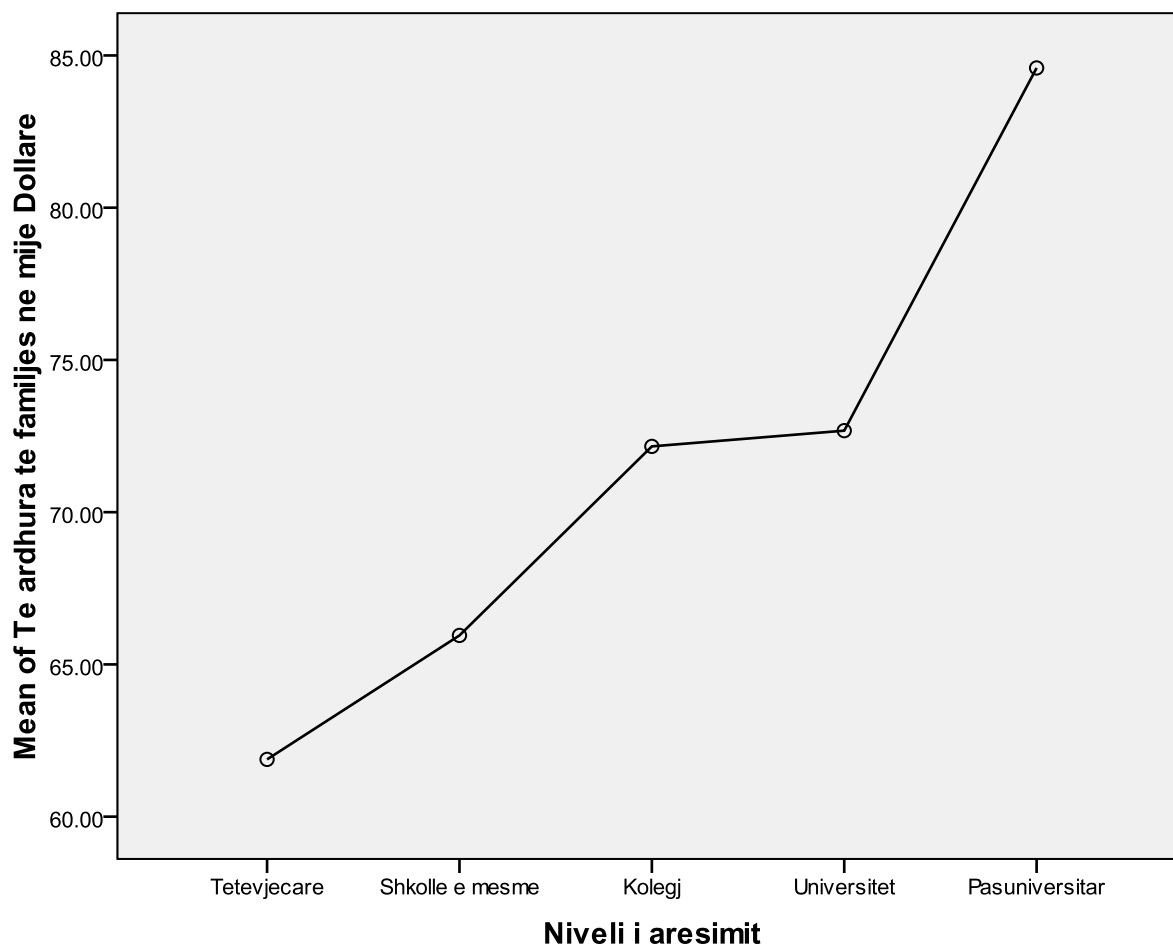
#### Multiple Comparisons

Të ardhura të familjes në mijë Dollarë

Tukey HSD

| (I) Niveli i aresimit | (J) Niveli i aresimit | Mean Difference<br>(I-J) | Std. Error | Sig.  | 95% Confidence Interval |             |
|-----------------------|-----------------------|--------------------------|------------|-------|-------------------------|-------------|
|                       |                       |                          |            |       | Lower Bound             | Upper Bound |
| Tetevjecare           | Shkolle e mesme       | -4.07088                 | 5.38817    | .943  | -18.7820                | 10.6403     |
|                       | Kolegj                | -10.28041                | 5.78992    | .388  | -26.0885                | 5.5276      |
|                       | Universitet           | -10.79669                | 5.91363    | .359  | -26.9425                | 5.3491      |
|                       | Pasuniversitar        | -22.70896*               | 6.85861    | .008  | -41.4348                | -3.9831     |
| Shkolle e mesme       | Tetevjecare           | 4.07088                  | 5.38817    | .943  | -10.6403                | 18.7820     |
|                       | Kolegj                | -6.20954                 | 5.45098    | .786  | -21.0922                | 8.6731      |
|                       | Universitet           | -6.72581                 | 5.58221    | .749  | -21.9668                | 8.5151      |
|                       | Pasuniversitar        | -18.63808*               | 6.57500    | .037  | -36.5896                | -.6866      |
| Kolegj                | Tetevjecare           | 10.28041                 | 5.78992    | .388  | -5.5276                 | 26.0885     |
|                       | Shkolle e mesme       | 6.20954                  | 5.45098    | .786  | -8.6731                 | 21.0922     |
|                       | Universitet           | -.51627                  | 5.97092    | 1.000 | -16.8185                | 15.7860     |
|                       | Pasuniversitar        | -12.42854                | 6.90807    | .374  | -31.2894                | 6.4324      |
| Universitet           | Tetevjecare           | 10.79669                 | 5.91363    | .359  | -5.3491                 | 26.9425     |
|                       | Shkolle e mesme       | 6.72581                  | 5.58221    | .749  | -8.5151                 | 21.9668     |
|                       | Kolegj                | .51627                   | 5.97092    | 1.000 | -15.7860                | 16.8185     |
|                       | Pasuniversitar        | -11.91227                | 7.01207    | .435  | -31.0571                | 7.2326      |
| Pasuniversitar        | Tetevjecare           | 22.70896*                | 6.85861    | .008  | 3.9831                  | 41.4348     |
|                       | Shkolle e mesme       | 18.63808*                | 6.57500    | .037  | .6866                   | 36.5896     |
|                       | Kolegj                | 12.42854                 | 6.90807    | .374  | -6.4324                 | 31.2894     |
|                       | Universitet           | 11.91227                 | 7.01207    | .435  | -7.2326                 | 31.0571     |

\*. The mean difference is significant at the 0.05 level.





## TEMA 9:

### TESTIMI I HIPOTEZAVE (RASTI I 'K' POPULLIMEVE): ANOVA (NJË DHE DY FAKTORIALE) , ANALIZA E MATJEVE TË PËRSËRITURA

#### Ushtrime

**Ushtrim 1:** Për të krahasuar jetëgjatësinë e gomave të makinave nga 3 prodhues të ndryshëm , një administrator , i cili është përgjegjës për furnizimin me goma të automjeteve të qendrës së tij spitalore, bëri një zgjedhje rasti dhe u përfatuan të dhënat e mëposhtme për kilometrat e përshkuar (në mijë) nga gomat deri në daljen e tyre jashtë përdorimit.

| Prodhuesi |    |    |
|-----------|----|----|
| A         | B  | C  |
| 41        | 47 | 47 |
| 38        | 42 | 41 |
| 44        | 44 | 42 |
| 42        | 34 | 49 |
| 36        | 42 | 40 |
| 39        | 43 | 45 |

Provoni nëse ka ndryshime të rëndësishme në jetëgjatësinë e gomave të të 3 prodhuesve me nivel rëndësie 0.05.( Supozohet, jetëgjatësia e gomave, ka shpërndarje normale)

**Ushtrim 2:** Nga një testim i njohurive të aplikanteve për menaxherë në 3 qytete të ndryshme, të dhënat mbi variancën për 11 aplikantë të zgjedhur rastësisht nga çdo qytet jepen në tabelën e mëposhtme ANOVA.

| Burimi i variacionit | Shuma e katroreve |
|----------------------|-------------------|
| Qytet                | 491.7             |
| Gabimi               | 1165.3            |
| Totali               | 1657.0            |

1. Plotësoni tabelën ANOVA dhe kontrolloni nëse ka diferenca të rëndësishme në nivelin e njohurive të aplikantëve të të tre qyteteve me nivel rëndësie 0.05.

2. Nëse numri mesatar i pikëve në çdo zgjedhje është  $\bar{x}_1 = 74$  ,  $\bar{x}_2 = 77.5$  ,  $\bar{x}_3 = 83.4$  kontrolloni me anën e NVR nëse ka ndryshime të rëndësishme në formimin e menaxhereve midis çdo dy qyteteve.

**Ushtrimi 3:** Analiza e variancës për një eksperiment 3 x 2 faktorial, me katër vërtetime për çdo trajtim dha tabelën e mëposhtme ANOVA.

| Burimi            | Shuma e katrorëve | Shkallët e lirisë | Katrori mesatar | F |
|-------------------|-------------------|-------------------|-----------------|---|
| Faktori A         | 100               |                   |                 |   |
| Faktori B         |                   | 1                 |                 |   |
| Bashkëveprimi A B |                   | 2                 | 3               |   |
| Gabimi            |                   |                   | 2               |   |
| Totali            | 800               |                   |                 |   |

1. Plotësoni tabelën ANOVA.
2. Testoni efektin e bashkëveprimit AB në rezultatet e eksperimentit, për  $\alpha = 0.05$ .
3. Testoni efektin e faktorit A, për  $\alpha = 0.05$ .
4. Testoni efektin e faktorit B, për  $\alpha = 0.05$ .

**Ushtrimi 4:** Zgjidhen rastësisht 7 pacientë të cilëve iu studiuan rezultatet pas një trajtimi mjeksor. U mbledhën të dhënat lidhur me të rrahurat e zemrës, mbështetur në 3 situata të ndryshme studimi: ( trajtim mjeksor në shtëpi, trajtim mjeksor I çastit në klinikë; tratim mjeksor spitalor disa ditë).

Aplikimi në SPSS i analizës së Matjeve të Përsëritura dha këto rezultate:

*(Tabela 1) Mauchly's Test of Sphericity*

Measure: MEASURE\_1

| Within Subjects Effect | Mauchly's W | Approx. Chi-Square | df | Sig. | Epsilon <sup>a</sup> |             |             |
|------------------------|-------------|--------------------|----|------|----------------------|-------------|-------------|
|                        |             |                    |    |      | Greenhouse-Geisser   | Huynh-Feldt | Lower-bound |
| Vendi                  | .375        | 4.902              | 2  | .086 | .615                 | .694        | .500        |

*(Tabela 2) Tests of Within-Subjects Effects*

| Source       |                    | Type III Sum of Squares | df    | Mean Square | F     | Sig. |
|--------------|--------------------|-------------------------|-------|-------------|-------|------|
| Vendi        | Sphericity Assumed | 974.000                 | 2     | 487.000     | 8.552 | .005 |
|              | Greenhouse-Geisser | 974.000                 | ..... | 791.289     | 8.552 | .018 |
|              | Huynh-Feldt        | 974.000                 | ..... | 702.067     | 8.552 | .014 |
|              | Lower-bound        | 974.000                 | ..... | 974.000     | 8.552 | .026 |
| Error(Vendi) | Sphericity Assumed | 683.333                 | 12    | 56.944      |       |      |
|              | Greenhouse-Geisser | 683.333                 | ..... | 92.525      |       |      |
|              | Huynh-Feldt        | 683.333                 | ..... | 82.092      |       |      |

| Source        |                    | Type III Sum of Squares | df    | Mean Square | F     | Sig. |
|---------------|--------------------|-------------------------|-------|-------------|-------|------|
| Vendi         | Sphericity Assumed | 974.000                 | 2     | 487.000     | 8.552 | .005 |
|               | Greenhouse-Geisser | 974.000                 | ..... | 791.289     | 8.552 | .018 |
|               | Huynh-Feldt        | 974.000                 | ..... | 702.067     | 8.552 | .014 |
|               | Lower-bound        | 974.000                 | ..... | 974.000     | 8.552 | .026 |
| Error(Vendi ) | Sphericity Assumed | 683.333                 | 12    | 56.944      |       |      |
|               | Greenhouse-Geisser | 683.333                 | ..... | 92.525      |       |      |
|               | Huynh-Feldt        | 683.333                 | ..... | 82.092      |       |      |
|               | Lower-bound        | 683.333                 | ..... | 113.889     |       |      |

1. Cilat janë supozimet e analizës së matjeve të Përsëritura.? A plotësohet në këtë rast supozimi i sfericitetit? ( $\alpha = 0.1$ )
2. Plotësoni vendet bosh te kolona e shkallëve të lirisë ( Tabela 2) dhe testoni nëse vendi i studimit ka efekt në rezultatet në provim për studentet e trajtuar? ( $\alpha = 0.1$ )
3. Llogaritni efektin e pjesshëm të përmasës dhe interpretoni

**Ushtrimi 5 :** Zgjidhen rastësisht 4 dyqane. Vrojtohen shitjet e tyre në fund të 3 ditëve: dita 1 ( mot diell), dita 2 (mot me vranësirë), dita 3 ( mot me shi) . U mbledhën të dhënat lidhur me të ardhurat nga shitja në mijë lekë në fund të çdo dite për secilin nga 4 dyqanet. Aplikimi në SPSS i analizës së Matjeve të Përsëritura dha këto rezultate:

*(Tabela 1) Mauchly's Test of Sphericity*

| Within Subjects Effect | Mauchly's W | Approx. Chi-Square | df | Sig. | Epsilon <sup>a</sup> |             |             |
|------------------------|-------------|--------------------|----|------|----------------------|-------------|-------------|
|                        |             |                    |    |      | Greenhouse-Geisser   | Huynh-Feldt | Lower-bound |
| moti                   | .375        | 1.963              | 2  | .375 | .615                 | .826        | .500        |

*(Tabela 2) Tests of Within-Subjects Effects*

#### Tests of Within-Subjects Effects

| Source |                    | Type III Sum of Squares | df  | Mean Square | F   | Sig. |
|--------|--------------------|-------------------------|-----|-------------|-----|------|
| moti   | Sphericity Assumed | 370.500                 | ... | ...         | ... | ..   |
|        | Greenhouse-Geisser | 370.500                 | ... | ...         | ... | ..   |

|             |                    |         |     |     |     |    |
|-------------|--------------------|---------|-----|-----|-----|----|
|             | Huynh-Feldt        | 370.500 | ... | ... | ... | .. |
|             | Lower-bound        | 370.500 | ... | ... | ... | .. |
| Error(moti) | Sphericity Assumed | ...     | ... | ... |     |    |
|             | Greenhouse-Geisser | ...     | ... | ... |     |    |
|             | Huynh-Feldt        | ...     | ... | ... |     |    |
|             | Lower-bound        | ...     | ... | ... |     |    |

1. Plotësoni vendet bosh te ( Tabela 2)
2. A plotësohet në këtë rast supozimi i sfericitetit? ( $\alpha = 0.05$ ) Nëse nuk plotësohet, teorikisht si duhet të veprojmë?
3. Testoni nëse faktori “mot” është statistikisht i rëndësishëm. ( $\alpha = 0.05$ )

**TESTIMI I HIPOTEZAVE (k popullime):**

**-Anova  
- Analiza e Matjeve të Përsëritura  
(METODA PARAMETRIKE)**

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

1

**Çështjet Kryesore që Trajtohen:**

**A. Testimi i hipotezave për mesataren e më shumë se dy popullimeve (ANOVA 1 faktoriale, 2 faktoriale)**

**B. Analiza e Matjeve të Përsëritura)**

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

2

**A. 1. Testimi i hipotezave për mesataren e më shumë se dy popullimeve (ANOVA 1 Faktoriale)**

Supozimet: Shpernd. Normale

Pavarësi

Homogjenitet variancash ( Testi Levene)

H0:  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$

H1: Të paktën një nga  $\mu$ -te e  $H_0$  e ndryshme nga të tjerat

H0  $\downarrow$  nqse  $F_v > F_{(\alpha, k-1; n-k)}$

| Burimi i Variacionit | Shuma e katrorëve | Shkallët e lirisë | Katrori mesatar   | F        |
|----------------------|-------------------|-------------------|-------------------|----------|
| Trajtimet            | SKTR              | k-1               | KMTR = SKTR/(k-1) | KMTR/KMG |
| Gabimi               | SKG               | n-k               | KMG = SKG/(n-1)   |          |
| Totali               | SKT               | n-1               |                   |          |

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

3

$$SKTR = \sum_{i=1}^k n_i (\bar{x}_i - \bar{\bar{x}})^2 \quad SKG = \sum_{i=1}^k s_i^2 (n_i - 1) \quad SKT = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2$$

|                                               |                                               |                                               |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| H0: $\mu_1 = \mu_2$<br>H1: $\mu_1 \neq \mu_2$ | H0: $\mu_1 = \mu_3$<br>H1: $\mu_1 \neq \mu_3$ | H0: $\mu_1 = \mu_k$<br>H1: $\mu_1 \neq \mu_k$ |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|

Testohen të gjitha çiftet e mundshme

(Procedura TUCKEY)

$$H0 \downarrow \text{nqse } (\bar{x}_{\max} - \bar{x}_{\min}) > NRT \quad NRT = q \sqrt{\frac{KMB}{n}}$$

E q gjendet nga tabela Tuckey

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

q (k, n-k,  $\alpha$ )

4

### Testimi i hipotezave për mesataren e më shumë se dy popullimeve (ANOVA 2 Faktoriale)

Faktori A (a nivele i=1...a), Faktori B (b nivele, j=1...b), Bashkëveprimi AB, r- nr. l vërtetimeve për çdo bashkëveprim k=1...r

| BV               | Shuma e katrorëve | Shkallët e lirisë | Katrori mesatar   | F        |
|------------------|-------------------|-------------------|-------------------|----------|
| Faktori A        | SKA               | a-1               | KMA = SKA/(a-1)   | KMA/KMG  |
| Faktori B        | SKB               | b-1               | KMB = SKB/(b-1)   |          |
| Bashkëveprimi AB | SKAB              | (a-1)(b-1)        | SKAB/[(a-1)(b-1)] | KMAB/KMG |
| Gabimi           | SKG               | ab(r-1)           | KMG = SKG/ab(r-1) |          |
| Totali           | SKT               | n-1               |                   |          |

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

5

$$SKA = br \sum_{i=1}^a (\bar{x}_i - \bar{\bar{x}})^2$$

$$SKB = ar \sum_{j=1}^b (\bar{x}_j - \bar{\bar{x}})^2$$

$$SKAB = r \sum_{I=1}^A \sum_{j=1}^b (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$$

$$SKT = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (x_{ijk} - \bar{\bar{x}})^2$$

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

6



## ANALIZA E MATJEVE TË PËRSËRITURA

### Çështjet Kryesore që Trajtohen:

- 1. HYRJE
- 2. AVANTAZHET DHE DISAVANTAZHET E ANALIZËS SË MATJEVE TË PËRSËRITURA
- 3. ANALIZA E MATJEVE TË PËRSËRITURA NJË FAKTORIALE
- 4. APLIKIMI NE SPSS I ANALIZËS SË MATJEVE TË PËRSËRITURA (AMP)- ANOVA
- 5. SUPOZIMET PËR ANALIZËN E MATJEVE TË PËRSËRITURA
- 6. EFEKTI I PËERMASËS
- 7. FUQIA STATISTIKORE

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

7

### HYRJE & Avantazhe e Disavantazhe të AMP

Në dallim nga Anova, tek AMP, nuk ekziston supozimi për zgjedhje të pavarura ose AMP përdoret kur ka zgjedhje të pavarur por elementët janë të koreluar.

AMP në vetvete është një ANOVA me blloqe, ku elementi konsiderohet si bllok.

#### Avantazhet e AMP:

1. Variabiliteti ndërmjet elementeve për shkak të ndryshimeve individuale largohet nga gabimi, ndaj AMP është me e fuqishme
2. Kërkohe më pak elemente në shqyrtim ( 3 trajtime x20 elementë =60 elementë te ANOVA; 20 elementë që shihen në tre faza të ndryshme të AMP)

#### Disavantazhet e AMP:

1. Duhet të tregohet kujdes në renditjen e elementëve
2. Ekziston mundësia e mbartjes se efekteve (rekomandohet kohë pauze midis trajtimeve)

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

8

### ANOVA dhe AMP Një- Faktoriale

1= Jo stres; 2= Dhimbje ; 3= Kushte pune    4= Gjendje irritimi.  
(Trajtimet janë kushtet emocionale)

| Elem(i)/Trajt(i) | 1       | 2       | 3    | 4    | Mesatarja(j) |
|------------------|---------|---------|------|------|--------------|
| 1                | 70      | 86      | 80   | 87   | 80.75        |
| 2                | 78      | 82      | 77   | 80   | 79.25        |
| 3                | 61      | 70      | 64   | 72   | 66.75        |
| 4                | 70      | 74      | 68   | 88   | 75           |
| 5                | 89      | 105     | 89   | 101  | 96           |
| 6                | 79      | 84      | 88   | 87   | 84.5         |
| Mesatarja        | 74.5    | 83.5    | 77.7 | 85.8 | <u>80.37</u> |
| (xi)             |         |         |      |      |              |
| SSb/ (SKTR)      | SSw/SKG | SSt/SKT |      |      |              |

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

**Testimi i hipotezave për mesataren e më shumë se dy popullimeve (AMP 1 Faktoriale)**

Supozimet: Shpernd. Normale  
Zgjedhje të varura  
Uniformitet & Sfericitet

H0:  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$

H1: Të paktën një nga  $\mu$ -te e  $H_0$  e ndryshme nga të tjerat

H0  $\downarrow$  nqse  $F_v > F_{\alpha, k-1; (k-1)(b-1)}$

| Burimi i Variacionit | Shuma e katrorëve | Shkallët e lirisë | Katrori mesatar        | F                |
|----------------------|-------------------|-------------------|------------------------|------------------|
| Trajtimet            | SKTR              | k-1               | KMTR = SKTR/(k-1)      | KMT<br>R/K<br>MG |
| Elementët (Bloqet)   | SKBL              | b-1               | KMBL = SKBL/(b-1)      |                  |
| Gabimi               | SKG               | (k-1)(b-1)        | KMG = SKG/((k-1)(b-1)) |                  |
| Totali               | SKT               | n-1               |                        |                  |

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

**Supozimet për Analizën e Matjeve të Përsëritura:**

**1. Uniformitet (simetria e përbërë)** (Variancat e popullimit për të gjitha trajtimet si edhe kovariancat përkatëse supozohet si të barabarta.)

**2. Sfericiteti (shpesh i njohur si rrumbullakësia).** (Sfericiteti kërkon vetëm që, variancat e ndryshmeve për çdo çift të matjeve të përsëritura duhet të jenë të barabarta)

“Nje matrice që ka simetri të përbërë, detyrimisht plotëson dhe supozimin e sfericitetit; por një matricë që është sferike jo domosdoshmërisht tregon për simetri të përbërë.”

-Sfericiteti testohet në SPSS nga **testi Mauchly** (i cili ka fuqi të vogël për nr. të vogël elementësh – rasti i AMP ndaj shpesh herë punohet sikur ky supozim çënohet).

Ka dy rugë për të eliminuar këtë :

(1) Axhustohen shkallët e lirisë (2) MANOVA

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

11

- Në rastin kur nuk plotësohet ose kur çënohet supozimi i sfericitetit, atëherë shkalla e devijimit të matricës së kovariancës nga sfericiteti reflektohet në parametrin e quajtur  $\epsilon$ .

- **Vlera më e madhe që ka  $\epsilon$  është 1, (rasti kur plotësohet supozimi i sfericitetit), vlera më e vogël që merr  $\epsilon$  është kur  $\epsilon = 1/(k-1)$ ; (ku k është numri i matjeve të përsëritura**

- Për të axhustuar situatën sugjerohet ndryshimi i shkallëve të lirisë nga (k-1) dhe (k-1)(b-1) në 1 dhe (b-1) ose shumëzohen me  $\epsilon$ .

- Duhet të kihet parasysh që vlerat relativisht të mëdha të  $\epsilon$ , të shoqëruara me test të rëndësishëm Mauchly , sugjerojnë se është fare pa rëndësi cënimi i supozimit të sfericitetit.

- Korrigjimi i shkallëve të lirisë nuk ka fare efekt në vlerën e llogaritur të F, i vetmi efekt i tyre është që mbi bazën e vlerave të reja të gjetura të shkallëve të lirisë do të afektohet vlera e F-së kritike.

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

12

### EFEKTI I PËRMASES

Tek **AMP** efekti i pjesshëm i përmasës është  $\eta^2$

$\eta^2$  i pjesshëm =  $SS_{\text{trajtimeve}} / (SS_{\text{trajtimeve}} + SS_{\text{error}}) = 488.458 / (488.458 + 284.782) = 0.632$

Kështu, në variacionin e të rrahurave të zemrës pasi janë larguar ndryshimet individuale të personave; rreth 63% shpjegohet nga ndryshime ndërmjet 4 kushteve të trajtimeve.

Tek **MANOVA** efekti i përmasës bazohet në  $\Lambda$  Wilks ku

$\eta^2 = 1 - \Lambda = 1 - 0.16 = 0.84$

Kështu 84% e variancës së mbetur (pasi variancës së të rrahurave të zemrës i është hequr ndryshimet individuale sistematike), shpjegohet nga tipet e stresit

### Përdorimi i Tabelave:

Kur njihet  $\eta^2$ :

1. Përcaktohet përmasa e pjesshme e efektit psh  $\eta^2 = 0.25$
2. Përcaktohen shkallët e lirisë ( $k-1$ )
3. Përcaktohet fuqia e dëshiruar statistikore,
4. Nga tabela gjendet sa duhet të jete  $n$  (numri total i elementëve të zgjedhur)

Kur nuk njihet  $\eta^2$ : atëherë në shumë fusha studimi merret një  $\eta^2$  në intervalet  $0.1 < \eta^2 < 0.2$

## **TEMA 10:**

### **ANALIZA E FUQISË**

#### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjëndje :

- Të kuptojnë ç'është fuqia e një testi statistikor
- Të vlerësojnë fuqinë para dhe pas një procesi testimi hipotezash në rastin e :
  - një popullimi
  - dy popullimeve ( zgjedhje të pavarura dhe të varuara)
  - më shumë se dy popullimeve ( ANOVA dhe AMP)
- Të vlerësojnë paraprakisht subjektet që nevojiten për një fuqi të dhënë
- Njohin dhe përdorin mënyrat e përmirësimit të fuqisë
- Vlerësojnë fuqinë në SPSS

#### **LITERATURA**

- “Applied Statistics”, Warner A.R, 2008
- “Intermediate Statistics- A modern Approach”, Stevens J, 2009
- Statistical for Business and Economics”, Newbold P, Carlson W, Thorne B, 2011
- “Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014

# **TEMA : ANALIZA E FUQISË**

## **1 HYRJE**

## **2 TESTI $t$ PËR ZGJEDHJET E PAVARURA**

## **3 VLERËSIMI I FUQISË PARA DHE PAS TESTIMIT**

## **4 VLERËSIMI I FUQISË PËR ANALIZËN E VARIANCËS ME NJË FAKTOR**

## **5 NJË VLERËSIM PARAPRAK I SUBJEKTEVE QË NEVOJITEN PËR NJË FUQI TË DHËNË**

## **6 MËNYRAT E PËRMIRËSIMIT TË FUQISË**

## **7 VLERËSIMI I FUQISË NË SPSS**

### **1. HYRJE**

Në një kontroll hipotezash; gabimi i llojit të parë (I) ( $\alpha$ ) ose niveli i rëndësisë, është probabiliteti që bëjmë kur hedhim poshtë hipotezën  $H_0$  kur ajo është realisht e vërtetë. Niveli i  $\alpha$  përcaktohet nga studiuesi, pra është një vendim subjektiv; ku shumica e studiuesve preferojnë  $\alpha = 0.05$  ose  $\alpha = 0.01$ .

Arësyeja pse  $\alpha$  vendoset në nivele kaq të ulëta është pikërisht që të minimizohet probabiliteti i bërjes së këtij gabimi. Megjithatë rekomandohet nga statisticienët, që për zgjedhje me vëllim më të vogël se 20, shpesh nuk është strategji e mirë vendosja e  $\alpha$  në nivele të ulëta si 0.05 ose 0.01.

Ekziston një tjetër lloj gabimi, që mund të bëhet gjatë një kontrolli statistikor hipotezash; që quhet gabimi i llojit të dytë. Gabimi i llojit të dytë (II), i njohur si  $\beta$ , është probabiliteti që të pranohet  $H_0$  kur ajo është realisht e gabuar. Duhet thënë se vetëm njëri nga këto dy gabime mund të ndodhë në një studim të caktuar; që do të thotë: ose në gabimisht pranojmë  $H_0$  ose gabimisht  $H_0$  e hedhim poshtë.

Kjo nënkulon faktin që jo vetëm njëri nga këto gabime mund të ndodhë; por për më tepër ato janë të lidhur në mënyrë inverse me njëri tjetrin. Pra nëse në përpqëmi të kontrollojmë  $\alpha$  në nivele të ulëta, gabimi i llojit të dytë  $\beta$  rritet.

Në statistikë jemi të interesuar për fuqinë statistikore, që është plotësuesi i gabimit të llojit të dytë. Në këtë mënyrë fuqia statistikore njihet si  $1 - \beta$ .

Në këtë këndvështrim faktorët që ulin riskun e gabimit  $\beta$ , rrisin fuqinë statistikore  $1 - \beta$ .

Më poshtë po trajtojmë në mënyrë cilësore sesi disa faktorë që ndikojnë në gabimin  $\beta$  afektojnë fuqinë  $1 - \beta$ .

Kështu në një kontekst të gjerë, fuqia varet nga shumë faktorë: 1) Niveli  $\alpha$  i vendosur nga studiuesi 2) volumi i zgjedhjes 3) Përmasat e efektit 4) testi statistikor i përdorur 5) projekti i kërkimit shkencor

Psh. Testi  $t$  për zgjedhjet e varura është më i fuqishëm sesa testi  $t$  për zgjedhjet e pavarura; një studim i matjeve të përsëritura është më i fuqishëm sesa ANOVA me një faktor etj si këto.

Fuqia statistikore varet kryesisht nga:

- 1) Niveli  $\alpha$  i vendosur nga studiuesi

- 2) Volumi i zgjedhjes
- 3) Përmasat e efektit

**1) Niveli  $\alpha$  i vendosur nga studiuesi.** Ekziston lidhje negative ndërmjet  $\alpha$  dhe  $\beta$ . Ndërsa  $\alpha$  rritet;  $\beta$  ulet dhe

$(1 - \beta)$  rritet. Megjithëse duke patur nivel të lartë të  $\alpha$ , mund të përfitohet nivel i lartë i fuqisë statistikore. Shumë studiues preferojnë të pranojnë nivele të ulta të  $\alpha$  duke synuar ndryshimin e elementëve të tjerë të studimit për përmirësimin e fuqisë.

**2) Volume i zgjedhjes dhe 3) Përmasat e efektit:** Kemi parasysh që në një kontroll hipotezash  $t$  llogaritet si më poshtë:

$$t = (\bar{x} - \mu_0) / (s / \sqrt{n}) = (\bar{x} - \mu_0) / s \times 1 / (1 / \sqrt{n}) = d \times \sqrt{n}$$

duke patur parasysh faktin që  $d$  quhet Cohen's  $d$  dhe është tregues i përmasës së efektit.

Duke supozuar volumin e zgjedhjes ( $n$ ) si konstante,  $t$  rritet ndërsa  $d$  rritet. Nqse përmasat e efektit ( $d$ ) mbahet konstante,  $t$  rritet ndërsa  $n$  rritet. Kështu duke sup. se volumi i zgjedhjes  $n$  mbahet konstante; ndërsa përmasat e efektit të popullimit të përfaqësuar nga Cohen's  $d$  rriten, pritet që vlera e  $t$  të tentojë të rritet. Në këto kushte ka të ngjarë të përfitohet një vlerë e madhe e  $t$  –së së llogaritur (mjaftueshëm për të plotësuar kushtin  $t_{vr} > t_{kr}$ ).

Kështu themi se nqse rritet  $d$  (kur faktorët e tjerë mbahen konstante), fuqia statistikore rritet, duke u ulur kështu gabimin e llojit të dytë.

Në mënyrë të ngjashme themi se nqse rritet  $n$  (kur faktorët e tjerë mbahen konstante), vlera e  $t$  synon të rritet, fuqia statistikore  $(1 - \beta)$  rritet, duke u ulur kështu gabimin e llojit të dytë ( $\beta$ )

Në shqyrtim mund të merret edhe se si ndryshon vlera e fuqisë, duke parë variabilitetin e devijimit standart ( $s$ )

Ka tabela të posaçme për llogaritjen e fuqisë (koordinatat janë  $n$  dhe  $d$ ) ku pohimet e mësipërme mund të konfirmohen duke iu referuar këtyre tabelave, psh: per  $\alpha=0.05$  kemi: (**Tabela 1**)

| <b>n</b> | <b>d</b> | <b>1- <math>\beta</math></b> |
|----------|----------|------------------------------|
| 11       | 0.25     | 0.1                          |
| 11       | 0.35     | 0.16                         |
| 15       | 0.25     | 0.13                         |

## **2. TESTI $t$ PËR ZGJEDHJET E PAVARURA**

Për rastin e zgjedhjeve të pavarura, kur krahasohen dy mesatare popullimesh kemi se:

$$d = (\mu_1 - \mu_2) / \sigma$$

ku  $\sigma$  nënkupton devijimin standart të popullimit.

Kjo përmasë efekti e popullimit ka si vlerësim pikësor :



$$d' = (\bar{x}_1 - \bar{x}_2) / s$$

ku  $\sigma'^2 = [(n_1-1)s_1^2 + (n_2-1)s_2^2] / (n_1+n_2-2) = s^2$  është vlerësimi i variancës së popullimit.

Jemi në kushtet e zgjedhjeve të pavarura. Vemë re se ekuacioni i  $d'$  shpreh diferencën ndërmjet mesatareve të 2 popullimeve të shprehur në njësi të devijimeve standarte.

Psh. nqse mesataret e zgjedhjeve do të ishin përkatësisht:  $\bar{x}_1 = 10$  dhe  $\bar{x}_2 = 4$ ,  $s = 15$ , atëherë  $d' = (10-4)/15 = 0.4$  që do të thotë se mesataret ndryshojnë ndërmjet tyre me 0.4 devijime standarte.

Duhet të kihet parasysh fakti se një përmasë efekti  $d \leq 0.2$  konsiderohet si e vogël;

$0.20 < d \leq 0.79$  konsiderohet e mesme dhe  $d > 0.80$  konsiderohet e madhe.

Përgjithësisht një përmasë efekti e mesme është më i përshtatshëm për studiuesit.

Në rastin e zgjedhjeve të vogla të pavarura ku kemi të njohura  $n_1$  dhe  $n_2$ ; për një përmasë efekti  $d$  lehtësisht të llogaritshme, ekzistojnë *tabela të posaçme të Cohen*, ku për çdo  $\alpha$  ka tabela të veçanta. Madje tabela të veçanta ka edhe për faktin nëse testet e kontrolit të hipotezave janë të njëanshme apo të dyanshme. Në rastin kur  $n_1 = n_2$  në tabela orientimi bëhet me njërën nga vlerat. (**Tabelat 2-5**)

Kështu psh shohim se nga këto tabela, vlerat e fuqisë për testin  $t$  ( $\alpha = 0.05$ , test i dyanshëm) janë shumë të vogla kur volumet e zgjedhjeve dhe/ose përmasat e efektit janë të vogla :

1.  $n_1 = n_2 = 15$ ;  $d = 0.50$ , fuqia = 0.25
2.  $n_1 = n_2 = 35$ ;  $d = 0.30$ , fuqia = 0.23

Fuqia mund të përmirësohet, edhe duke mbajtur volum të vogël zgjedhjesh, por vetëm nqse përmasa e efektit është e madhe : 3.  $n_1 = n_2 = 15$ ;  $d = 1$ , fuqia = 0.75

*Si vlerësohet n nqse.  $n_1 \neq n_2$  ?*

Literatura sugjeron në këtë rast përdorimin e mesatares harmonike. Dihet se mesatarja harmonike për dy vlera:  $n_1$  dhe  $n_2$  llogaritet si:  $2 n_1 n_2 / (n_1 + n_2)$ .

Kështu nqse kemi dy zgjedhje me vëllime përkatësisht 10 dhe 20; mesatarja harmonike e tyre llogaritet si:  $2(10)(20)/30 = 13.3$  dhe do të përdoret 13 si  $n$  për të vlerësuar fuqinë statistikore në tabelë.

Vlen të theksohet fakti se dallimi midis mesatares aritmetike dhe asaj harmonike, qëndron pikërisht se ndryshimi ndërmjet këtyre dy mesatareve është i vogël kur volumi i zgjedhjeve është pothuajse i barabartë; por nqse volumet e zgjedhjeve janë shumë të ndryshme atëherë diferenca ndërmjet këtyre mesatareve është e konsiderueshme, dhe ne do të përdorim mesataren harmonike për vlerën  $n$ , që do të përdorim në tabelat përkatëse; psh:

| $n_1$ | $n_2$ | Mesatare aritmetike | Mesatare harmonike |
|-------|-------|---------------------|--------------------|
| 10    | 15    | 12.5                | 12                 |
| 10    | 20    | 15                  | 13.3               |
| 10    | 30    | 20                  | 15                 |
| 10    | 40    | 25                  | 16                 |

Në rastin e kontrollit të hipotezave, testet e njëanshme aplikohen kryesisht kur për problemin që shqyrtohe ka evidenca empirike (studime të mëparshme) dhe /ose teori të ngjashme që shpjegojnë fenomenin që trajtohet. Avantazhi statistikor i testeve të njëanshme, krahasuar me ato të dyanshmet, është se të parat kanë fuqi statistikore më të madhe. Për të vënë në dukje këtë ndryshim të fuqisë në varësi të tipit të testit , si edhe reagimin e fuqisë kundrejt vlerës së  $\alpha$ , i referohemi gjetjeve të mëposhtme:

***Fuqia e testit t për zgjedhje të pavarura***

***( $n_1 = n_2 = 20$ )***

| <b>Niveli <math>\alpha</math> dhe tipi i testit<br/>efektit</b> | <b>Përmasa të moderuara të efektit</b> | <b>Përmasë e madhe e</b>    |
|-----------------------------------------------------------------|----------------------------------------|-----------------------------|
|                                                                 | <b>(<math>d=0.6</math>)</b>            | <b>(<math>d=0.8</math>)</b> |
| $\alpha=0.01$ , test i dyanshëm                                 | 0.21                                   | 0.44                        |
| $\alpha=0.05$ , test i njëanshëm                                | 0.58                                   | 0.80                        |
| $\alpha=0.1$ , test i njëanshëm                                 | 0.71                                   | 0.89                        |

Në nivelin tradicional  $\alpha=0.01$ , në testin e dyanshëm, fuqia është e vogël në të dy rastet; ndërsa në nivelin  $\alpha=0.1$ , me avantazhin e të paturit të një testi të njëanshëm, fuqia është e mirë në të dy rastet.

### **3. VLERËSIMI I FUQISË PARA DHE PAS TESTIMIT**

Nqse dikush vendos të shpenzojë kohë dhe para në ndërmarrjen e një studimi, atëherë ai/sjo duhet të ketë një siguri në masën 70-80% se do të gjejë një ndryshim ndërmjet mesatareve, nqs. vërtet ekziston një i tillë (që do të thotë se fuqia = 0.70 ose 0.80). Në këtë mënyrë një vlerësim paraprak i fuqisë i tregon studiuesit se sa do të duhet të jetë volumi i zgjedhjeve për çdo grup, në mënyrë që të përfitohet vlera e dëshiruar e fuqisë statistikore.

Vlerësimi i fuqisë pas testimit është gjithashtu shumë i rëndësishëm, sepse ndikon në faktin se si do të interpretohen rezultatet e studimeve tashmë të përfunduara. Për një fenomen të caktuar, duke iu referuar literaturës dhe studimeve të mëparshme; nqse nuk do ti kushtonim rëndësi fuqisë, atëherë nqse literatura sugjeron rezultate të parëndësishme statistikore (s’ka diferencë të rëndësishme ndërmjet mesatareve) , atëherë konkluzioni do të ishte se aplikimi i nje metode të caktuar nuk ka efekt në këto mesatare. Nqse do ti referohemi më tutje edhe fuqisë, atëherë, duhet të merren parasysh edhe volumet e zgjedhjeve nëpër studime, edhe përmasat e efektit sipas rasteve përkatëse. Nqse volumet e zgjedhjeve do të ishin të vogla në shumicën e studimeve me rezultate të parëndësishme (s’ka diferencë të rëndësishme ndërmjet mesatareve), atëherë mungesa e rëndësisë është për shkak të vlerës së vogël të fuqisë. Ose me fjalë të tjera themi se: “*Shumë studime me vlerë të vogël fuqie që raportojnë rezultate të parëndësishme (për të njejtën çështje) (s’ka diferencë të rëndësishme ndërmjet mesatareve), tregojnë pikërisht për ekzistencë të efektit.*”

Kështu, përmasat e efektit ( $d'$ ) për testin  $t$  mund të shprehen përkundrejt vlerës së statistikës  $t$  si më poshtë:

$$d' = t \sqrt{1/n_1 + 1/n_2}$$

ku  $n_1$  dhe  $n_2$  janë volumet respektive të zgjedhjeve. Ky ekuacion ndihmon në vlerësimin e fuqisë për studime tashmë të përfunduara; pasi në këtë ekuacion vlera e  $d'$  llogaritet shpejt dhe pastaj i referohemi direkt tabelave të Cohen sipas rastit.

#### **4 VLERËSIMI I FUQISË PËR ANALIZËN E VARIANCËS ME NJË FAKTOR**

Për të përcaktuar përmasën e efektit, për një ANOVA me një faktor, në fillim përcaktojmë një tregues të ndryshimit të mesatareve të grupeve (trajtimeve) nga mesatarja e përgjithshme, e cila është e pavarur nga volumi i zgjedhjes.

$$\sigma_m^2 = \frac{\sum n_i (\bar{x}_i - \bar{x})^2}{n}$$

ku  $\bar{x}$  është mesatarja e përgjithshme dhe  $n = \sum n_i$ . Pjestimi me  $n$  e bën treguesin të pavarur nga volumi i zgjedhjes. Për ta patur përmasën e efektit të zhveshur nga shkalla, përsëri pjestojmë me  $\sigma$ , siç edhe bëme për rastin e treguesit të efektit të testit  $t$ :

$$f = \sigma_m / \sigma$$

Ky tregues përfaqëson devijimin standart të mesatareve të standartizuara, që do të thotë, ndryshimin e rezultatit  $z$  të mesatareve të grupeve (trajtimeve) nga mesatarja e përgjithshme. Kështu: përmasa e efektit mund të shprehet duke u nisur nga statistika  $F$  si më poshtë:

$$f = \sqrt{(k-1)F/n}$$

Sipas Cohen;  $f=0.1$  tregon për një përmasë të vogël efekti;  $0.10 < f \leq 0.4$  tregon për një përmasë të mesme efekti; dhe  $f > 0.40$  tregon për një përmasë të madhe efekti.

Ekuacioni i mësipërm, është shumë i dobishëm për vlerësimin e fuqisë pasi është bërë studimi, për aq kohë sa nevojitet në këtë ekuacion vlera e  $F$ , e cila gjendet nga testimi.

Mbështetur në vlerën e përmasës së efektit të llogaritur në këtë mënyrë, në volumin e zgjedhjes për çdo grup (ose në volumin mesatar të zgjedhjes, nqse. grupet kanë volume të ndryshme zgjedhjeje); hyhet në tabelat e fuqisë Cohen për të llogaritur fuqinë e studimit. (**Tabelat 6-12**)

##### Shembull 1:

Albani ndërmori një studim me tre grupe mbi efikasitetin e metodave matematikore. Ai zgjodhi 10 elementë për çdo grup dhe përfitoi  $F = 4.38$ . Sa është fuqia e studimit për  $\alpha = 0.05$ ?

##### Zgjidhje:

$$f^* = \sqrt{(k-1)F/n} = \sqrt{2(4.38)/30} = 0.54 \text{ (përmase e madhe efekti)}$$

Referuar tabelave Cohen (për ANOVA me një faktor);  $\mu = k-1$  (në rastin tonë  $\mu = k-1 = 3-1=2$ ). Gjejmë që për  $f=0.50$ , fuqia është 0.64; ndërsa për  $f=0.60$ , fuqia është 0.81 (tabela 7). Duke interpoluar vlerat, për  $f=0.54$  rezulton që fuqia është 0.7. Nga studimi vlera F ishte statistikisht e rëndësishme në nivelin 0.05. Kështu Albani ka fuqinë statistikore të mjaftueshme (0.73) për të hedhur poshtë  $H_0$ , pavarësisht volumit të vogël të zgjedhjes, pasi përmasa e efektit ishte e madhe.

#### Shembull 2:

Konsiderojmë një studim me 4 grupe ku:  $n_1 = 15$ ;  $n_2 = 13$ ,  $n_3 = 20$ ;  $n_4 = 17$ . Studiuesi gjeti  $F=1.4$ . Sa është fuqia në nivelin  $\alpha=0.05$ ? po në nivelin  $\alpha=0.1$ ?

#### Zgjidhje:

$$f^* = \sqrt{(k-1)F/n} = \sqrt{2(1.4)/65} = 0.254 \text{ (përmasë e mesme efekti)}$$

Volumi mesatar i grupit është  $(15+13+20+17)/4 = 16.25$ , ne do përdorim 16.;  $\mu = k-1 = 4-1=3$ , për  $\alpha=0.05$ , fuqia është 0.34 (Tabela Cohen 7), ndërsa për  $\alpha=0.1$ , fuqia është 0.48 (Tabela 11). Në të dy rastet fuqia nuk është e mirë.

Nqse nga një studim rezulton se nuk ka ndryshim statistikisht të rëndësishëm mes mesatareve, dhe ku përmasa e efektit është e vogël (0.10), atëherë studiuesi duhet të vendosë nëse praktikisht ka apo jo ndryshim statistikisht të rëndësishëm. Nga ana tjetër, nëse nuk ka ndryshim statistikisht të rëndësishëm mes mesatareve, dhe nga analiza pas testimit rezulton një përmasë efekti e mesme ose e madhe; atëherë është e rëndësishme të përsëritet studimi me volum më të përshtatshëm zgjedhjeje.

### **5 NJË VLERËSIM PARAPRAK I SUBJEKTEVE QË NEVOJITEN PËR NJË FUQI TË DHËNË**

Për të realizuar këtë analizë, duhet një vlerë e pritshme e përmasës së efektit. Me qëllim që të futemi në tabelat e fuqisë Cohen, për të përcaktuar sa duhet të jetë volumi i zgjedhjes për çdo grup, në mënyrë që të përfitohet një fuqi e specifikuar në një nivel të caktuar  $\alpha$ .

Për këtë qëllim, duhet t'i referohemi përmasës mesatare të efektit duke iu referuar përmasave të efekteve të studimeve të mëparshme të ngjashme.

Nqse ka shumë studime të ngjashme, mesatarja do të ishte më përfaqësuese, por në raste më të rralla edhe kur mund të gjenden dy studime të ngjashme; do t'i referohemi përmasës mesatare të efektit të tyre. Tek e fundit është më mirë t'i referohemi paraprakisht disa vlerave për të gjetur volumin e përshtatshëm të zgjedhjes, sesa të mos e kesh fare idenë lidhur me këtë të fundit.

#### Shembull 3:

Supozojmë se studiuesi X ka gjetur 2 studime të ngjashme me të tijn:

Studimi 1: 3 grupe; n= 42; F=2.16

Studimi 2: 3 grupe , n =81; F= 1.42

$$f_1 = \sqrt{2(2.16)/42} = 0.32 \text{ dhe } f_2 = \sqrt{2(1.42)/81} = 0.187$$

Kështu vlera e pritshme e përmasës së efektit për këtë studiu është  $(0.3+0.187)/2= 0.25$

Studiuesi do të dijë se sa do të jetë vëllimi i zgjedhjes për çdo grup në mënyrë që të ketë një fuqi prej 0.70, për  $\alpha= 0.05$  dhe për nivelin  $\alpha= 0.10$ . Duke ju referuar tabelës 7, nën kolonën  $f= 0.25$  kërkojmë deri kur fuqia e kriterit të jetë të paktën 0.70 dhe gjejmë se 42 elementë për grup do të nevojiten të zgjidhen (për  $\alpha= 0.05$ ). Tani i referohemi tabelës 11 për nivelin  $\alpha= 0.1$  dhe gjejmë se 32 elementë duhet të zgjidhen për çdo grup.

Në shembullin më lart për gjetjen e efektit të pritshëm të përmasës, përdorëm mesataren ku pesha e çdo studimi ishte e njëjtë. Mund të ndodhë që nëse njëri nga studimet i përafrohet më shumë studimit në fjalë, atëherë përmasa e efektit të këtij studimi, mund të ketë peshë më të madhe, por kjo është në varësi të mendimit të studiuesit.

## **6 MËNYRAT E PËRMIRËSIMIT TË FUQISË**

Duke parë sa e vogël është vlera e fuqisë në zgjedhjet me volum më të vogël se 20 elementë për grup, studiuesi duhet të ketë parasysh katër mënyra se si mund të përmirësojë këtë fuqi:

1. Aplikoni një nivel më të butë  $\alpha$ , ndoshta  $\alpha= 0.1$  ose  $\alpha= 0.15$ .
2. Përdorni teste kontrolli të njëanshme, kur ka mundësi të përdoret një rast i tillë.
3. Konsideroni mundësitë e zvogëlimit të ndryshimeve brenda grupeve. Një mënyrë është nëpërmjet selektimit të zgjedhjeve: elementë më homogjenë, tentojnë të kenë më pak ndryshime të variablit të varur. Psh. përdorni vetëm meshkuj, në vend të meshkujve dhe femrave; ose vetëm fëmijë 6-7 vjeçarë, sesa fëmijë 6-9 vjeçarë. Një mënyrë tjetër e zvogëlimit të ndryshimeve është përdorimi i analizës së kovariancës. Elementët që kanë korelacion më të vogël me njëri tjetrin janë të pranueshëm. Një mënyrë tjetër është përdorimi i matjeve të përsëritura, me të cilat do merremi në kapitujt në vijim.
4. Sigurohuni që ka lidhje të fortë ndërmjet metodave që aplikohen dhe variablit të varur, dhe që këto metoda të jenë të shtrira mjaftueshëm në kohë, për të prodhuar një përmasë efekti të madhe ose të pranueshme .

Vlen të theksohet se çështja e fuqisë, varet shumë nga pasojat e të bërit të gabimit të llojit të I. Të diskutosh fuqinë përkundrejt  $\alpha$ , merr kuptim po ti referohemi një konteksti të dhënë psh: “nëse po testoj dy metoda mësimdhënieje, të cilat kushtojnë njëlloj; jam e interesuar për fuqinë. Nëse një metodë kushton 10 herë më shumë sesa metoda tjetër, atëherë jam e interesuar për gabimin  $\alpha$ . Ose psh, nëse po krahasoj dy medikamente A dhe B dhe njëri ka më shumë efekte anësore, atëherë jam më tepër e interesuar për  $\alpha$ .”

Në shembullin e mësimdhënies, nqse bëhet gabimi i llojit të I në konkluzionin që metoda që kushton 10 herë më shumë është më efektive; ky do të ishte një gabim shumë i kushtueshëm. Nqse gabimi i llojit të I bëhet tek shembulli me medikamentët; në konkludimin se medikamentit A (me efekte më shumë anësore) është më i miri, kur ai nuk është i tillë, kjo do të ketë pasoja të rënda shëndetësore për njerëz të tjerë që marrin medikamentin A.

Këto shembuj ishin për të thënë që ndonjëherë mënyra 1 e trajtuar më lart, për të marrë nivelin  $\alpha = 0.1$  ose 0.15 (pra në nivelë të ulëta) me qëllim rritjen e vlerës së fuqisë; mund të mos jetë zgjidhja më e mirë në disa raste.

## **7. VLERËSIMI I FUQISË NË SPSS**

Ekziston një program i veçantë (SPSS-MANOVA) me anë të të cilit gjendet Fuqia. Kështu në këtë program, nënkomanda POWER ndihmon për gjetjen e fuqisë. Fuqia mund të gjendet për çdo nivel  $\alpha$  ndërmjet 0 dhe 1. Nqse duam fuqinë në nivelin  $\alpha = 0.05$  për ANOVA me një faktor, thjesht fusim nënkomandën : POWER = F(0.05)/ Që në fillim përcaktohet nëse ke të bësh me test  $t$  apo me ANOVA me një faktor.

Dhe pastaj fuja e nënkomandës bëhet njëlloj. Një shembull i rezultuar nga printimi i rezultateve të një aplikimi të tillë jepet në **Tabelën 14**

Shembulli majtas ( i parë ) i referohet testit  $t$  për zgjedhjet e pavarura, dhe shembulli djathtas (i dytë) i referohet ANOVA me një faktor. Vlerat respektive të fuqive jepen në tabelë.

