GIZA ZIENTZIEI EGOKITUTAKO ESTATISTIKA

SPSS/PC+ programa-paketea erabiliz

JUAN ETXEBERRIA

udako euskal unibertsitatea

GIZA ZIENTZIEI EGOKITUTAKO ESTATISTIKA

SPSS/PC+ PROGRAMA-PAKETEA ERABILIZ

Juan Etxeberria

HEZKUNTZA, UNIBERTSITATE ETA IKERKETA SAILAK ONETSIA
(1.989-X)

© U.E.U.ko Matematika Saila
General Concha 25, 6. 48010 BILBO

I.S.B.N.: 84-86967-12-0
Lege-gordailua: BI-2167-89

Inprimategia: BOAN S.A.
Particular de Costa 12, 48010 BILBO
Azalaren diseinua: Julio Pardo

AURKIBIDEA

AURKIBIDEA ..... *V*

HITZAURREA ..... *XI*

1.- GIZA ZIENTZIAK ETA ESTATISTIKA *1*
(IÑAKI DENDALUZE)

1.- Istorio txiki bat *1*
2.- Giza-zientzietan Estatistika ematearen arrazoiak *2*
3.- Estatistikaren zenbait erabilpen Giza zientzietan *3*
3.1.- Ikerketa *3*
3.2.- Neurketa *6*
3.3.- Diagnostikoa *8*
3.4.- Ebaluazioa *9*
4.- Estatistika egoki erabiltzeko bete beharreko baldintzak *9*
5.- Heziketa estatistikoa *11*
6.- Informatika eta Estatistika *14*
7.- Ondorioak *16*
8.- Bibliografia *16*

2.- ESTATISTIKA APLIKATUA. KONTZEPTU OROKORRAK. MAIZTASUNEN BANAKETA *19*

1.- Aldagai kualitatibo edo nominalak *20*
2.- Aldagai kuantitatiboak *20*
3.- Banaketak *21*

3.- ESTATISTIKA DESKRIBATZAILEA *25*

1.- Zentralizazio-neurriak *25*
2.- Posizio-indizeak: koantilak *30*
3.- Sakabanatze-neurriak *33*
4.- Asimetri eta kurtosi-neurriak *39*
5.- Banaketa bidimentsionalak *42*

4.- KOERLAZIOA ETA ERREGRESIOA *45*

1.- Koerlazio-koefizientearen kalkulua *47*
1.1.- Aldagaiak kuantitatiboak dira *47*
1.2.- Aldagai bat kuantitatiboa eta bestea dikotomikoa edo dikotomizatua direnean: koerlazio-koefiziente biserial *48*
1.3.- Aldagai biak dikotomikoak dira: Φ koerlazio-koefizientea edo koerlazio--koefiziente laukoitza *50*
1.4.- Aldagai biak ordinalak dira: "r" Spearman-en koerlazio-koefizientea *51*
2.- Erregresio lineala *54*

5.- PROBABILITATEA *61*

1.- Probabilitatearen definizioak *62*
2.- Probabilitatearen propietateak *63*
3.- Bayes-en teorema *67*

6.- ALDAGAI ALEATORIOAK. PROBABILITATE ETA BANAKETA DISKRETUAK *69*

1.- Itxaropen matematikoa eta aldagai aleatorio baten bariantza *70*
2.- Banaketa diskretuak *74*
2.1.- Bernouilli-ren banaketa *74*
2.2.- Banaketa binomiala *75*
2.3.- Pascal-en banaketa *84*
2.4.- Poisson-en banaketa *85*

7.- BANAKETA JARRAIAK *89*

1.- Banaketa normala *89*
1.1.- Aldagaien tipifikazioa *91*
1.2.- Taulen erabilpenerako adibideak *92*
1.3.- Normalaren bidez, banaketa binomialen hurbilketa *96*
2.- Banaketa normalean oinarritutako banaketak *100*
2.1.- X^2-aren banaketa *100*
2.2.- Student-en T banaketa *103*
2.3.- Snedeckor-en F banaketa *106*

8.- NEURRIA GIZA ZIENTZIEN BARNEAN 111
(B. ORDEÑANA eta B. MUNARRIZ)

1.- Neurria. Sarrera 111
2.- Neurriaren mugak Giza zientzietan 113
3.- Estatistika eta psikometrien arteko erlazioa 114
4.- Tresna psikometrikoaren plangintza 116
5.- Laginaren aukeraketa 117
6.- Item-en analisiak 117
6.1.- Zailtasun-indizea 118
6.2.- Diskriminazio-indizea 119
6.3.- Baliozkotasun-indizea 120
6.4.- Erantzun ezuzenen (distraktoreen) analisia 120
7.- Testaren analisi teknikoa 121
7.1.- Testaren fidagarritasuna 121
7.1.1.- Fidagarritasun-koefizientea 122
7.2.- Testaren baliozkotasuna 125
8.- Testaren baremazioa 126
8.1.- Arau kronologikoak 126
8.2.- Gradu-arauak 126
8.3.- Pertzentilak 127
8.4.- Arau tipikoak 127
9.- Testaren eskuliburu teknikoa 127

9.- ESTATISTIKA ETA INFORMATIKA. SPSS/PC+ programa-paketea 129
(LUIS LIZASOAIN HERNANDEZ)

1.- SPSS/PC+ paketerako sarrera 130
2.- Idazkera-arau orokorrak 131
2.1.- Aginduen sintaxia 132
2.2.- Agindu era ezberdinak 132
2.2.1.- Eragiketa-aginduak 133
2.2.2.- Datuak definitu eta aldatzeko aginduak .. 133
2.2.3.- Prozedura-aginduak 134
2.3.- Prozedura era ezberdinak 134
2.3.1.- Lotez erabiltzeko prozedura 135
3.- Laguntasun-aginduak 136
3.1.- HELP agindua 136
3.1.1.-Laguntza orokorra 137
3.1.2.- Gai bati buruzko informazioa 137

3.1.3.- Agindu bati buruzko informazioa 138
3.1.4.-Azpiagindu baten informazioa 138
3.2.- DISPLAY agindua 138
4.- Sistemaren konfigurazio-aginduak: SET aginduak 139
5.- Datuen prestaketa 140
5.1.- Adibidearen aurkezpena 140
5.2.- Aldagaien izenak 142
5.3.- Balioen kodeketa 143
5.4.- Datu galdu edo ausenteen kodeketa 144
5.5.- Datu-matrizeen elaborazioa 145
6.- Datuak irakurtzeko aginduak 146
6.1.- Programaren izenburua: TITLE agindua 147
6.2.- Datu-matrizeen definizioa: DATA LIST agindua . 147
6.3.- Datu galdu edo ausenteen definizioa:
MISSING agindua 149
6.4.- Agindu eta balioen etiketak: VARRIABLE
LABELS eta VALUE LABELS aginduak 150
6.5.- Informazio honen grabaketa. SAVE agindua.
Grabatutako informazioa berrezkuratu.
GET agindua.
Grabatutako informazioa ikusi. LIST agindua ... 151
7.- Aldagaien birkodeketa. Aldagai berriak
sorterazteko erak 152
7.1.- Aldagaien balioen birkodeketa:
RECODE agindua 153
7.2.- Aldagai berrien sortzapena: COMPUTE agindua . 154
7.3.- Aldaketa baldintzatuak: IF agindua 155
8.- Datuak