Treguesit e përmasës së efektit në çdo rast në Tabelën 14, jepen nëpërmjet eta- në katror të pjesëshme ( $\eta_p^2$ ) ku:

$$\eta_p^2 = (df..F) / (df_h.F + df_e) \quad (a)$$

ku  $df_h$  tregon shkallët e lirisë për hipotezat dhe  $df_e$  tregon shkallët e lirisë për gabimin. Teoria pranon faktin që “ $\eta_p^2$  është një mbivlerësim i përmasës aktuale të efektit. Megjithatë, ajo është një tregues i mirë i përmasës së efektit dhe aplikohet për të gjitha testet  $F$  dhe  $t$ ” .

Le të vërtetojmë këtë mbivlerësim:

Në analizën ANOVA ekziston kjo lidhje ndërmjet  $\eta^2$  dhe përmasës së efektit  $f$ :

$$\eta^2 = f / (1 + f^2) \quad (b)$$

$$\text{Dimë se } f^2 = \sqrt{(k-1)F/n} \rightarrow f'^2 = (k-1)F/n$$

Kështu bëjmë zëvendësimin e  $f^2$  të ekuacioni (b) dhe kemi:

$$\eta^2 = \frac{(k-1).F}{(k-1).F + n} \quad (c)$$



Le të krahasojmë  $\eta^2$  që gjetëm me  $\eta_p^2$  për ANOVA me një faktor. Për ANOVA me një faktor  $df_h = (k-1)$  dhe  $df_e = (n-k)$ . Duke zëvendësuar këto te ekuacioni (a) kemi:

$$\eta_p^2 = (k-1).F / [(k-1).F + (n-k)] \quad (d)$$

Kështu i vetmi ndryshim midis  $\eta^2$  dhe  $\eta_p^2$  për ANOVA me një faktor është  $n$  përkundrejt  $(n-k)$  në emërues. Për aq kohë sa emëruesi në ekuacionin (d) është gjithmonë më i vogël se emëruesi në ekuacionin (c);  $\eta_p^2$  do të jetë një mbivlerësim i përmasës së efektit. Megjithatë për vlera të mëdha të  $n$  ( $n > 50$ ), ndryshimi midis  $\eta^2$  dhe  $\eta_p^2$  është shumë i vogël.

**(Bashkëngjitur Tabelat 1-13)**

## TEMA 10:

### ANALIZA E FUQISË

#### Ushtrime

**Ushtrim 1:** Albani ndërmori një studim me katër grupe mbi efikasitetin e një medikamenti. Ai zgjodhi 15 pacientë për çdo grup dhe përfitoi  $F = 3.46$ . Sa është fuqia e studimit për  $\alpha = 0.05$ ?

**Ushtrim 2:** Një studiues po planifikon një studim, i cili të përfshijë 4 grupe dhe të aplikojë teknikën ANOVA për analizimin e rezultateve. Nga studimi i literaturës së mëparshme, ai vlerëson një përmasë të pritshme efekti prej 0.35. Sa elementë do të duhet të përfshijë ai në këtë studim për të siguruar një fuqi prej 70% dhe besueshmëri 95%?

**Ushtrimi 3:** Një menaxher spitali krahason 4 grupe pacientësh dhe sheh sjelljen e llojit të sëmundjes perkundrejt nivelit të arësimimit. Pyetësi u dërgohet 100 pacientëve. Është përdorur shkalla e zakonshme me vlerësim nga 1 deri te 5, për të vlerësuar këtë sjellje. Një rezultat më I lartë I korespondon një sjelljeje më positive. Të dhënat e grupeve jepen si më poshtë:

|           | Grupi 1 | Grupi 1 | Grupi 1 | Grupi 1 |
|-----------|---------|---------|---------|---------|
| $n_i$     | 23      | 32      | 24      | 20      |
| X mesatar | 32      | 33.1    | 34      | 31      |
| $s_i$     | 7.09    | 7.6     | 7.8     | 7.49    |

1. Bëni analizën e variancës për këtë rast
2. Vlerësoni përmasën e efektit
3. Vlerësoni fuqinë për nivelin e rëndësisë 5 %

**Ushtrimi 4:** Një menaxher spitali I referohet raporteve të tre prej paraardhësve të tij në lidhje me krahasimin e çmimit mesatar të karburantit që ata kishin bërë për dy ofertues, duke u mbështetur në zgjedhjen e disa pikave të karburantit për secilin prej tyre (kuptohet periudhave të ndryshme kohore, dhe teste të dyanshme me nivel rëndësie 5%). Të dhënat janë si më poshtë:

| Menaxheri | Nr. i pikave të karburantit të ofertuesit 1 ( $n_1$ ) | Nr. i pikave të ofertuesit 2 ( $n_2$ ) | tv   |
|-----------|-------------------------------------------------------|----------------------------------------|------|
| 1         | 30                                                    | 25                                     | 1.46 |
| 2         | 15                                                    | 17                                     | 2.6  |
| 3         | 19                                                    | 13                                     | 0.45 |

1. Llogaritni përmasën e efektit për çdo studim
2. Llogaritni fuqinë për çdo rast

**3. Mbështetur në përmasën e efektit dhe fuqinë jepni opinionin tuaj për cilësinë e çdo studimi**

**Ushtrimi 5 :** Një administrator spitali do të testojë nëse niveli mesatar i njohurive të vartësve të tij në lidhje me motivimi në punë është mbi 70 pikë. Për këtë ai do të referohet një zgjedhjeje, mbi bazën e së cilës do të masë nivelin mesatar të pikëve të punonjësve. Nëse ai prët një “t” min 2.7, sa duhet të jetë volumi i zgjedhjes së tij, kur fuqia e pritshme e testit të tij është 80%? Çfarë rekomandoni të bëjë administratori për të ruajtur fuqinë që ai dëshiron por në pamundësi të realizojë një volum kaq të madh zgjedhjeje ? ( niveli i besueshmërisë 95%)

Tema:  
**ANALIZA E FUQISË**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

1

**Çështjet Kryesore që Trajtohen:**

- 1 HYRJJE**
- 2 TESTI  $t$  PËR ZGJEDHJET E PAVARURA**
- 3 VLERËSIMI I FUQISË PARA DHE PAS TESTIMIT**
- 4 VLERËSIMI I FUQISË PËR ANALIZËN E VARIANCËS ME NJË FAKTOR**
- 5 NJË VLERËSIM PARAPRAK I SUBJEKTEVE QË NEVOJITEN PËR NJË FUQI TË DHËNË**
- 6 MËNYRAT E PËRMIRËSIMIT TË FUQISË**
- 7 VLERËSIMI I FUQISË NË SPSS**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

2

|                             | <b>H0<br/>e vërtetë</b>                                                          | <b>H0 jo<br/>e vërtetë</b>                                             |
|-----------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------------|
| <b>Nuk e<br/>refuzoj H0</b> | Vendim i drejtë<br>(niveli i<br>besueshmërisë)<br><b><math>1 - \alpha</math></b> | Gabim i llojit të dytë<br><br><b><math>\beta</math></b>                |
| <b>E refuzoj<br/>H0</b>     | Gabim i llojit të parë<br><b><math>\alpha</math></b>                             | Vendim i drejtë<br>(Fuqia Statsitike)<br><b><math>1 - \beta</math></b> |

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

3

## 1 HYRJE

- Niveli i  $\alpha$  përcaktohet nga studiuesi, pra është një vendim subjektiv; ku shumica e studiuesve preferojnë  $\alpha = 0.05$  ose  $\alpha = 0.01$ .
- Vetëm njeri nga këto dy gabime mund të ndodhë në një studim të caktuar; që do të thotë: ose ne gabimisht pranojmë  $H_0$  ose gabimisht  $H_0$  e hedhim poshtë  $\alpha$  dhe  $\beta$  janë të lidhura në mënyrë inverse.
- Faktorët që ulin riskun e gabimit  $\beta$ , rrisin fuqinë statistikore  $1 - \beta$

### **Në një kontekst të gjerë, fuqia varet nga shumë faktorë:**

- Niveli  $\alpha$  i vendosur nga studiuesi
- Volumi i zgjedhjes
- Përmasat e efektit
- Testi statistikor i përdorur (Psh. Testi  $t$  për zgjedhjet e varura është më i fuqishëm sesa testi  $t$  për zgjedhjet e pavarura; një studim i matjeve të përsëritura është më i fuqishëm sesa ANOVA me një faktor etj)
- Projekti i kërkimit shkencor

### **1. Niveli $\alpha$ i vendosur nga studiuesi**

$\alpha \uparrow \rightarrow \beta \downarrow \rightarrow 1 - \beta \uparrow$

Shumë studiues preferojnë të pranojnë nivele të ulta të  $\alpha$  duke synuar ndryshimin e elementëve të tjerë të studimit për përmirësimin e fuqisë.

### **2. Volumi i zgjedhjes e 3. Përmasat e efektit**

$$t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} = (\bar{x} - \mu_0) / sx \left( \frac{1}{\sqrt{n}} \right) = dx \sqrt{n}$$

$$d = \frac{\bar{X} - \mu_0}{S}$$

$d$  quhet Cohen's  $d$  dhe është tregues i përmasës së efektit

Sup.  $n$  mbahet konstante; përmasat e efektit të popullimit (Cohen's  $d$ )  $\uparrow$ ,  $\rightarrow t \uparrow$  Ka të ngjarë të përfitohet një vlerë e madhe e  $t$ -së së llogaritur (mjaftueshëm për të plotësuar kushtin  $t_{vr} > t_{kr}$ ).

a) Nqse.  $d \uparrow$  (kur faktorët e tjerë mbahen konstantë)  $\rightarrow 1- \beta \uparrow \rightarrow \beta \downarrow$

b) Nqse  $n \uparrow$  rritet  $n$  (kur faktorët e tjerë mbahen konstantë),  $t \rightarrow t \uparrow \rightarrow 1- \beta \uparrow \rightarrow \beta \downarrow$

Ka tabela të posaçme për llogaritjen e fuqisë (koordinatat janë  $n$  dhe  $d$ ) ku pohimet e mësipërme mund të konfirmohen duke i referuar këtyre tabelave, psh: për  $\alpha=0.05$  (tabela

1)

## 2 TESTI t PËR ZGJEDHJET E PAVARURA

Në rastin e zgjedhjeve të pavarura:

$d = (\mu_1 - \mu_2) / \sigma$  (popullimi)

$d' = (x_1 - x_2) / s$  (zgjedhja)

Psh. nqse mesataret e zgjedhjeve do të ishin përkatësisht:  $x_1 = 10$  dhe  $x_2 = 4$ ,  $s = 15$ , atëherë  $d = (10 - 4) / 15 = 0.4$  që do të thotë se mesataret ndryshojnë ndërmjet tyre me 0.4 devijime standarte.

Përmasë efekti  $d \leq 0.2$  konsiderohet si e vogël;

$0.20 < d \leq 0.79$  konsiderohet e mesme

$d \geq 0.80$  konsiderohet e madhe.

Përgjithësisht një përmasë e efekti e mesme është më e përshtatshme për studiuesit.

Në rastin kur  $n_1 = n_2$ , për nivel të caktuar te  $d$ -se, ka tabela (Tabelat 2-5) për të cilat llogaritet fuqia. Psh:

$n_1 = n_2 = 15$ ;  $d = 0.50$ , fuqia = 0.25

$n_1 = n_2 = 35$ ;  $d = 0.30$ , fuqia = 0.23

Fuqia mund të përmirësohet, edhe duke mbajtur volum te vogël zgjedhjesh, por vetëm nqse përmasa e efektit është e madhe: 3.  $n_1 = n_2 = 15$ ;  $d = 1$ , fuqia = 0.75

**Si vlerësohet  $n$  nqse.  $n_1 \neq n_2$  ?**

në këtë rast përdorimin e mesatares harmonike. Dihet se mesatarja harmonike për dy vlera:  $n_1$  dhe  $n_2$  llogaritet si:  $2 n_1 n_2 / (n_1 + n_2)$ .

Vlen të theksohet fakti se dallimi midis mesatares aritmetike dhe asaj harmonike, qëndron pikërisht se ndryshimi ndërmjet këtyre dy mesatareve është i vogël kur volumi i zgjedhjeve është pothuajse i barabartë; por nqse volumet e zgjedhjeve janë shumë të ndryshme atëherë diferenca ndërmjet këtyre mesatareve është e kondiderueshme



Avantazhi statistikor i testeve të njëanshme, krahasuar me ato të dyanshmet, është se të parat kanë fuqi statistikore më të madhe

| $(n1 = n2 = 20)$                 | $(d=0.6)$ | $(d=0.8)$ |
|----------------------------------|-----------|-----------|
| $\alpha=0.01$ , test i dyanshëm  | 0.21      | 0.44      |
| $\alpha=0.05$ , test i njëanshëm | 0.58      | 0.80      |
| $\alpha=0.1$ , test i njëanshëm  | 0.71      | 0.89      |

### 3 VLERËSIMI I FUQISË PARA DHE PAS TESTIMIT

Një vlerësim paraprak i fuqisë (zakonisht 70-80%) i tregon studiuesit se sa do të duhet të jetë volumi i zgjedhjeve për cdo grup, në mënyrë që të përfitohet vlera e dëshiruar e fuqisë statistikore.

Kur s'ka diferencë të rëndësishme ndërmjet mesatareve ( $\mu_1 - \mu_2 = 0$ ), dhe :

**1.** nuk i kushtohet rëndësi fuqisë, atëherë konkluzioni do të ishte se aplikimi i një metode të caktuar nuk ka efekt në këto mesatare

**2.** duhet të merren parasysh edhe volumet e zgjedhjeve nëpër studime, edhe përmasat e efektit sipas rasteve perkatese. Nëse volumet e zgjedhjeve do të ishin te vogla atëherë mungesa e rëndësisë së diferencës së mesatareve është për shkak të vlerës së vogël të fuqisë.

Ndaj përcaktohet paraprakisht fuqia (70-80%) dhe përcaktohet më pas n

Pas përfundimit të testit: përcaktohet fuqia kur dihet:  $d'$ ,  $n_1$  e  $n_2$  ku

$$d' = t * \sqrt{1/n_1 + 1/n_2}$$

### 4 VLERËSIMI I FUQISË PËR ANALIZËN E VARIANCËS ME NJË FAKTOR

Sipas Cohen;

$$f = \sqrt{(k-1)F/n}$$

$f < 0.1$  tregon për një përmasë të vogël efekti;

$0.10 < f < 0.4$  tregon për një përmasë të mesme efekti;

$f > 0.40$  tregon për një përmasë të madhe efekti. (tabela 6-13)

Përmasa e efektit për Anova-n është  $f$

Ku:

$$\sigma_m^2 = \frac{\sum n_i (\bar{x}_i - \bar{x})^2}{n} = \frac{SKR}{n}$$

Ku del SKR . Dhe

$$F = \frac{SKR/(k-1)}{SKG/(n-1)} = \frac{SKR}{(k-1)\sigma^2} = \frac{\sigma_m^2 n}{(k-1)\sigma^2} = \frac{f^2 n}{(k-1)}$$

Në tabelat në rastin e ANOVA-s hyhet me **n** (kur n-të janë të barabarta) dhe me **f**.

Kur n-i t nuk janë të barabarta hyhet me **n** ku **n** është mesatarja e **n**-i-ve  
**Shembuj**

-Nqse nga një studim rezulton se nuk ka ndryshim statistikisht të rëndësishëm mes mesatareve, dhe ku përmasa e efektit është e vogël (0.10), atëherë studiuesi duhet të vendosë nese praktikisht ka apo jo ndryshim statistikisht të rëndësishëm.

- Nqse. nuk ka ndryshim statistikisht të rëndësishëm mes mesatareve, dhe nga analiza pas testimit rezulton një përmasë efekti e mesme ose e madhe; atëherë është e rëndësishme të përsëritet studimi me volum më të përshtatshëm zgjedhjeje

#### 5. NJE VLERESIM PARAPRAK I SUBJEKTEVE QE NEVOJITEN PER NJE FUQI TE DHENE

•Për këtë, duhet një vlerë e pritshme e përmasës së efektit; që të futemi në tabelat e fuqisë Cohen, për të përcaktuar sa duhet të jetë volumi i zgjedhjes për çdo grup, në mënyrë që të përfitohet një fuqi e specifikuar në një nivel të caktuar  $\alpha$ .

•Për këtë qëllim, duhet t'i referohemi përmasës mesatare të efektit duke iu referuar përmasave të efekteve të studimeve të mëparshme të ngjashme. Kështu është më mirë t'i referohemi paraprakisht disa vlerave për të gjetur volumin e përshtatshëm të zgjedhjes, sesa të mos e kesh fare idenë lidhur me këtë të fundit.

•Shembuj

#### 6 MËNYRAT E PËRMIRËSIMIT TË FUQISË

1. Përdorni teste kontrolli të njëanshme, kur ka mundësi të përdoret një rast i tillë  
2. Konsideroni mundësitë e zvogëlimit të ndryshimeve brenda grupeve. Një mënyrë është nëpërmjet selektimit të zgjedhjeve: elementë më homogjene, tentojnë të kenë më pak ndryshime të variablilit të varur. Psh. përdorni vetëm meshkuj, në vend të meshkujve dhe femrave

3. Aplikoni një nivel më të butë (më të lartë)  $\alpha$ , ndoshta  $\alpha = 0.1$  ose  $\alpha = 0.15$  (kujdes shembulli i mësimdhënies dhe i medikamenteve për të treguar që në disa raste rritja e gabimit të llojit të parë me qëllim rritjen e vlerës së fuqisë; mund të mos jetë zgjidhja më e mirë.)

7. VLERËSIMI I FUQISË NË SPSS

SPSS-MANOVA  
Komanda POWER

---

---

---

---

---

---

---

**Tema:**  
**ANALIZA E FUQISË**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

1

---

---

---

---

---

---

---

---

**Çështjet Kryesore që Trajtohen:**

- 1 HYRJE**
- 2 TESTI  $t$  PËR ZGJEDHJET E PAVARURA**
- 3 VLERËSIMI I FUQISË PARA DHE PAS TESTIMIT**
- 4 VLERËSIMI I FUQISË PËR ANALIZËN E VARIANCËS ME NJË FAKTOR**
- 5 NJË VLERËSIM PARAPRAK I SUBJEKTEVE QË NEVOJITEN PËR NJË FUQI TË DHËNË**
- 6 MËNYRAT E PËRMIRËSIMIT TË FUQISË**
- 7 VLERËSIMI I FUQISË NË SPSS**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

2

---

---

---

---

---

---

---

---

**Tema:**  
**ANALIZA E FUQISË**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

1

---

---

---

---

---

---

---

---

## **TEMA 11:**

### **TESTIMI I HIPOTEZAVE: METODAT JO PARAMETRIKE – TESTET ‘X2’**

#### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje, që edhe për rastet e popullimeve që nuk kanë supozimin e shpërndarjes normale të :

- Të përdorin testin chi-square për të gjykuar në lidhje me pavarësinë ose jo të fenomeneve
- Të përdorin testin e shenjave për rastin e zgjedhjeve të varura
- Të pëcaktjnë kur dhe si do të duhet të përdorin kriterin e shenjave të rangut Wilcoxon për rastin e zgjedhjeve të varura.
- Të përdorin testin e shenjave për rastin e testit të medianës së një popullimi.
- Të aplikojnë përafrimin normal për testin në të cilin aplikohet kriteri i shenjave Wilcoxon.
- Të dinë kur dhe si të performojnë kriterin Mann-Whitney U ( rasti i zgjedhjeve të pavarura)
- Të dinë kur dhe si të performojnë kriterin Kruskal Wallis ( rasti i testimit të hipotezave të më shumë se dy popullimeve)
- Të aplikojnë në SPSS dhe të interpretojnë outputin për rastin e testit të hipotezave mbi varësinë apo jo të variablave kategorikë

#### **LITERATURA**

- “ Applied Statistics”, Warner A.R , 2008
- Statistika”- Grup Autoresh, Shtëpia botuese Pegi, (2016)
- Statistical for Business nad Economics”, Newbold P, Carlson W, Thorne B, 2011
- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011

-“Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition,  
Belinda Barton, Jennifer Peat, September 2014



## KAPITULLI 11 METODAT JO PARAMETRIKE

Metodat statistikore që paraqiten deri ketu zakonisht quhen metoda parametrike. Metodat jo parametrike në përgjithësi nuk kërkojnë plotësimin në mënyrë të plotë të supozimeve mbi të cilat mbështeten metodat parametrike, të cilat lidhen me shkallën e matjes të të dhënave, me formën e shpërndarjes probabilitare të të dhënave të zgjedhjes etj. Një nga konsideratat për të përcaktuar secilat metodat mund të përdoren për analizën e të dhënave, është shkalla e matjes e përdorur për marrjen e vendimeve të tyre. Le të shikojmë edhe një herë përkufizimet e dhëna për shkallën e matjeve, me shembujt që u dhënë në kapitullin 1.

1. Shkalla e matjes quhet nominale, në qoftë se të dhënat janë thjeshtë shenja, ose kategori që përdoren për treguar një veti të elementi

2. Shkalla e matjes quhet e zakonshme, në qoftë se të dhënat kanë vetite e të dhënave nominale dhe përveç kësaj renditja e tyre është me kuptim.

3. Shkalla e matjes quhet intervalore, në qoftë se të dhënat kanë vetite e të dhënave të zakonshme dhe intervali midis vlerave shprehët në terma të njësive të matjes së tyre.

4. Shkalla e matjes quhet në formë raportit, në qoftë se të dhënat kanë vetite e shkallës intervalore dhe ka kuptim raportit i tyre.

Shumica e metodave statistikore që konsiderohen si parametrike, kërkojnë përdorimin e të dhënave intervalore ose të atyre në formë raportit. Për këtë nivel matje mund të llogariten mesataret, variancat, devijimet standarte etj., të cilat mund të përdoren më tej për analizën e të dhënave. Në rastin e të dhënave të shprehura në shkallën nominale, ose të zakonshme, llogaritja e këtyre treguesve nuk mund të bëhet, si rrjedhojë nuk mund të zbatohen as metodat parametrike. Kështu për analizën e të dhënave të shprehura në shkallën nominale ose të zakonshme mund të përdoren vetëm metodat jo parametrike.

Një konsideratë tjetër që përdoret për të konstatuar nëse për analizën e të dhënave mund të përdoren metodat parametrike ose ato jo parametrike, janë supozimet që bëhen për popullimin nga merren vlerat. Marrim p.sh. rastin kur kërkojmë të kontrollojmë diferencën midis mesatareve të dy popullimeve në kushtet e zgjedhjes së vogël. Siç dihet në këtë rast mund të përdoret statistika  $t$ , me kusht që dy zgjedhjet vijnë nga dy popullime me shpërndarje normale dhe varianca të barabarta. Në qoftë se ky supozim nuk plotësohet, atëherë kjo procedurë nuk mund të zbatohet megjithatë të dhënat janë në shkallë intervalore ose në formë raportit. Edhe në këtë rast analiza e të dhënave mund të bëhet vetëm me anë të metodave jo parametrike. Kjo është arsyeja që metodat jo parametrike quhen metoda të lira nga shpërndarjet. Në përgjithësi një metode statistikore quhet jo parametrike, kur ajo mund të zbatohet ndaj të dhënave nominale ose të zakonshme, apo kur mund të zbatohet ndaj të dhënave intervalore ose në formë raportit në rastin kur asnjë supozim nuk mund të bëhet për shpërndarjen probabilitare për popullimin. Në këtë kapitull do të paraqesim disa nga metodat më kryesore jo parametrike.

### 18.1 Kriteri i shenjave

Le të shikojmë zbatimin e kriterit të shenjave për një shembull nga studimi i tregut, ku për qëllime studimi kryhet një zgjedhje e rastit me  $n$  konsumatore, të cilëve u kërkojnë preferencat për një ose dy lloje të një produkti, si p.sh. kafe, pije të ftohta dhe detergjente. Të shprehjet për preferencat janë të dhëna nominale, me qëndrueshmëri se konsumatori vetëm emerton ose shpreh një preferencë. Qëllimi është që për të dhënat e vërtetuara të tregojmë nëse ekziston një ndryshim në preferencat lidhur me dy artikujt që krahasohen. Për të përgjigjur një pyetje të tillë mund të përdoret një procedurë statistikore që njihet me emrin kriteri i shenjave.

**Rasti i zgjedhjes së vogël** Rasti i zgjedhjes së vogël për kriterin e shenjave përdoret kur vëllimi i zgjedhjes në plotëson kushtin,  $n \leq 20$ . Le të zëmë se kërkojmë të njohim preferencat e konsumatoreve për dy lloje lengjesh portokalli. Për këtë zgjidhen 12 persona, të cilëve u jepen për të provuar të dy llojet e lengjeve, pa ditur llojin e lengut të dhënë në një radhë të rastit. Pas shijimit të dy llojeve personat, u pyeten se cilin prej llojeve të lengut të portokallit preferojnë më tepër. Qëllimi i studimit në këtë rast është të përcaktohet nëse konsumatorët preferojnë njërin lloj më shumë se tjetrin. Kështu në qoftë se quajmë me  $p$  përpjesën e popullimit të konsumatoreve që preferojnë llojin e hipotezës që kërkojmë të kontrollojmë janë.

$$H_0: p = 0,50$$

$$H_a: p \neq 0,50$$

Jo

$n \leq 20$

4



Ne qofte se hipoteza  $H_0$  nuk mund te hidhet poshte, atehere nuk ka argumenta per te thene se ekziston nje ndryshim ne preferencat e konsumatoreve per njerin nga llojet e lengut te portokallit. E kunderta ne qoftese hipoteza  $H_0$  hidhet poshte, themi se konsumatorët preferojne njerin lloj me shume se tjetri. Ne kete rast lloji me i preferuar nga shumica e konsumatoreve konsiderohet lloji me i mire i lengut te portokallit dhe si rrjedhoje do te tregetohet, ose prodhohet me shume nga firma.

Me poshte, duke zbatuar kriterin e shenjave per zgjedhjen e vogel, do te tregojme se si mund te kontrollojome hipoteza te tilla dhe te nxjerrim konkluzione per preferencat e konsumatoreve. Per regjistrimin e te dhenave te prefereneave te konsumatoreve per 12 pjesmarresit ne studim, do te shenojme me shenjen + ne rastin kur konsumatori shpreh preference per llojin e pare te lengut te portokallit dhe me shenjen - ne qofte se konsumatori shpreh preference per llojin tjetër te lengut te portokallit. Me anen e kesaj procedure do te kemi te regjistruara vetem shenja + dhe -, prandaj dhe ky kriter mban emrin kriteri i shenjave.

Me supozimin se hipoteza zero eshte e vertete (ne rastin konkret  $p = 0,50$ ), numri i shenjave te njejta ndjek nje shperndarje probabilitare binomiale me  $p = 0,50$ . Ne tabelen 5 te shtojces B, <sup>21/240</sup> per  $p=0,5$  dhe per vellimin e zgjedhjes  $n = 12$ , gjejme vleren e probabilitetit per numrin e shenjave plus ne kushtet kur hipoteza zero eshte e vertete. Keshtu shperndarja binomiale mund te perdoret per percaktimin e nje rregulli per hedhjen poshte te hipotezes  $H_0$ , e cila eshte analoge me metodën e kontrollit te hipotezave qe kemi pare ne kapitullin 9. P. sh. per nivelin e rendesise  $\alpha = 0,05$ , zona kritike do te jete e dyaneshme me siperfaqe te barabarte me 0,025 nga seicila ane e vlerave kritike. Duke filluar nga vlera skajore me e vogel e kesaj shperndarje, shohim se probabiliteti i marrjes te 0, 1, ose te 2 shenjave eshte i barabarte me  $0,0002 + 0,0029 + 0,0161 = 0,192$ . Ne duhet te ndalojme tek 2 shenja + sepse po te shtojme edhe probabilitetin e 3 shenjave + do te marrim nga ana e poshtme te vleres kritike nje zone me siperfaqe te barabarte me  $0,0192 + 0,0537 = 0,0729$ , e cila e kalon ne menyre te konsiderueshme zonen qe kerkojme me siperfaqe 0,025. Po ashtu nga ana tjetër ne pjesen e siperme te shperndarjes, gjejme te njehtin probabilitet te barabarte me 0,192 qe i korespondon 10, 11 dhe 12 shenjave +. Keshtu probabiliteti me i aferte me 0,05 dhe me i vogel se ai eshte  $0,0192 + 0,0192 = 0,0384$ . Si rrjedhoje mund te percaktojme kete rregull per hedhjen poshte te hipotezes.

Hipoteza zero hidhet poshte ne qofte se numri i shenjave eshte me i vogel se 3, ose me i madh se 9. Te dhenat mbi preferencat e konsumatoreve per llojin e lengut te portokallit jane paraqitur ne tabelen 18.1. Me qene se jane vrojtuar vetem 2 shenja +, atehere hipoteza zero hidhet poshte. Keshtu arrijme ne perfundimin se preferencat e konsumatoreve ndryshojne persa i perket llojeve te lengut te portokallit.

Tabela 18.1 Te dhenat per preferencat e 12 konsumatoreve

| Personat e pyetur | Preferenca per llojin | Te dhenat e regjistruara |
|-------------------|-----------------------|--------------------------|
| 1                 | Lloji 1               | -                        |
| 2                 | Lloji 1               | -                        |
| 3                 | Lloji 2               | +                        |
| 4                 | Lloji 1               | -                        |
| 5                 | Lloji 1               | -                        |
| 6                 | Lloji 1               | -                        |
| 7                 | Lloji 1               | -                        |
| 8                 | Lloji 1               | -                        |
| 9                 | Lloji 2               | +                        |
| 10                | Lloji 1               | -                        |
| 11                | Lloji 1               | -                        |
| 12                | Lloji 1               | -                        |

Ne rastin e shembullit tone te 12 personat ishin ne gjendje te jepnin nje preference per llojin e lengut te portokallit. Me gjithate ne disa raste, nje ose disa nga personat e zgjedhur nuk jane ne gjendje te japin nje preference te caktuar per ate qe u kerkohet. Ne keto raste pergjigjet e papercaktuara te personave duhet te largohen nga studimi dhe analiza te behet ne baze te nje vellimi me te vogel te zgjedhjes.

Shperndarja probabilitare binormale mund te perdoret per te nxjerre nje rregull te pergjithshem per cdo kriter shenjash per vellimin e zgjedhjes deri ne  $n = 20$ . Duke perdorur hipotezen zero  $p = 0,50$  dhe vellimin e zgjedhjes  $n$ , rregulli per marrjen e vendimit mund te vendoset per edo



nivel te rendesise. Rregulli i hedhjes poshte te hipotezes mund te ndertohet gjithashtu duke zgjedhur zonen kritike ne menyre te njeanshme, te majte ose te djathte. Tabela e shperndarjes probabilitare binomiale mund te perdoret per vellintin e zgjedhjes deri ne  $n=20$ . Per zgjedhjet me vellim me te madh se ne vend te shperndarjes binomiale perdoret shperndarja probabilitare normale.

- ① **Rasti i zgjedhjes se madhe** Per kontrollin e hipotezes zero ne kushtet e zgjedhjes se madhe, d.m.th. kur vellimi i zgjedhjes eshte me i madh se 20, shperndarja e zgjedhjes per numrin e shenjave + mund te perafrohet me shperndarjen normale si me poshte.

$$\text{Mesatarja: } \mu = 0,50n \quad 18.1$$

$$\text{Devijimi standart: } \sigma = \sqrt{0,25n} \quad 18.2$$

Forma e shperndarjes afersisht normale, per  $n > 20$

Marrim nje shembull per zbatimin e kriterit te shenjave kur kerkohet nje informacion paraprak, se cili nga dy kandidatet e nje partie politike ka me shume mbeshtetjen e votuesve ne nje zone elektorale. Le ta zeme se u muar mendimi i 200 votuesve nga te cilet 72 ishin per kandidatin e pare, 103 ishin per kandidatin e dyte dhe 25 nuk shikonin ndonje ndryshim midis kandidateteve. A tregon ky vrojtim se ekziston nje ndryshim i rendesishem i opinionit per te dy kandidatet?

Duke perdorur kriterin e shenjave shohim se  $n=200-25=175$  persona ishin ne gjendje te percaktonin kandidatin qe mund te ishte me i pershtatshem ne zgjedhjet e ardheshme. Me anen e 18.1 dhe 18.2 gjejm se shperndarja e zgjedhjes te numrit te shenjave + ka vetite:

$$\mu = 0,50n = 0,50(175) = 87,5$$

$$\sigma = \sqrt{0,25n} = \sqrt{0,25(175)} = 6,6$$

Perveç kesaj per nje vellim te madh te zgjedhjes  $n = 175$ , mund te supozojme se shperndarja e zgjedhjes eshte afersisht normale. Meqenese shperndarja eshte afersisht normale, per te ndertuar kriterin e hedhjes poshte te hipotezes mund te perdorim tabelen e shperndarjes probabilitare normale standarte. Per nivelin e rendesise  $\alpha = 0,05$  rregulli i hedhjes poshte te hipotezes per kriterin e dyanshem mund te shkruhet. Hipoteza zero hidhet poshte ne qofte se  $z < -1,96$ , ose  $z > 1,96$ . Ne rastin konkret per numrin e shenjave + ( $x = 72$ ) vlera e statistikes  $z$  rezulton.

$$z = \frac{x - \mu}{\sigma} = \frac{72 - 87,5}{6,6} = -2,3$$

Meqenese  $z = -2,35$  eshte me vogel se  $-1,96$ , per nivelin e rendesise  $\alpha = 0,05$  hipoteza se nuk ka ndryshim ne opinionin e elektoratit per kandidatet hidhet poshte. Keshtu bazuar ne kete studim do te ishte me mire qe si kandidat ne zgjedhjet e ardheshme te propozohet kandidati i dyte qe mbeshtetet nga me shume votues.

**Kontrolli i hipotezes per medianen** Kujtojm se mediana ndan popullimin ne dy pjese te barabarta, ku 50% e vlerave jane me te vogla se mediana dhe po kaq perqind e tyre jane me te medha se mediana. Per kete mund te perdorim kriterin e shenjave duke vendosur shenjen + per ato vlera qe jane mbi medianen dhe shenjen - per vlerat qe jane nden medianen. Vlerat qe jane te barabarta me medianen nuk merren ne konsiderate. Llogaritja e kriterit te shenjave behet krejtesisht njelloj si me pare. Le ta zeme se duhet te kontrollojme hipotezen per cmimin mediane te apartamenteve te reja me dy dhoma e nje kuzhine te ndertuara ne nje rreth te caktuar. Hipoteza qe duhet te kontrollojme eshte

$$H_0: \text{Mediana} = 3\,000\,000 \text{ leke}$$

$$H_a: \text{Mediana} \neq 3\,000\,000 \text{ leke}$$

Nga nje zgjedhje prej 62 apartamentesh te reja, 34 prej tyre kane patur nje cmim mbi 3 000 000 leke, 26 kane patur cmim nden 3 000 000 leke dhe 2 prej tyre jane shitur ekzakt me cmim 3 000 000 leke. Duke perdorur 18.1 dhe 18.2 dhe per numrin e apartamenteve te shitura me cmim te ndryshem nga 3 000 000 leke ( $n = 60$ ) kemi.

$$\mu = 0,50n = 0,50(60) = 30$$

$$\sigma = \sqrt{0,25n} = 0,25(60) = 3,87$$

Per numrin e shenjave te barabarte me  $x = 34$ , vlera e kriterit te shenjave eshte.

$$z = \frac{x - \mu}{\sigma} = \frac{34 - 30}{3,87} = 1,03$$

Per kriterin dyanshem te shenjave dhe nivelin e rendesise 0,05, hipoteza zero hidhet poshte ne qofte se vlera e  $z$  eshte me e vogel se  $-1,96$ , ose me e madhe se  $+1,96$ . Me qene se vlera



faktike e kriterit  $z = 1,03$  bie ne zonen e lejuar, hipoteza zero nuk mund te hidhet poshte. Ne baze te dhenave nuk mund te hedhim poshte supozimin se vlera mediane e cminut te apartamenteve te reja ndryshon nga 3 000 000 leke.

### 18.2 Kriteri i shenjave te rangut Wilcoxon

Kriteri i shenjave te rangut Wilcoxon eshte nje alternative tjeter jo parametrike e kriterit parametrik te perputhjes qe pame ne kapitullin 10. Ne situaten e zgjedhjes shoqeruese, cdo njesi eksperimentale gjeneron dy vrojtme te ciftuara qe lidhen midis tyre, nje nga popullimi i pare dhe nje nga popullimi i dyte. Ndryshimet midis vrojtmeve te ciftuara japin nje panye te brendshme te ndryshimeve qe ekzistojne midis dy popullimeve. Metodologjia e analizes parametrike te zgjedhjes te ciftuar (kriteri t per diferencat e cifteve) bazohet ne te dhenat intervalore dhe ne supozimin mbi shpermdajen normale te diferencave te cifteve te vrojtmeve te popullimit. Me kete supozim, kontrolli i hipotezes zero se nuk ekziston ndryshim midis mesatareve te popullimit mund te behet me anen e kriterit t. Ne qofte se supozimi mbi shpermdajen normale te diferencave nuk plotesohet, atehere ne vend te kriterit t mund te perdorim kriterin Wilcoxon te rangut me shenja. Zbatimin e kriterit t do ta shohim ne shembullin e krahasimit te dy metodave te prodhimit.

Ta zeme se nje firme prodhuese kerkon te percaktojne ne se ekziston ndonje ndryshim per kohezgjatjen e prodhimit te produktit per dy metoda prodhimi. Per kete zgjidhen 11 punetore, te cilet vrojtohen ne dy metoda prodhimi. Metoda e prodhimit me te cilen punetori punon ne fillim zgjidhet ne menyre te rastit. Keshtu cdo punetor ne zgjedhje gjeneron nje cift vrojtimesh, sic tregohet ne tabelen 18.2. Nje difference pozitive e vrojtuar midis kohes mesatare per prodhimin sipas metodave se pare dhe kohes mesatare per prodhimin sipas metodave se dyte tregon, se metoda e parem kerkon pak me shume kohe se metoda e dyte, kurse nje difference negative flet per te kunderten. A mund te nxjerrim perfundimin se dy metodat ndryshojne ne menyre te rendesishme midis tyre nga pikpamja e kohes se nevojshme per prodhimin e produktit.

Tabela 18.2 Kohezgjatja e prodhimit per dy metoda (ne minuta)

| Punetori | Metoda e prodhimit |      | Ndryshimi |
|----------|--------------------|------|-----------|
|          | I                  | II   |           |
| 1        | 10,2               | 9,5  | 0,7       |
| 2        | 9,6                | 9,8  | -0,2      |
| 3        | 9,2                | 8,8  | 0,4       |
| 4        | 10,6               | 10,1 | 0,5       |
| 5        | 9,9                | 10,3 | -0,4      |
| 6        | 10,2               | 9,3  | 0,9       |
| 7        | 10,6               | 10,5 | 0,1       |
| 8        | 10,0               | 10,0 | 0,0       |
| 9        | 11,2               | 10,6 | 0,6       |
| 10       | 10,7               | 10,2 | 0,5       |
| 11       | 10,6               | 9,8  | 0,8       |

Keshtu problemi qe shtrohet eshte te kontrollohet ne se te dy metodat kane ndryshim per kohen e nevojshme te prodhimit. Ne kete rast kemi dy popullime lidhur me kohen e prodhimit, nje per metodën e pare dhe nje per metodën e dyte. Hipotezat qe duhet te kontrollohen jane:

$H_0$ : Popullimet jane identike dhe  $H_a$ : Popullimet nuk jane identike

Ne qofte se hipoteza  $H_0$  nuk mund te hidhet poshte, atehere ekziston nje baze e mjaftueshme qe te arrijme ne perfundimin se te dy metodat ndryshojne midis tyre nga koha qe duhet per prodhimin e produktit.

Etapa e pare per zbatimin e kriterit te rangut me shenja Wilcoxon, eshte renditja e vlerave absolute te diferencave te cifteve te kohes sipas te dy metodave. Per te bere kete ne fillim largojme vlerat zero te diferencave dhe pastaj renditim te gjitha diferencat e mbetura ne rendin rrites, d.m.th. nga vlera me vogel ne vleren me te madhe te tyre me anen e vargut te numrave natyrale, d.m.th. cdo difference i veme perbri rangun perkates. Diferencave te njellojta u veme ne korespondence vleren mesatare te rangjeve qe jane ne dispozicion per vlerat perkatese. Rendi ose rang u vlerave absolute te diferencave per shembullin qe kemi marre tregohet ne tabelen 18.3. Veme re se diferenca 0 per punetorin e 8-te eshte larguar nga renditja, pastaj differences me te vogel 0, 1 i vihet rang 1. Renditja e vlerave absolute te diferencave vazhdon



deri në vlerën 0,9 që i përgjigjet punëtorit të gjashtë, të cilit i vëhet rangi 10. Diferencave absolute 0,4 për punëtorët 3 dhe 5 u vëhet mesatarja e rangjeve në dispozicion të tyre  $3,5 = (3 + 4) / 2$ . Nderkohe diferencave absolute 0,5 për punëtorët 4 dhe 10 u vëhet rangi 5,5.

Tabela 18.3 Vendosija e rangut me shenja për diferencat absolute të kohezgjatjes së prodhimit

| Punëtori                    | Diferenca | Vlera e differences | Rangu | Rangu me shenja |
|-----------------------------|-----------|---------------------|-------|-----------------|
| 1                           | 0,7       | 0,7                 | 8     | +8              |
| 2                           | -0,2      | 0,2                 | 2     | -2              |
| 3                           | 0,4       | 0,4                 | 3,5   | +3,5            |
| 4                           | 0,5       | 0,5                 | 5,5   | +5,5            |
| 5                           | -0,4      | 0,4                 | 3,5   | -3,5            |
| 6                           | 0,9       | 0,9                 | 10    | +10             |
| 7                           | 0,1       | 0,1                 | 1     | +1              |
| 8                           | 0,0       | 0,0                 | -     | -               |
| 9                           | 0,6       | 0,6                 | 7     | +7              |
| 10                          | 0,5       | 0,5                 | 5,5   | +5,5            |
| 11                          | 0,8       | 0,8                 | 9     | +9              |
| Totali i rangjeve me shenja |           |                     |       | +44             |

Pasi janë vendosur rangjet e diferencave absolute, përbri tyre vendosen shenjat e vetë diferencave të të dhënave të vrojtuesit. Për shembull differences 0,1 për punëtorin 7, të cilit i korrespondon rangi 1, i vëhet vlera +1 sepse diferenca e vrojtuar midis dy metodave është pozitive. Differences 0,2 që ka rangun 2 i jepet vlera -2, sepse diferenca e vrojtuar midis dy metodave për këtë person është negative. Lista e plotë e rangjeve me shenja së bashku me shumën e tyre paraqitet në shtyllën e fundit të tabelës 18.3.

Le të kthehemi tani tek hipoteza fillestare mbi barazimin e koheve të prodhimit për të dy metodat. Në qoftë se popullimet që paraqesin kohët e prodhimit për secilën nga të dy metodat janë identike, do të prisnim që rangjet pozitive dhe rangjet negative të eliminonin njëri tjetrin, kështu që shuma e rangjeve me shenja do të ishte afërsisht e barabartë me zero.

Në qoftë se quajmë me  $T$  shumën e rangjeve me shenja në kriterin Wilcoxon, atëherë mund të arëgohet se në qoftë se të dy popullimet janë identike dhe numri i cifteve shoqëruese të të dhënave janë më shumë se 10, shpërndarja e zgjedhjes së  $T$  mund të përafrohet

Mesatarja  $\mu_T = 0$  18.3

Devijimi standard  $\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{6}}$  18.4

Forma e shpërndarjes për  $n > 10$  është afërsisht normale.

Për shembullin tonë  $n = 10$ , mëqenëse një vrojtues ka kemi larguar (diferencën 0 për punëtorin 8). Kështu duke përdorur 18.4 për devijimin standard marrim.

$$\sigma_T = \sqrt{\frac{10(11)(12)}{6}} = 19,62$$

Me supozimet e bëra më lart shpërndarja  $T$  është normale, si rrjedhojë statistika  $z$  që paraqitet më poshtë është përsëri normale me mesatare 0 dhe devijim standard 1.

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{44 - 0}{19,62} = 2,24$$

Për kontrollin e hipotezës zero se nuk ekzistojnë ndryshime, në tabelën e shpërndarjes normale për nivelin e rëndësisë  $\alpha = 0,05$  gjejmë vlerën kritike të  $z = \pm 1,96$ . Meqenëse vlera e  $z = 2,24$  e kalon vlerën kritike 1,96, hipoteza zero hidhet poshtë dhe arrijmë në përfundimin se të dy popullimet ndryshojnë midis tyre dhe si rrjedhojë të dy metodat ndryshojnë midis tyre nga pikpamja e kohezgjatjes së prodhimit të produktit. Fakti se metoda e dytë kërkon një kohë më të shkurtër do të na çonte në konkluzionin se metoda e dytë është më efektive se e para.

### \* 18.3 Kriteri Mann-Whitney-Wilcoxon

Në këtë paragraf do të paraqesim një metodë tjetër jo parametrike, që mund të përdoret për të përcaktuar nëse ekziston një ndryshim midis dy popullimeve. Ky kriter në dallim nga kriteri i rangut me shenja nuk bazohet në zgjedhjen shoqëruese, por përdoren dy zgjedhje të rastit të pavarura të marra nga dy popullime.



Ne kapitullin 10 ndertuam nje kriter parametrik per kontrollin e differences midis mesatareve te dy popullimeve. Hipotezat e kontrolluara me anen e ketij krleri jane:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

Ne rastin e zgjedhjes se vogel metoda parametrike bazohej ne dy supozime.

1. Te dy popullimet kane shperndaje normale.
2. Variancat e dy popullimeve jane te barabarta.

Kriteri jo parametrik nuk kerkon plotesimin e asnjërit prej ketyre supozimeve, por qe shkalla e matjes per te dhenat e perfutuara nga dy zgjedhje te rastit te pavarura te jete te pakten nje shkalle e zakonshme. Ne vend te kontrollit te ndryshimit te mesatareve te dy popullimeve, kriteri MWW percakton ne se dy popullimet jane identike. Hipotezat per kriterin MWW jane:

$H_0$ : Dy popullimet jane identike

$H_a$ : Dy popullimet nuk jane identike

Ne fillim do te tregojme se si zbatohet kriteri MWW permes zbatimit ne rastin e zgjedhjes se vogel.

**Rasti i zgjedhjes se vogel** Rasti i zgjedhjes se vogel per kriterin perdoret kur vellimet e zgjedhjeve per te dy popullimet jane me te vogla ose te barabarta me 10. Zbatimin e kriterit MWW do ta shohim neperinjet shembullit te rezultatit te studenteve ne provlmin e matematikes te konkursit te pranimit ne universitet ne varesi nga lloji i shkolles se mesme qe kane kryer. Per te pare ndikimin e shkolles se mesme ne nivelin e pergatitjes te studenteve ne matematike, bejme nga nje zgjedhje te rastit te nje numri te vogel studentesh nga ata qe kane mbaruar shkollen e mesme te pergjithshme dhe nga ata qe kane mbaruar shkollat e tjera te mesme profesionale. Hipotezat qe duhet te kontrollojme shprehen

$H_0$ : Studentet qe vijne nga te dy kategorite e shkollave nuk ndryshojne nga njohurite ne matematike.

$H_a$ : Studentet qe vijne nga te dy kategorite e shkollave ndryshojne nga njohurite ne matematike.

Duke u bazuar ne regjistrat u zgjodhen 4 studente qe kane mbaruar shkollen e mesme te pergjithshme dhe 5 studente qe kane mbaruar shkollat e mesme profesionale. Te dhenat mbi vendin qe keta studente kane zene ne renditjen e pergjithshme te studenteve ne listat e rezultateve te konkursit paraqiten ne tabelen 18.4

Tabela 18.4 Piket e fituar ne konkurs sipas llojit te shkolles

| Shkolla e mesme e pergjithshme |                     | Shkolla e mesme profesionale |                     |
|--------------------------------|---------------------|------------------------------|---------------------|
| Studentet                      | Renditja ne konkurs | Studentet                    | Renditja ne konkurs |
| A1                             | 8                   | B1                           | 70                  |
| A2                             | 52                  | B2                           | 202                 |
| A3                             | 112                 | B3                           | 144                 |
| A4                             | 21                  | B4                           | 175                 |
|                                |                     | B5                           | 146                 |

Etapa e pare ne proceduren MWW eshte renditja e te gjithe studenteve ne baze te vendit qe kane zene ne konkurs, duke filluar nga vlera me vogel deri ne vleren me te madhe. Vlera me vogel 8 merr rangun 1 dhe vlera me madhe renditjes ne konkurs 202 merr rangun 9. Rangu ose renditja e 9 studenteve paraqitet ne tabelen 18.5

Tabela 18.5 Renditja e studenteve sipas vendit qe kane zene ne konkurs

| Studentet | Vendi ne konkurs | Rangu i kombinuar |
|-----------|------------------|-------------------|
| A1        | 8                | 1                 |
| A4        | 21               | 2                 |
| A2        | 52               | 3                 |
| B1        | 70               | 4                 |
| A3        | 112              | 5                 |
| B3        | 144              | 6                 |
| B5        | 146              | 7                 |
| B4        | 175              | 8                 |
| B2        | 202              | 9                 |

Etapa tjeter eshte gjetja e shumes se rangjeve per cdo zgjedhje me vehte. Keto llogaritje tregohen ne tabelen 18.6. Procedura mund te perdori per vleresim shumen e rangjeve per seicilen zgjedhje. Me poshte per kete do te perdorim shumen e rangjeve per te 4 studentet qe



vijne nga shkolla e mesme e pergjithshme. Kete shume e shenojme me simbolin T, ne rastin konkret ajo eshte e barabarte me  $T = 11$ .

Tabela 18.6 Shuma e rangjeve per studentet sipas llojit te shkolles se mesme

| Shkolla e mesme e pergjithshme |                  |       | Shkollat e mesme profesionale |                  |       |
|--------------------------------|------------------|-------|-------------------------------|------------------|-------|
| Studentet                      | Vendi ne konkurs | Rangu | Studentet                     | Vendi ne konkurs | Rangu |
| A1                             | 8                | 1     | B1                            | 70               | 4     |
| A2                             | 52               | 3     | B2                            | 202              | 5     |
| A3                             | 112              | 5     | B3                            | 144              | 6     |
| A4                             | 21               | 2     | B4                            | 175              | 8     |
|                                |                  |       | B5                            | 146              | 7     |
| Totali rangjeve                |                  | 11    |                               |                  | 34    |

Le te shikojme per nje moment vetite e shumes se rangjeve per zgjedhjen nga shkolla e mesme e pergjithshme. Meqenese ne zgjedhje jane 4 studente, atehere ata mund te ishin studentet e renditur ne vendet e para. Ne qofte se do te ndodhte ky rast, atehere shuma e rangjeve te tyre do te ishte  $T = 1+2+3+4 = 10$ , e cila eshte vlere me vogel e mundshme per T. Por keta studente mund te qendronin edhe ne vendet e fundit, si rezultat shuma e rangjeve te tyre do te ishte  $T = 6 + 7 + 8 + 9 = 30$ , e cila perben vleren me te madhe te mundshme qe mund te marri T. Keshtu vlere e T per zgjedhjen do te luhatej midis 10 dhe 30. Veme re se vlere e T afer 10 tregon se studentet qe vijne nga shkolla e mesme e pergjithshme jane me nga fillimi ne renditjen sipas rezultateve te provimit, ose kane nje rang me te larte se sa studentet qe vijne nga shkollat e tjera. Nderkohe vlere e T afer 30 tregon se studentet qe vijne nga shkolla e mesme e pergjithshme kane nje rang me te ulet se ata qe vijne nga shkollat e tjera profesionale. Keshtu ne qofte se dy popullimet e studenteve jane identike, atehere do te prisnim qe vlere e T te jete e aferte me vleren mesatare te dy vlerave te mesiperme  $(10 + 30) / 2 = 20$ .

Vlerat kritike te kriterit T jepen ne tabelen 10 te shtojces B per rastet kur vellimet e zgjedhjeve jane me te vogla ose te barabarta me 10. Ne keto tabela  $n_1$  shpreh vellimin e zgjedhjes qe i korespondon asaj zgjedhje, shuma e rangjeve te te ciles u perdor per kriterin e kontrollit. Vlere kritike e kriterit  $T_p$  (kufiri i poshtem) gjendet direkt ne tebele, kurse vlere e sipërme e tij  $T_s$  llogaritet me anen e formule 18.5.

$$T_s = n_1(n_1 + n_2 + 1) - T_p \quad 18.5$$

As vlere e poshtme dhe as vlere e sipërme e T nuk bie ne zonen kritike. Hipoteza zero mbi identitetin e dy popullimeve hidhet poshte, ne qofte se vlere faktike e T eshte ose me vogel se  $T_p$  ose me e madhe se  $T_s$ .

Per shembullin tone, ne tabelen 10 te shtojces B per nje nivel rendesie  $\alpha = 0.05$  me  $n_1 = 4$  dhe  $n_2 = 5$ , shohim se vlere kritike e poshtme e kriterit MWW eshte  $T_p = 12$ . Vlere e sipërme per statistiken MWW llogaritet  $T_s = 4(4 + 5 + 1) - 12 = 28$ .

Keshtu hipoteza zero hidhet poshte, ne qofte se vlere faktike e statistikes MWW eshte me e vogel se 12, ose me e madhe se 28. Ne rastin e shembullit tone vlere faktike e  $T = 11$  eshte me e vogel se kufiri i poshtem, prandaj hipoteza zero mbi identitetin e dy popullimeve hidhet poshte. Prej kendej rezulton se dy popullimet ndryshojne midis tyre, d.m.th. studentet qe vijne nga shkolla e mesme e pergjithshme jane me te pregatitur ne matematike se studentet qe vijne nga shkollat e mesme profesionale.

**Rasti i zgjedhjes se madhe** Ne rastin kur vellimet e te dy zgjedhjeve jane qe te dyja me te medha se 10, shpemdarda T mund te perafrohet me shperndarjen normale. Kete rast e ilustrojme me te dhenat e llogarive te dy filialeve te nje banke.

Dy zgjedhje te rastit mbi llogarite e dy filialeve te nje banke jane paraqitur ne tabelen 18.7. Shtrohet problemi a mundet qe ne baze te vrojtuimeve te arrijme perfundimin se bilancet e llogarive rrjedhese te dy filialeve te bankes nuk ndryshojne mes tyre?

Tabela 18.7 Bilanei i llogarive per dy filiale te nje banke

| Filiali 1           |                     | Filiali 2 |                     |
|---------------------|---------------------|-----------|---------------------|
| Llogaria e zgjedhur | Bilanci i llogarise | Llogaria  | Bilanci i llogarise |
| 1                   | 1095                | 1         | 885                 |
| 2                   | 955                 | 2         | 850                 |
| 3                   | 1200                | 3         | 915                 |
| 4                   | 1195                | 4         | 950                 |
| 5                   | 925                 | 5         | 800                 |

$n_1 = 4$  &  $n_2 = 5$



|    |      |    |      |
|----|------|----|------|
| 6  | 950  | 6  | 750  |
| 7  | 805  | 7  | 865  |
| 8  | 945  | 8  | 1000 |
| 9  | 875  | 9  | 1050 |
| 10 | 1055 | 10 | 935  |
| 11 | 1025 |    |      |
| 12 | 975  |    |      |

Etapa e pare ne zbatimin e kriterit MWW eshte renditja e te gjithë te dhenave nga vlerat me te vogla deri ne vlerat me te medha te tyre. Duke perdorur bashkesine e 22 te dhenave te vrojtuar te tabelës 18.7, bejme renditjen e tyre ose vendosim rangun perkates. Keshtu vleres me te vogel 750 ( njesia e 6 e zgjedhjes 2) i vihet rangu i barabarte me 1. Ne qofte se gjate vendosjes te rangut gjejmë disa vlera te barabarta, atyre u vihet rangu i barabarte me mesataren e rangjeve korepsondues te radhes. Keshtu p.sh. pas vleres 945 me rang 11 vijne dy vlera te barabarta me 950, ketyre u korespondojne rangjet 12 dhe 13, prandaj rangjet e tyre i marrim te barabarta me mesataren e tyre  $12.5 = (12+13)$ . Me pas vjen vlera 955, te ciles i vihet rangu i radhes 14 etj. Duke vazhduar ne kete menyre arrijme tek vlera 1200 te ciles i korespondon rang 22. Renditja e te dhenave jepet ne tabelen 18.8.

Tabela 18.8 Rangu i kombinuar per te dhenat e dy zgjedhjeve ne filialet e nje banke

| Filiali 1           |         |       | Filiali 2           |         |       |
|---------------------|---------|-------|---------------------|---------|-------|
| Llogaria e zgjedhur | Bilanci | Rangu | Llogaria e zgjedhur | Bilanci | Rangu |
| 1                   | 1095    | 20    | 1                   | 885     | 7     |
| 2                   | 955     | 14    | 2                   | 850     | 4     |
| 3                   | 1200    | 22    | 3                   | 915     | 8     |
| 4                   | 1195    | 21    | 4                   | 950     | 12,5  |
| 5                   | 925     | 9     | 5                   | 800     | 2     |
| 6                   | 950     | 12,5  | 6                   | 750     | 1     |
| 7                   | 805     | 3     | 7                   | 865     | 5     |
| 8                   | 945     | 11    | 8                   | 1000    | 16    |
| 9                   | 875     | 6     | 9                   | 1050    | 18    |
| 10                  | 1055    | 19    | 10                  | 935     | 10    |
| 11                  | 1025    | 17    | Shuma               |         | 83,5  |
| 12                  | 975     | 15    |                     |         |       |
| Shuma               |         | 169,5 |                     |         |       |

Etapa tjeter ne kriterin MWW ka te beje me llogaritjen e shumave te rangjeve per seicilin nga popullimet e dhena. Keto shuma i gjejmë ne tabelen 18.8. Procedura e kontrollit bazohet ne shumet e rangjeve per seicilen nga zgjedhjet. Le te marrim shumen e rangjeve per zgjedhjen nga filiali 1, e cila rezulton e barabarte me  $T = 169.5$ .

Per vellimin e zgjedhjes  $n_1 = 12$  dhe  $n_2 = 10$ , mund te perdorim peraftrimin normal te shperndarjes te zgjedhjes per shumen e rangjeve T. Kjo shperndarje mund te paraqitet.

$$\text{Mesatarja: } \mu_T = (1/2)n_1(n_1 + n_2 + 1) = 19,6$$

$$\text{Shmangla standarte: } \sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = 19,7$$

Forma e shpemdajes eshte normale per  $n_1$  dhe  $n_2 \geq 10$

Per filialin I kemi

$$\mu_T = (1/2)(12)(12+10+1) = 138$$

$$\sigma_T = \sqrt{\frac{(12)(10)(12+10+1)}{12}}$$

Duke vazhduar proceduren e zakonshme te kontrollit llogaritim statistiken z, per te pare ne se vlera e T vjen nga shpemdaja normale. Ne qofte se vlera e T nuk vjen nga kjo shperndade, atehere hipoteza zero hidhet poshte dhe arrijme ne perfundimin se dy popullimet nuk jane identike. Duke llogaritur statistiken e kontrollit kemi

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{169,5 - 138}{15,17} = 2,08$$

Per nivelin e rendesise  $\alpha = 0,05$  dime se per hedhjen poshte te hipotezes zero, vlera e z duhet te jete me e vogel se - 1,96, ose me e madhe se + 1,96. Me qene se vlera e z = 2.08, pra eshte me e



madhe se vlera kritike e saj, hipoteza zero mbi identitetin e dy popullimeve. Si perfundim bilancet e llogarive rrjedhese te dy filialeve nuk jane te njejta.

Ne perfundim japim etapat qe duhet te ndiqen me anen e kriterit MWW per kontrollin e hipotezes, se dy zgjedhje te pavarura jane marre nga dy popullime identike.

1. Renditim vrojtlmet e kombinuara te zgjedhjeve nga vlera me vogel ne vleren me te madhe, ku vlerave te barabarta u vendosim rangun e barabarte me mesataren e rangjeve ne dispozicion te tyre-

2. Ne rastin e zgjedhjes se madhe, kontrolli per diferencat e rendesishme midis dy popullimeve behet duke krahasuar vlerat e vrojtura te T me vlerat e shpemdardjes ne zgjedhje te tij per popullimet identike (ekuacionet 18.6 dhe 18.7). Vlerat e standartizuara te statistikes z do te perdoren per te gjykuar ne se hipoteza zero do te hidhet poshte ose jo. Ne rastin e zgjedhjes se vogel, per gjetjen e vlerave kritike te T ne vend te shpemdardjes normale perdorim vlerat e tabelës 10 te shtojces B.

#### 18.4 Kriteri Kruskal-Wallis

Ne paragrafin e meparshem pame se kriteri mund te perdoret per kontrollin e hipotezes qe dy popullimet jane identike. Ky kriter eshte zgjeruar nga Kruskal-Wallis per tu perdorur edhe ne rastin e tre ose me shume popullimeve. Hipotezat statistikore qe shtrihen per kontroll ne rastin kur numri i popullimeve eshte  $k \geq 3$ , mund te paraqiten ne formen e meposhtme.

Ho: Te gjitha popullimet jane identike dhe Ha: Jo te gjitha popullimet jane identike  
Kriteri Kruskal-Wallis bazohet ne analizen e zgjedhjeve te rastit te pavarura nga k popullime. Ne kapitullin 13 pame analizen e variaces (ANOVA) si nje procedure qe mund te perdoret per kontrollin e barazise se mesatareve per tre ose me shume popullime. Modeli ANOVA mund te zbatohet vetem per te dhenat intervalore ose per ato ne forme raporti, ku popullimet jane te shpemdara normalisht dhe variancat e tyre jane te barabarta.

Kriteri jo parametrik Kruskal-Wallis mund te perdoret per te dhenat e zakonshme, si edhe per te dhenat ne forme intervals apo ato ne forme raporti. Pervet kesaj ky kriter nuk kerkon te plotesohen ato kushte kufizuese qe kerkon zbatimi i analizes se variaces. Prandaj kurdohere qe te dhenat jane marre nga  $k \geq 3$  zgjedhje te rastit te pavarura te matur ne shkallen e zakonshme, ose kurdohere qe supozimet mbi normalitetin dhe barazine e variancave per popullimet nuk plotesohen, ky kriter perben nje procedure statistikore alternative per kontrollin e hipotezes qe popullimet jane identike. Me poshte paraqesim kriterin Kruskal-Wallis per kontrollin e hipotezes per identitetin e disa popullimeve permes nje shembulli.

Le ta zeme se kerkohet te kontrollohet qe numri i pikeve te fituara nga studentet ne provimin e konkursit te matematikes varet nga rezultati mesatar i arritur ne shkollen e mesme. Per kete shqyrtojme teresine e studenteve qe kane noten mesatare nga shkolla e mesme deri ne 7, teresine e studenteve qe kane noten mesatare nga 7 deri ne 9 dhe teresine e studenteve qe vijne nga shkolla e mesme me rezultat mesatar 9 deri ne 10. Studentet e zgjedhur nga seicila teresi me numrin perkates te pikeve te fituara ne konk-urs paraqiten ne tabelen 18.9.

Tabela 18.9 Piket e fituar ne konkurs nga studentet e kategorive te meposhtme

| Rezultati ne shkollen e mesme |       |           |
|-------------------------------|-------|-----------|
| 9 - 10                        | 7 - 9 | Deri ne 7 |
| 25                            | 50    | 60        |
| 70                            | 70    | 20        |
| 60                            | 60    | 30        |
| 85                            | 80    | 15        |
| 95                            | 90    | 40        |
| 90                            | 70    | 35        |
| 80                            | 75    |           |

Supozojme se jemi te interesuar te kontrollojme hipotezen se te tre popullimet jane identike lidhur me piket e fituara ne konkurs. Kriteri Kruskal-Wallis, i cili bazohet ne shumen e rangjeve per seicilen nga zgjedhjet paraqitet ne formen e meposhtme.

$$W = \frac{12}{n_T(n_T + 1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n_T + 1)$$

(4)

ku  $k$  = numri i popullimeve

$n_i$  = numri i njësive të zgjedhjes  $i$

$n_i = \sum n_i$  = numri total i njësive në të gjitha zgjedhjet

$R_i$  = shuma e rangjeve për zgjedhjen  $i$

Kruskal dhe Wallis kanë treguar se në kushtet e hipotezës zero, d.m.th. që popullimet janë identike, shpërndarja e zgjedhjes për  $W$  mund të përafrohet me anën e shpërndarjes hi-katror ( $\chi^2$ ) me  $k-1$  shkallë lirie. Ky përafrim në përgjithësi është i pranueshëm, në qoftë se vëllimet e zgjedhjeve janë me të mëdha ose të barabarta se 5.

Për llogaritjen e statistikës  $W$  për shembullin tonë, në fillim duhet të renditim të 20 të dhënat e tabelës 19-9. Vlera më e vogël 15 nga popullimi i studenteve me mesatare deri në 7 marrë rangun 1, ndërsa vlerën më të madhe 95 nga popullimi i studenteve me mesatare mbi 9 i jepet rangun 20. Vlerat e të dhënave me rangjet shoqëruar të tyre, si edhe shumën e rangjeve për të tre popullimet paraqiten në tabelën 18. 10. Vëmë re se ashtu si në rastet e mëparshme, vlerave të barabarta të të dhënave u vendosen rangje të barabarta sa mesatarja e rangjeve që janë në dispozicion të tyre. Të tilla janë për shembull vlerat e të dhënave 60, 70, 80 dhe 90. Vëllimet e zgjedhjeve janë:  $n_1 = 7$ ;  $n_2 = 7$ ;  $n_3 = 6$  dhe shuma e tyre është  $n_T = 7+7+6=20$ . Duke përdorur formulën 18.8 gjejmë statistikën  $W$  si më poshtë.

$$W = \frac{12}{20(21)} \left[ \frac{(95)^2}{7} + \frac{(88)^2}{7} + \frac{(27)^2}{6} \right] - 3(20+1) = 8,92$$

Tabela e shpërndarjes hi-katror (tabela 3 e shtojcës B) tregon që për  $k-1 = 2$  shkallë lirie dhe për nivelin e rëndësise 0,05, vlera kritike e kriterit hi-katror është  $\chi^2 = 5,99147$ . Meqenëse statistika  $W = 8,92$  është më e madhe se vlera që ndan kufirin e zonës për hi-katrorin, hipoteza zero hidhet poshtë dhe arrijmë në përfundimin se të tre popullimet nuk janë identike, d.m.th. se studentët që janë me përparim të lartë në shkollën e mesme arrijnë edhe rezultate të larta në provimin e konkursit të matematikës.

Tabela 18.10 Renditja e 20 studenteve

| 9 - 10       | Rangu     | 7 - 9 | Rangu     | Deri në 7 | Rangu     |
|--------------|-----------|-------|-----------|-----------|-----------|
| 25           | 3         | 50    | 7         | 60        | 9         |
| 70           | 12        | 70    | 12        | 20        | 2         |
| 60           | 9         | 60    | 9         | 30        | 4         |
| 85           | 17        | 80    | 15,5      | 15        | 1         |
| 95           | 20        | 90    | 18,5      | 40        | 6         |
| 90           | 18,5      | 70    | 12        | 35        | 5         |
| 80           | 15,5      | 75    | 14        |           |           |
| <b>Shuma</b> | <b>95</b> |       | <b>88</b> |           | <b>27</b> |



# TESTIMI I HIPOTEVAVE

## PER PAVARESINE

### 12.2 Kriteri i pavaresise: Tabela e perputhjes se rastit

Nje zbatim tjetër i rendesishem i shpërndarjes hi-katror është kontrolli i pavaresise te dy variablave te bazuar ne te dhenat e zgjedhjes. Per ilustrimin e kriterit te pavaresise shikojme studimin e preferencave te meshkujve dhe femrave per llojet e ndryshme te birres: birre e lehte, birre e zakonshme dhe birre e zeze. Ne qofte se preference per birren nuk varet nga seksi, nje firme tregtare do te bente nje porosi te pergjithshme per te gjitha llojet e birrave. Nga ana tjetër ne qofte se preferenca per birre do te varej nga seksi, atehere per tregje te ndryshme firma do te bente porosi te diferencuara. Problemi qe kerkohet te kontrollohet me anen e kriterit te pavaresise është ne se preferenca per birre e konsumatorit varet ose jo nga seksi i personit.

Situata e shembullit te mesiperme paraqitet ne tabelen 12.1. Pas ndarjes te popullimit ne grupe meshkuj dhe femra, kryhet nje zgjedhje e rastit dhe cdo individ pyetet per preferencat qe ai ka per llojet e birres. Cdo person mund te klasifikohet ne nje nga 6 kuadratet e tabelës. P.sh. nje person mund te jete mashkull qe preferon nje birre te zakonshme (kuadrati 2), nje femer qe preferon birre te lehte (kuadrati 4), femer qe preferon birre te zeze (kuadrati 6) etj. Meqenese kemi regjistruar te gjitha kombinimet e mundshme te preferencave per birre sipas seksit, ose me fjale te tjera te gjitha perputhjet e mundeshme, kjo table quhet table e perputhjes se rastit.

Tabela 12.1 Tabela e perputhjes per preferencen per birre sipas seksit

| Seksi   | Preferenca per birren |             |            |
|---------|-----------------------|-------------|------------|
|         | E lehte               | E zakonshme | E zeze     |
| Meshkuj | kuadrati 1            | kuadrati 2  | kuadrati 3 |
| Femra   | kuadrati 4            | kuadrati 5  | kuadrati 6 |

Le te supozojme tani se kemi marre nje zgjedhje te rastit prej 150 konsumatoresh. Pasi seicili prej tyre ka shijuar cdo lloj birre ne nje renditje te rastit, personat e zgjedhur u pyeten se cilen birre preferojne me teper. Pergjigjet e tyre jane permblodhur ne tabelen e perputhjes 12.2. Siç shihet ne table te dhenat e grumbulluara jane paraqitur ne termat e dendurive per cdo preference dhe cdo lloj seksi. P.sh. nga 150 persona te pyetur 80 jane meshkuj, nga te cilet 20 preferojne birre te lehte, 40 birre te zakonshme dhe 20 prej tyre preferojne birre te zeze:

Tabela 12.2 Rezultatet e zgjedhjes per preferencat per birre.

|       |         | Preferenca per birre |             |        | Totali |
|-------|---------|----------------------|-------------|--------|--------|
|       |         | E lehte              | E zakonshme | E zeze |        |
| Seksi | Meshkuj | 20                   | 40          | 20     | 80     |
|       | Femra   | 30                   | 30          | 10     | 70     |
|       | Totali  | 50                   | 70          | 30     | 150    |

Veme re se te dhenat e zgjedhjes permbajne denduri per seicilen nga 6 klasat ose kategorite. Ne qofte se percaktojme dendurite e pritura me supozimin e pavaresise te preferencave per birren lidhur me seksin, atehere per te percaktuar ne se ekziston ose jo ndryshim midis dendurive te vrojtuar dhe atyre te pritura mund te perdorim shperndarjen hi-katror.



Dendurite e pritura per klasat e tabelës te përpunimit bazohen në arsyetimin e mëposhtem. Në fillim supozojmë se hipoteza zero midis preferencave për birrë dhe seksit është e vërtetë. Pastaj shohim se të 150 konsumatorët janë shpërndarë sipas preferencave në këto mënyra: 50 preferojnë birrë të lehtë, 70 nga ata preferojnë birrë të zakonshme dhe 30 prej tyre preferojnë birrë të zeze. Të shprehura në terma të pjesës ndaj të tërës këto preferenca janë përkatesisht si vijon:  $50/150=1/3$ ;  $70/150=7/15$  dhe  $30/150=1/5$ .

Në qoftë se hipoteza mbi pavarësinë është e vërtetë, atëherë këto preferenca të shprehura në pjesë ndaj të tërës pritet të jenë të njëjta për të dy seksin. Kështu me supozimin e pavarësisë të preferencës nga seksi, mund të presim se 80 konsumatorët meshkuj janë shpërndarë sipas preferencave si më poshtë:  $1/3(80)=26,67$  preferojnë birrë të lehtë;  $7/15(80)=37,33$  kanë preferencë për birrë të zakonshme; dhe  $1/5(80)=16$  prej tyre për birrë të zeze. Duke përdorur të njëjtat pjesë për preferencat për birrë si për meshkujt edhe për femrat, mund të gjejmë denduritë e pritura përkatese si në tabelë.

Shenojmë me  $p_{ij}$  dendurinë e pritur për kategorinë në rrjeshtin  $i$  dhe shtyllën  $j$  të tabelës së përpunimit. Me këto shënime le të rishikojmë edhe një herë llogaritjen e dendurisë teorike  $P_{12}$  në rrjeshtin  $i=1$ , dhe shtyllën  $j=2$ , d.m.th. për konsumatorët meshkuj që preferojnë birrë të zeze.

Me lart pame se llogaritja e  $P_{12}$  u bë si vijon:  $P_{12} = (7/15)80 = 37,33$ . Këto llogaritje mund t'i paraqesim edhe në një mënyrë tjetër, konkretisht:  $P_{12} = (7/15)80 = (70/150)80 = (80)(70) / 150$ , d.m.th. denduria e pritur në rrjeshtin 1 dhe shtyllën 2 mund të paraqitet si produkt i dendurisë totale të rrjeshtit 1 me dendurinë totale të shtyllës 2 e pjesëtuar me vëllimin e zgjedhjes. Në përgjithësi denduria e pritur për klasën që ndodhet në rrjeshtin  $i$  dhe shtyllën  $j$  është e barabartë me produktin e dendurisë totale në rrjeshtin  $i$  me dendurinë totale në shtyllën  $j$  e pjesëtuar me vëllimin e zgjedhjes. Kështu denduria e pritur në rrjeshtin 2 dhe shtyllën 3 është  $e_{23}=(80)(30)/150=16,00$ . Këto denduri paraqiten në tabelën e mëposhteme.

**Tabela 12.3 Dendurite e pritura për preferencat për birrë**

|       |         | Prerencat për birrë |             |        | Totali |
|-------|---------|---------------------|-------------|--------|--------|
|       |         | E lehtë             | E zakonshme | E zeze |        |
| Seksi | Meshkuj | 26,67               | 37,33       | 16,00  | 80     |
|       | Femra   | 23,33               | 32,67       | 14,00  | 70     |
|       | Totali  | 50                  | 70          | 30     | 150    |

Procedura e kontrollit për krahasimin e dendurive të vërtetuara të tabelës 12.2 me denduritë e pritura të tabelës 12.3, bëhet në mënyrë analoge si në rastin e mëparshëm. Në mënyrë të përgjithshme vlera e  $\chi^2$  e bazuar në denduritë e vërtetuara dhe ato të pritura mund të llogaritet si më poshtë:

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - p_{ij})^2}{p_{ij}} \quad 12.3$$

ku  $f_{ij}$  = denduria e vërtetuar në rrjeshtin  $i$  dhe shtyllën  $j$  të tabelës së kontigjencës,  $p_{ij}$  = denduria e pritur në rrjeshtin  $i$  dhe shtyllën  $j$  të tabelës së përpunimit me supozim të bërë për pavarësinë.

Për tabelën e përpunimit me  $n$  rrjeshta dhe  $m$  shtylla, si dhe me kushtin që në çdo kategori denduritë janë me të njëjtën ose të barabartë me 5, statistika e kontrollit ka shpërndarje hi-katror me  $(n-1)(m-1)$  shkallë lirie.

Duke riparë denduritë e pritura të tabelës 12.3 shohim se ato janë me të njëjtën ose të barabartë me 5. Kështu mund të vazhdojmë të llogaritim vlerën e vërtetuar të kriterit:

$$\chi^2 = \frac{(20 - 26,67)^2}{26,67} + \frac{(40 - 37,33)^2}{37,33} + \dots + \frac{(10 - 14,00)^2}{14,00} = 1,67 + 0,19 + \dots + 1,14 = 6,13$$

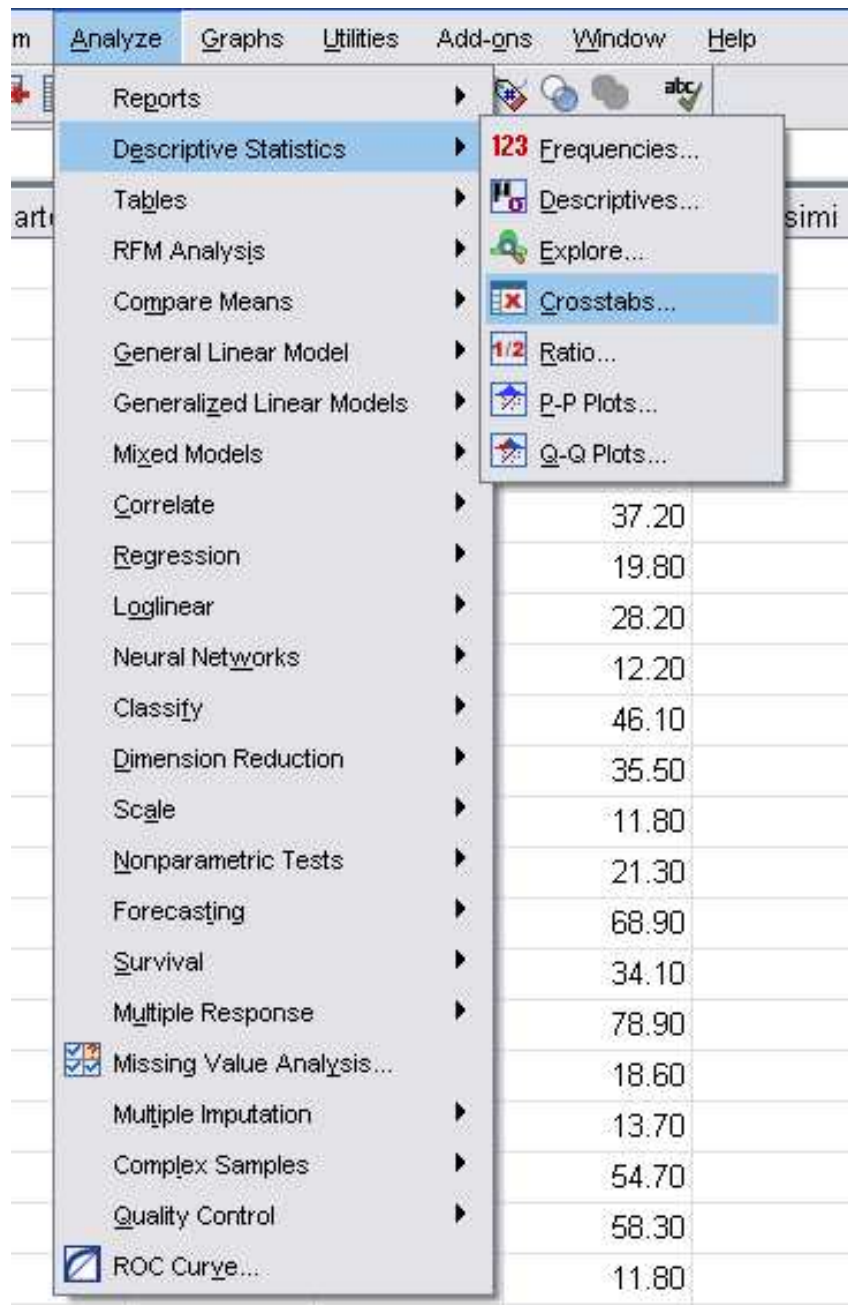


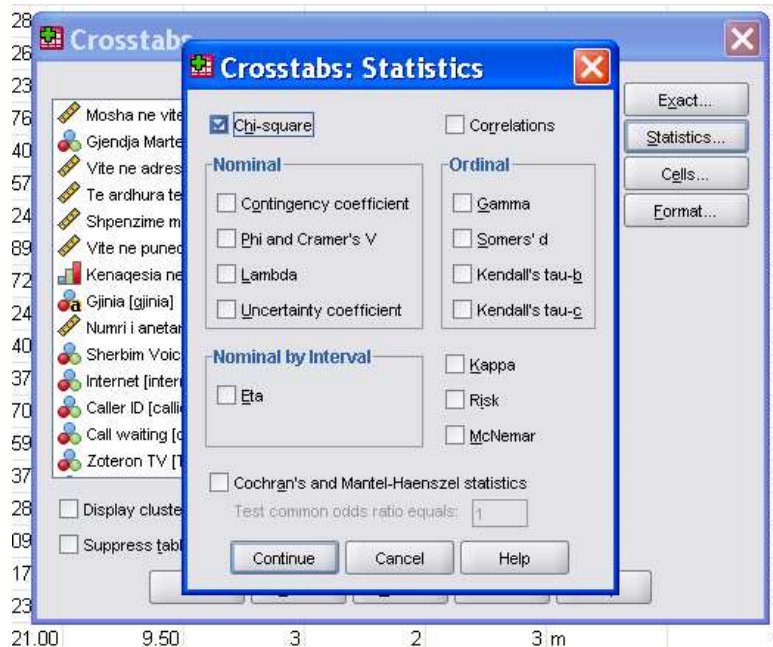
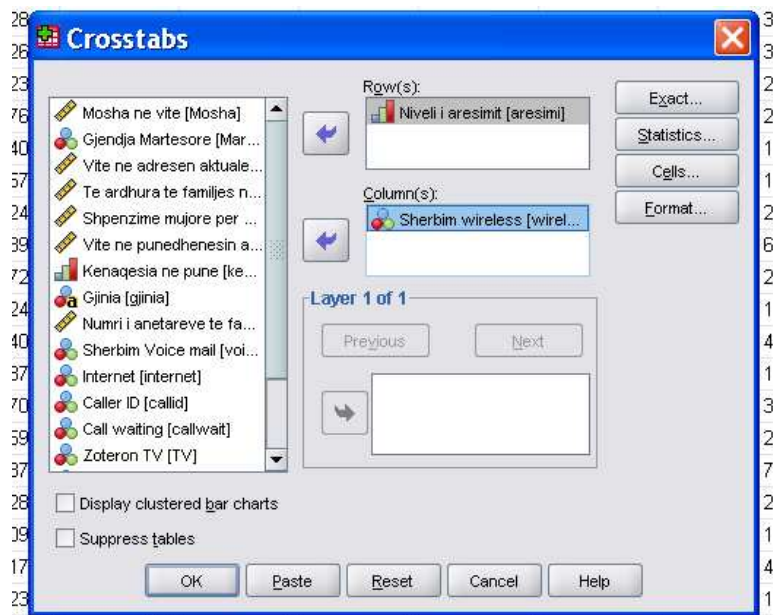
Numri i shkalleve te lirise per shperndarjen hi-katror eshte  $(2-1)(3-1) = 2$ . Ne baze te numrit te shkalleve te lirise 2 dhe nivelit te rendesise  $\alpha = 0,05$  ne tabelen e shperndarjes hi-katror gjejm vleren kritike te saj te barabarte me  $\chi^2 = 5,59$ . Shohim se perseri jemi duke perdorur zonen kritike te djathte, sepse hipoteza zero do te hidhet poshte ne qofte se ndryshimet midis dendurive te vrojtuar dhe teorike japin nje vlere te madhe per  $\chi^2$ . Ne shembullin tone vlera faktike e  $\chi^2 = 6,13$  eshte me e madhe se vlera kritike  $\chi^2 = 5,59$ , prandaj hipoteza zero mbi pavaresine hidhet poshte duke konkluduar se preferenca per birre nuk eshte e pavarur nga gjinia e konsumatorit.

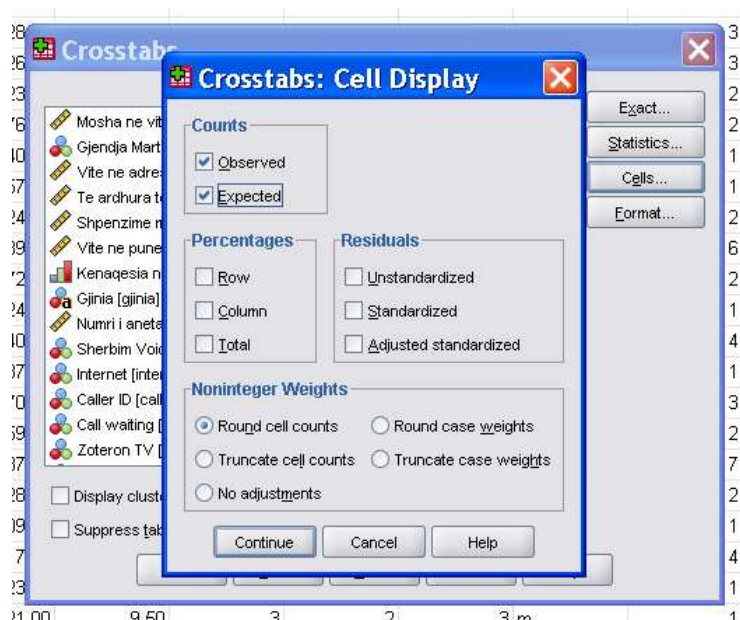
Megjithese kriteri i pavaresise lejon vetem nxjerrjen e konkluzionit te mesiperme, perseri mund te krahasojme dendurite perkatese te pritura dhe te vrojtuar per te marre nje ide me te plote se si eshte varesia midis preferences per birre dhe seksit te konsumatorit. Duke iu referuar tabelave 12.2 dhe 12.3 shohim se konsumatoret meshkuj kane denduri te vrojtuar me te medha se dendurite e pritura per llojet e birres se zakonshme dhe te zeze, ndersa konsumatoret femra kane denduri faktike me te medha se ato te pritura per llojin e birres se lehte. Keto krahasime na japin nje paraqitje me te plote per ndryshimet ne preferencat per birre te meshujve dhe femrave.

## Tema 11.

### Aplikimi ne SPSS I : Testi i hipotezave për pavarësi të variablave kategorikë.







**Niveli i arësimit \* Shërbim wireless Crosstabulation**

|                   |                 |                | Shërbim wireless |       | Total  |
|-------------------|-----------------|----------------|------------------|-------|--------|
|                   |                 |                | Jo               | Po    |        |
| Niveli i arësimit | Tetëvjecare     | Count          | 336              | 91    | 427    |
|                   |                 | Expected Count | 259.6            | 167.4 | 427.0  |
|                   | Shkollë e mesme | Count          | 392              | 164   | 556    |
|                   |                 | Expected Count | 338.0            | 218.0 | 556.0  |
|                   | Kolegj          | Count          | 218              | 192   | 410    |
|                   |                 | Expected Count | 249.3            | 160.7 | 410.0  |
|                   | Universitet     | Count          | 175              | 203   | 378    |
|                   |                 | Expected Count | 229.8            | 148.2 | 378.0  |
|                   | Pasuniversitar  | Count          | 95               | 134   | 229    |
|                   |                 | Expected Count | 139.2            | 89.8  | 229.0  |
|                   | Total           | Count          | 1216             | 784   | 2000   |
|                   |                 | Expected Count | 1216.0           | 784.0 | 2000.0 |

### Chi-Square Tests

|                              | Value                | df | Asymp. Sig. (2-sided) |
|------------------------------|----------------------|----|-----------------------|
| Pearson Chi-Square           | 158.519 <sup>a</sup> | 4  | .000                  |
| Likelihood Ratio             | 162.189              | 4  | <b>.000</b>           |
| Linear-by-Linear Association | 151.402              | 1  | .000                  |
| N of Valid Cases             | 2000                 |    |                       |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 89.77.

Ka varësi ndërmjet nivelit të arësimimit dhe aksesit kundrejt shërbimit wireles



## TEMA 11:

### TESTIMI I HIPOTEZAVE: METODAT JO PARAMETRIKE – TESTET 'X2'

#### Ushtrime

**Ushtrim 1:** Për një zgjedhje të rastit prej gjashtë vrojtimesh nga një popullimi i vijueshëm kemi të dhënat e mëposhtme :

18.2 21.3 20.5 19.4 19.6 17.7

Tregoni nëse mediana e popullimit është e ndryshme nga 20. Përdorni nivelin e rëndësisë 0.05

**Ushtrim 2:** 4. Nga një zgjedhje rasti për paciente që marrin vetëm shërbim ambulator dhe për ata që marrin shërbim spitalor, rezultoi se moshat e 10 pacientëve për secilin grup ishte:

|                                         |    |    |    |    |    |    |    |    |    |    |
|-----------------------------------------|----|----|----|----|----|----|----|----|----|----|
| Mosha e pacientëve me shërbim ambulator | 32 | 48 | 35 | 63 | 79 | 30 | 54 | 31 | 23 | 41 |
| Mosha e pacientëve me shërbim spitalor  | 58 | 54 | 83 | 67 | 40 | 51 | 42 | 29 | 45 | 55 |

Testoni nëse diferenca në moshë midis pacientëve me shërbim ambulator dhe atyre me shërbim spitalor është statistikisht e rëndësishme (përdorni kriterin Mann-Whitney me nivel rëndësie 0.05)

**Ushtrimi 3:** Planifikimi i një menuje nga kuzhina e spitalit përfshin dhe pyetjet e pacientëve mbi preferencat për dy lloje ushqimesh. Në 250 paciente të zgjedhur rastësisht u vu re se 140 ishin për një lloj ushqimi dhe 110 për ushqimin tjetër. Duke përdorur nivelin e rëndësisë 0.05 tregoni a ka diferenca në preferenca përsa i përket menuse ?

**Ushtrimi 4:** Në një qendër spitalore intervistohen drejtuesit më të lartë të financës dhe të shërbimit mjekësor për vendin e manaxherit të përgjithshëm. U zgjedhën rastësisht 10 mjekë dhe 14 financierë të cilët u ekspozuan ndaj intervistave intensive dhe testimeve nga një grup manaxherësh të qendrës. Kandidatët u vlerësuan nga një shkallë nga 1 (më i përshtatshëm për punësim) deri në 24, vlerësimet janë dhënë në tabelën e mëposhtme. Testoni hipotezën zero, që të dy profesionet janë njëlojtë të preferueshme ( $\alpha = 0,05$ )

|            |             |             |
|------------|-------------|-------------|
| 1. Financë | 9. Mjek     | 17. Mjek    |
| 2. Financë | 10. Mjek    | 18. Mjek    |
| 3. Mjek    | 11. Financë | 19. Financë |
| 4. Financë | 12. Financë | 20. Financë |
| 5. Financë | 13. Mjek    | 21. Financë |
| 6. Mjek    | 14. Financë | 22. Mjek    |
| 7. Financë | 15. Financë | 23. Mjek    |
| 8. Mjek    | 16. Financë | 24. Financë |

**Ushtrimi 5 :** Tabela e mëposhtëme tregon numrin e femrave dhe të meshkujve që shikojnë më shumë ose më pak se 2 1/2 orë TV në ditë. Si për femrat dhe meshkujt kemi zgjedhje të rastit të pavarura nga njëra tjetra. Testoni në nivelin 10% hipotezën zero që s'ka asnjë lidhje midis gjinise dhe numrit të orëve që ndiqet televizioni.

|         | Numri i orëve në ditë |                    |
|---------|-----------------------|--------------------|
|         | Më pak se 2 1/2       | 2 1/2 ose më shumë |
| Meshkuj | 18                    | 10                 |
| Femra   | 17                    | 13                 |

**Tema 11:**  
**TESTIMI I HIPOTEZAVE: METODAT JO PARAMETRIKE**  
**– TESTET 'X2'**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

1

---

---

---

---

---

---

---

---

**Çështjet Kryesore qe Trajtohen:**

- A. Karakteristika të testeve Joparametrike**
- B. Kushti: për shpërndarjen Normale të Variablit**
- C. Testimi i hipotezave për mesataren e një popullimi (testimi për medianën)**
- D. Testimi i hipotezave për mesataren e dy popullimeve: të pavarura: Kriteri Mann Whitney-Wilcoxon/ të varura: Kriteri i Shenjave të Rangut Wilcoxon**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

2

---

---

---

---

---

---

---

---

**E. Testimi i hipotezave për mesataren e më shumë se dy popullimeve (Kruskal Wallis)**

**F. Testi i hipotezave për pavarësi të variablave kategorikë.**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

3

---

---

---

---

---

---

---

---

**A: Karakteristika të testeve Joparametrike**

1. Nuk kërkon supozimin për shpërndarje normale të variablilit.

2. Aplikohet edhe për variabla në shkallë matje Ordinale.

**B. Kushti: për shpërndarjen Normale të Variablilit**

1. Ndërtimi i Histogramës (simetria dhe kulmi i saj- forma e këmbanës)

2. Grafiku i shpërndarjes normale (P-P ose Q-Q plot)- duhet të tregojë lidhje lineare

3. Kolmogorov-Smirnov Test ( $n > 50$ ) ose Shapiro-Wilk Test ( $n < 50$ ) (Sig "p"  $> 0.05$  tregon Shpërndaje Normale)

**C. Testimi i hipotezave për mesataren e një popullimi (testimi për medianën)**

$$H_0: Me = Me_0$$

$$H_1: Me \neq Me_0$$

$$H_0 \downarrow \text{nqse } |z_v| > z_{\alpha/2} \text{ ku } z_v = \frac{x - \mu}{\sigma}$$

$$\sigma = \sqrt{0.25 * n}$$

$$\mu = 0.5 * n$$

**D1. Testimi i hipotezave për mesataren e dy popullimeve: (të pavarura): Kriteri Mann Whitney-Wilcoxon)**

$$H_1: \mu_1 = \mu_2$$

$$H_0: \mu_1 \neq \mu_2$$

$$H_0 \downarrow \text{nqse } |z_v| > z_{\alpha/2} \quad T = \text{Shuma e rangjeve për zgjedhjet}(n_1)$$

Rangu: është numër radhitjeje që fillon me 1 për vlerën më të vogël e rritet, (2, 3 ...) me rritjen e vlerave. Për vlera të barabarta vendosen rangje mesatare (rangjet llogariten për të gjitha vlerat  $n_1 + n_2$ )

$$z_v = \frac{T - \mu_T}{\sigma_T} \quad \mu_T = \frac{n_1}{2} (n_1 + n_2 + 1) \quad \sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

## D2. Testimi i hipotezave për mesataren e dy popullimeve: (të varura): Kriteri i Shenjave të Rangut-Wilcoxon)

H1:  $\mu_1 = \mu_2$

H0:  $\mu_1 \neq \mu_2$

H0↓ nqse  $|z_v| > z_{\alpha/2}$  ku për çdo çift gjendet diferenca, pastaj gjendet vlera absolute e diferencave, më pas i vendosen rangje vlerave absolute; dhe në fund rangjeve i vendoset vetë shenja e diferencës duke konkluduar në fund me gjetjen e T që është shuma e rangjeve me shenjë.

n është nr i elementëve, për të cilët diferenca e vlerave është e ndryshme nga 0

$$Z_V = \frac{T - \mu_T}{\sigma_T} \quad \mu_T = 0 \quad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

## E. 1. Testimi i hipotezave për mesataren e më shumë se dy popullimeve (Kruskal Wallis)

H0:  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$

H1: Të paktën një nga  $\mu$ -te e  $H_0$  e ndryshme nga të tjerat

H0↓ nqse  $W > \chi_{(\alpha, k-1)}^2$

Ku: k = nr. i popullimeve;  $n_i$  = nr. i njësive në zgjedhjen "i";  $n_T$  = nr. total i elementëve;  $T_i$  = Shuma e rangjeve për zgjedhjen "i" ku rangjet fillojnë me 1 për vlerën më të vogël e rritet (2, 3 ...) me rritjen e vlerave. Për vlera të barabarta vendosen rangje mesatare (rangjet marrin parasysh  $n_T$  elementë)

$$W = \frac{12}{n_T(n_T + 1)} \sum_{i=1}^k \frac{T_i^2}{n_i} - 3(n_T + 1)$$

## F. Testi i hipotezave për pavarësi të variablave kategorikë.

H0: Leximi nuk varet nga niveli i arësimimit

H1: Leximi varet nga niveli i arësimimit

H0↓ nqse  $\chi_v > \chi_{(\alpha, (i-1)*(j-1))}^2$

$f_{ij}$  = dendruritë e vrojtuar në cross tab

$p_{ij}$  = denduritë e pritura në kushtet ku  $H_0$  është e vërtetë (kushti  $p_{ij} \geq 5$ )

| i/j            | Lexues | Jollexues |
|----------------|--------|-----------|
| Arësim I lartë | ...    | ...       |
| Arësim I mesëm | ...    | ...       |
| Arësim I ulët  | ...    | ...       |

$$\chi_v^2 = \sum_i \sum_j \frac{(f_{ij} - p_{ij})^2}{p_{ij}}$$



## **TEMA 12:**

### **Analiza e Regresionit mbështetur ne OLS (Ordinary Least square)**

#### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje që :

- Të kuptojnë kur dhe si përdoret metoda e katrorëve më të vegjël
- Të shpjegojnë modelin e thjeshtë të regresionit linear
- Të ndërtojnë dhe interpretojnë ekuacionin e thjeshtë linear për një set të dhënash
- Të përshkruajnë  $R^2$  si tregues të fuqisë shpjeguese të modelit të regresionit
- Të kuptojnë supozimet mbi të cilat mbështetet analiza e regresionit
- Të shpjegojnë masën e variacionit dhe të përcaktojnë nëse variabli I pavarur është statistikisht I rëndësishëm
- Të përdorin ekuacionin e regresionit për parashikime
- Të aplikojnë analizën e regresionit të shumfishtë në situata vendimmarrjeje.
- Të analizojnë dhe të interpretojnë outputin në SPSS të modeleve të regresionit të ndërtuara
- Të realizojnë testin e hipotezave për të gjithë koeficientët e regresionit të shumfshishtë linear dhe për secilin në mënyrë individuale.

#### **LITERATURA**

- “ Applied Statistics”, Warner A.R , 2008
- Statistika”- Grup Autoresh, Shtëpia botuese Pegi, (2016)
- Statistical for Business nad Economics”, Newbold P, Carlson W, Thorne B, 2011
- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011

-“Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition,  
Belinda Barton, Jennifer Peat, September 2014

## 15.1 Modeli i regresionit te shumefishte dhe supozimet per te

Modeli probabilitar per analizen e regresionit te shumefishte eshte nje zgjerim i drejtperdrejte i modelit te regresionit te thjeshte me nje variabel, per te cilin kemi folur me lart. Per te treguar kete analizojmë situaten qe ka te beje me shitjet e nje produkti te ri ne nje zone te caktuar. Le ta zeme se studiohen shitjet ( $y$ ) ne varesi te numrit te popullsisë ( $x_1$ ) dhe te te ardhurave te banoreve te asaj zone ( $x_2$ ) me anen e ekuacionit te regresionit linear si me poshte:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad 15.1$$

Varesia e treguar ne 15.1 eshte nje model i regresionit te shumefishte linear qe perfshin dy variabla te pavarur. Shohim se kur koeficienti  $\beta_2 = 0$ ,  $x_2$  nuk lidhet me  $y$  dhe modeli i regresionit te shumefishte kthehet ne nje model me nje variabel te pavarur, te cilin e kemi pare ne kapitullin 14. Modeli i regresionit te shumefishte mund kete me teper se dy variabla te pavarur. Ne forme te pergjithshme nje model i tille ka formen:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad 15.2$$

Supozimet e bera ne kapitullin 14 per termin e gabimit ( $\varepsilon$ ), gjithashtu kane vend edhe ne modelin e regresionit shumefaktorial.

1. Gabimi  $\varepsilon$  eshte nje variabel i rastit me mesatare, ose vlere te pritur te barabarte me zero, d. m.th.  $E(\varepsilon)=0$ . Rrjedhim: Per vlerat e dhena te variablave te pavarur  $x_1, x_2, \dots, x_p$  vlera e pritur, ose mesatarja e vlerave te  $y$  shprehet me anen e ekuacionit:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad 15.3$$

Ekuacioni 15.3 quhet ekuacioni i regresionit te shumefishte. Ne kete ekuacion  $E(y)$  paraqet mesataren e te gjithe vlerave te mundshme te  $y$  qe mund te shfaqen per vlerat e dhena te variablave,  $x_1, x_2, \dots, x_p$ .

2. Varianca e  $\varepsilon$  shenohet me  $\sigma^2$  dhe eshte e njejte per te gjitha vlerat e variablave te pavarur.

Rrjedhim. Varianca e  $y$  eshte gjithashtu e barabarte me  $\sigma^2$  dhe eshte e njejte per cdo bashkesi vlerash te variablave te pavarur.

3. Vlerat e  $\varepsilon$  jane te pavarura

Rrjedhim. Madhesia e gabimit per nje bashkesi vlerash te variablave te pavarur nuk lidhet me madhesine e gabimit per cdo bashkesi tjeter vlerash te variablave te pavarur.

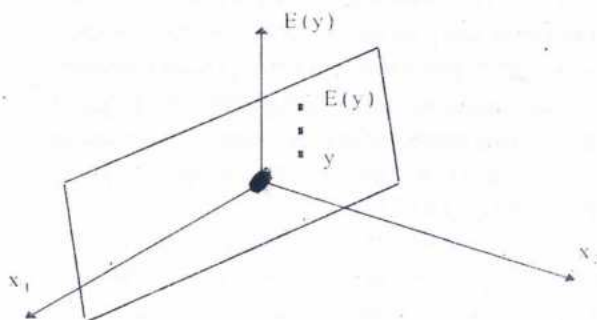
4. Gabimi  $\varepsilon$ , eshte nje variabel i rastit i shperndare normalisht, ai shpreh devijimet e vlerave te vrojtuar te  $y$  nga vlerat e pritura te tij te llogaritura me anen e ekuacionit te regresionit.

Rrjedhim. Me qene se  $\beta_0, \beta_1, \dots, \beta_p$  jane madhesi konstante, atehere per cdo bashkesi vlerash te variablave te pavarur, variabli i varur  $y$  eshte gjithashtu nje madhesi rasti e shpemdare normalisht. Per te patur nje paraqitje me te qarte mbi formen e lidhjes me anen e ekuacionit te regresionit, marrim rastin kur kemi vetem dy variabla te pavarur. Ne kete rast ekuacioni i regresionit mund te paraqitet:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad 15.4$$

Grafiku i ketij ekuacioni eshte nje plan ne hapsiren me tre dimensione. Figura 15.1 paraqet grafikisht kete ekuacion ku  $x_1$  dhe  $x_2$  vendosen ne boshtet horizontale, kurse  $y$  ne boshtin vertikal. Veme re se  $\varepsilon$  shprehet si difference midis vlerave faktike te  $y$  dhe vlerave te pritura te tij  $E(y)$ , per  $x = x_1$  dhe  $x = x_2$ .

Figura 15.1 Grafiku i ekuacionit te regresionit me dy variabla te pavarur





Ne kapitullin e mëparshëm pame se për vlerësimin e koeficientëve të ekuacionit të regresionit të thjeshtë linear përdorej metoda e katrorëve me të vegjël. Në mënyrë analoge kjo metodë mund të përdoret edhe për vlerësimin e parametrave në analizen e regresionit të shumëfishtë  $\beta_0, \beta_1, \dots, \beta_p$ . Vlerësimet e këtyre parametrave shënohen përkatesisht me  $b_0, b_1, \dots, b_p$ , kurse ekuacioni përkates i regresionit paraqitet:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad 15.5$$

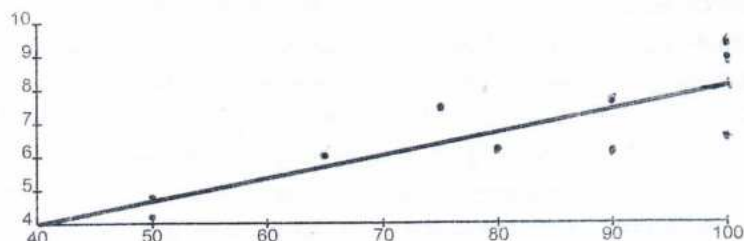
### 15.2 Gjetja e vlerësimit për ekuacionin e regresionit të shumëfishtë

Për të parë se si mund të gjendet vlerësimi për ekuacionin e regresionit me disa variabla le të marrim një shembull për një firmë transporti, për të cilën kërkojmë të gjejmë lidhjen midis kohës ditore të udhëtimit  $y$  në varësi të numrit të kilometrave të përshkuara  $x_1$  dhe numrit të dërgesave tek konsumatorët  $x_2$ . Ekuacionin e gjetur mund ta përdorim për parashikimin e kohës ditore të udhëtimit për shoferet e firmës, në varësi të kilometrave të përshkuara dhe numrit të dërgesave tek klientet.

Tabela 15.1 Të dhënat fillestare për problemin e transportit

| Dita | Kilometrat e përshkuara | Koha e udhëtimit (ore) |
|------|-------------------------|------------------------|
|      | $x_1$                   | $y$                    |
| 1    | 100                     | 9.3                    |
| 2    | 50                      | 4.8                    |
| 3    | 100                     | 8.9                    |
| 4    | 100                     | 6.5                    |
| 5    | 50                      | 4.2                    |
| 6    | 80                      | 6.2                    |
| 7    | 75                      | 7.4                    |
| 8    | 65                      | 6.0                    |
| 9    | 90                      | 7.6                    |
| 10   | 90                      | 6.1                    |

Figura 15.2 Paraqitja grafike e të dhënave për problemin e transportit



Në fillim le të shohim lidhjen midis kohës së udhëtimit  $y$  dhe numrit të kilometrave të përshkuara  $x_1$ . Të dhënat e nevojshme për këtë qëllim, të zgjedhura në mënyrë të rastit gjatë 10 ditëve, paraqiten në tabelën 15.1 dhe grafiku i tyre në figurën 15.2. Nga grafiku shohim se lidhja midis kohës së udhëtimit dhe numrit të kilometrave është në të njëjtin drejtim, e cila duket se mund të përafrohet me një ekuacioni linear.

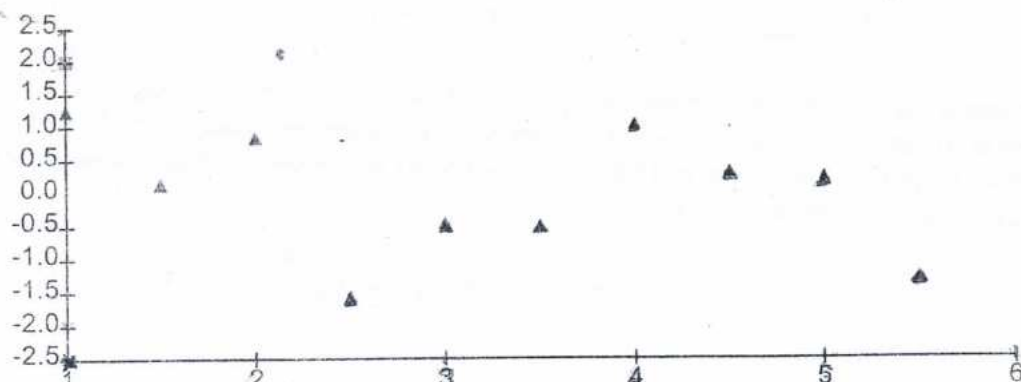
$$\hat{y} = \beta_0 + \beta_1x + \varepsilon \quad 15.6$$

Lidhur me të dhënat e tabelës 15.1 ekuacioni i regresionit rezulton  $\hat{y} = 1,27 + 0,0678 x_1$ . Lidhja midis  $y$  dhe  $x$  rezulton statistikisht e rëndësishme. Vlera  $t$  dhe  $F$  është 0,004. Por gjithashtu shohim se për koeficientin e përcaktueshmërisë  $r^2 = 0,664$  vetëm 66.4% e variacionit të kohës së udhëtimit mund të shpjegohet nga kilometrat e përshkuara. Megjithatë përpara se të arrijmë në përfundimin lidhur me dobësi të ekuacionit të regresionit të vlerësuar me anën e metodës së katrorëve me të vegjël, le të shohim grafikun e mbetjeve të standartizuara të paraqitur në figurën 15.4.

Në figurën 15.4 mbetjet e standartizuara janë vendosur në boshtin vertikal, kurse vlerat e parashikuara të  $y$  në boshtin horizontal. Grafiku i mbetjeve të standartizuara tregon se variacioni i tyre rritet me rritjen e  $y$  dhe se supozimi mbi variancën konstante të termit të gabimit në këtë rast nuk plotësohet. Me qenë se kontrolli për rëndësinë statistikore në analizen

e regresionit mbështetet në supozimet mbi termin e gabimit, konkluzionet e mëpërshme lidhur me rendesinë statistikore të lidhjes janë të dyshimta.

Figura 15.4 Grafiku i mbetjeve të standartizuara



Përveç kësaj vetëm 66,4% e variacionit të kohës së udhëtimit shpjegohet nga variacioni i kilometrave të përshkuara, d.m.th. rreth 33,6% e këtij variacioni mbetet e pa shpjeguar. Kështu ndoshta shkalla e përafrimit mund të rritet me tej në qoftë se në ekuacionin e regresionit shtohet edhe një variabël tjetër.

Tabela 15.2 Të dhënat fillestare mbi problemin e transportit

| Dita | Kilometrat e përshkuara | Numri i dergesave | Koha e udhëtimit |
|------|-------------------------|-------------------|------------------|
|      | $x_1$                   | $x_2$             | $y$              |
| 1    | 100                     | 4                 | 9,3              |
| 2    | 50                      | 3                 | 4,8              |
| 3    | 100                     | 4                 | 8,9              |
| 4    | 100                     | 2                 | 6,5              |
| 5    | 50                      | 2                 | 4,2              |
| 6    | 80                      | 2                 | 6,2              |
| 7    | 75                      | 3                 | 7,4              |
| 8    | 65                      | 4                 | 6,0              |
| 9    | 90                      | 3                 | 7,6              |
| 10   | 90                      | 2                 | 6,1              |

Duke u përpjekur për një alternativë tjetër, manaxheri mund të sugjerojë që ndoshta edhe numri i dergesave mund të përdoret për parashikimin e kohës së udhëtimit duke përmirësuar në këtë mënyrë modelin e regresionit. Të dhënat e nevojshme për këtë duke përfshirë edhe numrin e dergesave tregohen në tabelën 15.2, ku  $x_2$  paraqet numrin e dergesave.

**Regresioni i shumefishtë dhe metoda e katrorëve me të vegjël** Në kapitullin 14 pame përdorimin e metodës së katrorëve me të vegjël për gjetjen e ekuacionit të regresionit linear, që jep përafrimin më të mirë të varesisë midis variablit të varur dhe atij të pavarur. Kushti i katrorëve me të vegjël për këtë metodë është:

$$\min \sum (y_i - \hat{y}_i)^2 \quad 15.7$$

ku  $y_i$  = vlera faktike për variablin e varur në vërtetimin e  $i$ -të

$\hat{y}_i$  = vlera e gjetur me anën e ekuacionit të regresionit për variablin e varur për vërtetimin e  $i$ -të

Ky kriter zbatohet edhe në regresionin e shumefishtë. Në shembullin tonë, ekuacioni i regresionit që shpreh ndikimin e kilometrave të përshkuara dhe të numrit të dergesave në madhësinë e kohës ditore të udhëtimit është

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$



Duke zevendesuar ne kete ekuacion vlerat e variablave te pavarur per vrojtimin e i-te, mund te percaktojme vleren koresponduese te variablit te varur yi.

$$\hat{y} = b_0 + b_1 x_{1i} + b_2 x_{2i}$$

Kushti i katroreve me te vegjel per rastin e ekuacionit te regresionit me dy variabla te pavarur eshte

$$SKG = \sum (y - b_0 - b_1 x_{1i} - b_2 x_{2i})^2 = \min$$

Per minimizimin e shprehjes duhet te gjejme derivatet e pjesshme te SKG ne lidhje me parametrat  $b_i$ , per  $i=0,1,2$  dhe t'i barazojme ato me zero, pastaj per gjetjen e koeficienteve  $b_i$  te zgjidhim sistemin e ekuacioneve. Duke gjetur derivatet e pjesshme lidhur me koeficientet  $b_i$  dhe duke i barazuar me zero kemi.

$$\frac{\partial SKG}{\partial b_0} = -2 \sum (y - b_0 - b_1 x_{1i} - b_2 x_{2i})^2 = 0$$

$$\frac{\partial SKG}{\partial b_1} = -2 \sum x_{1i} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i})^2 = 0$$

$$\frac{\partial SKG}{\partial b_2} = -2 \sum x_{2i} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i})^2 = 0$$

Pas disa transformimesh algjebrike te thjeshta, sistemin e ekuacioneve mund paraqesim ne formen e tij normale.

$$\begin{aligned} nb_0 + (\sum x_{1i})b_1 + (\sum x_{2i})b_2 &= \sum y_i \\ (\sum x_{1i})b_0 + (\sum x_{1i}^2)b_1 + (\sum x_{1i}x_{2i})b_2 &= \sum x_{1i}y_i \\ (\sum x_{2i})b_0 + (\sum x_{1i}x_{2i})b_1 + (\sum x_{2i}^2)b_2 &= \sum x_{2i}y_i \end{aligned}$$

Ne rastin e shembullit tone per varesine e kohes se udhetimit nga kilometrat e pershkuara dhe numrin e dergesave, sistemi i ekuacioneve per percaktimin e koeficienteve  $b_i$  i ndertuar mbi bazen e te dhenave te tabelës, eshte si me poshte:

$$\begin{aligned} 10b_0 + 800b_1 + 29b_2 &= 67,0 \\ 800b_0 + 67,45b_1 + 2345b_2 &= 5594,0 \\ 29b_0 + 2345b_1 + 91b_2 &= 202,2 \end{aligned}$$

Nga zgjidhja e sistemit kemi  $b_0 = -0,8687$ ,  $b_1 = 0,0611$  dhe  $b_2 = 0,9234$ . Ekuacioni i regresionit ka formen:

$$y = -0,8687 + 0,0611x_1 + 0,9234x_2$$

Tabela e llogaritjes te koeficienteve per sistemin e ekuacioneve normale

| $y_i$ | $x_{1i}$ | $x_{2i}$ | $x_{1i}^2$ | $x_{2i}^2$ | $x_{1i}x_{2i}$ | $x_{1i}y_i$ | $x_{2i}y_i$ |
|-------|----------|----------|------------|------------|----------------|-------------|-------------|
| 9.3   | 100      | 4        | 10.000     | 16         | 400            | 930         | 37.2        |
| 4.8   | 50       | 3        | 2.500      | 9          | 150            | 240         | 14.4        |
| 8.9   | 100      | 4        | 10.000     | 16         | 400            | 890         | 35.6        |
| 6.5   | 100      | 4        | 10.000     | 16         | 400            | 650         | 26.0        |
| 4.2   | 50       | 2        | 2.500      | 4          | 100            | 210         | 8.4         |
| 6.2   | 80       | 2        | 6.400      | 4          | 160            | 496         | 12.4        |
| 7.4   | 75       | 3        | 5.625      | 9          | 225            | 555         | 22.2        |
| 6.0   | 65       | 4        | 4.225      | 16         | 260            | 390         | 24.0        |
| 7.6   | 90       | 3        | 8.100      | 9          | 270            | 684         | 22.8        |
| 6.1   | 90       | 2        | 8.100      | 4          | 180            | 549         | 12.2        |
| Shuma | 67.0     | 800      | 29         | 67.450     | 91             | 2.345       | 5.594       |



Nga zgjidhja per shembullin e varesise te kohes ditore te udhetimit nga kilometrat e pershkuara dhe numri i dergesave marrim vleresimin per ekuacionin e regresionit qe eshte:

$$y = -0,869 + 0,0611x_1 + 0,923x_2$$

Veme re se shuma e katroreve pergjate vijes se regresionit (SKR= 21,601) plus shumen e katroreve te gabimeve (SKG = 2,299) eshte e barabarte me shumen e katroreve totale (SKT = 23,9), d.m.th. SKR + SKG = SKT. Pervec kesaj shohim se numrat perkates te shkalleve te lirise jane ne te njejtin relacion, d.m.th. shuma e numrit te shkalleve te lirise per SKR (2) me numrin e shkalleve te lirise per SKG (7), eshte e barabarte numrin e shkalleve te lirise per SKT (9). Ne pergjithesi numri i shkalleve te lirise per shumen pergjate vijes se regresionit eshte i barabarte me numrin e variablave te pavarur (p). Numri i shkalleve te lirise per shumen totale eshte i barabarte me numrin e vrojtimeve pakesuar me 1. Si rrjedhoje numri i shkalleve te lirise per shumen e katroreve te gabimeve rezulton n-p-1. Veme re se katrori mesatar pergjate vijes se regresionit (KMR = 10,8) eshte i barabarte me raportin e SKR me numrin perkates te shkalleve te lirise (p = 2). Katrori mesatar i gabimeve (KMG = 0,328) eshte i barabarte me SKG i pjestuar me numrin perkates te shkalleve te lirise te kesaj shume (n - p - 1 = 10 - 2 - 1 = 7).

Le te krahasojme tani keto rezultate me ato te gjetura ne analizen e regresionit me nje variabel (kilometrat e udhetuara). Veme re se shuma e katroreve totale eshte e barabarte me 23,9, d.m.th. ka rezultuar e njejte ne te dy rastet. Per te kuptuar me mire rendesine e ketij rezultati kujtojmë se shuma e katroreve totale llogaritet me anen e formule:

$$SKR = \sum (\hat{y}_i - \bar{y})^2 \quad SKT = \sum (y_i - \bar{y})^2 \quad 15.8 \quad SKG = \sum (y_i - \hat{y}_i)^2$$

Me qe shuma e katroreve totale varet vetem nga vlerat e vrojtura te variabelit te varur, rrjedh se ajo nuk varet nga numri i variablave te modelit. Shuma e katroreve totale, qe mbetet e njejte per cdo model, eshte kyçi per te kuptuar se c'ndodh ne qofte se ne model perfshijme edhe variabla te tjere te pavarur. Keshtu me qe SKT=SKR+SKG, atehere kurdohere qe ne model shtojme variabla te pavarur, shuma e katroreve te gabimeve do te zvoglohet dhe do te rritet shuma e katroreve pergjate vijes se regresionit. Le te shqyrtojme me me hollesi interpretimin e koficienteve ne ekuacionin e regresionit te shumefishte.

**Koeficientet e ekuacionit te regresionit** Le te krahasojme tani te dy ekuacionet e regresionit, d.m.th. ekuacionin e thjeshte te regresionit ku si variabel i pavarur perfshihet vetem numri i kilometrave te pershkuara, me ekuacionin e regresionit ku pervec tij perfshihet edhe variabli tjetër i pavarur, numri i dergesave. Shohim se vlera e  $b_1$  nuk eshte e njejte per te dy rastet. Ne regresionin e thjeshte linear  $b_1$  interpretohet si madhesi e ndryshimit te  $y$  per cdo njesi te ndryshimit te  $x$ . Ne regresionin e shumefishte ky interpretim kerkon nje modifikim. Keshtu koficienti  $b_1$  prane variabelit te pavarur  $x_1$  (kilometrat e pershkuara) paraqet masen e ndryshimit te  $y$  qe i korespondon 1 njesie te ndryshimit te  $x$ , kur variablat e tjere te pavarur mbahen ne nivel te pandryshuar. Ne rastin e shembullit tone  $b_1 = 0,0611$ , d.m.th. me rritjen e distances se transportit me 1 km, koha e udhetimit rritet mesatarisht me 0,0611 ore, ne kushtet kur numri i dergesave mbetet i pandryshuar. Po ashtu me qene se  $b_2 = 0,923$ , kjo do te thote se me rritjen e numrit te dergesave me 1, koha e udhetimit rritet mesatarisht me 0,923 ore, ne kushtet kur numri i kilometrave te pershkuara nuk ndryshon.

### 15.3 Vleresimi i shkalles se perputhjes

Ne kapitullin 14 pame se per vleresimin e shkalles se perputhjes te ekuacionit te regresionit perdorej koficienti i percaktushmerise, i cili u llogarit si raport i shumes se katroreve pergjate vijes se regresionit me shumen e katroreve totale. Ne analizen e regresionit te shumefishte llogaritet i njejti raport i ngjashem, i cili quhet koficient i percaktushmerise se shumefisht. Koficienti i percaktushmerise llogaritet me formulën.

$$R^2 = \frac{SKR}{SKT} \quad 15.9$$

Ne qofte se ai shumëzohet me 100, koficienti i percaktushmerise paraqet perqindjen e variacionit te  $y$  qe percaktohet nga ekuacioni i regresionit te shumefishte. Ne shembullin qe kemi marre ne shqyrtim, koficienti i percaktushmerise rezulton i barabarte me  $R^2 = 21,601/23,900 = 0,9038$ . Si rrjedhoje 90,38% e variacionit te kohes se udhetimit percaktohet nga kilometrat e pershkuara dhe numri i dergesave.



I referohemi tani paragrafit 15.2. Veme re se modeli regresiv, qe perfshin vetem kilometrat e pershkuara si variabel te pavarur, e ka koeficientin e percaktushmerise te barabarte me 0,664. Si rrjedhoje perqindja e variacionit te variablit te varur qe mund te spjegohet me anen e ekuacionit e regresionit rritet nga 66.4% ne 90.38%. Ne pergjithesi eshte e vertete se me rritjen e numrit te variablave te pavarur qe perfshihen ne model, koeficienti i percaktushmerise vetem rritet. Kjo ndodh sepse me rritjen e numrit te variablave gabimi i parashikimit zvoglohet, me qene se edhe SKG zvoglohet. Duke patur parasysh se  $SKT = SKR + SKG$ , kur SKG zvoglohet SKR rritet dhe si rrjedhoje edhe koeficienti i percaktushmerise rritet. Ne disa raste rekomandohet qe koeficienti i percaktushmerise te korigjohet duke marre parasysh numrin e variablave te pavarur, me qellim qe te menjahet ai mbivleresim qe shkakton shtimi i variablave ne madhesine e variacionit te spjeguar. Korigjimi i koeficientit te percaktushmerie behet si me poshte:

$$R_k^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$$

15.10

Per shembullin tone ky koeficient rezulton:

$$R^2 = 1 - (1 - 0,9038) \frac{10-1}{10-2-1} = 1 - (0,0962)(1,2857) = 0,8763$$

Keshtu pas korigjimit per numrin e variablave te pavarur qe perfshihen ne model, vetem 87.63% e variacionit te y mund te percaktohet nga ekuacioni i regresionit.

#### 15.4 Kontrolli per rendesine e lidhjes

Ekuacioni i regresionit me dy variabla te pavarur ka formen e meposhteme:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Si rrjedhoje kriteri i pershtatshem per percaktimin ne se midis variablit te varur dhe variablave te pavarur ekziston nje lidhje e rendesishme eshte si me poshte:

$$H_0: \beta_1 = \beta_2 = 0$$

$H_a$ : te pakten nje nga parametrat eshte i ndryshem nga zero.

Ne qofte se hipoteza  $H_0$  hidhet poshte, atehere arrijme ne perfundimin se lidhja midis variablit te varur y dhe variablave te pavarur  $x_1$  dhe  $x_2$  eshte e rendesishme.

Kriteri qe perdoret ne analizen e regresionit te shumefishte per vleresimin e rendesise se lidhjes midis variablit te varur dhe variablave te pavarur eshte kriteri F, i cili paraqitet njelloj si ai qe kemi pare ne kapitullin 14 per modelet me nje variabel te pavarur. Ne fillim do te shohim perdorimin e kriterit F ne problemin e transportit dhe pastaj do te bejme pergjithesimin e tij per kontrollin e modeleve qe permbajne p variabla te pavarur.

**Kriteri F per problemin e transportit** Mund te provohet se ne qofte se hipoteza zero eshte e vertete dhe te kater supozimet per modelin e regresionit plotesohen, atehere raporti KMR / KMG ka shpemdaje F. Numri i shkalleve te lirise per numruesin eshte i barabarte me numrin e shkalleve te lirise qe i perket shumes se katroreve per regresionin, kurse numri i shkalleve te lirise per emruesin eshte sa numri perkates i shkalleve te lirise qe i perket shumes se katroreve te gabimit. Kujtojmë se shuma e katroreve per regresionin shpreh masen e variacionit te y qe spjegohet me anen e modelit te regresionit. Keshtu per te hedhur dyshime mbi vertetesine e hipotezes zero, per raportin F duhet te presim vlera te medha te tij. Ne praktike pikerisht keshtu veprohet; per vlerat e vogla te F hipoteza zero nuk ka baze te hidhet poshte, kurse vlerat e medha te F do te conin ne hedhjen poshte te hipotezes zero.

Le te kthehemi tani tek problemi i transportit per te kontrolluar rendesine e modelit te regresionit te shumefishte. Llogaritim raportin F per te dhenat faktike:  $F = KMR / KMG = 10,8 / 0,328 = 32,88$ . Per kontrollin e hipotezes qe  $\beta_1 = \beta_2 = 0$ , duhet te percaktojmë ne se 32,88 eshte nje vlere qe shfaqet ne nje zgjedhje te rastit nga shpemdaje F me 2 shkalle lirie per numruesin dhe 7 shkalle lirie per emruesin.

Supozojme se niveli i rendesise eshte 0,05. Vlera kritike per kriterin F (tabela 4 e shtojces B) eshte 4.74. Kjo do te thote se ne qofte se merret nje zgjedhje e rastit nga shpemdaje F, me shkalle lirie per numeruesin dhe emruesin perkatesi te barabarte me 2 dhe 7 ( sic do te ishte po qe se hipoteza  $H_0$  eshte e vertete), atehere vetem 5% e vlerave do te ishin me te medha se 4,74. Ne figuren 15.6 tregohet percaktimi i zones kritike te kriterit F.



Ne baze te analizës se mesiperme mund te formulome rregullin e hedhjes poshte te hipotezes per nivelin e rendesise te barabarte me 0.05 si me poshte.

Hipoteza zero hidhet poshte ne qofte se  $F = KMR / KMG > 4.74$ .

Me qene se vlera e raportit F eshte 32.88, pra me e madhe se vlera kritike e tij 4.74, hipoteza  $H_0$  hidhet poshte dhe arrihet ne perfundimin se midis kohes se udhetimit dhe dy variablave te pavarur ekziston nje lidhje e rendesishme. Keshtu ekuacioni i regresionit mund te perdoret per parashikimin e vlerave te  $y$  per vlerat e dhena te variablave te pavarura, por gjithnje brenda kufijve te variacionit te tyre.

**Tabela e analizës se variancës dhe kriteri F.** Tani qe dime se si mund te zbatohet kriteri F per vleresimin e modelit te regresionit te shumefishte me dy variabla, le ta pergjithesojme kete kriter ne rastin e  $p$  variablave te pavarura. Hipotezat qe duhet te kontrollojme per te percaktuar ne se ekziston nje lidhje midis variablit te varur dhe variablave te pavarur paraqiten si me poshte.

$$H_0: \beta_0 + \beta_1 = \beta_2 = \dots \dots \beta_p = 0$$

$H_a$ : Nje ose disa nga parametrat e ekuacionit te regresionit nuk jane te barabarte me zero. Edhe ne kete rast ne qofte se hipoteza zero hidhet poshte, konkludojme se midis variablave ekziston nje lidhje e rendesishme dhe se ekuacioni i regresionit eshte i pershtatshem per tu perdorur ne parashikimin e vlerave te variablit te varur  $y$ . Forma e pergjitheshme e tabelës per vleresimin e ekuacionit te regresionit te shumefishte me  $p$  variabla tregohet ne tabelen 15.6. I vetmi ndryshim nga rasti i dy variablave eshte persa i perket numrave te shkalleve te lirise per numruesin dhe emruesin e raportit F. Ne kete rast shuma e katroreve per regresionin ka  $p$  shkalle lirie, me qene se kemi  $p$  variabla te pavarur dhe si rrjedhoje katrori mesatar per regresionin llogaritet si me poshte.

$$KMR = SKR / p$$

Persa i perket shumës se katroreve per gabimin ajo ka  $n - p - 1$  shkalle lirie, keshtu qe mesatarja e katroreve perkatese llogaritet me anen e formules:

$$KMG = SKG / n - p - 1$$

Ne kete rast statistika F ka formen:  $F = KMR / KMG = (SKR / SKG) [(n - p - 1) / p]$ . Per gjejen e vlerës kritike ne tabelen e shpërndarjes se kriterit F, duhet te kemi parasysh se numruesi ka  $p$  shkalle lirie, kurse emruesi  $n - p - 1$  shkalle lirie. Sic u theksua me lart, hedhja poshte ose jo e hipotezes  $H_0$ , mund te behet duke krahasuar vleren e llogaritur te statistikes F me vleren kritike te tij.

**Tabela 15.6 Tabela ANOVA per modelin e regresionit te shumfishte me  $p$  variabla te pavarur**

| Burimet e variacionit | Shuma e katroreve | Shkallet e lirise | Katrori mesatar         | Kriteri F       |
|-----------------------|-------------------|-------------------|-------------------------|-----------------|
| Regresioni            | SKR               | $p$               | $KMR = SKR / p$         | $F = KMR / KMG$ |
| Gabimi                | SKG               | $n - p - 1$       | $KMG = SKG / n - p - 1$ |                 |
| Totali                | SKT               | $n - 1$           |                         |                 |

**Kriteri t per vleresimin e rendesise te parametrave individuale** Ne qofte se me anen e kriterit F konkludojme se lidhja e regresionit te shumefishte eshte e rendesishme (d.m.th. se te pakten nje nga parametrat  $\beta_i$  eshte e ndryshme nga zero), atehere eshte me interes te bejme nje kontroll tjeter per te vleresuar se cilet nga parametrat ( $\beta$ ) jane te rendesishem. Si metode statistikore per kontrollin e rendesise te parametrave individuale sherben kriteri t. Kriteri t eshte i njejte per te gjithë koeficientet e variablave te pavarur. Hipoteza qe kerkohet te kontrollohet eshte si me poshte:

$$H_0: \beta_i = 0 \quad H_a: \beta_i \neq 0$$

Shohim se hipoteza eshte e njejte si ne rastin kur kemi vetem nje variabel te pavarur. Per te kontrolluar nje hipoteze te tille llogaritim statistiken  $b_i / S_{b_i}$ , ku  $b_i$  eshte koeficienti prane variablit te pavarur  $x_i$  ne ekuacionin e regresionit te gjetur me anen e metodes se katroreve me te vegjel, kurse  $S_{b_i}$  eshte vleresimi i devijimit standart te shpërndarjes se zgjedhjes per  $b_i$ . Sic kemi pare me lart shpërndarja e zgjedhjes e  $b_i / S_{b_i}$  ka shpërndarje t me  $n - 2$  shkalle lirie. Keshtu per te kontrolluar hipotezen statistikore per rendesine e koeficientit  $b_i$  perdorim rregullin e meposhtem:



Hipoteza  $H_0$  hidhet poshte ne qofte se  $b_i / S_{b_i}$  eshte ne vlere absolute me e madhe se vlere kritike e kriterit t. Hipoteza  $H_0$  pranohet kur vlere e raportit eshte ne vlere absolute me vogel se vlere kritike e kriterit t. I vetmi ndryshim nga procedura e perdorur me pare eshte per numrin e shkalleve te lirse per shpemdarrjen t dhe per llogaritjen e shmangies standarte per  $b_i$ . Numri i shkalleve te lirse merret baras me ate te shumes se katroreve per gabimin, d.m.th.  $(n-p-1)$ , ku p eshte numri i variablave te pavarur.

Kthehemi tani tek shembulli i transportit per kontrollin e rendesise se parametrave  $\beta_0$  dhe  $\beta_1$ . Koeficientet e regresionit te gjetur kane rezultuar  $b_1=0,061135$  dhe  $b_2=0,9234$ , kurse shmangiet standarte  $S_{b_1}=0,009888$  dhe  $S_{b_2}=0,2211$ . Si rrjedhoje vlerat e llogaritura te kriterit F jane perkatesisht  $b_1 / S_{b_1} = 0,061135 / 0,009888 = 6,18$  dhe  $b_2 / S_{b_2} = 0,9234 / 0,2211 = 4,18$ . Per nivelin e rendesise 0,05 dhe numrin e shkalleve te lirse n,  $10-2-1=7$ , ne tabelen e vlerave te kriterit t gjejmë vleren kritike te tij  $t = 2,365$ . Me qene se vlerat e llogaritura te kriterit t jane qe te dyja me te medha se vlere kritike e tij, atehere hipotezat zero per koeficientet e ekuacionit te regresionit hidhen poshte dhe rrjedhimisht ata konsiderohen te rendesishem.

**Kolineariteti i shumefishte** Ne analizen regresive kemi perdorur termin variabel i pavarur me kuptimin e cdo variabli qe sherben per te parashikuar, ose spjeguar vlerat e variablit te varur. Kjo nuk do te thote se variablat e pavarur ne vetvete jane te pavarur ne cdo kuptim statistikor. Mund te ndodhe qe variablat e pavarur ne modelin e regresionit te shumefishte ne nje fare menyre jane te koreluar midis tyre. Keshtu ne shembullin e problemit te transportit qe perfshin dy variablat e pavarur, kilometrat e pershkuara dhe numrin e dergesave, per te percaktuar ne se keto dy variabla jane te lidhur midis tyre, te parin mund ta trajtojme si variabel te varur dhe te dytin si variabel te pavarur.

Per te percaktuar se sa ato jane te lidhur midis tyre mund te llogaritim koeficientin e korelacionit. Ky koeficient rezulton i barabarte me 0,28, pra ne kete menyre zbulojme nje fare korelacioni midis variablave. Ne analizen korelative te shumefishte per te treguar lidhjen midis variablave te pavarur perdoret termi kolineariteti i shumefishte, ose multikolineariteti.

Per te patur nje paraqitje me te qarte te problemeve qe mund te lindin ne analizen korelative per shkak te multikolinearitetit, le te marrim nje shembull tjeter. Ne se ne vend te numrit te dergesave do te mirrnim variablin per sasine e benzines te harxhuar, atehere eshte fare e qarte se sasine e benzines do te varej nga kilometrat e pershkuara dhe si rrjedhoje midis variablave te pavarur do te ekzistonte nje multikorelacion mjaft i forte.

Le ta zeme se kemi gjetur ekuacionin e regresionit te shumefishte  $Y = b_0 + b_1 x_1 + b_2 x_2$  dhe sipas kriterit F ka rezultuar se ky ekuacion eshte i rendesishem. Ne qofte se pas kesaj zbatojme kriterin t dhe konstatojme se hipoteza zero per koeficientin  $\beta_1, \beta_1 = 0$ , nuk mund te hidhet poshte, atehere shtrohet problemi a mundet qe nga kjo te nxjerrim perfundimin se midis kohes se udhetimit dhe kilometrave te pershkuara nuk ekziston lidhje? Ne asnje menyre jo, sepse dihet qe koha e udhetimit varet nga kilometrat e pershkuara. Kjo do te thote se me qene se ne model eshte perfshire edhe  $x_2$ , i cili eshte ne lidhje me y,  $x_1$  nuk jep nje kontribut te rendesishem ne percaktimin e vlerave te y. Per shembullin tone kjo ka kuptim, sepse duke njohur sasine e benzines se harxhuar ne mund te percaktojme kohen e udhetimit pa qene nevoja te kemi informacion tjeter per kilometrat e pershkuara. Po ashtu ne qofte se me anen e kriterit t kemi gjetur se per koeficientin  $B_2$  hipoteza zero nuk mund te hidhet poshte, kjo do te thote se duke patur ne model  $x_1$ , patja edhe e informacionit per  $x_2$  nuk ka ndonje peshe ne percaktimin e vlerave te y.

Veshtiresia e shkaktuar nga multikolineariteti ne zbatimin e kriterit t per vleresimin e rendesise te koeficienteve te regresionit, shfaqet ne rastin kur asnje prej variablave te pavarur nuk paraqitet i rendesishem, nderkohe kur ekuacioni i regresionit te shumefishte vete paraqitet i rendesishem. Ky problem mund te menjanohet kur variablat e pavarur jane ne lidhje te dobet midis tyre.

Zakonisht multikolineariteti nuk ndikon ne menyren e zbatimit te analizes se regresionit, apo te interpretimit te rezultateve te saj. Megjithate ne qofte se multikolineariteti eshte i larte, d.m.th. kur dy ose me shume variabla te pavarur jane te lidhura ngushte midis tyre, mund te lindin veshtiresi ne interpretimin e rezultateve te kriterit t per parametrat individuale. Pervec kesaj per problemin e tipit te mesiperme, ne rastet kur multikolineariteti eshte i larte, mund te ndodhe qe koeficientet e regresionit te gjetur me anen e metodës se katroreve me te vegjel te



kene rezultuar me shenje te kunderte me ate qe duhet te kishin. Keshtu ne rastet kur multikolineariteti eshte i larte, besimi ndaj koeficienteve te regresionit eshte i vogel.

Jane ndertuar disa kriteret per te percaktuar ne se multikolineariteti i larte mund te shkaktoj probleme te tilla. Nje nga keto kriteret pohon se multikolineariteti eshte problem kur koeficienti i korelacionit midis variablave te pavarur eshte me i madh se 0.7, per cdo dy variabla te pavarura. Ekzistojne edhe kriteret tjera per kete qellim, qe i kalojne caqet e ketij teksti. Duhet bere c'eshte e mundur per te menjanuar perfshirjen ne model te variablave te pavarura qe jane shume te koreluar midis tyre. Megjithese ne praktike kesaj nuk i permbahemi rigorozisht, gjithnje duhet patur parasysh se kur multikolineariteti eshte i larte ka mjaft veshtiresi ne ndarjen ne menyre te paster te efektit te variablave te pavarur ne nivelin e variablit te varur.

### 15.5 Vleresimi dhe parashikimi

Vleresimi i vleres mesatare te  $y$  dhe parashikimi i nje vlere te vecante te tij ne regresionin e shumefishte eshte i njejte me ate ne analizen e regresionit te thjeshte qe perfshin vetem nje variabel te varur. Ne kapitullin e 14-te pame se vleresimi pikesor i vleres se pritur te  $y$  per nje vlere te  $x$ , ishte i njejte me vleresimin pikesor te nje vlere individuale te  $y$ . Ne te dy rastet si vleresim pikesor perdorej  $\hat{y} = b_0 + b_1x$ .

Ne regresionin e shumefishte perdoret e njejta procedure si ne regresionin e thjeshte linear. D.m.th. vlerat e dhena te variablave te pavarur,  $x_1, x_2, \dots, x_p$ , zevendesohen ne ekuacionin e regresionit dhe vlerat e gjetura me anen e tij shërbejne si vleresime pikesore per  $y$ . Le ta zeme se ne problemin e transportit duam te vleresojme kohen e udhetimit  $y$ , ne varesi te kilometrave te pershkuara  $x_1$  dhe te numrit te dergesave  $x_2$  si me poshte.

1. Te gjejme intervalin e besimit per kohen mesatare te udhetimit per te gjithë automjetet qe udhetojne 100 kilometra dhe bejne 2 dergesa ose fumizime.

2. Te parashikojme kohen e udhetimit per nje automjet te vecante qe udheton 100 kilometra dhe ben vetem 2 fumizime.

Duke perdorur ekuacionin e regresionit  $y = -0,869 + 0,0611x_1 + 0,923x_2$ , per  $x_1 = 100$  dhe  $x_2 = 2$ , gjejme vleresimin pikesor per kohen e udhetimit,  $y = -0,869 + 0,0611(100) + 0,923(2) = 7,09$ . Keshtu vleresimi pikesor per kohen e udhetimit ne te dy rastet eshte i barabarte me rreth 7 ore.

Per te gjetur vleresimin intervalor te vleres mesatare te  $y$  dhe te nje vlere te individuale te tij, perdorim nje procedure analoge me ate te regresionit te thjeshte me nje variabel te pavarur. Theksojme se vleresimi intervalor per nje vlere individuale te  $y$  eshte me i gjere se vleresimi intervalor per vleren mesatare te  $y$ . Keto formula dalin jashte caqeve te ketij teksti, prandaj per kete nuk do te ndalemi me tej. Ne tabelen e meposhteme paraqiten llogaritjet per vleresimet intervalore te vlerave te kohes se udhetimit  $y$  ne varesi numrit te kilometrave te pershkuara  $x_1$  dhe te numrit te dergesave  $x_2$  ne te dy rastet, per vleren mesatare te  $y$  dhe per vleren individuale te tij.

Tabela 15.7 Vleresimi i intervalit 95%-esh te besimit dhe te parashikimit

| Vlerat e $x_1$ | Vlerat e $x_2$ | Intervali i besimit |                 | Intervali i parashikimit |                 |
|----------------|----------------|---------------------|-----------------|--------------------------|-----------------|
|                |                | Kufiri i poshtem    | Kufiri i siperm | Kufiri i poshtem         | Kufiri i siperm |
| 50             | 2              | 3,146               | 4,924           | 2,414                    | 5,656           |
| 50             | 3              | 4,127               | 5,789           | 3,368                    | 6,548           |
| 50             | 4              | 4,815               | 6,948           | 4,157                    | 7,607           |
| 100            | 2              | 6,258               | 7,926           | 5,500                    | 8,683           |
| 100            | 3              | 7,385               | 8,645           | 6,520                    | 9,510           |
| 100            | 4              | 8,135               | 9,742           | 7,362                    | 10,515          |

### 15.6 Perdorimi i variablave cilesore ne analizen e regresionit

Deri ketu variablat e perdorur per ndertimin e ekuacionit te regresionit ishin variabla sasiore. Megjithate ka raste kur disa nga variablat nuk marrin vlera numerike, ose sic thuhet ndryshe jane variabla cilesore. P.sh. le ta zeme se jemi te interesuar per parashikimin e shitjeve te nje produkti, i cili eshte pakeluar ne shishe ose kavanoza. Ketu variabli cilesor eshte lloji i enes, i cili do te influenconte ne vellimin e shitjeve. Vecoria e variablave cilesore eshte se ato nuk



mund te maten nga ana sasiore, ato shprehin vetem cilesi, e cila vetem mund te jete ose jo e pranishme.

Per te ilustruar se si mund te perdoren variablat cilesore ne analizen regressive, marrim rastin e nje firme sherbimi riparimi makinash. Per te percaktuar lidhjen midis kohes se sherbimit per riparimin e nje makine te vecante dhe numrit te muajve qe nga sherbimi i meparshem, jane mbledhur te dhena mbi keto karakteristika qe paraqiten ne tabelen 15.8. Duke pare grafikun e ketyre te dhenave ne figuren 15.7, shohim se lidhja midis kohes se sherbimit dhe numrit te muajve qe nga sherbimi i fundit mund te paraqitet ne formen lineare. Keshtu modeli i regresionit qe perdoret eshte:  $y = \beta_0 + \beta_1 x_1 + \varepsilon$  ku  $y$  eshte koha e riparimit, kurse  $x_1$  numri i muajve qe kane kaluar nga riparimi i fundit.

Tabela 15.8

|                                         |     |   |     |     |     |     |     |     |     |     |
|-----------------------------------------|-----|---|-----|-----|-----|-----|-----|-----|-----|-----|
| Koha e riparimit ne ore                 | 2.9 | 3 | 4.8 | 1.8 | 2.9 | 4.9 | 4.2 | 4.8 | 4.4 | 4.5 |
| Numri i muajve qe nga sherbimi i fundit | 2   | 6 | 8   | 3   | 2   | 7   | 9   | 8   | 4   | 6   |

Vleresimi per ekuacionin e regresionit eshte  $y = b_0 + b_1 x$ , i cili sipas metodës se katroreve me te vegjel rezulton  $y = 2,15 + 0,304x$ . Per vleresimin e rendesise te ekuacionit te regresionit, ose te koeficientit  $b_1$  marrim nivelin e rendesise te barabarte me 0,05. Probabiliteti qe i pergjigjet vleres faktike te kriterit F (ose te kriterit t), sepse sic dihet te dy keta kriterë ne rastin e regresionit te thjeshte jane te njellojte eshte i barabarte me 0,016. Me qene se ky probabilitet eshte me i vogel se niveli i rendesise, rezulton se ekuacioni i regresionit eshte i rendesishem. Megjithese numri i muajve nga sherbimi i fundit paraqitet si nje variabel i rendesishem per kohen e sherbimit. shohim se ai spjegon vetem 53,4 % te variacionit te kohes se sherbimit.

Pas nje shqyrtimi me te hollesishem te shkakut te defektit te makinerise, mekanik ose elektrik, le ta zeme se vendosim te perfshijme ne model edhe variablin lloji i defektit. Ky variabel paraqitet vetem ne dy gjendje, mekanik ose elektrik, pra ai eshte nje variabel cilesor. Per ta perfshire kete variabel ne model e paraqesim ate ne formen e meposhteme.

$x_2 = 0$ , ne se lloji i defektit eshte mekanik dhe 1 ne rast se lloji i defektit eshte elektrik. Nje variabel i futur ne menyre te tille ne analizen e regresionit quhet variabel fiktiv, ose variabel tregues. Duke perdorur variabla te tille mund te perfshijme ne analizen e regresionit jo vetem variabla sasiore, sic eshte numri i muajve nga sherbimi i fundit deri ne kerkesen per sherbim ( $x_1$ ), por edhe variabla cilesore si lloji i defektit qe ka ndodhur ( $x_2$ ). Ne kete rast ekuacioni i regresionit merr formen:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Per te pare se si mund te interpretohen parametrat ne kete ekuacion, le ta zeme se duhet te percaktojmë vleren e pritur te  $y$  ne rastin kur lloji i defektit eshte mekanik.

**Interpretimi i parametrave** Me qene se lloji mekanik i defektit i korespondon  $x_2 = 0$ , kemi:

$$E(y / \text{defekt mekanik}) = \beta_0 + \beta_1 x_1 + \beta_2 (0) = \beta_0 + \beta_1 x_1 \quad 15.14$$

Shprehja ne anen e majte te formule 15.14 duhet te kuptohet si vlere e pritur e  $y$  kur ka ndodhur defekt mekanik. Ekuacioni 15.14 tregon se koha e pritur e riparimit kur ka ndodhur defekt mekanik, eshte funksion linear i  $x_1$  (koha qe nga kerkesa e mepareslune per sherbim), me nderprerje te boshtit te ordinatave  $\beta_0$  dhe koeficient kendor te barabarte me  $\beta_1$ .

Po ashtu me qene se defektit elektrik i pergjigjet  $x_2 = 1$ , shprehja per vleren e pritur te  $y$  ne rastin e defektit elektrik eshte

$$E(y / \text{defektelektirik}) = \beta_0 + \beta_1 x_1 + \beta_2 (1) = \beta_0 + \beta_1 x_1 + \beta_2 = (\beta_0 + \beta_2) + \beta_1 x_1$$

Ky ekuacion tregon se koha e pritur e riparimit per nje defekt elektrik eshte gjithashtu nje funksion linear i  $x_1$ , por megjithese koeficienti kendor i ekuacionit eshte i perseri i barabarte me  $\beta_1$ , nderprerja me boshtin c ordinatave ka ndryshuar nga  $\beta_2$  ne  $\beta_0 + \beta_2$ . Keshtu per cdo vlere te  $x_1$  shohim se  $\beta_2$  paraqet ndryshimin midis kohes se riparimit per defektin mekanik dhe kohes se riparimit per defektin elektrik.

Ne qofte se  $\beta_2$  eshte pozitive koha e riparimit per defektin elektrik eshte me e madhe se sa koha e riparimit per defektin mekanik, kur  $\beta_2$  eshte negative koha e riparimit per defektin



elektrik është me vogël se koha e riparimit për defektin mekanik dhe në qoftë se  $\beta_2 = 0$ , atëherë koha e riparimit nuk varet nga lloji i defektit.

Interpretimi i  $\beta_2$  është i qartë po të zbrisim 15.14 (koha prituri e riparimit për defektin mekanik) nga 15.15 (koha e prituri e riparimit për defektin elektrik). Kështu ai paraqet ndryshimin midis kohës së riparimit për defektin elektrik me atë mekanik.

Në tabelën 15.9 paraqiten të dhënat mbi kohën e riparimit në varësi nga periudha që nga riparimi i fundit dhe llojin e defektit. Në tabelë lloji i defektit mekanik tregohet me vlerën e variablit  $x_2 = 0$  dhe lloji i defektit elektrik me vlerën e tij  $x_2 = 1$ . Për këtë variabël, ekuacionin e regresionit e marrim në formën

$$\hat{y} = b_0 + b_1x_1 + b_2x_2$$

Vlerësimi për ekuacionin e regresionit është  $y = 0.930 + 0.388x_1 + 1.26x_2$  ku  $x_2$  është variabli fiktiv për llojin e defektit. Kështu vlerësimi për  $\beta_2$ , që shpreh diferencën midis kohës së riparimit për defektin e llojit mekanik ( $x_2 = 1$ ) me atë për defektin e llojit elektrik ( $x_2 = 0$ ), është i barabartë me 1.26 ore.

Për vlerësimin e rendesisë të  $x_2$  për efektin e dhënë të  $x_1$  përdoren hipotezat e mëposhtme.

$$H_0: \beta_2 = 0$$

$$H_a: \beta_2 \neq 0$$

Për nivelin e rendesisë  $\alpha = 0.05$  vlera e probabilitetit që i korrespondon kriterit t për tipin është 0.05. Si rrjedhojë hipoteza zero hidhet poshtë dhe arrijmë në përfundimin se tipi i defektit është një faktor i rendesishëm për parashikimin e kohës së riparimit të automjetit, megjithatë për këtë është marrë parasysh edhe periudha e kohës që nga riparimi i fundit. Gjithashtu vëmë re se me përfshirjen edhe të variablit cilesor lloji i defektit, koeficienti i përcaktueshmërisë (81.9%) është rritur në krahasim me herën e parë kur më shumë në shqyrtim vetëm periudhën që ka kaluar që nga riparimi i fundit.

Tabela 15.9 Të dhënat mbi riparimin

| Kohë e riparimit<br>(ore) | Muajt nga<br>shërbimi fundit | Tipi<br>defektit |
|---------------------------|------------------------------|------------------|
| 2.9                       | 2                            | 1                |
| 3.0                       | 6                            | 0                |
| 4.8                       | 8                            | 1                |
| 1.8                       | 3                            | 0                |
| 2.9                       | 2                            | 1                |
| 4.9                       | 7                            | 1                |
| 4.2                       | 9                            | 0                |
| 4.8                       | 8                            | 0                |
| 4.4                       | 4                            | 1                |
| 4.5                       | 6                            | 1                |

Për të parë me mirë rolin e variablit cilesor të përfshirë në ekuacionin e regresionit, në figurën 15.10 janë paraqitur edhe njëherë të dhënat e vrojtueshme, duke dalluar pikat që i përgjigjen tipit të defektit mekanik të treguar me shigjetë dhe pikat që u përgjigjen defektit elektrik që nuk janë afër shigjetës. Nga figura duket se pikat që i përgjigjen defektit mekanik grumbullohen rreth një drejteze (pikat 3,6,9), kurse pikat që i përgjigjen defektit elektrik grumbullohen me mirë rreth një drejteze tjetër (pikat 2,4,6,7,8). Kjo dukuri vërtetohet edhe analitikisht, kështu ekuacioni i regresionit për të dhënat që u korrespondojnë defektit të tipit mekanik ( $x_2 = 0$ ) është

$$y = 0.930 + 0.388x_1$$

dhe ai për pikat që u korrespondojnë tipit të defektit elektrik ( $x_2 = 1$ ) është

$$y = 0.930 + 0.388x_1 + 1.26 = 2.190 + 0.388x_1$$

Kështu duke përdorur variablin fiktiv në ishëm ne gjenim dy ekuacione për të parashikuar kohën e riparimit në ore; njëri që i korrespondon tipit të defektit mekanik dhe tjetri të defektit elektrik.

### 15.7 Analiza e mbetjeve

Në kapitullin 14 pame se si me anën e analizës së mbetjeve mund të kontrollohej plotësimi i kushteve dhe i supozimeve për termin e gabimit dhe formën e modelit. Këto metoda mund të përdoren edhe në analizën e regresionit të shumëfishtë, për të parë plotësimin e



supozimeve te bera per modelin e propozuar, duke na dhene gjithashtu nje informacion shtese lidhur me pershtateshmerine e ekuacionit te gjetur me metoden e katroreve me te vegjel.

Kriteret statistikore qe pame ne paragrafin 15.4 bazohen ne supozimet e bera per termin e gabimit ne paragrafin 15.1 dhe ne supozimin mbi formen lineare te ekuacionit te regresionit per popullimin, i cili paraqitet ne formen  $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p$ . Me anen e analizes se mbetjeve mund te bejme gjykime per plotesimin e supozimeve, si edhe per formen e modelit dhe numrin e variablave qe mund te perfshihen ne model, per te paraqitur me mire lidhjen midis variablit te varur dhe variablave te pavarur.

Ne regresionin e thjeshte linear pame se me anen e grafikut te mbetjeve perkundrejt variablit x ose variablit y, mund te gjykonim per drejtesine e supozimeve per termin e gabimit  $\varepsilon$ . Ne regresionin e shumefishte grafikun i mbetjeve mund te perdoret ne menyre analoge, me te vetmin ndryshim qe ne fillim ai ndertohej perkundrejt variablit te varur y dhe pastaj perkundrejt te gjithe variablave te tjere.

Ne paragrafin 14 pame se ne shumicen e rasteve grafikun i mbetjeve ndertohej per mbetjet e standartizuara. Per regresionin e thjeshte linear mbetja e standartizuar e i-te percaktohet si me poshte:

$$z_{\varepsilon} = \frac{(y_i - \hat{y}_i)}{s\sqrt{1-h_i}} \quad 15.21$$

ku

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \quad 15.22$$

Komentet qe beme ne paragrafin e meparshem zbatohen krejtesisht njelloj edhe ne analizen e regresionit te shumfishte per grafikun e mbetjeve te standartizuara. Llogaritja e vlerave te parashikuara, te mbetjeve dhe mbetjeve te standartizuara eshte shume e veshtire te behet me dore. Duke perdorur paketat software, ato mund te llogariten mjaft lehte si pjese e programit te analizes regressive. Ne baze te mbetjeve te standartizuara te llogaritura mund te ndertohej grafikun i tyre perkundrejt vlerave te parashikuara.

Figura 15.11 paraqet tre forma te grafikut te mbetjeve te standartizuara kundrejt vlerave te parashikuara me anen e ekuacionit te regresionit. Grafikun a tregon tipin qe pritet kur supozimet e bera per modelin linear plotesohen. Grafikun b tregon llojin qe pritet ne rastin kur supozimi mbi variancen konstante nuk plotesohet. Sic shihet termi i gabimit rritet me rritjen e vlerave te y. Ne qofte se shtohet nje variabel tjetër, zakonisht ky problem korigjohet. Pikerisht ky lloj mbetjesh sugjeroi se me shtimin e nje variabli tjetër, perafrimi i kohes se udhimit me anen e ekuacionit te regresionit vetem mund te rritet.

Se fundi grafikun c i pergjigjet situates kur gabimet nuk jane te pavarura. Kur vlerat e llogaritura jane te vogla termi i gabimit eshte pozitiv, kur keto vlera jane te ndermjeme gabimet jane negative dhe kur vlerat e llogaritura jane te medha termi i gabimit eshte perseri pozitiv. Ne kete rast do te ishte me i pershtatshem modeli vijeperkule i regresionit. Ne figuren 15.12 shohim grafikun e mbetjeve te standartizuara kundrejt vlerave te parashikuara. Ky grafik nuk tregon asgje te pazakonte, si rrjedhoje konkludojme se supozimet e bera per modelin jane te aresyeshme.

**Vlerat e largeta.** Nje vlere e larget eshte nje nga te dhenat qe nuk ndjek prirjen qe paraqesin te dhenat qe mbeten pas largimit te saj. Ne kapitullin 14 pame se si mund te perdoren mbetjet e standartizuara per te percaktuar vlerat qe duhet te perjashtohen nga bashkesia e te dhenave. Ne qofte se vlera te tilla ekzistojne, atehere ekuacioni i regresionit nuk do te ishte aq i pershtatshem si ne rastin kur vlera te tilla nuk do te ekzistonin dhe si rrjedhoje gabimi standart do te jete me i madh. Me qe gabimi standart shfaqet ne emerues, me rritjen e s madhesia e mbetjeve te standartizuara do te zvoglohet. Keshtu edhe ne rastin kur nje mbetje eshte mjaft e madhe, do te ishte e veshtire qe me anen e mbetjeve te standartizuara te identifikohet nje vlere e larget. Per kete qellim mund te perdoret nje lloj tjetër mbetjesh te quajtura mbetje te sheshuara.



**Perdorimi i mbetjeve te sheshuara per zbulimin e vlerave te largeta.** Le ta zeme se vrojtimi i-te eshte larguar nga bashkesia e te dhenave dhe mbi bazen e tyre eshte ndertuar nje ekuacion i ri regresioni. Shenojme me  $s_{ij}$  gabimin standart te vleresimit te bazuar ne bashkesine e te dhenave qe nuk permban vrojtimin e i-te. Ne qofte se llogaritim mbetjet e standartizuara ne baze te  $s_{ij}$  dhe jo te  $s$ , atehere vlerat qe rezultojne nga llogaritjet quhen mbetje te sheshuara.

Ne qofte se vrojtimi i i-te eshte nje vlere e larget, duke e larguar kete vrojtim do te rezultojte nje vlere me e vogel per vleresimin e gabimit standart; d.m.th.  $s_{ij} < s$ . Me qene se mbetja e standartizuar e i-te llogaritet me ane te  $s_{ij}$ , vlera absolute per mbetjet e standartizuara te sheshuara do te jete me e madhe se sa vlera e mbetjeve te standartizuara.

Shume paketa software statistikore japin mundesi per gjetjen e mbetjeve te sheshuara si pjese e rezultateve te analizese se regresionit. Ne tabelen 15.15 jane paraqitur vlerat e mbetjeve te sheshuara per problemin e transportit qe perfshin 10 vrojtme dhe dy variabla te pavarur. Per identifikimin e vlerave te largeta ne kete bashkesi te dhenash duhet te shlikojme vlerat e medha absolute per mbetjet e sheshuara.

Per te pare se sa e madhe duhet te jete mbetja e i-te e sheshuar, qe vrojtimi i i-te konsiderohet si vlere e larget, perdoret shprehja  $t$ . Kujtojmë se  $p$  eshte numri i variablave te pavarur te perfshire ne model, kurse  $n$  tregon numrin e vrojtimeve ne bashkesine fillestare te te dhenave; keshtu ne qofte se largojme nga te dhenat vrojtimin e i-te, numri i vrojtimeve reduktohet ne  $n-1$  dhe si rrjedhoje shuma e katroreve te gabimeve ka  $n - p - 2$  shkalle lirie. Per problemin e transportit me 10 vrojtme dhe 2 variabla te pavarur numri i shkalleve te lrlise eshte  $10 - 2 - 2 = 6$ . Per nivelin e rendesise 0.05 dhe numrin e shkalleve te lrlise 6, vlera kritike e kriterit  $t$  eshte  $t = 2.447$ . Keshtu ne qofte se mbetja e standartizuar e sheshuar eshte me vogel se  $-2.447$ , ose me madhe se  $2.447$ , per nivelin e rendesise 0.05 arrijme ne perfundimin se vrojtmi i i-te eshte nje vlere e larget. Meqenese per problemin tone asnje nga mbetjet e sheshuara nuk e kalon ne vlere absolute vleren kritike te  $t$ , themi se ne te dhenat e vrojtuar nuk ekziston asnje vlere e larget.

**Vrojtimet me ndikim te madh** Ne kapitullin 14 pame se si mund te zbulojme me anen e "krahut" vrojtimet per te cilat vlera e variablit te pavarur ka nje ndikim te madh ne rezultatet e analizese se regresionit. Sjellim nder mend se krahu i nje vrojtmi ne rastin e ekuacionit te thjeshte te regresionit u percaktua ne fonnen e meposhtme.

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \quad 15.24$$

Edhe ne analizen e regresionit te shumfishte "krahu" mund te perdoret per identifikimin e vlerave me ndikim te madh ne koeficientet e ekuacionit te regresionit. Nje vrojtim konsiderohet me ndikim te madh ne qofte se  $h_i > 3(p+1)/n$ , ku  $p$  tregon numrin e variablave te pavarur dhe  $n$  numrin e vrojtimeve. Ne rastin e problemit te transportit  $n=10$  dhe  $p=2$ , si rrjedhoje vlera kritike per krahun eshte  $3(2+1)/10=0.9$ .

**Perdorimi i treguesit te distances per identifikimin e vrojtimeve me ndikim te madh.** Nje problem qe del gjate perdorimit te krahut per identifikimin e vrojtimeve me ndikim te madh, eshte konsiderimi i nje vrojtmi me ndikim te madh kur ai nuk eshte i tille, d.m.th. kur ai nuk ka influence ne ekuacionin e regresionit. P.sh. ne tabelen 15.7 tregohet nje bashkesi te dhenash prej 8 vrojtme me vlerat koresponduese te krahut. Me qene se vlera e krahut per vrojtimin e 8-te eshte  $0.91 > 0.86$  (vlera kritike e krahut per ekuacionin e thjeshte te regresionit), ky vrojtim identifikohet si vrojtim me ndikim te madh. Por perpara se te arrijme ne nje konkluzion te tille le te shohim situaten nga nje kendveshtrim tjeter.

Nga zgjidhja kemi gjetur ekuacionin e regresionit  $y = 18.2 + 1.39x$ . Ne qofte se tani heqim vrojtimin  $x = 15$ ,  $y = 39$  nga bashkesia e te dhenave dhe gjejme perseri ekuacionin e regresionit per te dhenat e mbetura, ai rezulton  $y = 18. + 1.42x$ .

Per kete mund te perdoret nje tregues tjeter i quajtur treguesi i distances, i cili per kete qellim perdor njekohosisht si vlerat e mbetjeve ashtu dhe vlerat e krahut te vrojtimeve. Treguesi i distances paraqitet



$$D_i = \frac{(y_i - \bar{y})^2}{(p+1)s^2} \left[ \frac{h_i}{(1-h_i)^2} \right]$$

15.25

ku  $p$  = numri i variablave te pavarur  
 $S^2$  = varianca e vleresimit (ne baze te  $n$  vrojttimeve)  
 $h_i$  = krahu i vrojtimit te  $i$ -te ( ne baze te  $n$  vrojttimeve)

Tabela 15.17 Te dhenat qe mund te perbejne problem ne perdorimin e krahut te forces per zbulimin e vlerave me ndikim

| x  | y  | h        |
|----|----|----------|
| 1  | 18 | 0,204170 |
| 1  | 21 | 0,204170 |
| 2  | 22 | 0,164205 |
| 3  | 21 | 0,138141 |
| 4  | 23 | 0,125977 |
| 4  | 24 | 0,125977 |
| 5  | 26 | 0,127715 |
| 15 | 39 | 0,909644 |

Tabela 15.18 Treguesi i distances

| X  | y  | $D_i$    |
|----|----|----------|
| 1  | 18 | 0,265344 |
| 1  | 21 | 0,230628 |
| 2  | 22 | 0,089443 |
| 3  | 21 | 0,113389 |
| 4  | 23 | 0,029850 |
| 4  | 24 | 0,004164 |
| 5  | 26 | 0,044397 |
| 15 | 39 | 0,055598 |

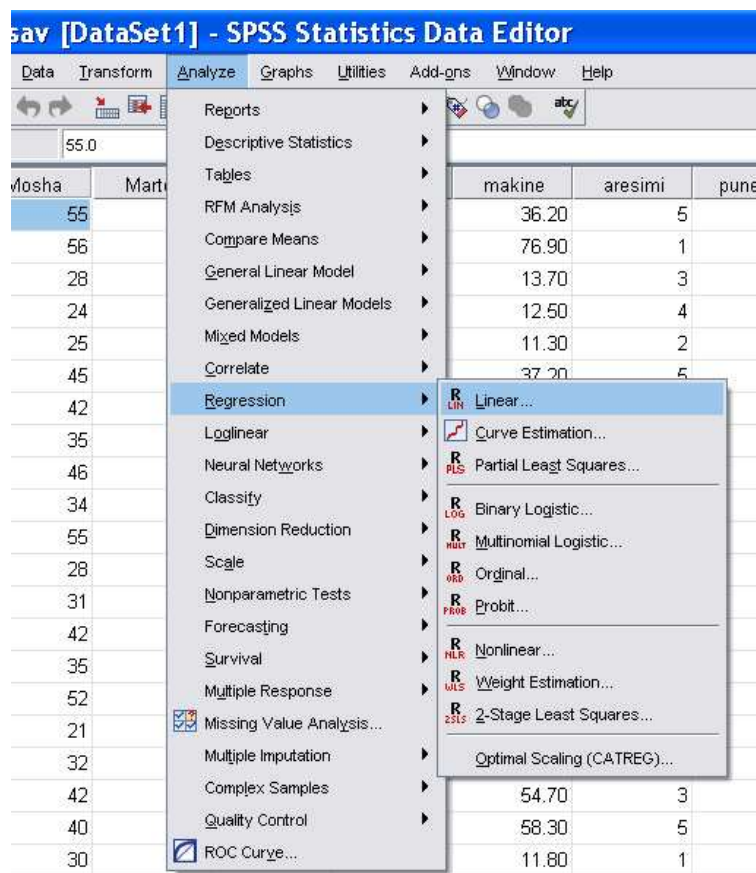
Shohim se vlera e  $D_i$  varet nga vlera e mbetjes se  $i$ -te ( $y_i - \bar{y}_i$ ) dhe nga krahu i vrojtimit te  $i$ -te ( $h_i$ ). Keshtu vrojtimi i  $i$ -te mund te jete me ndikim ne qofte se (a) ka mbetje te madhe dhe krah te vogel, (b) ka mbetje te vogel dhe vlere te madhe te krahut dhe (c) kur si mbetja ashtu dhe krah jane qe te dyja me vlere madhe. Per nje shikim te shpejte, vlerat e  $D_i > 4$  konsiderohen te mjaftueshme per te konkluduar se vrojtimi i  $i$ -te eshte me ndikim te madh, por ato duhet te shqyrtohen me tej. Ne tabelen 15.18 jane paraqitur vlerat e treguesit te distances per te dhenat e tabelës 15.17. Me qene se asnje nga vlerat nuk e kalon 1 arrijme ne perfundimin se bashkesia e te dhenave nek permban asnje vrojtim me ndikim.

## TEMA 12:

### Analiza e Regresionit mbështetur ne OLS (Ordinary Least square)

#### APLIKIMI NË SPSS

#### 12/1- Regresioni I thjeshtë Linear



**Linear Regression**

Dependent: Shpenzime mujore per karburant ne E...

Block 1 of 1

Previous Next

Independent(s): Te ardhura te familjes ne mije Dollare ...

Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options...

Mosha ne vite [Mosha]  
 Gjendja Martesore [...]  
 Vite ne adresen aktu...  
 Te ardhura te familje...  
 Niveli i aresimit [aresi...  
 Vite ne punedhenesi...  
 Kenaqesia ne pune [...]  
 Gjinia [gjinia]  
 Numri i anetareve te ...  
 Sherbim wireless [w...  
 Sherbim Voice mail [...  
 Internet [internet]  
 Caller ID [callid]  
 Call waiting [callwait]  
 Zoteron TV [TV]  
 Zoteron CD player [c...  
 Zoteron kompjuter [k...

**Model Summary**

| Model | R                 | R Square    | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|-------------|-------------------|----------------------------|
| 1     | .751 <sup>a</sup> | <b>.563</b> | .563              | 14.34199                   |

a. Predictors: (Constant), Të ardhura të familjes në mijë Dollarë

**ANOVA<sup>b</sup>**

| Model |            | Sum of Squares | df   | Mean Square | F        | Sig.                    |
|-------|------------|----------------|------|-------------|----------|-------------------------|
| 1     | Regression | 530451.561     | 1    | 530451.561  | 2578.855 | <b>.000<sup>a</sup></b> |
|       | Residual   | 410973.981     | 1998 | 205.693     |          |                         |
|       | Total      | 941425.542     | 1999 |             |          |                         |

a. Predictors: (Constant), Të ardhura të familjes në mijë Dollarë

# ANOVA<sup>b</sup>

| Model |            | Sum of Squares | df   | Mean Square | F        | Sig.                    |
|-------|------------|----------------|------|-------------|----------|-------------------------|
| 1     | Regression | 530451.561     | 1    | 530451.561  | 2578.855 | <b>.000<sup>a</sup></b> |
|       | Residual   | 410973.981     | 1998 | 205.693     |          |                         |
|       | Total      | 941425.542     | 1999 |             |          |                         |

a. Predictors: (Constant), Të ardhura të familjes në mijë Dollarë

b. Dependent Variable: Shpenzime mujore për karburant në Euro

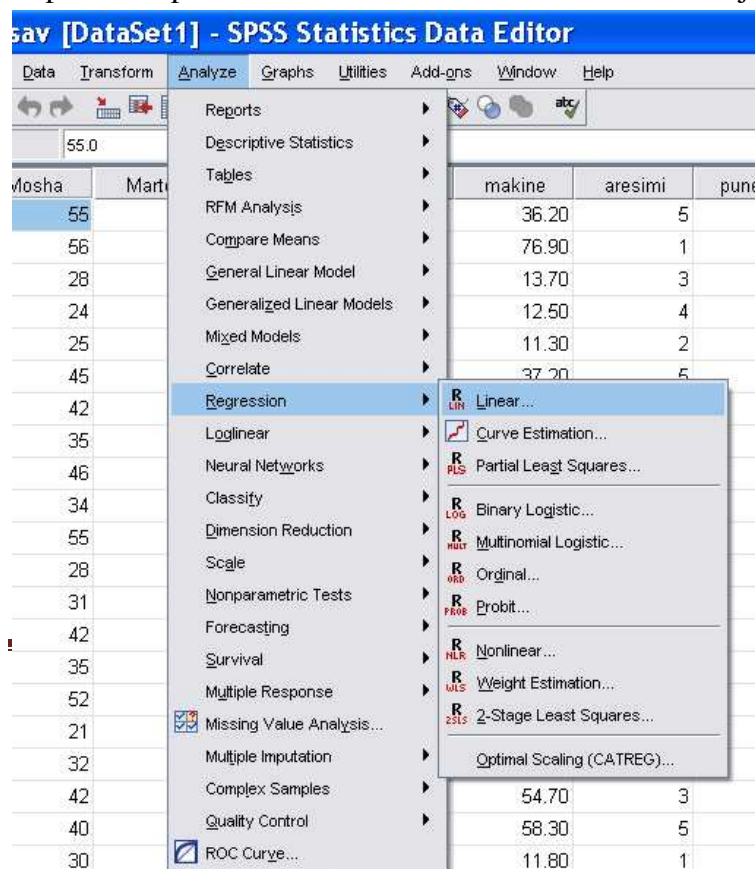
Lidhja statistikisht e rëndësishme (testi për rëndësinë statistikore të R<sup>2</sup>)

# Coefficients<sup>a</sup>

| Model |                                        | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig.        |
|-------|----------------------------------------|-----------------------------|------------|---------------------------|--------|-------------|
|       |                                        | B                           | Std. Error | Beta                      |        |             |
| 1     | (Constant)                             | 16.520                      | .417       |                           | 39.610 | .000        |
|       | Te ardhura te familjes ne mije Dollare | .194                        | .004       | .751                      | 50.782 | <b>.000</b> |

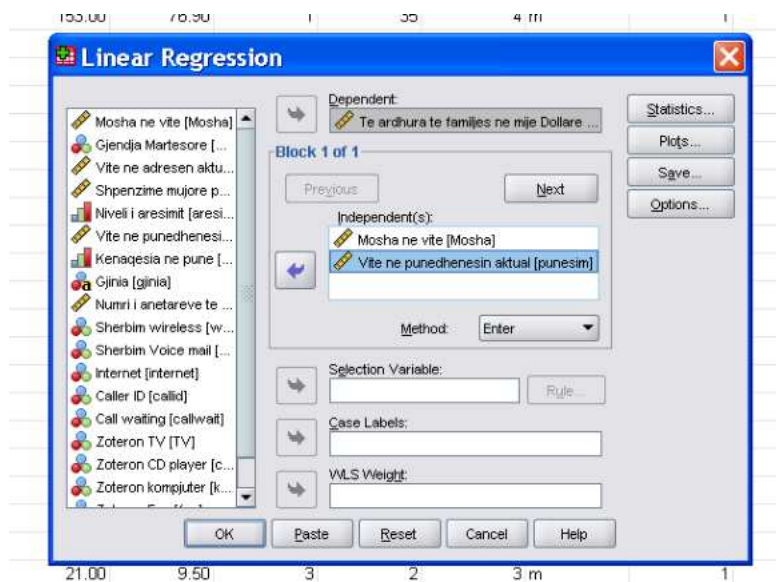
a. Dependent Variable: Shpenzime mujore për karburant në Euro

Shpenzimet për karburant = 16.52 + 0.194 të ardhura mujore të familjes



Koefiçienti I regresionit statistikisht I rëndësishëm

## 12/2- Regresioni I shumfishtë Linear



**Model Summary**

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | .570 <sup>a</sup> | .325     | <b>.324</b>       | 68.98327                   |

a. Predictors: (Constant), Vite ne punedhenesin aktual, Moshë ne vite

**ANOVA<sup>b</sup>**



| Model |            | Sum of Squares | df   | Mean Square | F       | Sig.                    |
|-------|------------|----------------|------|-------------|---------|-------------------------|
| 1     | Regression | 4576012.989    | 2    | 2288006.495 | 480.806 | <b>.000<sup>a</sup></b> |
|       | Residual   | 9503107.290    | 1997 | 4758.692    |         |                         |
|       | Total      | 1.408E7        | 1999 |             |         |                         |

a. Predictors: (Constant), Vite në punëdhënësin aktual, Mosha në vite

b. Dependent Variable: Të ardhura të familjes në mijë Dollarë

Lidhja statistikisht e rëndësishme

**Coefficients<sup>a</sup>**

| Model |                             | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig.        |
|-------|-----------------------------|-----------------------------|------------|---------------------------|--------|-------------|
|       |                             | B                           | Std. Error | Beta                      |        |             |
| 1     | (Constant)                  | 24.848                      | 5.933      |                           | 4.188  | <b>.000</b> |
|       | Mosha ne vite               | -.209                       | .164       | -.030                     | -1.276 | <b>.202</b> |
|       | Vite ne punedhenesin aktual | 5.129                       | .206       | .589                      | 24.839 | <b>.000</b> |

a. Dependent Variablë: Të ardhura të familjes në mijë Dollarë

Të ardhura = 24.84 -0.209 Mosha në vite +5.128 Vite në punëdhënësin aktual

Koefiçienti I Moshës Statistikisht I parëndësishëm

Koefiçienti I Viteve Statistikisht I rëndësishëm

## TEMA 12:

### ANALIZA E REGRESIONIT MBËSHTETUR NE OLS (ORDINARY LEAST SQUARE)

#### Ushtrime

**Ushtrim 1:** Në një zgjedhje rasti të 7 klinikave mjeksore u përftuan të dhënat e mëposhtme për vëllimin e te ardhurave dhe shpenzimet për reklamë në milion lekë.

|                        |    |    |    |    |    |    |    |
|------------------------|----|----|----|----|----|----|----|
| Shpenzimet për reklamë | 1  | 2  | 4  | 3  | 4  | 6  | 8  |
| Vëllimi i te ardhurave | 15 | 20 | 29 | 25 | 30 | 35 | 40 |

1. Gjeni ekuacionin e regresionit të thjeshtë linear.
2. Sa do të jetë vëllimi i mundshëm i te ardhurave në se shpenzimet për reklamë parashikohen 5 milion lekë ?
3. Verifikoni rëndësinë e koeficientit të regresionit  $b_1$  për  $\alpha = 0,05$ .
4. Llogaritni koeficientin e përcaktueshmërisë dhe interpretoni kuptimin e tij.

**Ushtrim 2:** Një administrator i një qendre spitalore dëshiron të vlerësojë një regresion linear të thjeshtë mes pacienteve dhe orëve të punës jashtë kohës normale. Rezultatet e regresionit tregojnë:  $SKG = 12745$  dhe  $SKR = 13500$ .

1. Llogaritni koeficientin e përcaktueshmërisë në këtë regresion dhe interpretoni kuptimin e tij duke njohur vëllimin e zgjedhjes të barabartë me 10-të.
2. A besoni se ky model regresioni do të jetë një mjet i rëndësishëm për parashikimin e orëve jashtë orarit duke u bazuar në numrin e pacientëve? Spjegoni.

**Ushtrimi 3:** Nga një regresion i thjeshtë linear i të ardhurave të një qendre përkundrejt te ardhurave të një klinike tjetër u përftuan të dhënat e mëposhtme:  $n = 128$ ,  $SKG = 3343$ ,  $SKR = 45678$ .

1. Llogaritni shumën e katrorëve totalë, katrorët mesatarë, raportin F dhe koeficientin e përcaktueshmërisë.
2. Bëni një tabelë ANOVA për këtë regresion. Spjegoni lidhjen mes të ardhurave të të dy qendrave?

**Ushtrimi 4:** Një prodhues programesh kompjuterash dëshiron të studiojë lidhjen mes numrit të kompjuterave në përdorim në zona të ndryshme dhe numrit të programeve që kompania shet në ato zona. Një analizë e regresionit të thjeshtë linear në 21 zona dha këto rezultate:  $b_0 = 12.43$ ,  $b_1 = 1.076$ ,  $s_{b_0} = 13.65$ ,  $s_{b_1} = 0.083$ ,  $SKG = 1076.11$ ,  $\sum x^2 - (\sum x)^2 / n = 72641$ ,  $\bar{x} = 12453$  kompjutera.

1. Shkruani ekuacionin e regresionit dhe interpretonitë regresionit.
2. Llogaritni koeficientin e përcaktueshmërisë; provoni rëndësinë e tij.

**Ushtrimi 5 :** Paga e personave që kanë dy punë, një të rëndësishme dhe tjetër më pak të rëndësishme, mund të ndikohet nga shumë faktorë.

Për të analizuar këtë problem u bë një zgjedhje prej 318 personash dhe u ndërtua një model regresioni:

$$P_g = 37.07 + 0.403p_1 - 90.06r + 75.51u + 47.33H + 113.64z + 2.26m \quad R^2 = 0.34$$

(0.062)      (24.47) (21.60) (23.42) (27.62) (0.94)

shkallët e lirisë = 311

ku, numrat në kllapa janë devijimet standarte të vlerësimeve të parametrave përkatës,

$P_g$  = Paga e punonjësit të dypunësuar

$p_1$  = Paga e punës së parë

$r$  = raca: 0 - i, e bardhë dhe 1 - me ngjyrë

$u$  = zona e venbanimit : 0 - urban dhe 1 - jo-urban

$H$  = me shkollë të lartë ose jo: 0 - pa shkollë të lartë dhe 1 - me shkollë të lartë

$z$  = zona e vendlindjes: 0 - veri dhe 1 - jo-veri

$m$  = mosha në vite

1. Analizoni rezultatet e arritura.
2. Verifikoni rëndësinë e secilit variabël në model për  $\alpha = 5\%$ .

# **TEMA 13:**

## **ANALIZA KORELATIVE: KORELACIONI PEARSON DHE TË TJERË KOEFICIENTË KORELACIONI**

### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje që :

- Të kuptojnë ç'është dhe përse shërben analiza korelative ndërmjet variablave
- Të dinë ç'është në vëtvete koeficienti i korelacionit, cilat janë vlerat e mundshme të tij që tregojnë për një marrëdhënie të fortë apo të dobët mes variablave
- Të dinë cili koeficient korelacioni është i përshtatshëm të përdoret në varësi të variablave me shkallë matje të ndryshme
- Të llogarisin, të interpretojnë dhe të testojnë rëndësinë statistiore të koeficientit të korelacionit Pearson
- Të llogarisin, të interpretojnë dhe të testojnë rëndësinë statistiore të koeficientit të korelacionit të Spirmanit
- Të llogarisin dhe interpretojnë koeficientët e tjerë të korelacionit
- Të përdorin ekuacionin e regresionit për parashikime

Të analizojnë dhe të interpretojnë outputin në SPSS të korelacioneve të lartpërmendura

### **LITERATURA**

- “ Applied Statistics”, Warner A.R , 2008
- Statistika”- Grup Autoresh, Shtëpia botuese Pegi, (2016)
- Statistical for Business nad Economics”, Newbold P, Carlson W, Thorne B, 2011
- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011
- “Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014

## TEMA 13

### ANALIZA KORELATIVE

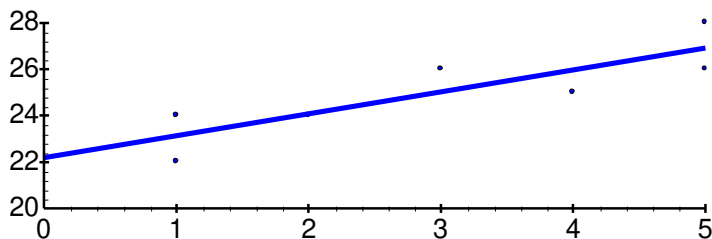
#### 1. Përcaktimi i Koeficientit të Korelacionit PEARSON

Në disa raste marrësi i vendimit nuk është i interesuar aq për ekuacionin që lidh dy variablat, se sa për matjen e shkallës së fortësisë së lidhjes midis tyre. Për matjen e fortësisë së lidhjes midis variablave përdoret një procedurë statistikore, që mban emrin analiza korelative. Si rezultat i studimit korelativ merret i ashtuquajturi koeficient i korelacionit. Nga mënyra se si përcaktohet ai, vlerat e koeficientit të korelacionit janë midis -1 dhe +1. Vlera +1 e tij tregon se të dy variablat janë në lidhje lineare të plotë dhe në të njëjtin drejtim. D.m.th. të gjitha pikat shtrihen në një vijë të drejtë me koeficient këndor pozitiv. Vlera -1 e koeficientit të korelacionit tregon se variablat janë në lidhje lineare të plotë dhe me drejtim të kundërt. Në këtë rast të gjitha pikat shtrihen në një vijë të drejtë me koeficient këndor negativ.

Si shembull ilustrimi të analizës korelative marrim rastin e shitjeve në një magazinë industriale. Manaxheri i saj kërkon të vërtetojë nëse ekziston një lidhje korelative midis numrit të shitësve  $x$  dhe vëllimit të shitjeve  $y$ . Të dhënat e vrojtuar paraqiten në tabelën e mëposhtme.

*Tabela 1. Të dhënat e zgjedhjes për vëllimin e shitjeve dhe numrin e shitësve dhe më poshtë paraqitja grafike e tyre*

|                    |    |    |    |    |    |    |    |
|--------------------|----|----|----|----|----|----|----|
| Numri i shitësve   | 2  | 5  | 1  | 3  | 4  | 1  | 5  |
| Vëllimi i shitjeve | 24 | 28 | 22 | 26 | 25 | 24 | 26 |



Nga grafiku duket se ekziston një lidhje lineare pozitive midis variablave. Për matjen e shkallës së lidhjes lineare midis  $x$  dhe  $y$ , në fillim përcaktojmë një masë të shoqërimit linear të këtyre variablave e njohur me emrin kovariancë.



**Kovarianca** Kovarianca e zgjedhjes përcaktohet si më poshtë:

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (1)$$

*Tabela 1. Llogaritja e kovariancës për të dhënat e zgjedhjes*

| $x_1$ | $y_1$ | $x_1 - \bar{x}$ | $y_1 - \bar{y}$ | $(x_1 - \bar{x})(y_1 - \bar{y})$ |
|-------|-------|-----------------|-----------------|----------------------------------|
| 2     | 24    | -1              | -1              | 1                                |
| 5     | 28    | 2               | 3               | 6                                |
| 1     | 22    | -2              | -3              | 6                                |
| 3     | 26    | 0               | 1               | 0                                |
| 4     | 25    | 1               | 0               | 0                                |
| 1     | 24    | -2              | -1              | 2                                |
| 5     | 26    | 2               | 1               | 2                                |
| 21    | 175   | 0               | 0               | 17                               |

Në këtë formulë çdo vlerë e x çiftëzohet me një vlerë të y. Pas kësaj shumojmë produktet e shmangieve të vlerave individuale të variablave nga vlerat mesatare të tyre dhe rezultatin e gjetur e pjestojmë me n-1.

Për të matur forcën e lidhjes lineare midis numrit të shitësve dhe vëllimit të shitjeve, llogaritim kovariancën e zgjedhjes sipas formulës së mëposhtme:

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{17}{6} = 2,8333$$

Siç shihet nga formula, në vend të mesatareve të variablave për popullimin janë përkatësisht mesataret e tyre për zgjedhjen. Kështu kovarianca e zgjedhjes është një vlerësim për kovariancën e popullimit, të bazuar në treguesit mesatarë të zgjedhjes dhe në vëllimin e saj n.

### **Interpretimi i kovariancës**

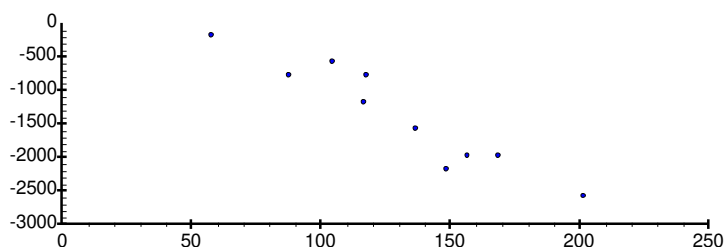
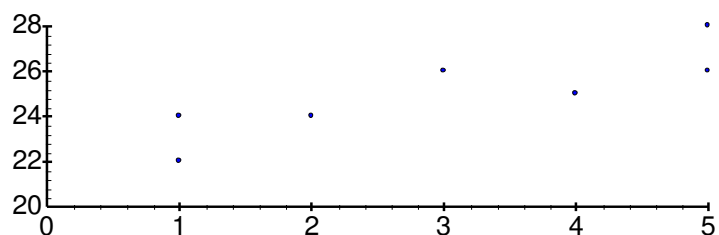
Në qoftë se vlera e  $s_{xy}$  është pozitive, pikat që kanë patur efekt më të madh tek ajo do t'i takojnë kuadrantit të parë dhe të tretë. Kështu vlera pozitive e  $s_{xy}$  shërben si tregues i një lidhjeje lineare midis x dhe y, d.m.th. kur vlerat e x rriten edhe vlerat e y gjithashtu rriten. Në

qoftë se  $s_{xy}$  është negative, atëherë peshë më të madhe në të shumtën e produkteve do të kenë pikat që bien në kuadrantin e dytë dhe të katërt. Kështu vlera negative e  $s_{xy}$  tregon për një lidhje lineare me drejtim të kundërt midis  $x$  dhe  $y$ , d.m.th. kur vlerat e  $x$  rriten vlerat e  $y$  zvogëlohen. Së fundi, në qofte se pikat janë të shpërndara uniformisht në të katër kuadrantet, vlera e  $s_{xy}$  është afër zeros, duke treguar se nuk ekziston asnjë lidhje lineare midis  $x$  dhe  $y$ . Në figurën 1 tregohen vlerat që mund të priten për  $s_{xy}$  në rastin e tre diagramave të tipeve të ndryshme.

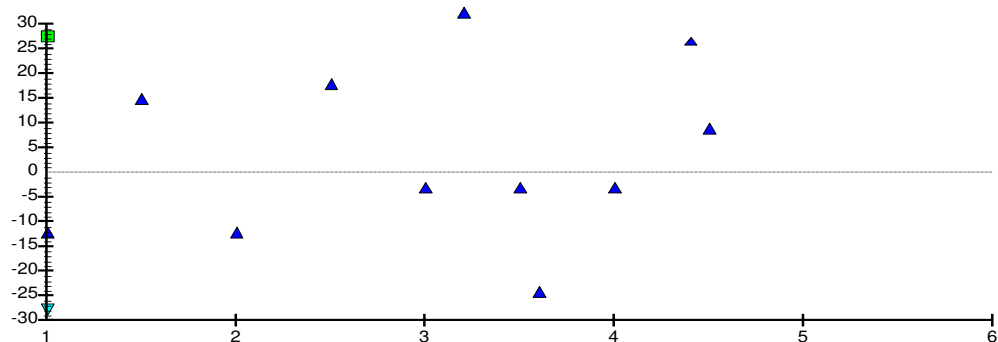
Nga diskutimi i mësipërm duket se një vlerë pozitive e madhe për kovariancën është tregues për një lidhje lineare të fortë, kurse një vlerë negative e madhe e saj flet për një lidhje lineare të fortë por më drejtim të kundërt. Megjithatë një problem që mund të dalë për matjen e fortësisë së lidhjes me përdorimin e kovariancës, është se vlerat e saj ndikohen nga njësiti fizike të shprehjes të karakteristikave. P. sh. le ta zemë se jemi të interesuar për lidhjen midis lartësisë  $x$  dhe peshës  $y$  për individët. Në qofte se lartësia matet në mm, atëherë për shumën e produkteve të shmangieve do të gjejmë një vlerë më të madhe se sa kur ajo matet në cm. Si rrjedhojë do të gjenim një vlerë më të madhe për kovariancën, ndërkohë kur nuk ka asnjë ndryshim për lidhjen që është si më parë. Treguesi që mat lidhjen duke mënjënuar njësitet e matjes të variablave është koeficienti i korelacionit.

*Figura 1 Interpretimi i kovariancës së zgjedhjes*

$s_{xy} > 0$ . Lidhje lineare me drejtim të njëjtë,  $s_{xy} < 0$  Lidhje lineare me drejtim të kundërt



$s_{xy} = 0$  Nuk ekziston lidhje



**Koeficienti i korelacionit** Koeficienti i korelacionit për zgjedhjen shprehet me anën e formulës:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (2)$$

ku,  $r_{xy}$  koeficienti i korelacionit të zgjedhjes,  $s_{xy}$  kovarianca e zgjedhjes,  $s_x$  shmangia standarte e zgjedhjes për x,  $s_y$  shmangia standarte e zgjedhjes për y.

Ekuacioni (2) tregon se koeficienti i korelacionit të zgjedhjes llogaritet duke pjesëtuar kovariancën e zgjedhjes me produktin e devijimeve standarte të variablave x dhe y. Le të llogaritim koeficientin e korelacionit për të dhënat e shembullit për varësinë e vëllimit të shitjeve nga numri i shitësve

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{18}{6}} = 1,7321 \quad s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{22}{6}} = 1,9149$$

Me qenë se  $s_{xy} = 2,8333$ , për koeficientin e korelacionit kemi:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{2,8333}{(1,7321)(1,9149)} = 0,854$$

Për llogaritjen e koeficientit të korelacionit e mësipërme transformohet në një formë tjetër, e cila mund të zbatohet drejtëpërdrejtë mbi bazën e të dhënave fillestare. Ajo ka formën e mëposhteme:

$$r_{xy} = \frac{\sum x_i y_i - (\sum x_i \sum y_i) / n}{\sqrt{\sum x_i^2 - (\sum x_i)^2 / n} \sqrt{\sum y_i^2 - (\sum y_i)^2 / n}} \quad (3)$$

*Tabela 2*

| $x_1$ | $y_1$ | $(x_1 y_1)$ | $x_1^2$ | $y_1^2$ |
|-------|-------|-------------|---------|---------|
| 2     | 24    | 48          | 4       | 576     |
| 5     | 28    | 140         | 25      | 784     |
| 1     | 22    | 22          | 1       | 484     |
| 3     | 26    | 78          | 9       | 676     |
| 4     | 25    | 100         | 16      | 625     |
| 1     | 24    | 24          | 1       | 576     |
| 5     | 26    | 130         | 25      | 676     |
| 21    | 175   | 542         | 81      | 4397    |

Algjebrikisht formulat (2) dhe (3) janë ekuivalente. Në tabelën 2 jepen llogaritjet e nevojshme për përdorimin e formulës (3). Duke përdorur këto llogaritje gjejmë.

$$r_{xy}=17/19,8997=0.854$$

Kështu shohim se vlera e gjetur për koeficientin e korelacionit me formulën (3) është e njëjtë sa vlera e tij e gjetur me anën e formulës (2). Formula e llogaritjes të koeficientit të korelacionit të popullimit, që shënohet me  $\rho$  paraqitet si më poshtë:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (4)$$

Duke krahasuar këtë formulë me atë të zgjedhjes shohim se ato janë të ngjashme, me të vetmin ndryshim se në vend të treguesve të zgjedhjes janë ato të popullimit. Nga kjo rrjedh se koeficienti i korelacionit për zgjedhjen është një vlerësim për koeficientin e korelacionit të popullimit.

**Interpretimi i koeficientit të korelacionit** Në fillim le të marrim një shembull që ilustron konceptin e një lidhjeje lineare pozitive dhe të plotë. Le të jenë tre pikat e një lidhje të drejtpërdrejtë lineare  $A_1(1,10)$ ,  $A_2(2,30)$  dhe  $A_3(3,50)$ . Këto pika ndodhen në një vijë të drejtë, gjë që tregon se lidhja midis  $x$  dhe  $y$  është lineare pozitive dhe e plotë. Llogaritjet e nevojshme për koeficientin e korelacionit tregohen në tabelën 3. Duke përdorur vlerat e kësaj table gjejmë se  $r_{xy}=40/40=1$ .

*Tabela 3. Llogaritjet për koeficientin e korelacionit të barabartë me 1*

| $x_1$ | $y_1$ | $x_1 y_1$ | $x_1^2$ | $y_1^2$ |
|-------|-------|-----------|---------|---------|
| 1     | 10    | 10        | 1       | 100     |
| 2     | 30    | 60        | 4       | 900     |
| 3     | 50    | 150       | 9       | 2500    |
| 6     | 90    | 220       | 14      | 3500    |

Kështu shohim se vlera e koeficientit të korelacionit të zgjedhjes është e barabartë me 1. Në përgjithësi mund të tregohet se në qoftë se pikat e një bashkësie të dhënash bien në vijën e drejtë që ka koeficient këndor pozitiv, atëherë vlera e koeficientit të korelacionit është e barabartë me 1. D.m.th. vlerës së koeficientit të korelacionit të barabartë me 1 i korespondon një lidhje e plotë. Nga ana tjetër, në qoftë se pikat e një bashkësie të dhënash ndodhen në një vijë të drejtë me koeficient këndor negativ, vlera e koeficientit të korelacionit është e barabartë me -1. Kështu vlerës së koeficientit të korelacionit të barabartë me -1 i përgjigjet një lidhje e plotë si në rastin e parë, por me drejtim të kundërt.

Le të supozojmë tani se për disa të dhëna ekziston një lidhje lineare pozitive midis  $x$  dhe  $y$ , por që nuk është lidhje e plotë. Në këte rast vlera e koeficientit të korelacionit do të jetë më e vogël se 1, gjë që shfaqet edhe në figurë, ku të gjitha pikat nuk bien mbi vijën e drejtë. Sa më shumë që pikat të shmangen nga vija e drejtë, aq më shumë do të zvogëlohen vlerat e koeficientit të korelacionit. Vlera e koeficientit të korelacionit e barabartë me zero tregon që midis  $x$  dhe  $y$  nuk ekziston lidhje lineare, kurse vlera e koeficientit të korelacionit afër zeros tregon për një lidhje të dobët midis tyre. Kujtojmë se për të dhënat për lidhjen midis vëllimit të shitjeve në varësi nga numri i shitësve, koeficienti i korelacionit ka rezultuar i barabartë me 0,854. Me qenë se ky është afër njëshit lidhja midis tyre është e fortë, ndërsa qenia pozitive e tij tregon se lidhja është në të njëjtin drejtim. Në këtë rast themi se me rritjen e numrit të shitësve rritet edhe vëllimi i shitjeve.

Më lart kemi thënë se vlerat e koeficientit të korelacionit në madhësi absolute afër 1, tregojnë për një lidhje të fortë lineare midis variablave, kurse vlerat e koeficientit të korelacionit afër zeros flasin për një lidhje lineare të dobët midis tyre. Megjithatë në rastin kur koeficienti i korelacionit është i barabartë me zero duhet të jemi të kujdesshëm dhe të mos arrijmë menjëherë në përfundimin se midis variablave nuk ekziston asnjë lidhje. Në raste të tilla mund të themi se nuk ekziston lidhje lineare midis variablave dhe jo në përgjithësi që nuk ekziston lidhje, sepse mundet që midis tyre të ekzistojë një lidhje vijë përkulët apo ndonjë lidhje tjetër funksionale dhe madje e plotë.



Në mbyllje të kësaj pjese për korelacionin theksojmë se, ndërsa një koeficient korelacioni afër + 1 ose -1 tregon për një lidhje të fortë lineare midis dy variablave, ai nuk flet kurresesi për shkakun dhe pasojën e lidhjes. Përfundimet mbi shkakun dhe pasojën e lidhjes duhet të bëhen vetëm në bazë të gjykimit të studiuesit.

**Përcaktimi i koeficientit të korelacionit të zgjedhjes në bazë të rezultateve të analizës së regresionit** Le ta zemë se kemi gjetur ekuacionin e regresionit me anën e metodës të katrorëve më të vegjël,  $y = b_0 + b_1x$ . Në këtë rast koeficienti i korelacionit të zgjedhjes mund të llogaritet duke përdorur njërën nga formulat e mëposhtme.

$$r_{xy} = (\text{sign. } b_1) \sqrt{\text{koeficienti i përcaktueshmerise}} = \pm \sqrt{r^2} \quad (5)$$

$$r_{xy} = b_1 (s_x / s_y) \quad (6)$$

ku  $b_1$  = koeficienti këndor i vijës së ekuacionit të regresionit;  $s_x$  = devijimi standart i zgjedhjes për x;  $s_y$  = devijimi standart i zgjedhjes për y

Theksojmë se shenja e koeficientit të korelacionit të zgjedhjes është e njëjtë me shenjën e koeficientit  $b_1$ .

**Kontrolli i rëndësisë së lidhjes** Koeficienti i korelacionit të zgjedhjes është një vlerësim pikësor i koeficientit të korelacionit të popullimit. Kriteri statistikor për vlerësimin e rëndësisë së lidhjes lineare midis x dhe y, ka të bëjë me kontrollin e hipotezës për vlerën zero të koeficientit të korelacionit të popullimit. Hipotezat që kërkohet të kontrollohen janë si më poshtë:

$$H_0 = \rho_{xy} = 0$$

$$H_a = \rho_{xy} \neq 0$$

Mund të tregohet se kontrolli i këtyre hipotezave është ekuivalent me kontrollin e hipotezave për rëndësinë e koeficientit të regresionit  $b_1$ .

Hipotezat që duhet të kontrollohen në këtë rast janë.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

Me qenë se për problemin e kioskave kemi provuar se koeficienti i regresionit është i rëndësishëm, pra kemi hedhur poshtë hipotezën zero, atëherë edhe hipoteza zero për koeficientin e korelacionit hidhet poshtë. Megjithatë kontrolli i hipotezës zero për koeficientin e korelacionit mund të bëhet edhe në mënyrë të drejtpërdrejtë, pa qenë nevoja e kryerjes më parë të analizës së regresionit.

Për këte mund të tregohet se në qofte se hipoteza zero është e vërtetë, atëherë statistika:

$$r_{xy} \sqrt{\frac{n-2}{1-r^2}} \sim T(n-2) \quad (7)$$

Për rastin e të dhënave të shembullit tonë për nivelin e rëndësisë ( $\alpha=0,05$  dhe numrin e shkallëve të lirisë  $n-2 = 10 - 2 = 8$ , shohim se vlera e t në tabelë është e barabartë me 2,306. Kështu në qofte se vlera e formulës (7) e kalon 2,306, ose është më vogël se - 2,306, atëherë hipoteza zero hidhet poshtë, pra arrijmë në përfundimin se koeficienti i korelacionit ndryshon në mënyrë të rëndësishme nga zero. Për koeficientin e korelacionit  $r = 0,95$  vlera e fomulës (7) është 8,63

Meqenëse  $8,63 > 2,306$ , atëherë hipoteza zero hidhet poshtë dhe arrijmë në përfundimin se lidhja midis variablave ka koeficient korelacioni të rëndësishëm. Shohim se ky krititer çon në të njëjtin rezultat me atë që kemi arritur kur vlerësua rëndësinë e koeficientit të regresionit.

### Outputi nga SPSS i koeficientit të korelacionit Pearson

Komanda ne SPSS për Illogaritjen e ketij koeficienti: <Analyze>, < Correlate>, < Bivariate>, < Pearson>

### Correlations

|            |                     | Correlations |            |
|------------|---------------------|--------------|------------|
|            |                     | Arsimi       | TeArdhurat |
| Arsimi     | Pearson Correlation | 1            | .932*      |
|            | Sig. (2-tailed)     |              | .021       |
|            | N                   | 5            | 5          |
| TeArdhurat | Pearson Correlation | .932*        | 1          |
|            | Sig. (2-tailed)     | .021         |            |
|            | N                   | 5            | 5          |

\*. Correlation is significant at the 0.05 level (2-tailed).

## 2. Koeficienti i korelacionit të rangut ( Spearman)

Koeficienti i korelacionit që trajtuam deri tani shërben si një masë e lidhjes midis dy variablave , për matjen e të cilit nevojiten të dhëna të cilat maten në shkallë intervalore ose raporti. Në këtë paragraf do të shohim matjen e korelacionit midis dy variablave të cilët mund të shprehen në

shkallën e zakonshme. Për këtë qëllim mund të përdorët koeficienti i korelacionit të rangut, ose i Sperman, i cili llogaritet si në formulën (8) :

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \quad (8)$$

Ku n është numri i njëisve që renditen , d është diferenca e rangjeve të njësisve të variablave. Ky koeficient i merr vlerat nga -1 ne 1, dhe ka pothuaj të njëjtin shpjegim dhe kuptim për vlerat e tij si dhe korelacioni i Pearson.

Si shembul për llogaritjen e koeficientit të korelacionit te rangut kemi marrë renditjen e pesë studentëve sipas pikëve të marra në provimin e dy lëndëve A dhe B të cilat jepen si më poshtë, tabela 4:

Tabela 4

| Studenti | A | B | C | D | E |
|----------|---|---|---|---|---|
| Lenda A  | 1 | 2 | 3 | 4 | 5 |
| Lenda B  | 5 | 3 | 1 | 4 | 2 |

Mbështetur në llogaritjet e bëra sipas tabelës 5, gjemë korelacionin e rangut :

Tabela 5

| Lenda A | Lenda B | X-Y          | (X-Y) <sup>2</sup> |
|---------|---------|--------------|--------------------|
| X       | Y       | d            | d <sup>2</sup>     |
| 1       | 5       | -4           | 16                 |
| 2       | 3       | -1           | 1                  |
| 3       | 1       | 2            | 4                  |
| 4       | 4       | 0            | 0                  |
| 5       | 2       | 3            | 9                  |
|         |         | $\sum d^2 =$ | 30                 |

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 * 30}{5(5^2 - 1)} = -0.5$$

Korelacioni i rangut të Spearman ka shënje negative që do të thotë se lidhja midis variablit X dhe Y është negative dmth, se me rritjen e rangjeve të njerit variable , rangjet për variablin tjetër

zvogëlohen, Koeficienti i korelacionit të rangut afër  $\pm 1$  tregon përputhje të renditjeve me drejtim të njëjtë ose të kundërt, që do të thotë një lidhje e fortë midis variablave. Nëse koeficienti i korelacionit të rangut është afër zeros pavarësisht nga shenja, tregon se përputhja e rangjeve është e vogël dhe rangjet në të dy popullimet janë të shpërndara në mënyrë të rastit. Në rastin tonë kemi një vlerë të koeficientit të korelacionit të rangut afër zeros, që tregon për një lidhje të moderuar midis renditjeve të studentëve në të dyja lëndët.

Ashtu si në shumë procedura të tjera statistikore që kemi mesuar, mund të duam që me anën e rezultateve të bëjmë vlerësimin për koeficientin e korelacionit të rangut të popullimit. Në rastin tonë koeficienti i korelacionit të rangut për popullimin do të merrej për të gjithë studentet e kursit. Megjithatë, për të mënjeluar grumbullimin e të gjithë të dhënave statistikore, të cilat mund të kërkonin shumë kohë dhe shpenzime, do të bëjmë vlerësimin e koeficientit të korelacionit të rangut të tërësisë nëpërmjet të dhënave të zgjedhjes. Për këtë duhet të bëjmë kontrollin e hipotezave të mëposhtme:

$$H_0: \rho_s = 0$$

$$H_a: \rho_s \neq 0$$

Në kushtet e hipotezës zero kur nuk ekziston korelacioni rangor, rangjet janë të pavarura dhe shpërndarja e zgjedhjes për koeficientin e korelacionit të rangut të zgjedhjes është:

$$\mu_{r_s} = 0 \quad (9)$$

Dhe

$$\sigma_{r_s} = \sqrt{\frac{1}{n-1}} \quad (10)$$

Forma e shpërndarjes është afërsisht normale, për  $n \geq 10$ , koeficienti i korelacionit të rangut përse i përket renditjes së studentëve në dy lëndët rezultoi -0.5. Le të kontrollojmë për këtë vlerë rëndësishme të këtij koeficienti.

Dime që:

$$\mu_{r_s} = 0, \text{ llogarisim } \sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{5-1}} = 0.5$$

Përse i përket statistikës:

$$z = \frac{r - \mu}{\sigma} = \frac{-0.5 - 0}{0.5} = -1$$

Siç dihet për nivelin e rëndësisë 5%, hipoteza zero hidhet poshtë nëse  $z < -1.96$ , ose  $z > +1.96$ . meqenëse Z i vrojtuar në rastin tonë  $+1$ , hipoteza zero për mungesën e korelacionit të rangut nuk mund të hidhet poshtë, ndaj arrijmë në përfundimin se renditja e pikëve të studentëve në të dyja lëndët nuk ndryshon në mënyrë statistikisht të rëndësishme.

### Outputi nga SPSS i koeficientit të korelacionit Spearman

Komanda në SPSS për llogaritjen e këtij koeficienti: <Analyze>, < Correlate>, < Bivariate>, < Spearman>

#### Nonparametric Correlations

##### Correlations

|                |        |                         | LendaA | LendaB |
|----------------|--------|-------------------------|--------|--------|
| Spearman's rho | LendaA | Correlation Coefficient | 1.000  | -.500  |
|                |        | Sig. (2-tailed)         | .      | .391   |
|                |        | N                       | 5      | 5      |
|                | LendaB | Correlation Coefficient | -.500  | 1.000  |
|                |        | Sig. (2-tailed)         | .391   | .      |
|                |        | N                       | 5      | 5      |



### TEMA 13:

#### ANALIZA KORELATIVE: KORELACIONI PEARSON DHE TË TJERË KOEFICIËNTË KORELACIONI

##### Ushtrime

**Ushtrim 1:** Në një zgjedhje rasti të 7 klinikave mjeksore u përfatuan të dhënat e mëposhtme për vëllimin e të ardhurave dhe shpenzimet për reklamë në milion lekë.

|                        |    |    |    |    |    |    |    |
|------------------------|----|----|----|----|----|----|----|
| Shpenzimet për reklamë | 1  | 2  | 4  | 3  | 4  | 6  | 8  |
| Vëllimi i të ardhurave | 15 | 20 | 29 | 25 | 30 | 35 | 40 |

1. Duke supozuar se plotësohen supozimet e analizës regressive lineare, llogaritni koeficientin e korelacionit ndërmjet këtyre dy variablave.
2. Verifikoni rëndësinë e fortësisë së lidhjes për  $\alpha = 0,05$ .

**Ushtrim 2:** Ekuacioni i mëposhtëm i regresionit tregon lidhjen midis numrit të pacientëve dhe koston totale të qëndrimit në një qendër spitalore në një zgjedhje rasti të 20 qendrave spitalore:

$$\hat{y} = 60 + 3.2x$$

Devijimi standard për pacientët është 100 dhe devijimi standard për koston është 350 Euro.

1. Llogaritni koeficientin e korrelacionit.
2. Testoni hipotezat  $H_0 : R_{xy} = 0$  dhe  $H_a : R_{xy} \neq 0$  për  $\alpha = 0.05$

**Ushtrimi 3:** Një menaxher klinike spitalore kërkon të marrë informacion mbi cilësinë e shërimit të mjekëve duke u bazuar në opinionet e pacientëve aktuale dhe atyre të para një viti . Për 10 mjekë të kësaj qendre kemi rezultatet e mëposhtme:

| Mjekët | Rangjet sipas pacientëve aktualë | Rangjet sipas ish pacientëve |
|--------|----------------------------------|------------------------------|
| 1      | 4                                | 6                            |
| 2      | 6                                | 8                            |
| 3      | 8                                | 5                            |
| 4      | 3                                | 1                            |
| 5      | 1                                | 2                            |
| 6      | 2                                | 3                            |
| 7      | 5                                | 7                            |
| 8      | 10                               | 9                            |
| 9      | 7                                | 4                            |
| 10     | 9                                | 10                           |

Ka ndonjë lidhje midis preferencave për të dy grupet ? Përdorni nivelin e rëndësisë 0.05.

**Tema 13**  
**ANALIZA KORELATIVE: KORELACIONI PEARSON DHE**  
**TË TJERË KOEFICIËNTË KORELACIONI**

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

1

**Çështjet Kryesore që Trajtohen:**

**A. Analiza korelative: Koeficienti i Korelacionit PEARSON dhe rëndësia e fortësisë së lidhjes.**

**B. Korelacioni Joparametrik: Korelacioni i rangut të Spearman-it**

**C. Koeficientë të tjerë Korelacioni**

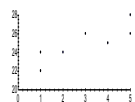
Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

2

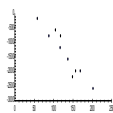
**A. Analiza korelative: Koeficienti i Korelacionit dhe**  
**rëndësia e fortësisë së lidhjes**

Koeficienti i Korelacionit (Pearson correlation)

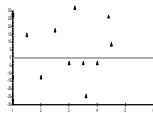
$$-1 \leq r_{xy} \leq +1 \quad r_{xy} = \frac{s_{xy}}{s_x s_y} = b_1 \frac{s_x}{s_y} = \text{sign}(b_1) \sqrt{r^2}$$



$r \approx +1$



$r \approx -1$



$r \approx 0$

Përgatiti: Prof. Dr. BESA SHAHINI    Tiranë    2018

3

$$\text{Ko varianca i logaritetsi } s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

### A. Analiza korelative: Koefficienti i Korelacionit dhe rëndësia e fortësisë së lidhjes

$$H_0: \rho_{xy} = 0$$

$$H_1: \rho_{xy} \neq 0$$

$$H_0 \downarrow \text{ nqse } |t_v| > t_{\alpha/2, n-2} \text{ ku}$$

$$t_v = r_{xy} \sqrt{\frac{n-2}{1-r^2}}$$

### B. Korelacioni Joparametrik: Korelacioni i rangut të Spearman-it

Përdoret kur dy variablat janë në shkallë matje ordinale, ose kur ato mund të derivohen si të tillë.

Korelacioni i matur kështu quhet: Korelacioni i Rangut të Spearman-it.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad \text{ku } d_i = x_i - y_i$$

$$H_0: \rho_s = 0$$

$$H_1: \rho_s \neq 0$$

$$H_0 \downarrow \text{ nqse } |z_v| > z_{\alpha/2} \text{ ku}$$

$$z = \frac{r - \mu}{\sigma}$$

$$\mu_{rs} = 0$$

$$\sigma_{rs} = \sqrt{\frac{1}{n-1}}$$

### C. Koefficientë të tjerë të Korelacionit

| Variabli Y/X | X Sasior                  | X- E Zakonshme          | X – Nominale              |
|--------------|---------------------------|-------------------------|---------------------------|
| Y- Sasior    | <u>Person</u> $r_{xy}$    | Biserial $r_b$          | Biserial pikesor $r_{pb}$ |
| Y- Zakonshme | Biserial $r_b$            | <u>Sperman</u> $r_s$    | Biserial rangi $r_{rp}$   |
| Y Nominale   | Biserial pikesor $r_{pb}$ | Biserial rangi $r_{rb}$ | Ø, L, C, LAMBDA           |

## REGRESIONI

Regresioni është një teknike statistikore që tregon sa ndryshon mesatarisht një variabël kur variabli tjetër ndryshon me një njësi. (tregon efektin që ka një variabël i pavarur mbi një variabël të varur).

### Regresioni i shumëfishtë

#### *Standartizimi i koeficientit të regresionit*

Shumë e autorë e paraqesin koeficientin e regresionit si të nevojshëm për t'u përdorur në subjektet individuale për të marrë vlerën korresponduese të y.

Por vlera e koeficientit të regresionit nuk mund të përdoret për të parë se cilët variabël janë më të rëndësishëm, pasi vlera e tij lidhet dhe me faktin se cila është shkalla me të cilën janë matur variabël, njësoj si në regresionin e thjeshtë linear.

Kështu, në shembullin e llogaritjes së arqindjes së dhjamosjes, variabli raca është koduar 1 për racën e bardhë dhe 0 për racën e zezë, ndërsa variabli moshë është koduar si numër vitesh për moshën e secilit subjekt. Nëse raca dhe mosha janë njëlooj të rëndësishëm në predicting subsequent depression, koeficienti i regresionit për racën duhet të jetë shumë më i madh se koeficienti i regresionit për moshën, kështu që e njëjta shuëm i duhet shtuar përqindjes së parashikuar të dhjamosjes. Këta koeficientë korelacioni quhen të pastandartizuar dhe nuk mund të përdoren për të nxjerrë konkluzione lidhur “me rëndësinë” e variabëlve, por vetëm për lidhjen mes variabëlve, nëse ajo është pozitive apo negative. Një rrugë që mund të përdoret për të eliminuar efektin e shkallëzimit është standartizimi i koeficientëve të regresionit. Standartizimi realizohet duke zbritur mesataren aritmetike të vlerës së x pjesëtuar me SD para analizës, kështu të gjithë variabël do të kenë një mesatare 0 dhe SD të barabartë me 1. tani është tërësisht e mundur të krahasohet magnitudo e koeficientëve të regresionit dhe të nxirren konkluzionet se cilët variabël luajnë një rol të rëndësishëm. Gjithashtu është e mundur që net ë llogarisim koeficientët e korrelacionit të standartizuar dhe pasi ne kemi ndërtuar modelin e ekuacionit të regresionit. Sa më i madh koeficienti i regresionit i standartizuar, aq më e madhe vlera e testit F. Koeficientët e regresionit të standartizuar njihen zakonisht si beta ( $\beta$ ) koeficientë. E meta më e madhe e koeficientëve të standartizuar të regresionit është se ato nuk mund të përdoret për të parashikuar vlerat e rezultatëve. Duke përdorur koeficientët e regresionit të standartizuar për të

analizuar moshën dhe gjininë themi se moshë ka influencën apo efektin më të madh në përqindjen e dhjamosjes, sepse 0,2109 e moshës është më e madhe se (- 0,4934) e gjinisë.

### ***Multiple R***

Multiple R është një regresion multipël, i ngjashëm me koeficientin e korrelacionit  $r$ , të Pearson. Ai quhet gjithashtu koeficienti I determinimit të shumëfishtë. Në shëmbullin tonë, supozojmë se përqindja e dhjamosjes është llogaritur për çdo person dhe është parë lidhja mes përqindjes së parashikuar të dhjamosjes dhe përqindjes aktuale të saj. Ky koeficient korrelacioni është multiple R. Nëse R e ngremë në katror ( $R^2$ ) atëherë ne masim se sa prej variacioneve të vlerave, përdoren për të marrë informacion lidhur me variablet e përfshirë në ekuacionin e regresionit. Tetsi statistikor që përdoret për  $R^2$  është shpërndarja F dhe jo ajo t. kështu në shëmbullin më lart  $R^2 = 0,80$ , tregon se 80% e variabilitit të vlerave të dhjamosjes, janë përdorur për të njohur BMI, racën dhe moshën. Meqë  $R^2$  është më e vogël se një ( $<1$ ), ne kuptojmë se Edhe faktorë të tjerë, përveç këtyre të përfshirë në ekuacion, luajnë rol në llogaritjen e përqindjes së dhjamosjes trupore.

### ***Zgjedhja e variableve për modelin e regresionit***

Në shëmbullin më lart për të parë lidhjen mes BMI dhe përqindjes së dhjamosjes, autorët përdorën analizën e regresionit të shumëfishtë për të parë sesi karakteristika të vecanta kontribuojnë në lidhjen mes variableve. Gjithshtu ata duan të dinë se si karakteristikat ndërveprojnë me njëra tjetrën si psh pasha dhe raca. Disa pyetje të tjera kërkimore fokusohen në parashikimin e rezultateve; ata përdorin ekuacionin e regresionit për të parashikuar përqindjen e masës trupore në subjektet e tjerë në përgjithësi.

Për të vendosur se cilët variable shërbejnë si prediktorët më të mirë të procesit ne I referohemi **ndërtimit të modelit**. Zgjedhja e variableve për modelin e regresionit mund të bëhet në disa mënyra.

Mënyra I. Fusim të gjithë variablet në ekuacionin e regresionit, me metodën “enter” për procedurën e regresionit multipël në SPSS. Në vazhdim të kësaj procedurë, nëse ndonjë nga variablet e përfshirë nuk është sinjifikant, ai përjashtohet nga ekuacioni I regresionit. Ky



ekuacion përsëritet ose më saktë rillogaritet duke përdorur variablet e mbetur, sepse koeficienti i regresionit merr vlera të ndryshme në varësi të variableve të përzgjedhur.

Mënyra II. Programet kompjuterike në rutinën e tyre përmbajnë procedura për të përzgjedhur setin e variableve shpjegues. Një nga këto procedura quhet **forward selection**. Kjo metodë fillon duke futur fillimisht një variabël në ekuacionin e regresionit dhe më pas kjo metodë I merr në analizë variablet e tjerë një nga një derisa të përfshihen në ekuacionin e regresionit të gjithë variablet që janë statistikisht të rëndësishëm në ekuacionin e regresionit. I pari në listën e variableve është ai që ka koeficientin e regresionit të madh (pra ai që e rrit  $R^2$  me një vlerë më të madhe).

Mënyra III. Kjo mënyrë quhet procedura **backward elimination**. Në këtë metodë, përfshihen në ekuacion, që në fillim, të gjithë variablet. Variabli X që e rrit më pak  $R^2$ , hiqet nga ekuacioni. Procedura e eliminimit të variableve vazhdon derisa variabli x që mendohet të lëvizet nga ekuacioni i regresionit, të shkaktojë zvogëlimin sinjifikant të  $R^2$ . Ky variabël X rikthehet në ekuacion dhe ekuacioni i regresionit llogaritet sërish.

Kur përdoren të dyja metodat së bashku, si procedura **forward selection** ashtu dhe ajo **backward elimination**, atëherë kemi metodën **stepwise regression**. Në përfundim, në ekuacionin e regresionit, variablet x që mbeten janë vetëm ata që shkaktojnë reduktim sinjifikant të  $R^2$ . Disa herë, variablet në një ekuacion regresioni multipël, grupohen sipa sfunksionit që ata kanë, psh, variablet demografike. Këta shtohen në ekuacionin e regresionit si një grup, apo bllok duke realizuar një proces që quhet regresioni hierarkik. Avantazhet e kësaj përjasje logjike për të ndërtuar modelin e ekuacionit të regresionit është ky: rezultatet tentojnë të bëhen gjithnjë e më shumë të qëndrueshme dhe sa më shumë të përsëritshme në studime të ngjashme.

## Regresioni logjistik

Regresioni logjistik përdoret atëherë kur variablet e pavarur mund të jenë nominalë ose ordinale, kurse variabli i varur është dikotomik apo binary.

Modeli I regresionit logjistik jep probabilitetin që një rezultat, të jetë funksion eksponencial I variableve të pavarur. Modeli I regresionit logjistik për tri variable të pavarur është:

$$p_x = \frac{1}{1 + \exp[-(b_0 + b_1x_1 + b_2x_2 + b_3x_3)]}$$

Ku  $b_0$  është y-intercepti dhe  $b_1$ ,  $b_2$  dhe  $b_3$  janë koeficientët e regresionit dhe exp tregon bazën e logaritmit natyral (2,718). Ekuacioni I regresionit ndërtohet duke u bazuar në metodën e selektimit, të ngjashëm me regresionin e shumëfishtë. Për të përcaktuar cilët variable do të përfshihen në ekuacionin e regresionit përdoret testi Hi-katror (dhe jo testi t apo testi F).

Le ta sqarojmë këtë situatë me një shembull të Soderstrom dhe kolegë, në 1997, donin të zhvillonin një model për të parë lidhjen mes aksidenteve rrugore dhe përqëndrimit të alkoolit në gjak mbi 50 mg/dl. Pasi mblodhën të dhënat, studjuesit I koduan ato vlerat e variableve të pavarur si 0 dhe 1.

| Variablet        | Vlera | Frekuenca |
|------------------|-------|-----------|
| Mosha            |       |           |
| < 39 vjec        | 0     | 3514      |
| > 40 vjec        | 1     | 1534      |
| Ora e ditës      |       |           |
| 6.00-18.00       | 0     | 2601      |
| 18.00-6.00       | 1     | 2447      |
| Ditet e javës    |       |           |
| H-E              | 0     | 2642      |
| P-D              | 1     | 2406      |
| Gjinia           |       |           |
| femra            | 0     | 1457      |
| meshkuj          | 1     | 3591      |
| Raca             |       |           |
| e bardhe         | 0     | 1758      |
| me ngjyre        | 1     | 3290      |
| Demtime          |       |           |
| te qellimshme    | 0     | 3966      |
| te paqellimshme  | 1     | 1082      |
| Sasia e alkoolit |       |           |
| < 50 mg/dl       | 0     | 4067      |

|            |   |      |
|------------|---|------|
| > 50 mg/dl | 1 | 1465 |
|------------|---|------|

Kjo praktikë kodimi bën më të lehtë interpretimin e odds ratio. Gjithashtu, nëse qëllimi është të vendosim një sistem pikëzimi, koeficienti që shoqëron një variabël të dhënë, patjetër duhet të jetë me sistem pikësh, vetëm nëse pacienti ka vlerën 1 për këtë variabël. P.sh: nëse pacientët kanë nivelin e alkoolit në >50mg/dl, kryesisht në fundjavë, pikët e marra në ditët e tjera të javës nuk përfshihen në dëmtimet që ndodhin në ditët e tjera të javës.

Studjuesit e llogaritën regresionin logjistik në katër grupe:

Meshkuj me dëmtimë të qëllimshme;

Meshkuj me dëmtimë të paqëllimshme;

Femra me dëmtimë të qëllimshme; dhe

Femra me dëmtimë të pa qëllimshme.

Rezultatet e analizës në meshkuj me dëmtime të paqëllimshme jepen në tabelën më poshtë.

| Variablet         | Koeficienti i regresionit | SE   | OD    | Vlera p | CI 95%      |
|-------------------|---------------------------|------|-------|---------|-------------|
| intercept         | -0,796                    | 0,11 |       | 0,001   |             |
| ora e dites       | -1,844                    | 0,1  | 0,158 | 0,001   | 0,128-0,194 |
| dita e javes      | 0,66                      | 0,09 | 1,93  | 0,001   | 1,601-2,374 |
| raca              | 0,27                      | 0,11 | 1,32  | 0,013   | 1,059-1,646 |
| mosha mbi 40 vjec | -0,119                    | 0,1  | 0,88  | 0,268   | 0,717-1,096 |

Ne duam të dimë cili nga variantet është shënuar me zero dhe me një, në mënyrë që të interpretojmë rezultatet. Psh koha e paraqitjes në urgjencë ka një koeficient regresioni negative.

Intervali I kohës nga 6-18 është koduar me 1 dhe meshkujt që vinë në urgjencë në këtë interval me dëmtime të paqëllimshme kanë më shumë mundësi të kenë nivelin e alkoolit < 50 mg/dl, se burrat që vijnë me dëmtime të paqëllimshme por në mesnatë. Sic shihet nga tabela, variabli moshë është I parëndësishëm (  $p= 0,268$ ). Duke interpretuar ekuacionin e regresionit për variablet e tjerë , mund të themi se meshkujt që paraqiten në departamentin e urgjencës natën dhe me dëmtime të paqëllimshme, të racës me ngjyrë, kanë nivele të alkoolit më të larta.

Regresioni logjistik mund të përdoret për të llogaritur probabilitetin për cdo individ. Psh: le të gjejmë probabilitetin që një 27 vjecar, me ngjyrë që vjen në departamentin e urgjencës në 14.00, të ditës së enjtë, të ketë nivelin e alkoolit në gjak më shumë se 50mg/dl (>50 mg/dl). Nga tabela nxjerrim ekuacionin e regresionit:

$$Y = -0,80 - (1,84 \cdot \text{orën e paraqitjes në urgjencë}) + (0,06 \cdot \text{ditën e javës}) + (0,28 \cdot \text{racën}) - (0,11 \cdot \text{moshën})$$

$$Y = -0,80 - 1,84 + 0,28 = -2,36$$

Zëvendësojmë këtë vlerë në ekuacionin e regresionit logjistik dhe kemi:

$$p_x = \frac{1}{1 + \exp[-(b_0 + b_1x_1 + b_2x_2 + b_3x_3)]} =$$

$$= \frac{1}{1 + \exp[-(-2,36)]} = \frac{1}{1 + 10,59} = 0,0863$$

Pra shansi që një burrë 27 vjec, me ngjyrë të paraqitet në urgjencë në një ditë të javës, në drekë dhe që të ketë nivel të alkoolit në gjak > 50 mg/dl është një në dhjetë.

### ***Përparësitë e regresionit logjistik janë:***

regresioni logjistik nuk kërkon të dijë për formën e shpërndarjes së variableve të pavarura;

koeficientët e regresionit interpretohen bazuar në RR, kur është fjala për studimet kohort ose në OD kur është fjala për studimet rast-kontroll.

Në shembullin tonë RR që një mashkull të paraqitet ditën në departamentin e urgjencës me dëmtime të paqëllimshme, por që të ketë nivele të alkoolit > 50 mg/dl është  $\exp(-1,845) = 0,158$ . RR që ky person të paraqitet natën është reciproku I tij, pra  $1/0,158 = 6,33$ , që do të thotë se meshkuj që paraqiten natën në shërbimin e urgjencës, me dëmtime të paqëllimshme, kanë gjashtë herë më shumë shanc që të kenë nivele të alkoolit > 50 mg/dl.

Po për OD c'mund të themi? Ne dimë se nëse CI95% nuk e përmban 1, ne mund të jemi të sigurt në masën 95%, se faktori shoqërues me këtë OD, ka një risk sinjifikant, ose siguron një nivel sinjifikant mbrojtjeje.

Tërësia e rezultateve të marra nga regresioni logjistik mund të testohen me testin e përshtatshmërisë së mirë Hosmer dhe Lemeshow. Ky test bazohet në shpërndarjen Hi-katror. Vlera e probabilitetit  $p < 0,05$ , tregon se modeli i vlerësimit përshtat të dhënat në një nivel të pranueshëm.

### Analiza log-lineare

Analiza log-lineare është e njëjtë me modelin e regresionit në të cilin të gjithë variablet, të varur dhe të pavarur, janë matur në një shkallë nominalë. Teknika është quajur log-lineare, sepse ajo përdor logaritmin e frekuencave të observuara në një tabelë kontigjence.



## REGRESIONI LOGJISTIK



Veshtrim i pergjithshem dhe interpretimi i rezultateve

---

---

---

---

---

---

---

---

## Tipet e testeve statistikore

|                                 |                                 | Variable te varur Dependent            |                                                               |
|---------------------------------|---------------------------------|----------------------------------------|---------------------------------------------------------------|
| Variable te pavarur Independent | Te vazhdueshem (Quantitative)   | Regression<br>Correlation              | Kategorik Diskret (Qualitative)<br><b>Logistic Regression</b> |
|                                 | Kategorik Diskret (Qualitative) | T-test<br>ANOVA<br>Regression (ANCOVA) | Chi-square<br><b>Logistic regression</b>                      |

---

---

---

---

---

---

---

---

## Variablet e varur Dichotomik

- Rezultatet jane shenuar me 1 (kur ngjarja ndodh) ose 0 (kur ngjarja nuk ndodh)
- Regresioni Multiple linear nuk mund te perdoret. **PERSE?** Sepse vlerat e pritshme mund te mos kene kuptim; ekuacion i regresionit linear mund te parashikoje vlera te ndryshme nga rangu 1/0 – p.sh., vlera negative ose vlera me te medha sesa 1

---

---

---

---

---

---

---

---

## Regresioni Logistic

- Zgjidhja e problemit realizohet permes zgjidhjes se ekuacionit per te parashikuar rezultatet mes 0 dhe 1, duke perdorur 'logit transformation', te percaktosh logaritmin (log) natyror te odds per nje rast qe te jete 0 or 1:

$$L_i = \ln(p / 1-p),$$

dhe koeficientet e regresionit jane odds ratio per cdo I.V.

I.V. llogarit log (odds)?

---

---

---

---

---

---

---

---

## Ekuacioni i regresionit Logistic...

Na jep ne, rezultate te pershtatshme per nje perfundim dichotomik:

- Per cdo variabel predictor, drejtimin e influences se variabelit te varur (- ose +)
- Odds ratios per secilin nga variabellet predictore
- Logs odds per secilin nga variabellet predictore [te perdorshme ne nje equation shites, me qellim qe te merret probabiliteti i ngjarjes (eventit) ☺ ]

---

---

---

---

---

---

---

---

## Shembull

- Niveli i ndryshem i edukimit a influencon ne probabilitetin per te perdorur metoda moderne kontraceptive?

Ho: niveli i edukimit nuk ndikon ne perdorimin e metodave moderne kontraceptive

Ha: Gratë me nivel te larte edukimi jane me te prirura per te perdorur metoda moderne kontraceptive

Duke kontrolluar per moshe, rezidence (U/R), rajon, gjendje martesore

---

---

---

---

---

---

---

---

Si mund te shohim nese pedorimi aktual I  
kontraceptiveve ndryshon me edukimin?

- **Testi Chi-katror:** dhe krahaso perdorimin e kontraceptiveve mes niveleve te edukimit duke perdorur testin chi-katror

| Edukimi | Perdorimi  |        |
|---------|------------|--------|
|         | Po         | Jo     |
|         | Secondar + | Primar |
|         | A          | B      |
|         | C          | D      |

$$\text{ODDS RATIO} = \frac{A/C}{B/D}$$

---

---

---

---

---

---

---

---

---

---

## Perfitimet e regresionit logistik

- Mund te masim magnituden e efektit
  - Grate me nivel te ulet edukimi jane 2.7 here me te prirura te perdorin kontraception, sesa grate me nivel me te larte edukimi (>sekondar)
- Na tregon ne nese rezultati eshte statistikisht sinjifikant
  - Ne jemi 95% te sigurt, qe rezultati nuk shkaktohet nga shanci
- Mund te perfshije shume variable prediktore

---

---

---

---

---

---

---

---

---

---

## Regresioni Logistik

- Variabli i varur  $y$  eshte gjithmone binar (dichotomik)
- Variablet e pavarur mund te jene ( $x$ ) te vazhdueshem, kategorik ose binar.
  - Njesoj si per regresionin linear

$$y = \begin{cases} 1 & \text{Aktualisht perdorin kontraception} \\ 0 & \text{Nuk perdorin kontraception} \end{cases}$$

---

---

---

---

---

---

---

---

---

---

## Linear vs. Logjistik: variabli i varur

- **Regresion Linear**
  - Variabli i varur  $y$  është i vazhdueshëm
- **Regresioni Logjistik**
  - Variabli i varur  $y$  është binar (dichotomik)
- E vetmja pyetje që një kërkues duhet të bëjë kur zgjedh regresionin si metode të analizës, është:
  - “Si duket variabli im i varur?”

---

---

---

---

---

---

---

---

## Linear vs. Logistic: Interpretation

|                                | Linear                      | Logistic                   |
|--------------------------------|-----------------------------|----------------------------|
| Rezultati:                     | $\beta$ : coefficient       | $\exp(\beta)$ : Odds Ratio |
| Rangu i vlerave për rezultatet | Numra Pozitive dhe negative | Vetem numra Pozitive       |
| Lidhje Negative                | $<0$                        | $<1$                       |
| Lidhje Positive                | $>0$                        | $>1$                       |
| Mungese lidhje                 | 0                           | 1                          |

---

---

---

---

---

---

---

---

## Odds ratios

- Një gjë “e bukur” e regresionit logistik është se rezultati mund të interpretohet si odds ratios
- Odds i një ngjarje mund të definohet si:

$$\text{odds} = \left[ \frac{p}{1-p} \right]$$

- Ku  $p$  = probabiliteti që një ngjarje të ndodhë
  - Për shembull, probabiliteti që një grua të perdore kontraceptive.

---

---

---

---

---

---

---

---

Odds Ratio = odds/odds

| Përdorimi aktual i<br>kontraseptive | Residenca |       |       |       |
|-------------------------------------|-----------|-------|-------|-------|
|                                     | Urban     | Rural | Total |       |
|                                     | Po        | 44.4% | 36.0% | 37.3% |
|                                     | Jo        | 55.6% | 64.0% | 62.7% |
|                                     | Total     | 100%  | 100%  | 100%  |

---

---

---

---

---

---

---

---

Odds Ratio = odds/odds

| Përdorimi aktual i<br>kontraseptive | Residenca |       |       |       |
|-------------------------------------|-----------|-------|-------|-------|
|                                     |           | Urban | Rural | Total |
|                                     | Po        | 44.4% | 36.0% | 37.3% |
|                                     | Jo        | 55.6% | 64.0% | 62.7% |
|                                     | Total     | 100%  | 100%  | 100%  |

Odds per perdorim contracep Urban:

---

---

---

---

---

---

---

---

Odds Ratio = odds/odds

| Përdorimi aktual i<br>kontraseptiv | Residenca |       |       |       |
|------------------------------------|-----------|-------|-------|-------|
|                                    |           | Urban | Rural | Total |
|                                    | Po        | 44.4% | 36.0% | 37.3% |
|                                    | Jo        | 55.6% | 64.0% | 62.7% |
|                                    | Total     | 100%  | 100%  | 100%  |

Odds per perdorim contracep Urban:       $44.4\% \div 55.6\% = 0.80$

---

---

---

---

---

---

---

---



Odds Ratio = odds/odds

| Perdorim aktual i<br>contraceptive | Residenca |       |       |       |
|------------------------------------|-----------|-------|-------|-------|
|                                    |           | Urban | Rural | Total |
|                                    | Po        | 44.4% | 36.0% | 37.3% |
|                                    | Jo        | 55.6% | 64.0% | 62.7% |
|                                    | Total     | 100%  | 100%  | 100%  |

Odds per perdorim contracep Urban:  $44.4\% \div 55.6\% = 0.80$

Odds per perdorim contracept Rural:

---

---

---

---

---

---

---

Odds Ratio = odds/odds

| Perdorim aktual i<br>contraceptive | Residenca |       |       |       |
|------------------------------------|-----------|-------|-------|-------|
|                                    |           | Urban | Rural | Total |
|                                    | Po        | 44.4% | 36.0% | 37.3% |
|                                    | Jo        | 55.6% | 64.0% | 62.7% |
|                                    | Total     | 100%  | 100%  | 100%  |

Odds per perdorim contracep Urban:  $44.4\% \div 55.6\% = 0.80$

Odds per perdorim contracept Rural:  $36\% \div 64\% = 0.56$

---

---

---

---

---

---

---

Odds Ratio = odds/odds

| Perdorim aktual i<br>contraceptive | Residenca |       |       |       |
|------------------------------------|-----------|-------|-------|-------|
|                                    |           | Urban | Rural | Total |
|                                    | Po        | 44.4% | 36.0% | 37.3% |
|                                    | Jo        | 55.6% | 64.0% | 62.7% |
|                                    | Total     | 100%  | 100%  | 100%  |

Odds per perdorim contracep Urban:  $44.4\% \div 55.6\% = 0.80$

Odds per perdorim contracept Rural:  $36\% \div 64\% = 0.56$

Raporti i odds per Urban contracept ÷  
Rural contracep

---

---

---

---

---

---

---

## Odds Ratio = odds/odds

| Perdorim aktual i<br>kontraceptive | Residence |       |       |
|------------------------------------|-----------|-------|-------|
|                                    |           | Urban | Rural |
|                                    | Po        | 44.4% | 36.0% |
|                                    | Jo        | 55.6% | 64.0% |
| Total                              |           | 100%  | 100%  |

Odds per perdorim kontracep Urban:  $44.4\% \div 55.6\% = 0.80$

Odds per perdorim kontracept Rural:  $36\% \div 64\% = 0.56$

Raporti i odds per Urban kontracept ÷  
Rural kontracep  **$0.80 \div 0.56 = 1.42$**

---

---

---

---

---

---

---

---

## Raporti Odds (Ratio) = odds/odds

| Perdorim aktual i<br>kontraceptive | Residenza |       |       |
|------------------------------------|-----------|-------|-------|
|                                    |           | Urban | Rural |
|                                    | Po        | 44.4% | 36.0% |
|                                    | Jo        | 55.6% | 64.0% |
| Total                              |           | 100%  | 100%  |

Odds per perdorim kontracep Urban:  $44.4\% \div 55.6\% = 0.80$

Odds per perdorim kontracept Rural:  $36\% \div 64\% = 0.56$

Ratio i odds per Urban kontracept ÷  
Rural kontracep  **$0.80 \div 0.56 = 1.42$**

**Odds ratio per perdorimin e kontracep duke krahasuar grate urbane me rurale=1.42**

---

---

---

---

---

---

---

---

## Modeli i Regresionit Logistic

**Modeli matematikor transformohet :**

$$\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1$$

"funksioni logistik"

**Modeli tregon sesi odds e variablit te varur (y)  
ndryshojne ne varesi te vlerave te variableve te  
pavarur (xs). variable I varur jepet ne shkalle log.**

---

---

---

---

---

---

---

---

## Logistic Regression Model

$$\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1$$

□ **p** = probabiliteti per perdorim metodash kontraceptive

□ **x<sub>1</sub>** = vendbanimi urban/rural

□ **β<sub>0</sub>** dhe **β<sub>1</sub>** jane koeficientet e regresionit

□ Me fjale te tjera,

log (ODDS per perdorim kontrcept.) =

= constant + β<sub>1</sub>\*vendbanimi

## Per te gjetur odds ratios...

□ **β<sub>1</sub>** vlereson ndryshimin ne log odds per ndryshimin me nje njesi te **x**.

□ Per te gjetur odds ratio, bejme exponentin me baze natyrore te **β<sub>1</sub>**:

□ **exp(β<sub>1</sub>)** eshte odds ratio ndryshimin me nje njesi te **x**.

SPSS ne output, siguron si logs odds dhe odds ratios (exponencialin e log odds)

## SPSS na jep log(odds) dhe odds

Categorical Variables Codings

|      |                            | Frequency                 | Parameter coding (1) |
|------|----------------------------|---------------------------|----------------------|
| V025 | Type of place of residence | 1 Urban 2949 2 Rural 7844 | 1.000 .000           |

Variables in the Equation

|           |          | B     | S.E. | Sig. | Exp(B) | 95.0% C.I. for EXP(B) |
|-----------|----------|-------|------|------|--------|-----------------------|
| Step 1(a) | V025(1)  | .352  | .054 | .000 | 1.422  | 1.280 1.579           |
|           | Constant | -.577 | .022 | .000 | .562   |                       |

a. Variable(s) entered on step 1: V025.

B = β, **log odds** per perdorim kontrceptiv

Exp(B) = exp(β), **odds** per perdorim kontrceptiv

## Interpretimi i regresionit Logistic

- Pergjithesisht ne publikime raportohen me shume Odds ratios sesa log odds, per arye se jane me intuitive
- Konstantja ka te njejtin interpretim si per regresionin linear
- Odds ratios jane raportet e odds: odds e perdorimit te kontraceptiveve per grate ne zone urbane krahasuar me ato ne zone rurale.

---

---

---

---

---

---

---

---

## Interpretimi i rezultateve

Categorical Variables Codings

|                                 | Frequency    | Parameter coding |
|---------------------------------|--------------|------------------|
| V025 Type of place of residence | 1 Urban 2949 | (1)              |
|                                 | 2 Rural 784  | .000             |

Variables in the Equation

|           | B     | S.E. | Sig. | Exp(B) | 95.0% C.I. for EXP(B) |       |
|-----------|-------|------|------|--------|-----------------------|-------|
|           |       |      |      |        | Lower                 | Upper |
| Step 1(a) |       |      |      |        |                       |       |
| V025(1)   | .352  | .054 | .000 | 1.422  | 1.280                 | 1.579 |
| Constant  | -.577 | .022 | .000 | .562   |                       |       |

a. Variable(s) entered on step 1: V025.

(1) Ne tregon ne se rezultatet jane per perfundimin 1 = urban vs. rural. Grate ne zone rurale jane "reference group".

Odds per perdorimin e kontraceptive eshte 1.4 here me i larte per grate ne zone urbane, krahasuar me ato ne zone rurale.

---

---

---

---

---

---

---

---

## Interpretimi i rezultateve nga analiza SPSS

Variables in the Equation

|           | B     | S.E. | Sig. | Exp(B) | 95.0% C.I. for EXP(B) |       |
|-----------|-------|------|------|--------|-----------------------|-------|
|           |       |      |      |        | Lower                 | Upper |
| Step 1(a) |       |      |      |        |                       |       |
| V025(1)   | .352  | .054 | .000 | 1.422  | 1.280                 | 1.579 |
| Constant  | -.577 | .022 | .000 | .562   |                       |       |

a. Variable(s) entered on step 1: V025.

**Sig:** testohet hipoteza qe  $\beta=0$ .  
-Nese  $<0.05$  = aty ka me pake se 5% shanc qe  $\beta = 0$ .  
-Ne jemi 95% te sigurt qe  $\beta \neq 0$ .

**CI 95%: Rangu i vlerave te mundshme te  $\exp(\beta)$**   
-Ne regresionin logistic : nese CI95% nuk e perfshin 1, rezultati eshte statistikisht i rendesishem

---

---

---

---

---

---

---

---

## Interpretimi i Odds Ratios (OR)

- $OR > 1$ : **rritet** odds per grupin 1 krahasuar me grupin 2; grupi 1 **ka me shume gjasa** per patur ngjarjen.
- $OR = 1$ : nuk ka difference ne odds mes grupit 1 krahasuar me grupin 2.
- $OR < 1$ : **ulet** odds per grupin 1 krahasuar me grupin 2; grupi 1 **ka me pak gjasa** per te patur ngjarjen.

---

---

---

---

---

---

---

---

## Interpretimi

- **OR per perdorim te kontraceptiveve, krahasuar grate urbane vs. rurale = 1.42, CI95% = (1.28-1.58)**
- **Ne mund te themi:**
  - Ne pergjithesi, gjasat per te perdorur kontraceptive jane 1.42 here me te larta ne grate urbane krahasuar me grate ne zone rurale
  - Gjasat per te perdorur kontraceptive jane 42% here me te larta ne grate urbane krahasuar me grate ne zone rurale [logaritmet : 1-1.42]
  - Lidhja mes perdorimit te kontraceptiveve dhe vendbanimit te grave urbane/rurale eshte statistikisht e rendesishme
  - Ne jemi 95% te bindur qe odds per perdorimin e kontraceptiveve ne grate ne zone urbane vs. rurale eshte mes 1.28 dhe 1.58

---

---

---

---

---

---

---

---

## Burimet per te mesuar me shume

- **Interpreting Odds Ratios:**
  - [http://www.ats.ucla.edu/stat/mvlt/pkg/faq/general/odds\\_ratio.htm](http://www.ats.ucla.edu/stat/mvlt/pkg/faq/general/odds_ratio.htm)
- **A nice, very clear explanation of what odds ratios are:**
  - <http://globalhealth.wordpress.com/2008/06/03/what-is-an-odds-ratio/>
- **Annotated SPSS logistic regression output**
  - <http://www.ats.ucla.edu/stat/spss/output/logistic.htm>

---

---

---

---

---

---

---

---



## Transformimi i ekuacionit logistik

□ Cfare eshte nje odds ratio:

**Hapi 1.**  $\text{Log(odds)} = \beta_0 + \beta_1 * x_1 + \dots + \beta_k * x_k$

**Hapi 2.**  $\text{Odds} = \exp(\beta_0 + \beta_1 * x_1 + \dots + \beta_k * x_k)$

**Hapi 3.**  $\text{Probabiliteti per nje ngjarje} = \frac{\exp(\beta_0 + \beta_1 * x_1 + \dots + \beta_k * x_k)}{(1 + \exp(\beta_0 + \beta_1 * x_1 + \dots + \beta_k * x_k))}$

---

---

---

---

---

---

---

## Regresioni logjistik

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

---

### Njohuri...

- Regresioni logjistik përdoret atëherë kur variablet e pavarur mund të jenë nominale ose ordinale, kurse variabli i varur është **dikotomik** apo **binar**.
- Modeli i regresionit logjistik jep probabilitetin që një rezultat, të jetë funksion eksponencial i variableve të pavarur.

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

---

### Formula

- Modeli i regresionit logjistik për tre variable të pavarur është:

$$p_x = \frac{1}{1 + \exp \left[ - (b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3) \right]}$$

- Ku  $b_0$  është y-intercepti dhe  $b_1, b_2$  dhe  $b_3$  janë koeficientët e regresionit dhe exp tregon bazën e logaritmit natyral (2,718).

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

---

- Ekuacioni i regresionit ndërtohet duke u bazuar në metodën e selektimit, të ngjashëm me regresionin e shumëfishtë.
- Për të përcaktuar se cilët variable do të përfshihen në ekuacionin e regresionit përdoret testi Hi-katror (dhe jo testi t apo testi F).

Elizana PETRELA

9/11/2018

## Shembull

- Soderstrom dhe kolegë, në 1997, donin të zhvillonin një model për të parë lidhjen mes aksidenteve rrugore dhe përqëndrimit të alkoolit në gjak në nivelet mbi 50 mg/dl.
- Pasi mbledhën të dhënat, studjuesit i koduan ato (vlerat e variableve të pavarur si 0 dhe 1).

Elizana PETRELA

9/11/2018

| Variablet               | Vlera | Frekuenca |
|-------------------------|-------|-----------|
| <b>Mosha</b>            |       |           |
| ≤ 39 vjeç               | 0     | 3514      |
| ≥ 40 vjeç               | 1     | 1534      |
| <b>Ora e ditës</b>      |       |           |
| 6.00-18.00              | 0     | 2601      |
| 18.00-6.00              | 1     | 2447      |
| <b>Ditet e javës</b>    |       |           |
| H-E                     | 0     | 2642      |
| P-D                     | 1     | 2406      |
| <b>Gjinia</b>           |       |           |
| femra                   | 0     | 1457      |
| meshkuj                 | 1     | 3591      |
| <b>Raca</b>             |       |           |
| e bardhe                | 0     | 1758      |
| me ngjyre               | 1     | 3290      |
| <b>Demtime</b>          |       |           |
| te qellimshme           | 0     | 3966      |
| te paqellimshme         | 1     | 1082      |
| <b>Sasia e alkoolit</b> |       |           |
| < 50 mg/dl              | 0     | 4067      |
| ≥ 50 mg/dl              | 1     | 1465      |

- Kjo praktikë **kodimi** bën **më të lehtë interpretimin** e odds ratio.
- Gjithashtu, nëse qëllimi është të vendosim një sistem pikëzimi, koeficienti që shoqëron një variabël të dhënë, patjetër duhet të jetë me sistem pikësh, vetëm nëse pacienti ka vlerën 1 për këtë variabël.
- P.sh: nëse pacientët kanë nivelin e alkoolit në  $>50\text{mg/dl}$ , kryesisht në fundjavë, pikët e marra në ditët e tjera të javës nuk përfshihen në dëmtimet që ndodhin në ditët e tjera të javës.

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

---

- Studjuesit e llogaritën regresionin logjistik në katër grupe:

- Meshkuj me dëmtime të qëllimshme;
- Meshkuj me dëmtime të paqëllimshme;
- Femra me dëmtime të qëllimshme; dhe
- Femra me dëmtime të pa qëllimshme.

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

---

## Tabela e rezultateve

| Variablet      | b      | SE   | OD    | Vlera p | CI 95%      |
|----------------|--------|------|-------|---------|-------------|
| intercept      | -0,796 | 0,11 |       | 0,001   |             |
| ora e dites    | -1,844 | 0,1  | 0,158 | 0,001   | 0,128-0,194 |
| dita e javes   | 0,66   | 0,09 | 1,93  | 0,001   | 1,601-2,374 |
| raca           | 0,27   | 0,11 | 1,32  | 0,013   | 1,059-1,646 |
| mosha mbi 40vj | -0,119 | 0,1  | 0,88  | 0,268   | 0,717-1,096 |
| Gjinia (1)     | 0,986  | 0,1  | 2,34  | 0,000   | 1,98-2,84   |

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

---

- Ne duam të dimë cili nga variantet është shënuar me zero dhe me një, në mënyrë që të interpretojmë rezultatet.
- Psh: koha e paraqitjes në urgjencë ka një koeficient regresioni negativ.
- Intervali I kohës nga 18-6 është koduar me 1 dhe meshkujt që vinë në urgjencë në këtë interval me dëmtime të paqëllimshme **kanë më shumë mundësi** të kenë nivelin e alkoolit < 50 mg/dl, se burrat që vijnë me dëmtime të paqëllimshme, por në mesnatë.

Elizana PETRELA

9/11/2018

## Tabela e rezultateve

| Variablet         | b      | SE   | OD    | Vlera p      | CI 95%      |
|-------------------|--------|------|-------|--------------|-------------|
| intercept         | -0,796 | 0,11 |       | 0,001        |             |
| ora e dites       | -1,844 | 0,1  | 0,158 | 0,001        | 0,128-0,194 |
| dita e javes      | 0,66   | 0,09 | 1,93  | 0,001        | 1,601-2,374 |
| raca              | 0,27   | 0,11 | 1,32  | 0,013        | 1,059-1,646 |
| mosha mbi 40 vjec | -0,119 | 0,1  | 0,88  | <b>0,268</b> | 0,717-1,096 |

Elizana PETRELA

9/11/2018

- Sic shihet nga tabela, variabli moshë është i parëndësishëm ( $p=0,268$ ).
- Duke interpretuar ekuacionin e regresionit për variablet e tjerë, mund të themi se **meshkujt** që paraqiten në departamentin e urgjencës **natën** dhe **me dëmtime të paqëllimshme**, të racës **me ngjyrë**, kanë **nivele** të alkoolit **më të larta**.

Elizana PETRELA

9/11/2018



## Tabela e rezultateve

| Variablet        | b      | SE   | OD    | Vlera p | CI 95%      |
|------------------|--------|------|-------|---------|-------------|
| intercept        | -0,796 | 0,11 |       | 0,001   |             |
| ora e dites (0)  | -1,844 | 0,1  | 0,158 | 0,001   | 0,128-0,194 |
| dita e javes     | 0,66   | 0,09 | 1,93  | 0,001   | 1,601-2,374 |
| Raca (1)         | 0,27   | 0,11 | 1,32  | 0,013   | 1,059-1,646 |
| mosha mbi 40vj   | -0,119 | 0,1  | 0,88  | 0,268   | 0,717-1,096 |
| Gjinia (1)       | 0,986  | 0,1  | 2,34  | 0,000   | 1,98-2,84   |
| Nivele akool (1) | 0,24   | 0,11 | 2,32  | 0,013   | 1,059-1,646 |

Elizana PETRELA

9/11/2018

- Regresioni logjistik mund të përdoret për të llogaritur probabilitetin për cdo individ.
- Psh: le të gjejmë probabilitetin që një 27 vjecar, me ngjyrë, që vjen në departamentin e urgjencës në 14.00, të ditës së enjte, të ketë nivelin e alkoolit në gjak më shumë se 50mg/dl (>50 mg/dl). Nga tabela nxjerrim ekuacionin e regresionit:

Elizana PETRELA

9/11/2018

$$y = -0,80 - (1,84 * \text{orën e paraqitjes në urgjencë}) + (0,06 * \text{ditën e javës}) + (0,28 * \text{racën}) - (0,11 * \text{moshën})$$

- $y = -0,80 - 1,84 + 0,28 = -2,36$

Elizana PETRELA

9/11/2018

## Probabiliteti

$$p_x = \frac{1}{1 + \exp[-(b_0 + b_1x_1 + b_2x_2 + b_3x_3)]} =$$
$$= \frac{1}{1 + \exp[-(-2,36)]} = \frac{1}{1 + 10,59} = 0,0863$$

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

- Pra shanci, që një burrë 27 vjec, me ngjyrë të paraqitet në urgjencë në një ditë të javës, në drekë dhe që të ketë nivel të alkoolit në gjak  $> 50$  mg/dl është një në dhjetë.

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

## Përparësitë e regresionit logjistik janë:

- regresioni logjistik nuk kërkon të dijë për formën e shpërndarjes së variableve të pavarura;
- koeficientët e regresionit interpretohen bazuar në RR, kur është fjala për studimet kohort ose në OD kur është fjala për studimet rast-kontroll.

Elizana PETRELA

9/11/2018

---

---

---

---

---

---

---

- Në shembullin tonë RR që një mashkull të paraqitet ditën në departamentin e urgjencës me dëmtime të paqëllimshme, por që të ketë nivele të alkoolit > 50 mg/dl është  $\exp(-1,845) = 0,158$ . RR që ky person të paraqitet natën është inversi i tij, pra  $1/0,158 = 6,33$ , që do të thotë se meshkuj që paraqiten natën në shërbimin e urgjencës, me dëmtime të paqëllimshme, kanë gjashtë herë më shumë shanc që të kenë nivele të alkoolit > 50 mg/dl.

---

---

---

---

---

---

---

---

- Po për OD c'mund të themi? Ne dimë se nëse CI95% nuk e përmban 1, ne mund të jemi të sigurt në masën 95%, se faktori shoqërues me këtë OD, ka një risk sinjifikant, ose siguron një nivel sinjifikant mbrojtjeje.
- Tërësia e rezultateve të marra nga regresioni logjistik mund të testohen me testin e përshtatshmërisë së mirë Hosmer dhe Lemeshov. Ky test bazohet në shpërndarjen Hi-katror.
- Vlera e probabilitetit  $p \leq 0,05$ , tregon se modeli i vlerësimit përshtat të dhënat në një nivel të pranueshëm

---

---

---

---

---

---

---

---

## Analiza log-lineare

- Analiza log-lineare është e njëjtë me modelin e regresionit në të cilin të gjithë variabellet, të varur dhe të pavarur, janë matur në njësi nominale. Teknika është quajtur log-lineare, sepse ajo përdor logaritmin e frekuencave të observuara në një tabelë kontigjence.

---

---

---

---

---

---

---

---

## Literatura

- Interpreting Odds Ratios:
  - [http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/odds\\_ratio.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/odds_ratio.htm)
- A nice, very clear explanation of what odds ratios are:
  - <http://globalhealth.wordpress.com/2008/06/03/what-is-an-odds-ratio/>
- Annotated SPSS logistic regression output
  - <http://www.ats.ucla.edu/stat/spss/output/logistic.htm>

---

---

---

---

---

---

---

---

## **TEMA 15:**

### **ANALIZA DISKRIMINANTE**

#### **Objektivat e temës:**

Në përfundim të kësaj teme studentët duhet të jenë në gjendje që :

- Të kuptojnë ç'është dhe përse shërben analiza diskriminante
  - Të dinë të vlerësojnë nëse një rast studimor plotëson apo jo supozimet e Analizës diskriminante
  - Të kuptojnë ekuacionet e analizës diskriminante dhe koefiçientët përkatës
- Të analizojnë dhe të interpretojnë outputin në SPSS të Analizës Diskriminante

■

#### **LITERATURA**

- “ Applied Statistics”, Warner A.R , 2008
- Statistical for Business and Economics”, Newbold P, Carlson W, Thorne B, 2011
- “The Essentials of Biostatistics for Physicians, Nurses, and Clinicians”, Michael R. Chernick, September 2011
- “Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal”, 2nd Edition, Belinda Barton, Jennifer Peat, September 2014



# ANALIZA DISKRIMINANTE

## I: ÇËSHTJE QË LIDHEN ME ANALIZËN DISKRIMINANTE

Deri më tani kemi raste kur variabli i varur (y) është sasior , dhe variablat e pavarur mund të jenë sasiorë dhe cilësorë. Analiza që përdoret në këto raste ishte analiza regressive.

Mund të ndodhë që jemi të inetersuar për parashikimin e një variabli të varur cilësor, ku variablat e pavaruar mund të jenë sasiorë dhe cilësorë. Variabli i varur (Y) kategorik, mund të jetë dikotom (të marrë vetëm dy vlera si kode) ose dhe më shumë së dy kategori, në këto kushte analiza që përdoret është Analiza Diskriminante (AD).

Pra dallimi midids një analize regressive dhe një analize diskriminante , është se te regresioni, variabli i varur është sasior, kurse te AD variabli i varur është kategorik. Në analizën regressive, ne vlerësojmë saktësinë e parashikimit duke iu referuar koeficientit të korelacionit dhe madhësisë së SKG , në AD, do të ekzaminojmë tipe të tjera të informacioneve, të tilla si përqindja e rasteve për të cilat anëtarësimi në një kategori përputhet me anëtarësimin aktual në këtë kategori.

Në të dy llojet e analizave, qëllimi është të gjendet `optimali i kombinimit linear me koeficientat përkatës”

Në analizën regressive, koeficientët që shoqërojnë çdo variabël x, i referohen të dhënave bruto, dhe përfaqësojnë pjerrësine e vijës regressive, dhe janë në formë të pastandartizuar (  $b_i$ )

$$\text{Psh: } Y = b_0 + b_1 x_1 + b_2 x_2 + \dots b_k x_k$$

Në formë të standartizuar, analiza regressive paraqitet nëpërmjet koeficientëve të standartizuar ose z:

Psh.  $z_y = \beta_1 z_{x1} + \beta_2 z_{x2} + \dots \beta_k z_{xk}$  ku madhësia dhe shenja e çdo koeficienti  $\beta$  përpara çdo variabli  $z_{xk}$  tregon se sa një rritje në çdo  $z_{xi}$  influencon rezultatin përfundimtar të parashikueshëm  $z_y$ . Kështu çdo  $\beta_i$  është koeficient më i vogël se 1, dhe sa më e madhe vlera absolute e  $\beta_i$  aq më i madh është efekti i variablit “i” tek variabli i varur. (shenja + tregon efekt pozitiv, shenja -, efekt negativ), duke çuar kështu në një korelacion më të madh me variablin respektiv “i”

Në analizën regressive, qëllimi është të gjenden ato koeficientë të regresit që sigurojnë kombinimi linear optimal të variablave të pavarur x, kombinim linear ky, që i referohet një  $SS_{\text{residual}}$  sa më të vogël të mundshme.

Në mënyrë analoge, qëllimi i AD, është të gjenden koeficientë të analizës diskriminante që çojnë dhe këto në një kombinim linear të variablave të pavarur x. Kombinimet lineare që rezultojnë nga analiza diskriminante quhen “ funksione diskriminante” , dhe janë optimale në sensin që sigurojnë një SKM sa më të lartë dhe një minimum  $SS_{\text{within}}$  (  $SS_{\text{within}}$  i referohet shumës se katrorëve brenda grupeve të përcaktuara nga kategoritë e variablit Y). Një tjetër sens i optimalitetit të AD, është se këto koeficientë mund të përdoren për të klasifikuar rastet sipas grupeve , duke bërë kështu një klasifikim për çdo grup më përqindjen më të lartë të mundshme të korrektesisë.

AD mund të përdoret kështu:

1. Për të bërë parashikime në lidhje me rezultate të anëtarësisë në grupe të caktuara. Psh. AD mund të përdoret për të parashikuar nëse një pacient ka më shumë gjasa të jetë anëtar i Grupit 1 (pacientë që kanë atak zemre) ose i Grupit 2 (pacientë, dhembja e gjoksit të cilëve nuk ka të bëjë me atak zemre) bazuar në rezultatet e secilit individ në lidhje me variablat që vlerësohen në fazën e intervistës së pacientëve dhe në rezultatet laboratorike, të tilla si: mosha, gjinia, tensioni i lartë, tensioni i ulët, kolesterolit, trigliceridet, duhanpirja, intensiteti i dhimbjeve të gjoksit, dhe tensioni i gjakut.

Për këtë mund të përdoret një numër pacientësh, diagnoza e të cilëve dihet për të zhvilluar një formulë statistikore parashikuese nëpërmjet të cilës për çdo pacient të mundshëm tjetër (të ndryshëm nga ata që u morën në studim) duke ditur variablat e pavarur, të mund të gjykojmë në kushte parashikuese nëse pacienti i përket potencialisht Grupit 1 apo Grupit 2.

E njëjta gjë mund të bëhet edhe për pranimin apo jo të studentëve në fakultet duke u nisur nga variabla parashikues të tillë si: mesatarja, mosha, gjinia, etj .

2. Për të përshkruar natyrën e ndryshimeve ndërmjet grupeve. Nëse ne mund të identifikojmë një grup pacientësh që kanë patur atak zemre (Grupi 1) dhe i krahasojmë ata me një grup tjetër pacientësh që nuk kanë patur atak zemre (Grupi 2) , mund të përcaktojmë se cilat cifte rezultatesh të tilla si: Mosha, tensioni i lartë, tensioni i ulët, kolesterolit, trigliceridet, duhanpirja, ankthi dhe presioni i gjakut përshkruanë ndryshimet ndërmjet këtyre dy grupeve të pacientëve.

AD është një ndër disa alanalizat që mund të përdoren për të bërë parashikime në lidhje me anëtarësinë potenciale në një nga kategoritë e variablit të varur (y). Kur të gjithë variablat të përfshirë në analizë (X dhe Y) janë kategorike, analiza log-lineare është më e përshtatshme. Një tjetër vlerësim i kësaj analize që përdoret shumë sot në parashikimin e anëtarësimt në një nga grupimet e variablit kategorik është Regresioni logjistik. Regresioni logjistik përfshin vlerësimin e raporteve të rezultateve të mundshme të anëtarësimt në grupet përkatëse, psh. Nëse një pacient ka 5 herë më shumë shanse të vdesë sesa të jetojë, raporti i vdekjeve përshkruhet si 5:1. Regresioni logjistik shpesh preferohet më shumë se AD, sepse kërkon supozime më pak kufizuese në lidhje me shpërndarjen e rezultateve. Për më tepër, rezultatet që derivojnë nga analiza e regresionit logjistik mund të interpretohet si informacion në lidhje me ndryshimet në ngjashmëritë e rezultateve negative kur rriten vlerat e variablave të pavarur (zakonisht më tepër përdoret regresioni logjistik binar (me dy vlera të Y) ndërkohë që në AD, Y merr dy ose më shumë vlera.)

AD siguron informacion në lidhje me natyrën e diferencës midis grupeve. Funksionet diskriminante mund të përdoren për të përshkruar dimensionet (që do të thotë kombinimin e koeficientëve të variablave përkatës) mbi të cilat ndryshojnë grupet. Në AD, çështjet e kërkimit që trajtohen janë: A mund të parashikojmë anëtarësinë në një grup në një përqindje të caktuar, bazuar në rezultatet e variablave diskriminante (x) të participantëve? Cilat cifte të rezultateve të variablave X parashikojnë anëtarësimin në secilin grup? Cili variabël parashikues (x) siguron informacion vendimtar (unik) në lidhje me anëtarësimin në secilin grup, kur korelacionet me variablat e tjerë janë statistikisht të kontrollueshme?

Në disa aplikime të AD, tipe të ndryshme të **gabimeve të klasifikimit** kanë pasoja ose kosto të ndryshme. Psh një klasifikim i gabuar i një personi të sëmurë në grupin tjetër dhe anasjelltas sjell pasoja fatale për individin. Një tip informacioni, që mund të sigurohet gjatë analizës së AD, është edhe një tabelë që tregon numrin dhe tipin e gabimeve të klasifikimit (që do të thotë numër individësh për të cilët anëtarësia e parashikuar në grup nuk është e njëjtë me anëtarësinë aktuale të grupit)

Në AD, sikurse edhe Në analizën regresive, supozohet se variablat  $x$  janë pothuaj të pavarur nga njëri tjetri, ose janë të koreluar deri në një farë shkalle. Ndaj është e rëndësishme të kontrollohet statistikisht interkorelimi ndërmjet variablove  $x$ , kur ndërtojmë kombinimin optimal linear të variablave diskriminante (dmth kur ndërtojmë funksionet diskriminante)

Ashtu si analiza regresive, Ekuacioni i funksionit diskriminant mund të jepet në terma bazë, bruto të variablave ose në terma të standartizuar ( $z$ ). Kështu ekuacioni për një funksion të standartizuar diskriminant  $D_1$  që përdoret për të parashikuar anëtarësimin në grup jepet si me poshtë “ $p$ ” është numri i variablave diskriminante.  $D_1 = d_{11} z_1 + d_{12} z_2 + \dots + d_{1p} z_p$  ku  $z_1, z_2, \dots, z_p$  janë versione të standartizuara të variablave  $x_1, x_2, x_p$ .

Më vonë do të shohim rastin kur, variabli kategorik  $Y$  ka “ $k$ ” grupe, do të mund të llogaritim deri në  $(k-1)$  funksione diskriminate të ndryshme (numri i funksioneve diskriminante që mund të prodhojë AD, kufizohet gjithashtu edhe nga numri “ $p$ ” i variablave të pavarur në model, ose më sakte nga numri i variablave parashikues që nuk mund të parashikohen tërësisht nga variablat e tjerë parashikues) Nëse variabli  $Y$  është dikotom, atëherë mund të llogaritet vetëm një funksion diskriminant. Në përgjithësi nëse  $Y$  ka “ $k$ ” grupe dhe “ $p$ ” variabla diskriminante (dmth  $i = 0, 1, 2, \dots, p$  për  $x_i$ ), atëherë numri i funksioneve diskriminante që mund të përfitohet është minimumi i këtyre dy vlerave:  $(k-1)$  dhe  $p$ . Njësitë e vlerave specifike të rezultateve në funksionet diskriminate  $D$  nuk kanë rëndësi, vlera  $D$  përgjithësisht janë programuar të kenë një mesatare 0 dhe variancë brenda grupeve të  $=1$ , njëllor si rezultatet e standartizuara  $z$ . Ajo çfarë ka rëndësi është marrëdhënia ndërmjet rezultateve diskriminante të anëtarësimit në grup.

Kur  $Y$  është dikotom ( $k=2$ ), qëllimi i AD është të gjendet kombinimi linear (ekuacioni i funksionit diskriminant  $D_1$  si më lart) që ka variancën më të madhe të mundshme midis grupeve dhe variancën më të vogël të mundshme brenda grupeve), kështu kur performohet ANOVA në këtë rast raporti  $F$  është maksimali i mundshëm.

Kur  $Y$  ka më shumë se dy grupe, psh nëse  $k=3$ , atëherë duhet më shumë se një funksion diskriminant për të bërë diferencën ndërmjet grupeve, psh: ekuacioni i parë bën dallimin midis anëtarësimit në Grupin 1 dhe anëtarësimit në Grupet 2 e 3, dhe funksioni i dytë bën dallimin midis anëtarësimit në Grupin 3 nga Grupet 1 e 2.

### **Supozimet e AD:**

1. Zgjedhje e rastit dhe vërtetimet e pavarura
2. Çdo variabël i pavarur ka shpërndarje normale
3. çdo klasifikim i elementëve në kategoritë e variablave të varur është kryer në mënyrë korrekte
4. Çdo element përfaqësohet në një dhe vetëm një grupim, ku supozohet se  $Y$  është i ndarë në të paktën dy grupe.
5. Çdo kategori është e mirëpërcaktuar, qartësisht e dallueshme nga kategoria tjetër dhe natyrale. Përdorimi i medianës për të krijuar kategori nuk është formë natyrale grupimi. Variabli  $Y$  është mirë të

jetë cilësor , por kur është sasior dhe duhet të krijohen kategori, atëherë duhet të ketë kuptim natyral krijimi i grupeve: psh. Tre grupe të cilat i referohen tre niveleve të huamarrjes për shtëpi.

6. Kategoritë duhet të përcaktohen para se të mblihdhen të dhënat.

7. Kriteret mbi të cilat bëhet ndarja e grupeve duhet të jetë qartësisht diskriminante ndërmjet grupeve , në mënyrë që të mos ketë elementë që të mos i përkasin asnjë grupi dhe që të mos ketë mbivendosje (ose të paktën kjo të jetë minimale)

8. Numri i elementëve për çdo grupi, duhet të jetë shumë i ndryshëm nga njeri tjetri dhe të paktën 5 herë më i madh se numri i variablave të pavarur në model.

### **Cështjet e kërkimit për një AD mund të përfshijnë një apo të gjitha si më poshtë:**

1. ***Së pari, ekziston një test për rëndësinë e gjithë modelit:*** - është testi omnibus për të vlerësuar nëse kur konsiderohen të gjithë variablat dhe funksionet diskriminante, modeli parashikon anëtarësimin sipas grupeve në nivele statsitkisht të rëndësishme më të mira sesa thjesht rastësia. Për të gjithë modelin është statsitika e përputhjes (goodness of fit statsitics) që quhet Wilks's lamda ( $\Lambda$ ) , e cila interpretohet si pjesa e variancës në një apo disa funksione diskriminante) që nuk përputhet me anëtarësimin në grupim. Në fakt Wilks's lamda ( $\Lambda$ ), mund të konsiderohet si masë e mospërputhjes , sepse është një vlerësim i pjesës së pashpjeguar të variancës , analoge me  $(1 - \eta^2)$  të ANOVA dhe  $(1 - R^2)$  te regresioni.

***Ka tre tipe të shkallëve të lirisë që lidhen me Wilks's lamda ( $\Lambda$ ) :*** N= numri total i elementëve në model, k= numri i grupeve të variablit Y, p= numri i variablave parashikues (të pavarur) në model. Programi SPSS e konverton Wilks's lamda ( $\Lambda$ ) në  $\chi^2$ . Pasi  $\Lambda$  konvertohet në një test statistikor me të përdorshëm, përdoren tabela statistikore për të vlerësuar rëndësinë statistikore të modelit AD. SPSS paraqet gjithashtu edhe vlerën e p-së për efekt të testit chi-square, që i korespondon  $\Lambda$ . Hipoteza  $H_0$  për vlerësimin e rëndësisë së AD është: Mesataret e funksioneve diskriminante për çdo grup në popullim janë të barabarta.

Një tjetër mënyrë për të shprehur  $H_0$  do të ishte: Në popullim, anëtarësim në kategoritë përkatëse nuk lidhet me (dmth. nuk mund të parashikohet nga) rezultatet e funksioneve diskriminante.

2. ***Çështja e dytë kërkimore në AD është: Sa funksione diskriminante janë të nevojshme për diferencimin ndërmjet grupeve.*** Kur kemi k=2 grupe, ne mund të sigurojmë një funksion diskriminant. Megjithatë, kur janë më shumë se dy grupe, atëherë do të ketë më tepër se një funksion diskriminant. Numri i funksioneve diskriminante do të jete sa minimumi ndërmjet (k-1) dhe 'p', ku p është numri i variablave të pavarur (x) në model. Funksionet diskriminante janë të tilla që rezultatet në funksionin e parë diskriminant  $D_1$  përmban diferencat më të mëdha të mundshme ndërmjet grupeve. Rezultatet në funksionin e dytë diskriminant  $D_2$ , duhet të jenë jo të koreluara me rezultatet e  $D_1$  dhe rezultatet e  $D_2$  duhet të shpjegojnë (parashikojnë) variancën më të madhe të mundshme midis grupeve, e cila nuk është shpjeguar në  $D_1$ .

Funksionet diskriminante janë rreshtuar nga 1,2,3,4, etj. Renditen nga ai me shumë parashikuesi deri tek ai më pak parashikuesi i anëtarësisë në grupe. Rezultatet e çdo funksioni diskriminatat nuk janë të koreluar me njëri tjetri. Nëse k=5, atëherë do kemi 4 funksione diskriminante. Pyetja që shtrohet është: Sa funksione discriminante duhen për të patur parashikime të sakta të anëtarësimt në grupe? SPSS teston rëndësinë e setit të funksioneve discriminante në tërësi, më mirë se ç'mund të testojë rëndësinë për secilin funksion. Psh në rastin tonë, SPSS do të siguronte testet e rëndësisë për setet e mëposhtme të funksioneve:

Set1:  $D_1, D_2, D_3, D_4$ .

Set 2:  $D_2, D_3, D_4$  (pra është eliminuar  $D_1$ )

Set 3:  $D_3$ ,  $D_4$

Set 4:  $D_4$

Pjesa e variancës midis grupeve që shoqëron çdo set të funksioneve diskriminante zvogëlohet ndërsa në zvogëlojmë numrin e funksioneve diskriminante të përfshira në setet përkatëse. Nëse  $S_2$  është i rëndësishëm, atëherë Set 1 është i rëndësishëm gjithashtu, sepse Set 1 përfshin të gjitha funksionet e Seteve. Ndaj rekomandohet të lexohet rëndësia e seteve nga poshtë lart. Nëse Set 4 është statistikisht i parëndësishëm atëherë heqim  $D_4$  nga interpretimi. Nëse Set 3 është i parëndësishëm, atëherë heqim  $D_3$  nga interpretimi. Nëse Set 2 është i rëndësishëm, atëherë duhet diskutuar për të përfshirë dy funksionet diskriminante  $D_1$  dhe  $D_2$  në interpretim dhe në diskutimin e diferencave të grupeve. Kur ne vendosim të përfshijmë 1 apo 2 funksione diskriminante në interpretimin tonë mbështetur në një set të madh funksionesh diskriminante atëherë kjo quhet 'reduktim dimensionali', kjo do të thotë se na intereson vetëm informacioni për diferencat e grupeve që rrjedhin nga funksioni i parë apo i dytë diskriminant, ndërkohë që të tjerët nuk japin informacione shtesë për diferencën ndërmjet grupeve ose nuk çojnë në përpimërsime në parashikimin e anëtarësimit në grup.

**3. *Një tjetër çështje që shtrohet ka të bëjë me vlerësimin e kontributit individual të variablave të pavarur në model . Cilët variabla janë të nevojshëm në parashikimin e anëtarësimit në grupe?***

SPSS për AD ndryshe nga analiza regressive, nuk ofron teste direkte të rëndësisë statistikore për kontributin e variablave individuale të pavarur. AD paraqet tre pjesë të informacionit në lidhje me kontributin individual , së bashku këto 3 informacione na ndihmojnë të gjykojmë nëse çdo variabël i pavarur ndikon në mënyrë statistikisht të rëndësishme në anëtarësimin e grupeve.

Kështu në fillim përcaktohet sa funksione diskriminante do të interpretohen. Pastaj vetëm për ato funksione vazhdohet me pyetjen: Cilët variabla të pavarur kanë korelacion të lartë (në vlerë absolute) për çdo funksion?

Madhësia e korelacioneve ndërmjet variablave të pavarur dhe funksioneve diskriminante gjykohet në varësi të krahasimeve relative me disa vlera: psh. Një korelacion që kalon vlerën 0,5 mund të interpretohet si një tregues që variabli 'x' është fort i lidhur me funksionin diskriminant .

Se dyti, mund të ekzaminohet koeficienti i standartizuar i funksionit diskriminant (analogu i  $\beta$  Në analizën regressive) për të vlerësuar se cilët variabla kanë efekt më të madh në secilin prej funksioneve diskriminante. Për të parë nëse koeficienti para variablit x lidhet me anëtarësimin në grup, duhen patur parasysh dy situata: *së pari*: a shoqërohet një rezultat i lartë i X me një rezultat të lartë apo të ulët në D? *Së dyti*: secili grup tenton të ketë rezultat të lartë apo të ulët në funksionin D? Të dy këto informacione na ndihmojnë të kuptojmë nëse rezultat i  $X_i$  ndryshon ndërmjet grupeve. *Së treti*: ANOVA-t për çdo rast tregojnë sesi rezultatet për çdo variabël X ndryshojnë ndërmjet grupeve, në rastet kur korelacionet ndërmjet variablave nuk merren në konsideratë. Për secilin variabël X, ka një raport F i cili tregon nëse mesataret ndryshojnë në mënyrë statistikisht të rëndësishme ndërmjet grupeve, duhet patur parasysh në këtë rast se vlerat e p-së që shfaqen në SPSS nuk janë të korigjuara për riskun e inflatuar nga gabimi i llojit të parë.

Nëse të tre keto informacione japin një panoramë konsistente në lidhje me fuqinë dhe natyrën e marrëdhënieve ndërmjet anëtarësimit në grupe dhe variablave X , atëherë mund të shkruhet lehtë në lidhje me faktin sesi rezultatet e  $X_i$  ndikojnë në parashikimin e anëtarësimit nëpër grupe. Ngjashmëria me analizën regressive është edhe në AD, ku: koeficienti i funksionit diskriminant  $d_i$  është koeficienti i standartizuar për  $x_i$  (si  $\beta_i$  e koef i standartizuar për  $b_i$ ) ka një shenjë përpara që i referohet marrëdhënies se  $x_i$  me y, dhe kjo shenjë dhe vlera përkatëse jepet në ANOVA-n përkatëse. Kur rezultatet e AD (në të cilën vlerësimi për vlerat e çdo variabli x që i bashkëngjiten anëtarësimit në grup bëhet nëpërmjet kontrollit të variablave të tjerë X) ndryshojnë cilësisht nga



rezultate e ANOVA-s përkatëse të çdo  $x$  ( në të cilat mesataret e çdo  $x_i$  ndërmjet grupeve vlerësohen në mënyrë të pavarur për çdo variabël  $x$ , pa marrë pasqysh ndërkorelimin ndërmjet  $x_i$  në shqyrtim), atëherë kërkuesi do ti duhet të mendojë me kujdes në lidhje me përshkrimin e ndryshimit ndërmjet grupeve.

4. ***Çilat grupe mund (ose jo) të diskriminohen?*** Ashtu si tek ANOVA, nëse janë më shumë se dy grupe në AD,  $H_0$  jonë është: Për funksionet diskriminate mesataret ndërmjet grupeve janë të barabarta. Nëse gjemë se e gjithë AD është statistikisht e rëndësishme, dhe janë më tepër se dy grupe, atëherë testet follow-up janë të nevojshme për të vlerësuar se cilat kontraste të varuara (çiftore) ndërmjet grupeve janë statistikisht të rëndësishme. Një test i përgjithshëm i AD shpesh do të thotë që të paktën një grup ndryshon rëndësishëm nga të tjerët (ashtu si tek ANOVA, që do të thotë se ka të paktën një nga mesataret e populimeve të ndryshme nga të tjerat). Megjithatë, një test i përgjithshëm i AD nuk do të thotë domosdoshmërisht që të gjitha grupet janë të ndryshme. SPSS i AD paraqet tipe të ndryshme informacionesh në lidhje me ndryshimet e grupeve. Tabela e kontigjencave në fund të përmbledhjes së outputit të AD, përmbledh informacionin në lidhje me numrin e personave të cilët në mënyrë korrekte janë klasifikuar si anëtarë të secilit grup dhe paraqet gjithashtu numrin e klasifikimeve të gabuara. Kjo tabelë mundëson për të parë nëse ka grupe anëtarët e së cilave janë në mënyrë frekvente të keqklasifikuar tek grupet e tjera. Gjithashtu mundësohet edhe 'harta territoriale' për të parë nëse klasterat e rezultateve të funksioneve  $D_1$  dhe  $D_2$  janë qartësisht të ndara apo ato janë të miksuara me njëra tjetrën.

5. ***Çfarë interpretimi kualitativ (në se ka një të tillë) mund ti japim çdo funksioni diskriminant?*** Është e rëndësishme shpesh ti vemë një emër funksionit diskriminant, ky emërtim duhet të bazohet duke konsideruar se cilët variabla  $X$  kanë korelacion të lartë dhe cilat grupe tregojnë diferencat më të mëdha në funksionin përkatës. Emërtimi në këtë rast është analog me emërtimin që i jepet faktorëve apo komponentëve Në analizën faktoriale ka shumë rëndësi ky moment, veçanërisht për studimet bazuar në teori të mirëfillta, pasi funksionet diskriminate konsiderohen si 'dimensione' që bëjnë dallimet ndërmjet grupeve. Psh në një rast hipotetik mund të quajmë  $D_1$  ( funksion i cili tregon korelacion të lartë me ankthin dhe depresionin) dimensionin i 'stresi të përditshëm psikologjik', grupi 'neurotik' ka rezultate të larta në këtë funksion diskriminant, pra rezultate të larta në ankth dhe depresion, krahasuar me dy grupet e tjera: 'pa diagnozë' dhe 'psikotik' Funksioni i dytë diskriminant  $D_2$ , ishte fort i koreluar me deluzionet dhe hallucinacionet, dhe mund të quhet dimensionin 'fokusi në psikoterapi' ose 'çrregullime deluzionale', në këtë dimension grupi psikotik pati rezultate më të larta se grupi pa diagnozë dhe ai neurotik.

Megjithatë në praktike, variablat që i jepet më tepër peshë në formimin e funksioneve diskriminante nuk mund të interpretohen lehtë. Nëse ka vështirësi ose nuk ka sens të mendohet në lidhje me masat që japin kontribut më të madh në një funksion diskriminant sesa në një tjetër funksion, nuk duhen bërë sforcime për interpretime. Në studime ku zgjedhja e variablat  $x$  bëhet nga teoria, ka më tepër kuptim që funksionet diskriminante të interpretohen si dimensione, mbi bazën e të cilave grupet ndryshojnë.

6. ***Cilat janë tipet e gabimeve të mundshme të klasifikimit, të cilat bëjnë sës?*** Një gabim klasifikimi ndodh kur anëtarësimi i parashikuar në një grup nuk përputhet me anëtarësimin real të elementit në shqyrtim. A duhet të modifikohen rregullat për të zvogëluar mundësinë e ndodhjes së tipeve të ndryshme të gabimeve parashikuese të cilat mund të jënë shumë të kushtueshme? Psh. AD, mund të përdoret si një metode statistikore e diagnostifikimit mjeksor (bazuar në raste të mëparshme të kombinimit të simptomave dhe diagnozave përfundimtare ose rezultateve mjeksore finale) ose në vendimarrje që lidhen me marrje në punë ose pranim në universitete (bazuar në të kaluarën apo në kredenciale të mëparshme të aplikantëve dhe suksesin/dështimin në punë apo në programin

akademik). Në situata të tilla aplikative, qëllimi i kërkimit është të gjenerohet formula që të mund të përdoret për parashikime në të ardhmen në lidhje me rastet individuale për të cilat gjendja mjeksore finale, apo rezultati në punë apo në shkollë ende nuk dihen.

Le të shohim rastin kur pacienti ka atak zemre (grupi 1) ose nuk ka (grupi 2) Bazuar në faktorët e lartpërmendur, mjekut i duhet të diagnotifikojë si anëtarë potencial të njerit nga grupet një pacient të cilit i janë bërë matjet e presionit të gjakur, kolesterolit, triglicerideve etj. Mirëpo bërja e këtyre testeve, ose e ekzaminimeve paraprake kërkon kohë dhe testet ndonjëherë janë dhe evazive... Ndaj ka dy tipe gabimesh klasifikimi në këte rast: pozitiv fals (mjeku mendon se ai është në grupin 2, kur në realitet është e kundërta) dhe negativ fals (mjeku mendon se ai është në grupin 1, kur në realitet është e kundërta) Kostot dhe pasojat e këtyre dy tipe keqklasifikimi secili mund ti interpretojë dhe janë të mëdha, deri fatale në humbje jete. Në rastet e mjeksisë, merren masa të rrepta për të limituar sa të munden këto tipe keqklasifikimesh. Një mënyrë do të ishte përfshirja në model e më shumë variabla X, një tjetër mënyrë për të shmangul gabimin `negativ fals` do të ishte ndryshimi i kufijve të variabla. Psh. nëse do zvogëlohej `negativ fals` kjo do bëhej nëpërmjet konsiderimit si hipertension jo ata me presion më të madh se 145, por ulja e kufirit në më të madh se 130, por kujdes, kjo të rriste gjithashtu gabimin `pozitiv fals`.

Në studimet teorike kjo çështje nuk ka shumë rëndësi, por në çështjet mjeksore kjo merr rëndësi të konsiderueshme.

## **2. PREZANTIMI I NJË RASTI EMPIRIK**

Shembulli ynë empirik do ti referohet rastit ku janë zgjedhur disa studentë, dhe janë pyetur në lidhje me preferencën (variabli Y) : Biznes (grupi 1), Mjeksi (grupi 2), Sociologji (grupi 3). Gjithashtu për secilin prej studentëve janë marrë informacion në lidhje me variablat (X) si: Gjinia, Pikët në testin e: Anglishtes, Leximit, Mekanikës, Krijimtarisë letrare, Matematikës Menaxhimit. Hedhja e të dhënave në SPSS bëhet njëlloj si për regresionin.

## **3. SUPOZIMET E ANALIZES DISKRIMINANTE**

AD përdor elementë të analizës regresive (konsideron interkorelacionet ndërmjet variabla x për të ndërtuar ekuacionet optimale për cdo funksion diskriminant) dhe të ANOVA-s (vlerëson ndryshimet e mesatareve ndërmjet grupeve ose kategorive). E thënë shkurt supozimet e AD janë ato të analizës regresive dhe të Anova-s, por ka edhe disa të tjera. Ashtu si edhe te regresioni, edhe në AD, variablat sasiore X duhet të kenë shpërndarje normale, pa ekstreme (outliers), duhet të lidhen në mënyrë lineare me njëri tjetrin pa vlera më të largëta (kjo shihet nga scatterplots). Ashtu si tek ANOVA, numri i grupeve duhet të jetë më i madh se një vlerë e caktuar minimale (nëse numri i rasteve brenda një grupi është më i vogël se numri i variabla të pavarur, atëherë nuk mund të bëhen vlerësime të pavarura të të gjithë vlerave të variancave dhe kovariancave brenda grupeve, duhen përmasa më të mëdha të zgjedhjes që të rritet fuqia statistikore).

Një tjetër supozim shtesë, ka të bëjë me natyrën e multivariabla të një supozimi tashmë të njohur: Kur realizojmë një AD, duhet të llogaritet Shuma e produkteve të kryqëzuara (SCP- Sum of Cross products). Kështu kemi SCPi për çdo grup i. Elementët e matricës së zgjedhjes SCP jepen si më poshtë: matrica SCP llogaritet në mënyrë të veçantë për çdo grup, dhe gjithashtu edhe të gjithë setin e të dhënave, ku kombinohen të gjitha grupet:

**SCP =**

SS (X<sub>1</sub>)

SCP (X<sub>1</sub>,X<sub>2</sub>) ...

SCP (X<sub>1</sub>, X<sub>p</sub>)

$$\begin{array}{ccccccc} \text{SCP} (X_2, X_1) & \text{SS} (x_2) & & \dots & & \text{SCP} (X_2, X_p) & \\ \dots & \dots & & & & \dots & \\ \text{SCP} (X_p, X_1) & \text{SCP} (X_p, X_2) & & \dots & & \text{SS} (X_p) & \end{array}$$

Elementët diagonal të cdo SCP në matricë janë shuma të katrorëve të  $x_1, x_2, \dots, x_p$ . Çdo term SS llogaritet si  $\text{SS} (x_i) = \sum (X_i - M_{X_i})^2$ . Vlerat diagonale të SCP të matricës llogaritet si  $\sum$  e produkteve të kryqëzuara për të gjitha çiftet e mundshme të variablave X. Për  $X_i$  dhe  $X_j$  kemi  $\text{SCP} = \sum (X_i - M_{X_i})(x_j - M_{X_j})$ . Tek ANOVA kemi supozimin e homogjenitetit të variancës, pra variancë e Y supozohet e barabartë për çdo grup. Në AD, supozimi analog është që elementët e variancës/kovariancës të paraqitura në matricë janë homogjenë ndërmjet grupeve.

Nga matrica SCP kalohet në matricën e mëposhtme S, e cila përftohet duke pjestuar çdo element të SCP me numrin e përshtatshëm të shkallëve të lirisë ose `n` duke konvertuar çdo SS në variancë dhe çdo SCP në kovariancë si më poshtë:

**SS =**

$$\begin{array}{ccccccc} \text{Var} (X_1) & & \text{Cov} (X_1, X_2) & & \dots & & \text{Cov} (X_1, X_p) \\ \text{Cov} (X_2, X_1) & \text{Var} (x_2) & & \dots & & \text{Cov} (X_2, X_p) & \\ \dots & \dots & & & & \dots & \\ \text{Cov} (X_p, X_1) & \text{Cov} (X_p, X_2) & & \dots & & \text{Var} (X_p) & \end{array}$$

Në kushtet e zgjedhjes kemi matricën S, në kushtet e popullimit kemi matricën  $\Sigma$ , ku çdo element i diagonales së matricës  $\Sigma$  i korespondon variancave të popullimit,  $\delta_1^2, \delta_2^2 \dots \delta_p^2$  dhe çdo element jo diagonal i korespondon kovariancave të popullimeve përkatëse. Supozimi i AD është që këto  $\Sigma_i$ , variancë apo kovarianca të popullimeve janë të njëjta ndërmjet grupeve, duke përfshirë kështu edhe supozimin që variancat ndërmjet grupeve janë homogjene. Ky supozim i homogjenitetit të variancave/kovariancave ndërmjet grupeve shkruhet si më poshtë:  $H_0: \Sigma_1 = \Sigma_2 = \dots \Sigma_p$ . Pra variancat për çdo variabël janë të barabarta ndërmjet grupeve, dhe kovariancat ndërmjet çdo çifti të variablave X supozohen të barabarta ndërmjet grupeve.

Në SPSS testi Box M, tregon nëse plotësohet apo jo ky supozim. Ka dy probleme që lidhen me testin Box M:

1. Është shumë i ndjeshëm ndaj shpërndarjes jonormale të rezultateve .
2. Kur numri i elementëve në AD është i madh, testi Box M mund të tregojë cënim statistikisht të rëndësishëm të supozimit të homogjenitetit të variancave/kovariancave, edhe kur kjo si çështje nuk është shumë problematike.

Nga ana tjetër kur numri i rasteve në AD është i vogël, testi Box M mund të mos rezultojë statistikisht i rëndësishëm për  $\alpha = 0,05$ , edhe kur cënimi i supozimit të homogjenitetit është problematik si çështje. Për të kompensuar këtë problem, si edhe në analiza të tjera edhe në SD, përdoret një  $\alpha$  më e vogël, psh  $\alpha = 0,01$  ose  $\alpha = 0,001$  kur numri i elementëve është i madh, dhe një  $\alpha$  më e madhe psh  $\alpha = 0,1$  kur numri i elementeve është më i vogël. Nëse rezultojnë që cënohet supozimi i homogjenitetit, atëherë ndiqen teknikat tashmë të njohura si psh: transformimi logaritmik në rastin e shpërndarjes jonormale, ose largimi i vlerave jonormale, ose largimi i vlerave ekstreme duke eliminuar ose variabla të veçantë ose elementë të veçantë ose edhe grupe të veçanta.

#### **4. ÇËSHTJE NË PLANIFIKIMIN E STUDIMIT**

Një çështje e rëndësishme veç atyre të trajtuara deri tani, është numri i elementëve për çdo grup në një AD. Si në çdo analizë, do ishte shumë mirë të kishim numër të njëjtë të elementëve për çdo grup. Nëse nuk është kështu, atëherë studiuesi duhet të zgjedhë ndërmjet dy metodave të

klasifikimit të rezultateve në grupet e parashikuara. 1. Metoda klasike e klasifikimit në SPSS për AD, supozon që çdo model duhet të klasifikojë të njëjtin proporcion në çdo grup pavarësisht anëtarësisë aktuale. Psh. Nëse janë 4 grupe ,  $k = 4$ , atëherë 25% e elementëve klasifikohen të secili grup. 2. Një tjetër metodë në SPSS është quajtur Priors= Size,e cila përdor proporcionet e elementëve në zgjedhje si bazë për rregullin e klasifikimit. Psh. Nëse  $k=4$ , dhe  $n_i= 40, 25, 25, 10$  atëherë rreth 40% e rasteve do klasifikohen në grupin 1, 25% në grupin 2, 25% në grupin 3 dhe 10% në grupin 4. Numri i elementëve brenda çdo grupi duhet ta tejkalojë numrin e variablave të pavarur, ky është minimumi ; por sa më i madh numri, aq më e madhe fuqia statistikore e AD. Kjo sepse nëse numri i elementëve do të ishte më i vogël, atëherë nuk kalohej dot nga matrica SCP në matricën S, pasi s' do kishte mjaftueshëm shkallë lirie . Mirëpo në SPSS, kjo realizohet edhe në këto kushte në një analizë AD, ndaj ne duhet të jemi vetë të ndërgjegjshëm paraprakisht për këtë çështje.

Një numër i vogël i elementëve për çdo grup ka dhe një tjetër problem, që vështirëson vlerësimin e supozimit për homogjenitetin e variancave/kovariancave, pasi zvogëlon fuqinë statistikore të testit Box M.

Një tjetër çështje ka të bëjë me zgjedhjen e variablave të pavarur. Në një kërkim të aplikuar do zgjidhen si të tillë që të jenë lehtësisht të matshëm pa shumë kosto. Në kërkimet teorike, këto variabla i dikton baza teorike. Kujdes futja në model e variablave `plehra` pa lidhje me çështjen që po trajtohet, dobëson fuqinë statistikore të modelit, duke zvogëluar kështu kontributin e variablave të tjerë dhe duke vështirësuar integrimin e tyre në model.

## 5. EKUACIONET E AD

AD prodhon shumë informacion. Një pjesë e informacionit është e ngjashme me ANOVA, një pjesë tjetër me analizën e regresionit. Testi statistikor omnibus, i cili tregon sa mirë grupet mund të dallojnë me njëri tjetrin, duke përdorur të gjithë variablat dhe të gjithë funksionet AD quhet Wilks  $\Lambda$ . Wilks  $\Lambda$  interpretohet si proporcion i variancës në rezultatin e funksioneve diskriminante që nuk parashikohet nga anëtarësimi në grup. Kështu Wilks  $\Lambda$  është analog me  $(1 - \eta^2)$  të ANOVA dhe  $(1 - R^2)$  të regresioni.

Tek ANOVA kishim  $SS_{total} = SS_{between} + SS_{within}$ , ku  $SS_{between}$  e  $SS_{within}$  i referoheshin setit të variancave dhe kovariancave ndërmjet dhe brenda grupeve.

Në AD rolin e këtyre seteve e luan seti i SCP përkatësisht:

$$SCP_{total} = SCP_{between} + SCP_{within} \quad (ek.1)$$

Ku SCP total i referohet totalit të shumave të katrorëve për  $X_s$  (në elementët diagonal) dhe shumave totale të produkteve kryq për të gjithë çiftet e variablave X (per elementët jo diagonal) për të gjithë bazën e të dhënave për të gjitha grupet e kombinuara. Brenda çdo grupi (i) ne kemi  $SCP_i$ , elementët e  $SCP_i$  përbëhen nga shumat e katrorëve të  $X_s$  vetëm për rezultatet Brenda grupit `i`. Sigurojmë kështu një vlerësim të termit të gabimit Në analizën tonë  $SCP_{within}$ , duke interpoluar këto matrica  $SCP_i$  ndërmjet grupeve (ky veprim merr sens nëse plotësohet supozimi i homogjenitetit  $\Sigma$  ndërmjet grupeve. Kështu:

$$SCP_{within} = (SCP_1 + SCP_2 + \dots + SCP_k) / (n_1 + n_2 + \dots + n_k - k) \quad (ek. 2)$$

Ku k është numri i grupeve.

Kjo është anloge me  $SS_{within}$  të ANOVA ku  $SS_{within} = SS_1 + SS_2 + \dots + SS_k$ .

Tek ANOVA për të gjykuar në lidhje me madhësinë e variancës së shpjegueshme nga ajo e pashpjegueshme i referohesmi  $F = MS_{between} / MS_{within}$ . Ne AD ne i referohemi:

$$\Lambda = ISCP_{within} / ISCP_{total} \quad (ek. 3)$$

Në një AD me vetëm një variabël të pavarur do kishim:

$$\Lambda = SS_{\text{within}} / SCP_{\text{total}} \quad (\text{Ek. 4})$$

$\Lambda$  është ekuivalentja e  $(1 - \eta^2)$  të ANOVA ku  $\eta^2 = SS_{\text{between}} / SS_{\text{total}}$ . Statistikisht F të ANOVA ka dy shkallë lirie (within, between). Statistika Wilks  $\Lambda$  ka tre shkallë lirie në AD:  $n$ = numri total i elementëve,  $k$ = numri i grupeve,  $p$ = numri i variablave të pavarur

Megenëse është e vështirë të përdoren tabela statistikore 3 dimensionale atëherë  $\Lambda$  konvertohet në  $\chi^2$ . SPSS e konverton  $\Lambda$  me  $\chi^2$  me shkallë lirie  $p(k-1)$  ku

$$\chi^2 = -\{n - [(p+k)/2] - 1\} \ln \Lambda \quad (\text{ek. 5})$$

Së fundmi mund të përcaktojmë efektin e përmasës për analizën AD:

$$\eta^2 = 1 - \Lambda \quad (\text{ek.6})$$

Gjithë sa u tha më lart, i referohet testit omnibus të AD. Nëse Wilks  $\Lambda$  është e vogël, dhe nëse F e omnibus koresponduese është statistikisht e rëndësishme, mund të konkludojmë që anëtarësim në grup mund të parashikohet në nivel më të mirë statistikor sesa kjo mund të bëhet duke konsideruar të gjithë funksionet diskriminante si set.

Çështja tjetër në interes është: Cilët nga variablat e pavarur sigurojnë informacion të vlefshëm në lidhje me anëtarësimin në grupe? Për këtë i referohemi vlerave dhe shenjave të koeficientëve para variablave në funksionet diskriminante. Funksionet diskriminante jepen si:  $D_1, D_2, D_3 \dots D_{k-1}$ . Për secilin funksion diskriminant  $D_i$  ka koeficientë si më poshtë:  $D_i = d_{i1}z_1 + d_{i2}z_2 + \dots d_{ip}z_p$

Pra, çdo funksion diskriminant  $d_i$  ndërtohet nga kombinimi linear i rezultateve  $z: z_1, z_2, \dots z_p$ . Për çdo funksion diskriminant që ndryshon statistikisht midis grupeve (dmth. parashikon anëtarësimin në grupe) Studiuesi mund të dojë të interpretojë kuptimin e funksionit duke treguar se cili nga variablat ka peshën më të madhe (vlerën e koeficientit) në llogaritjen e funksionit diskriminant. Koeficientët  $d$  janë analoge me koeficientët  $\beta$  Në analizën regresive. Ata sigurojnë kombinimin optimal linear të variablave  $X$  për të shpjeguar  $Y$ . Pra koeficientet  $d$  janë një set vlerash, të cilët së bashku me rezultatet  $z$  të variablave të pavarur bëjnë të mundur që çdo funksion diskriminant  $D_i$  të ketë vlerën më të lartë të mundshme  $SS_{\text{between}}$  dhe vlerën më të vogël të mundshme  $SS_{\text{within}}$ . Koeficientët e funksionit diskriminant për një AD sigurohen nga një llogaritje algjebre matricore që s'ë kemi hasur më parë: Problemi vlereseigen/vektoreigen. Vektoreigen është një vektor ose një listë koeficientësh, të cilët i korespondojnë variablave të pavarur kur ne formojmë kombinimet lineare të variablave nga seti i  $X_s$ . Elementët e vektorit eigen duhet të rishkallëzohen për tu përdorur si koeficientë të funksionit diskriminant. Çdo vektoreigen ka një vlerëeigen koresponduese. Në AD, vleraeigen ( $\lambda$ ) për një funksion të caktuar diskriminant i korespondon vlerës së përmasës së përputhjes  $SS_{\text{between}}/SS_{\text{within}}$ . Me fjalë të tjera vleraeigen që i korespondon funksionit diskriminant na jep raportin  $SS_{\text{between}}/SS_{\text{within}}$  që rezulton nga ANOVA e funksionit diskriminant respektiv. Kur rishkallëzohet vektoreigen jep koeficientët  $d$  për një funksion diskriminant, njëllë si koeficienti  $\beta$  Në analizën regresive, ku ne i referohemi rezultateve  $z$ .

Vleraeigen ( $\lambda$ ) çon në një përmasë efekti (korelacioni kanonikal  $r_c$ ) që tregon sesa fort rezultatet e funksionit diskriminant lidhen me anëtarësimin në grup. Termi 'kanonikal' tregon që korelacioni është ndërmjet koeficientëve që përfshin shumë variabla të pavarur. Korelacioni kanonikal ( $r_c$ ) ndërmjet rezultateve të funksionit diskriminant dhe anëtarësimt në grup është analog me  $\eta$  të ANOVA. Indeksi

$$\sqrt{\frac{\lambda_1}{1 + \lambda_1}}$$



$\lambda_i$  i referohet numrit të funksioneve diskriminante, psh.  $R_{c2}$  do të thotë korelacion kanonikal i  $D_2$ . Vlerat e  $\lambda_i$  për secilin funksion diskriminant mund të përdoret për të llogaritur  $r_{ci}$  psh:

$$R_{c1} = \quad \quad \quad (\text{ek.7})$$

Algjebra matricore për problemin e vlerësive gjeneron rezultatet me karakteristikat si më poshtë: 1) Rezultatet janë renditur sipas madhësisë së vlerësive, që do të thotë se vlera e parë i korrespondon funksionit diskriminant me  $\lambda_1$  dhe  $r_{c1}$  më të madhe të mundshme 2) E dyta i korrespondon vlerës së dytë më të madhe të mundshme  $\lambda_2$  dhe  $r_{c2}$  e kështu me radhë.

Rezultatet në funksionet diskriminante duhet të mos jenë të koreluar me njëri tjetrin, dmth. koeficientët e  $D_1$  të mos jenë të koreluar me ato të  $D_2$  dhe me asnjë nga koeficientët e çdo funksioni diskriminant.

Meqënëse rezultatet janë të pakoreluar është më lehtë të kombinohet varianca e shpjeguar midis funksioneve diskriminante dhe të përcaktohet kështu se sa mirë një set i funksioneve diskriminante lidhet me parashikimin e anëtarësimit në grupe.

Një mënyrë që ndiqet për të përcaktuar sesa funksione diskriminante do të përdoren për interpretim është të vendoset një nivel  $\alpha = 0,05$ ; mund të shikohet tek vlerat p në SPSS tek seti i funksioneve diskriminante, dhe të largohen ato funksione diskriminante nga interpretimi, të cilat nuk janë statistikisht të rëndësishme në kontributin që japin për parashikimin e anëtarësimit në grupe, sikurse u përmend më parë në reduktimin dimensional. Ekuacioni që përdoret për koeficientët e standartizuar të funksionit diskriminant është:

$$D_i = d_{i1}z_1 + d_{i2}z_2 + \dots + d_{ip}z_p \quad (\text{ëk. 8})$$

Ku i tregon numrin e funksioneve diskriminante 1,2, ... k-1 dhe numri 1,2, ... p tregon variablin.-

Këto koeficientë  $d_i$  si dhe koeficientët  $\beta$  tek regresioni marrin vlera nga -1 ne +1 (por mund të ketë dhe raste përjashtimore). Edhe interpretimi i koeficientëve  $d_i$  është i ngjashëm me  $\beta$ :  $\beta$  tregon se sa peshë i jepet variablit të standartizuar të  $X$  përkatës kur të gjithë variablat e tjerë  $x$  janë statistikisht të kontrolluar. SPSS jep koeficientë të papërpunuar të funksioneve diskriminante, por këto nuk përdoren për të interpretuar natyrën e ndryshimit të grupeve, por mund të përdoren për të gjeneruar një formulë parashikuese, që mund të përdoret për të parashikuar anëtarësimin në grup nga të dhënat bazë të realitetit për elementët e tjerë të papërfshirë në model. SPSS raporton gjithashtu koeficientët e strukturës që do të thotë se raportohet për çdo variabël të pavarur korelacioni i tij me secilin funksion diskriminant.

Disa studiues i referohen në interpretime madhësisë së koeficientëve të standartizuar të funksioneve diskriminante, të tjerë i referohen interpretimeve bazuar në setin e koeficientëve të strukturës.

## **7. BAZA KONCEPTUALE: FAKTORËT QË NDIKOJNË NË MADHËSINË E WILK A**

Nëse grupet në AD kanë variancë të madhe midis tyre (dhe relativisht të vogla brenda tyre) për një ose më shumë variabla  $x$ , ka të ngjarë që Wilk  $\Lambda$  do të jetë e vogël. Megjithatë ka të tjerë faktorë që influencojnë përmasën e Wilks  $\Lambda$ , si psh. fuqia dhe shenjat e korelacioneve ndërmjet variablave  $x$ ,

dhe fakti nëse kombinimet e variablave diferencojnë grupet më shumë efektivitet se variabli i vetëm apo jo.

## 8. PËRMASA E EFEKTIT

Një pyetje që lidhet me AD është: Sa mirë mund të bëhet parashikimi i anëtarësisë nëpër grupe, duke përdorur të gjithë variablat dhe të gjithë funksionet diskriminante? Përgjigja lidhet me Wilks  $\Lambda$  për të gjetur përmasën korresponduese të efektit  $\eta^2$  ku

$$\eta^2 = 1 - \Lambda \quad (\text{ek. 9})$$

Për çdo funksion diskriminant, mund të llogaritni indeksin e përmasës së efektit (korelacionin kanonik  $r_c$ ) nga vlerat eigen  $\lambda_i$ , ku

$$R_{ci} = (\text{ek.10}) \sqrt{\frac{\lambda_i}{1 + \lambda_i}}$$

Nuk ka indeks për përmasën e efektit për çdo variabël individual  $x$  që të mund të interpretohet si porcion i variancës. Mund të vlerësohet rëndësia relative e variablave  $x$  duke iu referuar koeficientëve të strukturës, koeficientëve të standartizuar të funksionit diskriminant dhe vlerave  $F$  (nga ANOVA për secilin variabël  $X$ ). Megjithatë këto krahasime janë cilësore dhe deklarata për të gjykuar se cili  $X$  jep kontribut të fortë në parashikim është arbitrare.

## 9. TESTET `` FOLLOW UP `` PËR TË VLERËSUAR S E CILAT ÇIFTE VARIABLASH DIFERENCOJNË MË MIRË GRUPET

Nëse gjithë modeli AD është statistikisht i parëndësishëm (nëse  $\chi^2$  që i referohet Wilks  $\Lambda$  për të gjithë setin e funksioneve diskriminante është statistikisht i parëndësishëm) atëherë nuk ka kuptim të vazhdoen analiza të mëtejshme follow up për të vlerësuar se cilat çifte të grupeve ndryshojnë statistikisht më njëri tjetrin dhe cilët variabla  $X$  japin informacion të mjaftueshëm për këtë.

Nëse ka vetëm 2 grupe në AD, nuk ka sens të realizohen teste të mëtejshme për të vlerësuar kontrastet. Megjithatë nëse ka më tepër se dy grupe, krahasimet post-hoc i grupeve do të ndihmonte për të trajtuar se cilat çifte grupesh ndryshojnë statistikisht në një apo disa funksione diskriminante. Kjo mund të bëhet nga MANOVA. Për të parë se cilët nga variablat  $X$  ndikojnë më shumë në parashikimin e anëtarësisë në grupe, mund të bëhet një analizë cilësore: Variablat me koeficientë (në vlerë absolute) relativisht të madh janë më shumë parashikues në kushtet kur merren në konsideratë korelacionet e variablit  $x$  me variablat e tjerë.

Mund të ekzaminohet gjithashtu edhe ANOVA në të cilën mesataret e çdo variabli  $X$  krahasohen ndërmjet grupeve. (po nuk kontrollon ndërkorelacionet midis variablave  $x$ ), por këto ANOVA duhet të shikohen me kujdes për dy arsye: **Së pari**: rëndësia statistikore e vlerës  $p$  që gjeneron SPSS për këto ANOVA nuk është i korigjuar për të limituar riskun e gabimit të llojit të parë që ndodh kur realizohet një numër i madh testesh statistikore. Në raste të tilla është mirë të përdoret korigjimi Bonferroni për

të eliminuar këtë problem (gabimi i llojit të parë të eksperimentit  $\alpha_e = 0,05$ . Për një set testesh q, kemi  $\alpha = 0,05/q$ .

**Së dyti:** këto teste nuk sigurojnë kontroll të interkorelacionit ndërmjet variablave x. Kemi vetëm korelacione të pjesshme ndërmjet  $X_i$  dhe Y por pa marrë parasysh ndërvarësinë e  $X_i$  me njëri tjetrin. Siç edhe është thënë më lart: Vlerësimi i dobisë së variablit X në model bëhet nëpërmjet 3 informacioneve: **1.** Madhësia e shenjave të koeficientëve të standartizuar para x përkatës në funksionin diskriminant. **2.** Madhësia dhe shenja e koeficientit të strukturës **3.** Rëndësia statistikore e raportit F për çdo variabël x si parashikues. Konkluzioni në lidhje me rëndësinë e variablit  $X_i$  si parashikues, qoftë në shenjë qoftë në madhësi mund të rezultojë i ndryshëm nisur nga 3 informacionet e përmendura. Është më lehtë të deklarosh konkluzionin nëse të 3 opsionet janë konsistente në përfundim.

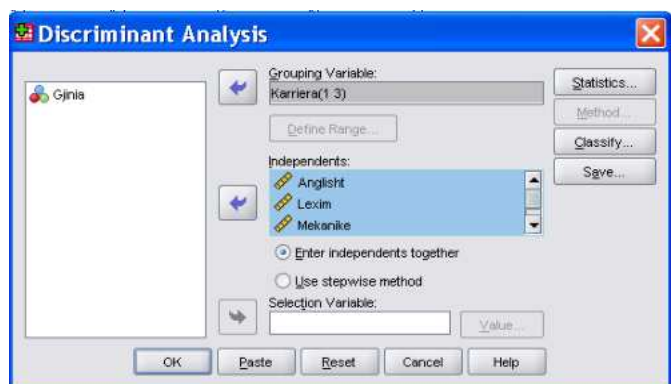
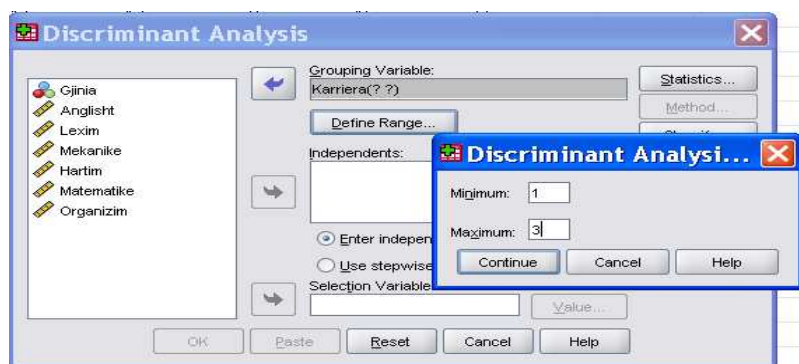
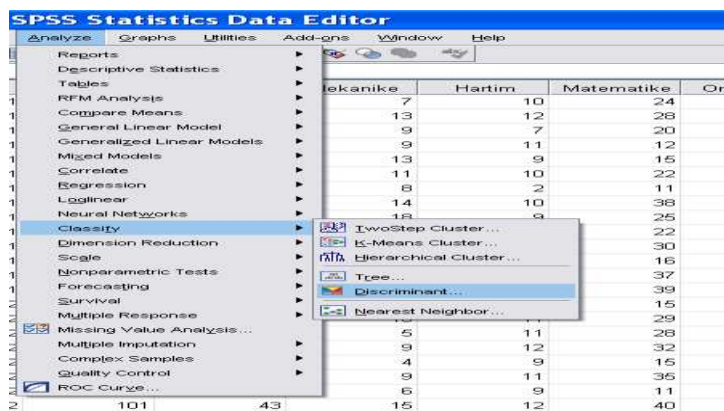
Kur tre opzionet që dhamë janë jo konsistente kjo do të thotë që roli i  $X_i$  si parashikues ndryshon kur shihet  $x_i$  ndaj Y dhe variablat e tjerë  $X_i$  injorohen (ANOVA) nga rasti kur shihet  $X_i$  ndaj Y dhe variablat e tjerë  $X^i$  janë statistikisht të kontrolluar (koeficientët e funksionit diskriminant). Nëse asnjëri nga X nuk është statistikisht i rëndësishëm, atëherë as edhe kombinimi linear i  $x_s$  i shfaqur në formën e funksioneve diskriminante nuk mund të përdoret për parashikim të anëtarësisë në grupe.

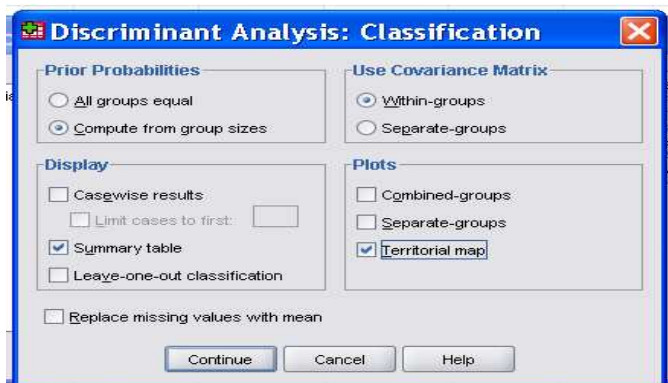
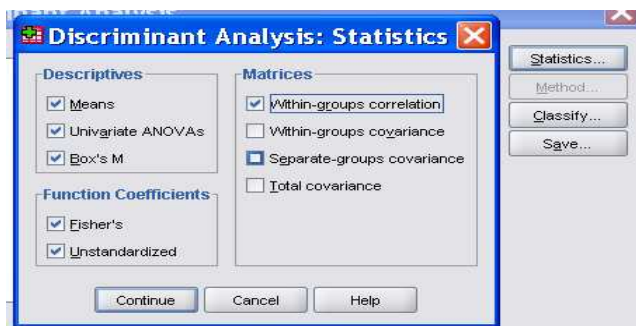
Megjithatë, mund të ndodhë edhe që të mund të ndërtohet modeli duke kombinuar 2 ose më shumë variabla x, edhe nëse asnjëri në veçanti nuk është statistikisht i rëndësishëm për parashikimin e Y. Nga ana tjetër nëse AD në tërësi rezulton statistikisht i parëndësishëm atëherë në të shumtën e rasteve kjo nënkupton që asnjë nga variablat X nuk është statistikisht i rëndësishëm për parashikimin e anëtarësisë në grupe. Por ka edhe raste përjashtimore kur rezulton që AD është statistikisht i parëndësishëm në tërësi, mirëpo disa nga X e përfshirë në model janë në mënyrë individuale statistikisht të rëndësishme. Ky është sinjal që variablat e tjerë `plehra` në model ulin shkallën e lirisë së gabimit duke çuar në ulje të fuqisë statistikore.

## **10. REZULTATET**

Rasti studimor i zgjedhur bazohet në një zgjedhje të 169 studentëve që janë pyetur në lidhje me preferencat e tyre në lidhje me degën e studimit: 1=Biznes; 2= Mjeksi; 3= Sociologji. Për këtë janë parë dhe rezultatet e gjithsecilit në 6 teste: Anglisht, Lexim, Mekanikë, Hartim, Matematikë, Organizim. Këto konsiderohen si variabla x të pavarur. Qëllimi është të vlerësohet nëse zgjedhja e karrieres është e parashikueshme nga këta variabla.

Të dhënat janë hedhur në SPSS, dhe komandat jepen hap pas hapi si më poshtë për realizimin e AD në SPSS:





Rezultatet e outputit SPSS janë :

***Tab. 1: Statistika deskriptive të AD***

### Group Statistics

|        |          | Mean  | Std.<br>Deviation | Valid N (listwise) |          |
|--------|----------|-------|-------------------|--------------------|----------|
|        |          |       |                   | Unweighted         | Weighted |
| Biznes | Anglisht | 88.29 | 8.682             | 100                | 100.000  |
|        | Lexim    | 32.41 | 8.424             | 100                | 100.000  |
|        | Mekanike | 9.19  | 3.681             | 100                | 100.000  |
|        | Hartim   | 9.10  | 2.452             | 100                | 100.000  |



|            |            |       |        |     |         |
|------------|------------|-------|--------|-----|---------|
|            | Matematike | 20.07 | 7.089  | 100 | 100.000 |
|            | Organizim  | 28.26 | 9.146  | 100 | 100.000 |
| Mjeksi     | Anglisht   | 86.90 | 9.634  | 29  | 29.000  |
|            | Lexim      | 32.34 | 12.364 | 29  | 29.000  |
|            | Mekanike   | 9.59  | 5.565  | 29  | 29.000  |
|            | Hartim     | 9.62  | 2.541  | 29  | 29.000  |
|            | Matematike | 26.03 | 10.425 | 29  | 29.000  |
|            | Organizim  | 18.17 | 9.681  | 29  | 29.000  |
| Sociologji | Anglisht   | 88.38 | 13.335 | 40  | 40.000  |
|            | Lexim      | 34.95 | 8.817  | 40  | 40.000  |
|            | Mekanike   | 9.53  | 3.063  | 40  | 40.000  |
|            | Hartim     | 9.45  | 3.537  | 40  | 40.000  |
|            | Matematike | 24.23 | 9.308  | 40  | 40.000  |
|            | Organizim  | 20.70 | 8.489  | 40  | 40.000  |
| Total      | Anglisht   | 88.07 | 10.073 | 169 | 169.000 |
|            | Lexim      | 33.00 | 9.302  | 169 | 169.000 |
|            | Mekanike   | 9.34  | 3.919  | 169 | 169.000 |
|            | Hartim     | 9.27  | 2.751  | 169 | 169.000 |
|            | Matematike | 22.08 | 8.604  | 169 | 169.000 |
|            | Organizim  | 24.74 | 10.017 | 169 | 169.000 |

**Tab. 2: Rezultatet e ANOVA 1 Faktoriale per AD**

**Tests of Equality of Group Means**

|            | Wilks' Lambda | F      | df1 | df2 | Sig. |
|------------|---------------|--------|-----|-----|------|
| Anglisht   | .997          | .237   | 2   | 166 | .789 |
| Lexim      | .986          | 1.154  | 2   | 166 | .318 |
| Mekanike   | .998          | .173   | 2   | 166 | .841 |
| Hartim     | .994          | .509   | 2   | 166 | .602 |
| Matematike | .916          | 7.586  | 2   | 166 | .001 |
| Organizim  | .814          | 19.018 | 2   | 166 | .000 |

**Tab. 3: Matrica e korelacioneve brenda grupeve**

### Pooled Within-Groups Matrices

|                      | Anglisht | Lexim | Mekanike | Hartim | Matematike | Organizim |
|----------------------|----------|-------|----------|--------|------------|-----------|
| Correlation Anglisht | 1.000    | .602  | .193     | .490   | .530       | .049      |
| Lexim                | .602     | 1.000 | .421     | .455   | .511       | -.079     |
| Mekanike             | .193     | .421  | 1.000    | .352   | .510       | -.234     |
| Hartim               | .490     | .455  | .352     | 1.000  | .473       | -.017     |
| Matematike           | .530     | .511  | .510     | .473   | 1.000      | -.192     |
| Organizim            | .049     | -.079 | -.234    | -.017  | -.192      | 1.000     |

***Tab. 4: Testi Box M per barazine e variancave/kovariancave***

### Log Determinants

| Karriera             | Rank | Log Determinant |
|----------------------|------|-----------------|
| Biznes               | 6    | 19.429          |
| Mjeksi               | 6    | 21.886          |
| Sociologji           | 6    | 21.160          |
| Pooled within-groups | 6    | 20.725          |

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

### Test Results

|           |           |
|-----------|-----------|
| Box's M   | 78.939    |
| F Approx. | 1.746     |
| df1       | 42        |
| df2       | 23705.058 |
| Sig.      | .002      |

Tests null hypothesis of equal population covariance matrices.

***Tab. 5: Permbledhje e funksioneve diskriminnate kanonike***

Eigenvalues

| Function | Eigenvalue        | % of Variance | Cumulative % | Canonical Correlation |
|----------|-------------------|---------------|--------------|-----------------------|
| 1        | .347 <sup>a</sup> | 95.3          | 95.3         | .508                  |
| 2        | .017 <sup>a</sup> | 4.7           | 100.0        | .130                  |

a. First 2 canonical discriminant functions were used in the analysis.

#### Wilks' Lambda

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---------------------|---------------|------------|----|------|
| 1 through 2         | .730          | 51.484     | 12 | .000 |
| 2                   | .983          | 2.783      | 5  | .733 |

**Tab. 6: Koeficientët e standartizuar të funksionit diskriminant kanonikal**

|            | Standardized Discriminant Coefficients | Canonical Function |
|------------|----------------------------------------|--------------------|
|            | Function                               |                    |
|            | 1                                      | 2                  |
| Anglisht   | .410                                   | -.023              |
| Lexim      | -.041                                  | 1.228              |
| Mekanike   | .455                                   | -.196              |
| Hartim     | -.111                                  | -.299              |
| Matematike | -.746                                  | -.418              |
| Organizim  | .750                                   | -.189              |

**Tab. 7: Matrica strukture: Korelacioni i çdo variabli X me funksionet diskriminante (psh. Rezultatet e testit të Organizimit korelohen me +0,8 me rezultatet e funksionit diskriminant 1**

#### Structure Matrix

|            | Function           |       |
|------------|--------------------|-------|
|            | 1                  | 2     |
| Organizim  | .812 <sup>*</sup>  | -.156 |
| Matematike | -.513 <sup>*</sup> | -.008 |
| Hartim     | -.133 <sup>*</sup> | -.015 |
| Mekanike   | -.077 <sup>*</sup> | .042  |

|          |       |       |
|----------|-------|-------|
| Lexim    | -.093 | .796* |
| Anglisht | .061  | .300* |

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

\*, Largest absolute correlation between each variable and any discriminant function

***Tab. 8: Koeficientët bruto të funksionit diskriminant***

**Canonical Discriminant  
Function Coefficients**

|            | Function |        |
|------------|----------|--------|
|            | 1        | 2      |
| Anglisht   | .041     | -.002  |
| Lexim      | -.004    | .132   |
| Mekanike   | .116     | -.050  |
| Hartim     | -.040    | -.108  |
| Matematike | -.090    | -.050  |
| Organizim  | .082     | -.021  |
| (Constant) | -4.186   | -1.060 |

Unstandardized coefficients

***Tab. 9: Funksionet Lineare diskriminante Fisher (Një funksion për cdo grup)***

**Classification Function Coefficients**

|            | Karriera |        |            |
|------------|----------|--------|------------|
|            | Biznes   | Mjeksi | Sociologji |
| Anglisht   | 1.175    | 1.119  | 1.135      |
| Lexim      | -.317    | -.333  | -.282      |
| Mekanike   | 1.004    | .851   | .880       |
| Hartim     | -.325    | -.250  | -.310      |
| Matematike | -.420    | -.285  | -.343      |
| Organizim  | .278     | .167   | .193       |

|            |         |         |         |
|------------|---------|---------|---------|
| (Constant) | -50.117 | -45.649 | -47.209 |
|------------|---------|---------|---------|

Fisher's linear discriminant functions

**Tab. 10: Mesataret e funksioneve diskriminante 1 dhe 2 për secilin grup**

**Functions at Group Centroids**

|            | Function |       |
|------------|----------|-------|
|            | 1        | 2     |
| Karriera   |          |       |
| Biznes     | .472     | -.024 |
| Mjeksi     | -.928    | -.197 |
| Sociologji | -.508    | .204  |

Unstandardized      canonical  
discriminant          functions  
evaluated at group means

**Tab. 11: Probabilitetet paraprake për Grupet (Bazuar në numrin e rasteve në secilin grup sepse zgjedhëm metodën: Priors=Size)**

**Prior Probabilities for Groups**

|            | Prior | Cases Used in Analysis |          |
|------------|-------|------------------------|----------|
|            |       | Unweighted             | Weighted |
| Karriera   |       |                        |          |
| Biznes     | .592  | 100                    | 100.000  |
| Mjeksi     | .172  | 29                     | 29.000   |
| Sociologji | .237  | 40                     | 40.000   |
| Total      | 1.000 | 169                    | 169.000  |

**Tab. 12: Tabela e rasteve origjinale (aktuale) përkundrejt anëtaresimt të parashikuar**

**Classification Results<sup>a</sup>**

|                |            |  | Predicted Group Membership |        |            | Total |
|----------------|------------|--|----------------------------|--------|------------|-------|
|                |            |  | Biznes                     | Mjeksi | Sociologji |       |
| Original Count | Biznes     |  | 91                         | 6      | 3          | 100   |
|                | Mjeksi     |  | 16                         | 8      | 5          | 29    |
|                | Sociologji |  | 26                         | 5      | 9          | 40    |
| %              | Biznes     |  | 91.0                       | 6.0    | 3.0        | 100.0 |
|                | Mjeksi     |  | 55.2                       | 27.6   | 17.2       | 100.0 |
|                | Sociologji |  | 65.0                       | 12.5   | 22.5       | 100.0 |

a. 63.9% of original grouped cases correctly classified.



Tabelat 1 dhe 2 na ndihmojnë për të vlerësuar nëse grupimet sipas karrierës ndryshojnë sipas çdo variabli, duke e izoluar këtë variabël nga variablat e tjerë në një AD. Psh. Nëse statistikisht nuk ushtrohet kontroll mbi 5 variablat e tjerë, cila është natyra e diferencës në rezultatin e testit të anglishtes ndër 3 grupet në shqyrtim? Kur rezultate e anglishtes vlerësohen në izolim, nuk ka ndryshim statistikisht të rëndësishëm ndërmjet 3 grupeve të karrierës. :  $F(2,166)=0,237$ ,  $p=0.789$ .

Sikurse dhe u përmend më lart, testet e paraqitura në tabelën 2, kanë risk të madh të gabimit të llojit të I, aq më tepër që p-respektive nuk janë të korigjuara për këtë problem. Ndaj në këto raste do të ishte mirë të përdorej korrigjimi Bonferroni. Ku  $\alpha_c=0.05$  dhe meqenëse janë 6 teste F, atëherë  $\alpha = \alpha_c/q = 0.05/6 = 0,008$ . Kështu, në testet F, në këtë rast (tabela 2) do të konsiderohen statistikisht të rëndësishëm vetëm nëse  $p < 0.008$ . Sipas këtij kriteri vetëm dy nga variablat individuale (Matematika dhe Organizimi) çojnë në ndryshime statistikisht të rëndësishme ndërmjet 3 grupeve të karrierës me respektivisht :  $F(2,166)=7.586$ ,  $p=0.001$  dhe  $F(2,166)=19.018$ ,  $p=0.000$ . Wilks  $\Lambda$  për testin e matematikës është 0.916, nga ku mund të llogaritim  $\eta^2=1-\Lambda=1-0.916=0.084$ . Dmth rreth 8.4% e variancës në rezultatin e matematikës i atribuohet diferencës ndërmjet grupeve. Po ti kthehemi tabelës 1, ku gjendet mesatarja e Matematikës për të tre grupet, shohim se mesataren më të lartë e ka grupi i Mjeksisë ( $M_2=26.03$ ) ( $M_1=20.07$  dhe  $M_3=24.23$ ). Në mungesë të informacioneve të tjera, dhe pa kontrolluar ende statistikisht korelacionin ndërmjet matematikës dhe variablave të tjerë, në arrijmë në mendimin se një student me rezultate të ulta në matematikë mund të zgjedhë më shumë Biznes si karrierë sesa mjeksi apo soicologji.

Me të njëjtin aresyetim do të gjykonim edhe për testin e organizimit, i cili dhe ky nga tabela 2 rezultoi statistikisht i rëndësishëm. Ku paraprakisht mendojmë se vetëm duke përdorur testin ANOVA, dhe duke iu referuar tabelës 1 ( $M_1=28.26$ ,  $M_2=18.17$ ,  $M_3=20.07$ ) studentët me rezultate më të larta në organizim ka më shumë gjasa të jenë në grupin 1 (biznes)

Këto teste ANOVA, shpesh herë nuk janë pjesë e AD. Ato sigurojnë një bazë për krahasime, por janë koeficientët e funksioneve diskriminante që na tregojnë se sa kontribuon çdo variabël X në parashikimin e anëtarësisë në grupe, pasi kontrollohet statistikisht për korelacionet me variablat e tjerë X. Ndërkohë rezultatet e analizave një faktoriale (ANOVA) tregojnë se sa çdo variabël X kontribuon në anëtarësimin në grup, por kur nuk kemi kontroll ndaj variablave të tjerë.

Faza tjetër e analizës lidhet me faktin nëse natyra e diferencës ndërmjet grupeve ndryshon apo mbetet e njëjtë, dhe kjo bëhet nëpërmjet kontrollit të korelacionit ndërmjet variablave X (Tabela 3).

Tabela 3 paraqet korelacionet e mesatarizuara (pooled). Që do të thotë për çdo grup është ndërtuar tabela e korelacioneve për të 6 variablat, pastaj është krijuar tabela 3 e përgjithshme duke i interpoluar korelacionet përkatëse për të tre grupet e karrierës. Është shumë e ngjashme me tabelën e korelacioneve të regresionit. Qëllimi është të mos ketë korelacione të larta ndërmjet variablave X, sepse nëse ndodh e kundërta kjo do të thotë se kontributi individual i variablit X nuk mund të përcaktohet. Por nga ana tjetër nëse variablat x janë fare të pakoreluar me njëri tjetrin, AD nuk do na japë rezultate të ndryshme nga seti i ANOVA-ve 1 faktoriale. AD është më e përshtatshme kur ka korelacione të dobëta ndërmjet variablave. Në rastin konkret (tabela 3) asnjë korelacion në vlerë absolute nuk e kalon 0.7 dhe themi se nuk ka problem serioz në multikolinearitet.

Tabela 4 tregon rezultatet e testit Box M. Hipoteza nul që testohet lidhet me supozimin e homogjenitetit të shumës së matricave të produkteve të kryqëzuara ndërmjet grupeve.  $H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_p$ , dhe do të

preferonim që testi të rezultonte statistikisht i parëndësishëm ( $p > \alpha$ ,  $H_0$  të qëndrojë) Në të kundërtën rezultatet AD do të ishin të painterpretueshme. Në rastin tonë:  $B_0 M = 78.939$  dhe  $F(42, 23705.058) = 1.746$ ,  $p = 0.002$ . në rastin e  $B_0 X M$ ,  $\alpha = 0.001$ , pasi shkallët e lirisë në një test të tillë janë shumë të larta, siç shihet. Në këto kushte të  $p = 0.002 > 0.001$  themi se nuk ka dhunim statistikisht të rëndësishëm të supozimit në lidhje me homogjenitetin e variancave dhe mund të vazhdohet më tej me analiza. (Dilema ekziston dhe këtu : sa më i madh numri i rasteve aq më pak problematike është çështje e dhunimit të supozimit për homogjenitetin e variancave. Megjithatë, sa më i madh numri i elementëve, aq më shumë fuqi statistikore kemi për të gjykuar në lidhje me supozimin, ndaj duhet  $\alpha$  sa më e vogël, ndaj në testin Box M ne kemi  $\alpha = 0.001$ )

Nëse do kishim dhunim të supozimit të homogjenitetit (rezultat i testit Box M) atëherë studiuesi duhet të konsiderojë disa opsione: **1.** Aplikimi i transformimeve log për variablat që nuk kanë shpërndarje normale , mund të ishte një mënyrë efektive për të zvogëluar diferencat në variancat e X midis grupeve dhe impaktin e vlerave ekstreme. **2.** Një tjetër alternativë do të ishte zgjedhja e një analize statistikore me më pak kufizime në lidhje me çiftet e të dhënave ( si regresioni logjistik binar) do të ishte i preferuar në këte rast. Në rastin tonë nuk ka një dhunim të tillë dhe ne do vazhdojmë me AD.

Qëllimi është të vlerësohet nëse zgjedhja e karrierës është e parashikueshme nga këta 6 variabla, që janë marrë në shqyrtim. AD e parë realizohet duke marrë në konsideratë të 6 variablat e përfshirë në model në një hap të vetëm. Rezultatet shfaqen në tabelën 5. Në këtë tabelë me  $k=3$  (numri i grupeve) dhe  $p=6$  (numri i variablave x) ka dy funksione diskriminante (min. i  $(k-1)$  dhe  $p$ ). Wilks  $\Lambda$  tregon testin e rëndësisë për setin e funksioneve diskriminante. Rreshti i parë ( 1 through 2) teston nëse anëtarësia në grup mund të parashikohet duke përdorur  $D_1$  dhe  $D_2$  të kombinuar- kjo do të thotë se këtu testohet rëndësia statistikore e gjithë modelit , duke përdorur të gjithë funksionet diskriminate dhe të gjithë variablat X. Për këtë test Omnibus,  $\Lambda = 0.730$  (keshtu  $\eta^2 = 1 - \Lambda = 1 - 0.730 = 0.27$ ) . që do të thotë se rreth 27% e variancës në dy funksionet e para diskriminante (duke i marrë të dy së bashku) i atribuohet diferencës ndërmjet grupeve. Kjo konvertohet në  $\chi^2(12) = 51.484$ ,  $p = 0.000 < 0.001$ , ndaj modeli i përgjithshëm , duke përfshirë të dy funksionet diskriminante 1 dhe 2 , parashikon në mënyrë statistikisht të rëndësishme anëtarësimin në grupe.

Për t'iu përgjigjur pyetjes nëse do të na duhet të përfshijmë të gjithë funksionet diskriminante në interpretimin tonë, apo vetëm disa prej tyre, le të shohim në testet në vijim. Në rastin tonë, ka vetëm një mundësi që duhet marrë në konsideratë: në tabelën 5, në rreshtin e dytë , jepet testi i rëndësisë vetëm për funksionin 2. Për  $D_2$ ,  $\Lambda = 0.983$ ,  $\chi^2 = 2.783$ ,  $p = 0.733$ . kështu funksioni  $D_2$  vetë nuk është statistikisht i rëndësishëm në parashikimin e anëtarësisë në grupe. Ne duhet ta limitojmë interpretimin dhe diskutimin e diferencave të grupeve , atyre ndryshimeve që përfshijnë  $D_1$ . Kujdes, nuk ka një test direkt që të vlerësojë rëndësinë statistikore të vetëm  $D_1$  në një analize AD.

Informacion shtesë në lidhje me vlefshmërinë e parashikimeve të  $D_1$  dhe  $D_2$  jepet në tabelën `eigenvalues` të (tabela 5) ku për  $D_1$ , vleraeigen  $\lambda_1 = 0.347$  (kjo i korespondon  $SS_{\text{between}}/SS_{\text{within}}$  te ANOVA kur krahasohen grupet për rezultatet e  $D_1$ ). Vleraeigen mund të konvertohet në korelacionin kanonikal:

$r_c$  thamë, është analog me korelacionin  $r_c = \frac{0.347}{\sqrt{0.347 + 0.017}} = 0.52$  ndërmjet anëtarësimt në grupe dhe rezultateve për variablin sasior tek ANOVA. Nëse  $r_c$  ngrihet në kampion, atëherë sigurohet informacion i ngjashëm me  $\eta^2$ , që është proporcioni i variancës në rezultate e  $D_1$  i cili i atribuohet diferencës (ndryshimit) ndërmjet grupeve.

Është e qartë që në këtë rast  $D_1$  është i lidhur fort me anëtarësimin në grupe (Y) ( $r_{c1}=0.52$ ) ndërsa  $D_2$  është shumë i dobët (statistikisht i parëndësishëm) në lidhje me anëtarësimin në grupe ( $r_{c2}= 0.04$ ,

Varianca totale që mund të parashikohet nga kombinimi i  $D_1$  dhe  $D_2$  realizohet duke mbledhur vleraveigen ( $0.347+0.017=0.364$ ) Mënyra se si kjo variancë e parashikuar duhet të shpërndahet ndërmjet  $D_1$  dhe  $D_2$  vlerësohet duke pjestuar çdo vlerëigen me këtë shumë të vleraveigen, që do të thotë me variancën që mund të parashikohet nga  $D_1$  dhe  $D_2$  të kombinuar ( $0.364$ ) pra  $0.347/0.364=0.953$  ose 95.3% e kësaj variance është për shkak të  $D_1$ . Është e qartë në këtë rast se,  $D_1$  parashikon në mënyrë statistikisht të rëndësishëm anëtarësinë,  $D_2$  siguron pak informacion , dhe në këtë mënyrë  $D_2$  nuk duhet ti kushtohet shumë rëndësi në interpretim.

Në rezultate që do japim në vijim, janë të përfshira vetëm ato të  $D_1$ . Për të parë se çfarë informacioni përfshihet në  $D_1$ , shohim tabelën 6. Vlerat në këtë tabelë i korespondojnë koeficientëve d në ekuacionin  $D_1 = d_{11} z_1 + d_{12} z_2 + \dots d_{1p} z_p$  ku  $z_1, z_2, \dots z_p$  janë versione të standartizuara të variablave  $x_1, x_2, x_p$ , që kemi përmendur. në fillim . Kolona 1 jep koeficientët që sigurojnë kombinimin linear optimal të z-ve që përdorën për të ndërtuar rezultatet e  $D_1$ , Kolona 2 jep të njëjtin informacion për  $D_2$ . Nuk ka teste statistikore për këto koeficientë. Studiuesi, në këto raste vendos një vlerë kufi (arbitrate), ku: variabli konsiderohet i dobishëm nëse koeficienti i standartizuar i funksionit diskriminant të tij e tejkalon këtë vlerë arbitrare (në vlerë absolute). Vlera arbitrare në këte shembull është 0.5. Kështu në rastin tonë, vetëm dy variablave X i është dhënë peshë e rëndësishme në llogaritjen e  $D_1$ : Matematika ( $d_{1matem}= -0.746$ ) dhe Organizimi ( $d_{1organizim}= +0.75$ )

Zakonisht përcaktohet një vlerë më e ulët arbitrare si sph: 0.4 ose 0.3, por nëse do vepronim kështu në këtë rast, atëherë i bie të konsiderohej Anglishtja dhe Mekanikja si të rëndësishëm , kur ndërkohë në ANOVA- 1 faktoriale ato rezultuan statistikisht të parëndësishëm.

Në tabelën 7 jepet një tjetër informacion në lidhje me marrëdhënien ndërmjet variablave x dhe secilit koeficient të funksionit diskriminant, kjo quhet `matrica e strukturës` Vlerat në tabelë tregojnë në lidhje me korelacionin midis variablave X dhe funksioneve diskriminate. Psh. Për  $D_1$ , 2 variablat që kanë korelacione më të mëdha me këtë funksion diskriminant janë Organizimi dhe Matematika. Korelacioni ndërmjet rezultateve të Organizimit dhe rezultateve të  $D_1$  janë +0.812 dhe -0.513 për Matematikën. Këto korelacione nuk kanë teste formale të rëndësisë statistikore por mund të gjykojmë për rëndësinë duke iu referuar vlerës arbitrare 0.5. Dy variabla si më lart kanë rezultate më të > 0.5, pra janë të rëndësishëm.

Tabela 8 i referohet koeficientëve që janë përdorur për të ndërtuar funksionin diskriminant nisur nga rezultatet aktuale të variablave x. Këto koeficientë janë të ngjashëm me koeficientët bazë ( $b_i$ ) Në analizën regressive. Keto koeficientë nuk interpretohen si evidenca për fortësinë dhe aftësinë parashikuese të çdo variabli x; ato mund të përdoren për të gjeneruar rezultatet e funksionit diskriminant, të çilat mund të përdorën për të parashikuar anëtarësimin në grupe mbështetur në rezultatet bazë.

Tabela 9 paraqet koeficientët e klasifikimit të Fisherit. Përdoren kryesisht në ato studime, ku qëllimi kryesor është parashikimi i grupeve. Kujdes se këto ndryshojnë nga funksionet diskriminante të trajtuara më parë, këtu kemi funksion klasifikimi fisher për secilin nga k grupet. Në rastet kur numri i grupeve është >2, koeficientët e klasifikimit fisher ofrojnë një mënyrë me konveniente për parashikimin e anëtarësisë në grup (dhe gjenerojnë vlerësime të probabiliteteve të anëtarësimit në

grupe, nisur nga vlerat e variablave  $x$ ). Zakonisht koeficientët  $f$  nuk përdoren për të interpretuar natyrën e ndryshimeve në grupe.

Tabela 11 raporton probabilitetet paraprake të klasifikimit në secilin nga 3 grupet. Në këtë rast u zgjedh si metodë e numrit të grupeve ( $\text{prior}=\text{size}$ ) ndaj, probabilitetet paraprake të klasifikimit në secilin nga 3 grupet i korespondojnë përpjesës së rasteve për çdo grup.

Deri tani kemi 3 informacione në lidhje me çështje që po trajtojmë: Fisherat nga tabelat Anova 1 faktoriale; koeficientët e funksioneve diskriminante kanonikale të standartizuar; dhe koeficientët e strukturave. Në këtë rast nuk ka rëndësi kujt informacioni i referohemi, pasi në të tre rastet arrijmë në të njëjtin konkluzion në lidhje me rezultatin e zgjedhjes së karrierës nisur nga vlerat e variablave. Nga të 6 variablat, vetëm rezultatet e Matematikës dhe Organizimit janë të lidhura me parashikimin anëtarësimi në grupe, 4 variablat e tjerë nuk ndikojnë në parashikim. Ashtu sikurse koeficientët beta mund të ndryshojnë në shenjë nga korelacionet Pearson ndërmjet një  $x$  dhe  $Y$ , dhe ashtu sikurse mesataret e grupeve në një ANCOVA mund të ndryshojnë rangun kur ne korrigjojmë për kovariancën, ashtu edhe mund të gjemë në një AD, që natyra e ndryshimit të grupeve për një variabël mund të duket e ndryshme kur nuk kemi nën kontroll 5 variablat e tjerë (ANOVA) nga rasti kur i kemi nën kontroll ato (rasti i koeficientëve të standartizuar të funksionit diskriminant dhe koeficientët e strukturës). Në situata të tilla, diskutimi i rezultateve duhet të përfshijë një diskutim sesi këto dy sete rezultatesh ndryshojnë ndërmjet tyre.

Deri tani dimë që një rezultat i lartë në  $D_1$  do të thotë rezultat i lartë në Organizim dhe i ulët në Matematikë. Sa rezultate në  $D_1$  lidhen me anëtarësimin në grupe? Kjo mund të vlerësohet duke parë në vlerat e funksioneve e te qendrat e grupeve (tabela 10) dhe nga harta territoriale (shiko në SPSS). Një centroid është rezultati mesatar në secilin nga funksionet diskriminante për anëtarët e grupit 1. Për funksionin e parë diskriminant, grupi i Biznesit ka një mesatare relativisht të lartë të  $D_1$  (0.472) ndërsa dy grupet e tjera kanë mesatare të ulta (-0.928 për mjeksinë dhe -0.508 për Sociologjinë). Kështu grupi i biznesit ka rezultate më të larta në  $D_1$ , anëtarët e këtij grupi arrijnë rezultate më të larta në Organizim dhe më të ulta në Matematikë krahasuar me anëtarët e dy grupeve të tjera. Vini re se, mesataret (ose centroidet) për  $D_2$  janë shumë më pranë njera-tjetrës sesa mesataret e  $D_1$  (centroidet në  $D_2$  ishin -0.024; -0.197; 0.204). Kjo është konsistente me informacionin e mëparshëm (që lidhej me përmasën e vogël të  $r_{c2}$ ) që grupet nuk ishin shumë të ndryshëm në  $D_2$ .

Harta territoriale, ilustron grafikisht vendimin ose rregullin e klasifikimit, që përdoret për të parashikuar anëtarësimin në grup për individët, bazuar në rezultatet e tyre për funksionet diskriminante  $D_1$  dhe  $D_2$ . Rezultatet e  $D_1$  paraqiten te boshti  $X$ , kurse rezultatet e  $D_2$  te boshti  $Y$ . Strings 1s i korespondojnë kufijve të rajonit për të cilin ne klasifikojmë njerëzit si anëtarë të mundshëm të grupit 1, e kështu me radhë edhe për grupet e tjera. Kështu, çdo kush me rezultate pozitive në  $D_1$  është klasifikuar si anëtarë i mundshëm i grupit 1, Biznes. Grupet e tjera diferencohen (jo shumë mirë) nga rezultatet e  $D_2$ , rezultate më të larta në  $D_2$  na sugjerojnë të klasifikojmë njerëzit si anëtarë të grupit 3; rezultate më të ulta në  $D_2$  na ndihmon të klasifikojmë njerëzit si anëtarë të grupit 2. Megjithatë siç edhe duket nga tabela e anëtarësisë aktuale dhe asaj të parashikuar, në tabelën e fundit të AD, anëtarët e këtyre dy grupeve nuk mund të diferencohen në mënyrë të besueshme nga kjo analizë.

Bashkë me informacionin e mëparshëm, harta territoriale, ndihmon të interpretojmë kuptimin e  $D_1$ . Rezultate të larta në  $D_1$  i korespondojnë rezultateve të larta në Organizim dhe rezultateve të ulta në Matematikë, dhe rezultate të larta në  $D_1$  parashikojnë anëtarësi në grupin 1. Centroidet në grupe

shfaqen si astetriks (me yje) në hartën territoriale, në SPSS, kurse në printim ato shfaqen si M1, M2 dhe M3. Nëse AD bën një punë të mirë në dallimin ndërmjet grupeve këto centoride duhet të jenë larg njëri tjetrit, dhe çdo centroid duhet të jetë brenda rajonit për të cilin njerëzit klasifikohen përkatësisht. Në këtë shembull centoridet e grupit 1 janë larg centorideve të grupeve 2 e 3, kjo tregon që anëtarët e grupit 1 dallon nga 2 e 3. Megjithatë centroidet e grupeve 2 e 3 janë pranë njëri tjetri, gjë që konfirmon faktin që në këtë AD, këto dy grupe nuk janë dallueshëm të diferencuar. Ka raste që ndodh që asnjë element nuk klasifikohet në një grup të caktuar, dhe kur ndodh kështu nuk ka rajon në hartë që ti korespondojë grupeve.

Rezultatet në Tabelë 12, lidhen me konsistencën ndërmjet anëtarësimit bazë dhe atij të parashikuar nga analiza, për grupet. Nëse klasifikimi i bazuar në rezultatet e funksionit diskriminant është i saktë, atëherë shumica e rasteve do të bien në qelizat diagonale të kësaj tablele (duhet të jetë identik parashikimi me realitetin). Në fakt vetëm 63.9 % e elementëve janë klasifikuar korrekt. Anëtarët e grupit 1 janë klasifikuar më shumë korrekt (91 nga 100 ose 91%) . Megjithatë kjo analizë nuk ka bërë një punë të mirë për të identifikuar anëtarët e grupeve 2 dhe 3, faktikisht shumica e anëtarëve të këtyre grupeve janë keqklasifikuar. Kështu, megjithëse modeli i përgjithshëm është statistikisht i rëndësishëm (bazuar në chi- square për  $D_1$  dhe  $D_2$  e kombinuar) , anëtarësimi në grup nuk është parashikuar mirë për të gjithë anëtarët e grupeve 2 dhe 3. Vetëm 2 nga 6 variablat e trajtuar ishin të vlefshëm për diferencimin e grupeve, rezultatet e larta në Organizim dhe të ulta në Matematikë parashikojnë preferencë për Biznesin, por këto rezultate nuk sigurojnë informacion në lidhje me zgjedhjen ndërmjet mjeksisë apo sociologjisë.

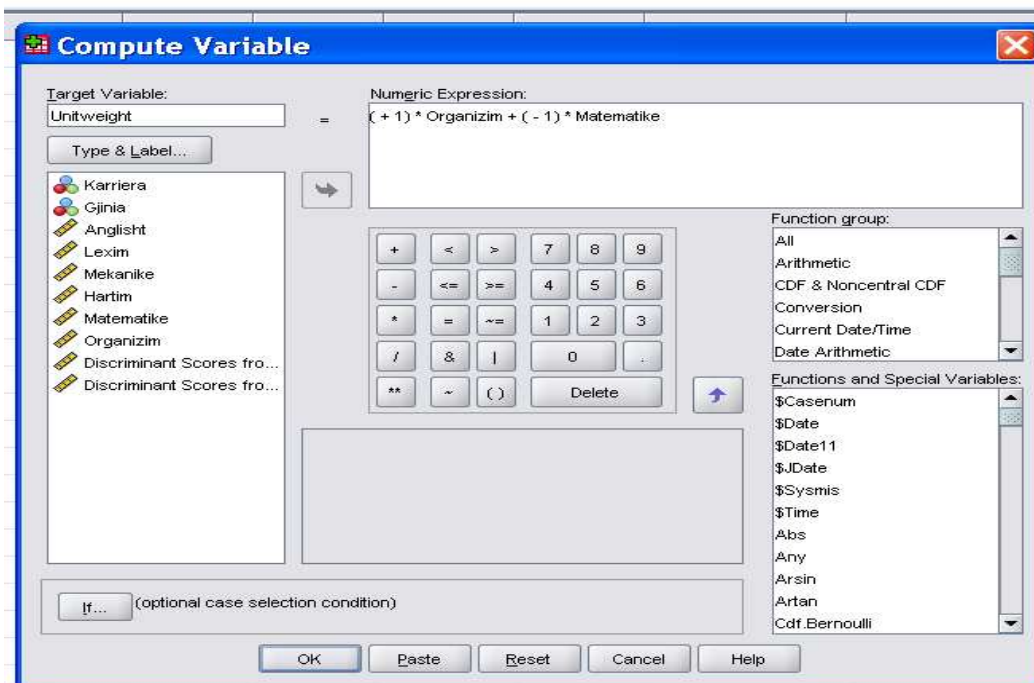
## **11. ANOVA NJË FAKTORIALE PËR FUNKSIONET DISKRIMINANTE**

Një hap final në përfundimin e analizës për të kuptuar më mirë AD, por që nuk shkruhet për të kur bën AD në një artikull shkencor lidhet me një tjetër komandë që jepet në SPSS, dhe që është `Save` sipas kësaj ruhen rezultatet e llogaritura të funksionit diskriminant për secilin person për  $D_1$  dhe  $D_2$  , si variabla të rinj në faqen e SPSS. Llogaritet kështu një variabël i ri për çdo subjekt në këtë database, që është rezultati i përbërë i interpoluar i dy variablave që sipas AD ishin variablat më të mirë në parashikimin e anëtarësisë në grupe:

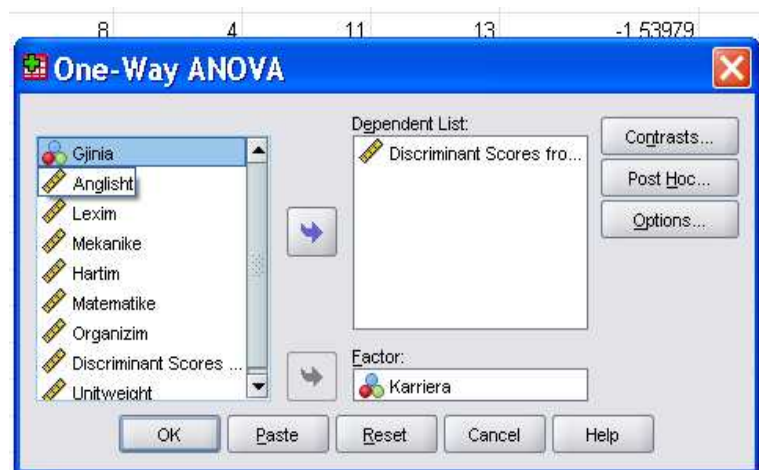
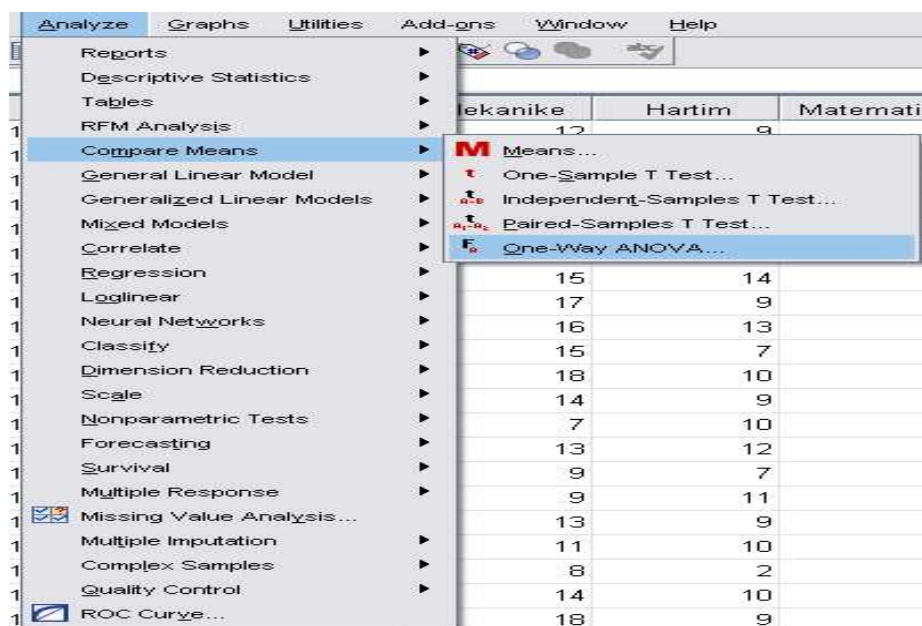
Përbërësi i njësisë së interpoluar=  $U = (+1 \times \text{Organizim}) + (-1 \times \text{Matematikë})$  (ek. 11)

U e llogaritur në këtë formë është diferenca ndërmjet rezultateve të çdo personi në Organizim dhe Matematikë. Në figurën më poshtë tregohet se si rezultatet e këtij variabli të ri (përbërësinjësiësë) ose U llogaritet në SPSS duke përdorur procedurën e llogaritjes së variablit.





Pas kësaj u realizua ANOVA për të vlerësuar sesi 3 grupet e karrierës ndrsyshojnë në mesatare për të 6 variablat: D<sub>1</sub>, rezultati i ruajtur nga funksioni diskriminant i AD së mëparshme, dhe U e llogaritur nga ek. 11. Rezultatet shfaqen në tabelën 13.



**Tab. 13: Tabela ANOVA- 1 faktoriale për të krahasuar mesataret e tre grupeve për AD (D<sub>1</sub>)**

## Descriptives

Discriminant Scores from Function 1 for Analysis 1

|        | N   | Mean      | Std. Deviation | Std. Error | 95% Confidence Interval for Mean |             | Minimum  | Maximum |
|--------|-----|-----------|----------------|------------|----------------------------------|-------------|----------|---------|
|        |     |           |                |            | Lower Bound                      | Upper Bound |          |         |
| Biznes | 100 | .4723253  | .92674142      | .09267414  | .2884397                         | .6562109    | -2.22806 | 2.04321 |
| Mjeksi | 29  | -.9276900 | 1.15194359     | .21391055  | -1.3658660                       | -.4895141   | -3.11468 | 1.10857 |

|            |     |          |            |            |           |           |          |         |
|------------|-----|----------|------------|------------|-----------|-----------|----------|---------|
| Sociologji | 40  | -        | 1.05997835 | .16759729  | -.8472355 | -.1692405 | -        | 2.42252 |
| Total      | 169 | .5082380 | .0000000   | 1.15366655 | .08874358 | -.1751963 | .1751963 | 2.42252 |
|            |     |          |            |            |           |           | 3.11468  |         |

## ANOVA

### Discriminant Scores from Function 1 for Analysis 1

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 57.599         | 2   | 28.800      | 28.800 | .000 |
| Within Groups  | 166.000        | 166 | 1.000       |        |      |
| Total          | 223.599        | 168 |             |        |      |

Për  $D_1$  ka rëndësi statistikore ndërmjet grupeve,  $F(2,166) = 28.8$ ,  $p < 0.001$ . Vlera koresponduese e eta square (proporcioni i variancës që i atribuohet diferencave midis grupeve =  $SS_{\text{between}} / SS_{\text{total}}$ ) është 0.2576 ose 25.76%. Duke iu referuar tabelës 5, Wilks  $\Lambda$  për të gjithë modelin, pra për setin e  $D_1$  dhe  $D_2$  të kombinuar është 0.73 (megjithatë funksioni  $D_2$  nuk dha kontribut të rëndësishëm në diferencën ndërmjet grupeve, ndaj shumica e variancës së shpjeguar i atribuohet  $D_1$  në këtë rast). Kemi parasysh që  $1-\Lambda$  i korespondon eta square, në rastin tonë  $1-\Lambda = 1-0.73 = 0.27$ . Kështu 27% e variancës së rezultateve për të dy funksionet së bashku nuk i atribuohet anëtarësimit në grupe. Në këtë shembull, përpjesa e variancës për shkak të diferencave ndërmjet grupeve në variablin  $D_1$  në Anova 1 faktoriale (0,2576) është pak më e vogël se  $1-\Lambda = 0.27$  për gjithë modelin AD. Këto vlera do kishin qenë më afër njëra tjetrës nëse funksioni i dytë diskriminant do të kishte një fuqi edhe më të vogël parashikimi. Eta square për ANOVA në  $D_1$  do të kishte qenë më e vogël se  $1-\Lambda$  nëse funksioni i dytë diskriminant do të kishte fuqi më të madhe parashikuese, ky është rasti sepse ANOVA, ndryshe nga AD vetëm mori në konsideratë rezultatet e  $D_1$  dhe jo të  $D_2$ .



**Tab. 14: Tabela ANOVA- 1 faktoriale për të krahasuar mesataret e tre grupeve për U**

### Descriptives

#### Unitweight

|            | N   | Mean    | Std. Deviation | Std. Error | 95% Confidence Interval for Mean |             | Minimum | Maximum |
|------------|-----|---------|----------------|------------|----------------------------------|-------------|---------|---------|
|            |     |         |                |            | Lower Bound                      | Upper Bound |         |         |
| Biznes     | 100 | 8.1900  | 12.62809       | 1.26281    | 5.6843                           | 10.6957     | -31.00  | 32.00   |
| Mjeksi     | 29  | -7.8621 | 15.36397       | 2.85302    | -13.7062                         | -2.0179     | -34.00  | 17.00   |
| Sociologji | 40  | -3.5250 | 13.88042       | 2.19469    | -7.9642                          | .9142       | -38.00  | 29.00   |
| Total      | 169 | 2.6627  | 14.98209       | 1.15247    | .3875                            | 4.9379      | -38.00  | 32.00   |

### ANOVA

#### Unitweight

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 7798.962       | 2   | 3899.481    | 21.641 | .000 |
| Within Groups  | 29910.813      | 166 | 180.186     |        |      |
| Total          | 37709.775      | 168 |             |        |      |

Le ti referohemi tashmë tabelës 14. Për këtë ANOVA, tek seksioni i U,  $\eta^2=0.2068$ . Rreth 20.68 % e variancës në Përbërësi i njësisë së interpoluar për vetëm dy rezultate në Matematikë dhe Organizim, i atribuohet anëtarësisë në 3 grupet. Kjo ishte më pak se 27% ( $1-\Lambda=0.27$ ) i variancës së  $D_1$  dhe  $D_2$  që i atribuohet anëtarësisë në grupe në AD. Kështu, tregohet por nuk provohet që një kompozim linear i peshave (të derivuar nga AD) sigurojnë një parashikim më të mirë të anëtarësisë në grupe sesa

Përbërësi i njësisë së interpoluar për dy variablat më të mirë X. (qëlloi rasti me këto të dhëna por jo gjithmonë mund të jetë kështu)

Më fjalë të tjera, AD siguron kombinimin më të mirë linear të rezultateve për variablat X. Për këto të dhëna, rezultatet e  $D_1$ , që kanë diferencat më të mëdha të mundshme ndërmjet grupeve (dhe minimumin e mundshëm të variancës brenda grupeve) i korespondojnë:

$$D_1 = +0.41Z_{\text{anglisht}} - 0.0411Z_{\text{lexim}} + 0.44Z_{\text{mekanikë}} - 0.111Z_{\text{hartim}} - 0.746Z_{\text{matematikë}} + 0.75Z_{\text{organizim}}$$

(këto koeficientë të standartizuar i referohen tabelës 6) këto koeficientë krijuan funksionin  $D_1$  që ka minimumin e mundshëm  $\Lambda$  dhe maksimumin e mundshëm të  $\eta^2$  dhe F.

Megjithatë ne arritëm të tregonim diferenca të rëndësishme (me rëndësi më të vogël së AD) ndërmjet grupeve duke ekzaminuar Përbërësin e njësisë së interpoluar, i cili u krijua duke i vendosur peshën +1 çdo variabli me koeficient funksioni diskriminant të lartë pozitiv dhe -1 çdo variabli me koeficient funksioni diskriminant të lartë negativ, dhe peshën 0, çdo variabli me koeficient funksioni diskriminant të vogël). Megjithëse diferencat ndërmjet grupeve për U ishin më të vogla së diferencat ndërmjet grupeve për funksionin diskriminant  $D_1$  (në këte rast), ende këtu (tek U) u vërtetua diferencë e rëndësishme ndërmjet grupeve.



**TEMA 15:**  
**ANALIZA DISKRIMINANTE**

**Ushtrime**

**Ushtrim 1:** Përshkruani dy situata kërkimi kur mund të përdoret Analiza Diskriminante: një situatë ku interesi primar të jetë klasifikimi i grupeve dhe një tjetër situatë ku interesi primar të jetë përshkrimi i natyrës së ndryshimeve mes grupeve.

**Ushtrim 2:** Çfarë supozimesh për të dhënat duhet të “kënaqë” analiza diskriminante? Çfarë tip “data screening” duhet të realizohet për të vlerësuar supozimet që nuk vërtetohen?

**Ushtrimi 3:** Nëse “p” është numri I variablave parashikues dhe “k” është numri I grupeve për variablin rezultat në një AD, sa funksione të ndryshme diskriminante mund të sigurohen për të bërë diferencimet mes grupeve?

**Tema 15**  
**ANALIZA DISKRIMINANTE**

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

1

**Çështjet Kryesore që Trajtohen:**

**A. Kur mund të përdoret Analiza Diskriminante ( AD)**

**B. Supozimet e AD**

**C. Çështjet e kërkimit për një AD**

**D. Rasti empirik – ekuacionet e AD**

**E. Përmasa e efektit**

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

2

**A. Kur mund të përdoret Analiza Diskriminante**

1. Për të bërë parashikime në lidhje me rezultate të anëtarësisë në grupe të caktuara. Psh. AD mund të përdoret për të parashikuar nëse një pacient ka më shumë gjasa të jete anëtar i Grupit 1 (pacientë që kanë atak zemre) ose i Grupit 2 (pacientë, dhembja e gjoksit të të cilëve nuk ka të bëjë me atak zemre) bazuar në rezultatet e secilit individ në lidhje me variablat që vlerësohen në fazën e intervistës së pacientëve dhe në rezultatet laboratorike, të tilla si: moshë, gjinia, tensioni i lartë, tensioni i ulët, kolesterolli, trigliceridet, duhanpirja, intensiteti i dhimbjeve të gjoksit, dhe tensioni i gjakut.

2. Për të përshkruar natyrën e ndryshimeve ndërmjet grupeve. Nëse ne mund të identifikojmë një grup pacientësh që kanë patur atak zemre (Grupi 1) dhe i krahasojmë ata me një grup tjetër pacientësh që nuk kanë patur atak zemre (Grupi 2) , mund të përcaktojmë se cilat çifte rezultatesh të tilla si: Moshë, tensioni i lartë, tensioni i ulët, kolesteroli, trigliceridet, duhanpirja, ankthi dhe presioni i gjakut përshkruajnë ndryshimet ndërmjet këtyre dy grupeve të pacientëve.

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

3

## B. Supozimet e AD

1. Zgjedhje e rastit dhe vrojtimit e pavarura
2. Çdo variabël i pavarur ka shpërndarje normale
3. Çdo klasifikim i elementëve në kategoritë e variablit të varur është kryer në mënyrë korrekte
4. Çdo element përfshihet në një dhe vetëm një grupim, ku supozohet se Y është i ndarë në të paktën dy grupe.
5. Çdo kategori është e mirëpërcaktuar, qartësisht e dallueshme nga kategoria tjetër dhe natyrale . Përdorimi i medianës për të krijuar kategori nuk është forme natyrale grupimi . Variabli Y është mirë të jetë cilësor , por kur është sasior dhe duhet të krijohen kategori, atëherë duhet të ketë kuptim natyral krijimi i grupeve: psh. Tre grupe të cilat i referohen tre niveleve të huamarrjes për shtëpi.
6. Kategoritë duhet të përcaktohen para se të mbliiden të dhënat.
7. Kriteret mbi të cilat bëhet ndarja e grupeve duhet të jetë qartësisht diskriminante ndërmjet grupeve , në mënyrë që të mos ketë elementë që të mos i përkasin asnjë grupi dhe që të mos ketë mbivendosje (ose të paktën kjo të jetë minimale)
8. Numri i elementeve për çdo grupi, duhet të jetë shumë i ndryshëm nga njëri tjetri dhe të paktën 5 herë më i madh se numri i variablove të pavarur në model.

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

4

## C. Çështjet e kërimit për një AD

1. *Së pari, ekziston një test për rëndësinë e gjithë modelit:*
2. *Çështja e dytë kërimore ne AD është: Sa funksione diskriminante janë të nevojshme për diferencimin ndërmjet grupeve.*
3. *Një tjetër çështje që shtrihet ka të bëjë me vlerësimin e kontributit individual të variablove të pavarur në model . Cilët variabla janë të nevojshëm në parashikimin e anëtarësimit në grupe?*
4. *Cilat grupe mund (ose jo) të dikriminohen?*
5. *Çfarë interpretimi kualitativ (në se ka një të tillë) mund ti japim cdo funksioni diskriminant?*
6. *Cilat janë tipet e gabimeve të mundshme të klasifikimit, të cilat bëjnë sens?*

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

5

## D . Rasti empirik – Ekuacionet e AD

$$SCP_{total} = SCP_{between} + SCP_{within} \quad SCP_{within} = (SCP_1 + SCP_2 + \dots SCP_k) / (n_1 + n_2 + \dots n_k - k)$$

$\Lambda = SS_{within} / SCP_{total}$  (Ek. 4)  
 $\Lambda$  është ekuivalentja e  $(1 - \eta^2)$  të ANOVA ku  $\eta^2 = SS_{between} / SS_{total}$ . Statistikisht F te ANOVA ka dy shkallë lirie (within, between) . Statistika Wilks  $\Lambda$  ka tre shkallë lirie ne AD: n= numri total i elementëve, k= numri i grupeve, p= numri i variablove të pavarur

Meqenëse është e vështirë të përdoren tabela statistikore 3 dimensionale atëherë  $\Lambda$  konvertohet në  $\chi^2$ . SPSS e konverton  $\Lambda$  me  $\chi^2$  me shkallë lirie p (k-1) ku  
 $\chi^2 = -[n \cdot ((p+k)/2) - 1] \ln \Lambda$  (ek. 5)  
Së fundmi mund të përcaktojmë efektin e përmasës për analizën AD:  
 $\eta^2 = 1 - \Lambda$  (ek.6)  
Gjithë sa u tha më lart, i referohet testit omnibus të AD. Nëse Wilks  $\Lambda$  është e vogël, dhe nëse F e omnibus korresponduese është statistikisht e rëndësishme, mund të konkludojmë që anëtarësimi në grup mund të parashikohet në nivel më të mirë statistikor sesa kjo mund të bëhet duke konsideruar të gjithë funksionet diskriminante si set.

Përgatiti: Prof. Dr. BESA SHAHINI Tiranë 2018

6

## E . Përmasa e Efektit

Një pyetje që lidhet me AD është: Sa mirë mund të bëhet parashikimi i anëtarësisë nëpër grupe, duke përdorur të gjithë variablat dhe të gjithë funksionet diskriminante? Përgjigja lidhet me Wilks'  $\Lambda$  për të gjetur përmasën korresponduese të efektit  $\eta^2$  ku  $\eta^2 = 1 - \Lambda$  (ek. 9)

Për çdo funksion diskriminant, mund të llogaritet indeksi i përmasës së efektit (korelacionin kanonikal  $r_j$ ) nga vlera eigen  $\lambda_j$ , ku

$$R_{cj} = \sqrt{\frac{\lambda_j}{1 + \lambda_j}} \quad (\text{ek.10})$$

Nuk ka indeks për përmasën e efektit për çdo variabël individual  $x$  që të mund të interpretohet si porcion i variancës. Mund të vlerësohet rëndësia relative e variablove  $x$  duke iu referuar koeficientëve të strukturës, koeficientëve të standartizuar të funksionit diskriminant dhe vlerave  $F$  (nga ANOVA për secilin variabël  $X$ ). Megjithatë këto krahasime janë cilësore dhe deklarata për të gjykuar se cili  $X$  jep kontribut të fortë në parashikim është arbitrare.



*Ky publikim dhe programi i Masterit në Menaxhim Shëndetësor  
janë mbështetur nga Agjencia Zvicerane për Zhvillim dhe Bashkëpunim (SDC)  
nëpërmjet projektit të saj Shëndet për të Gjithë (HAP)*



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

**Agjencia Zvicerane për Zhvillim  
dhe Bashkëpunim SDC**

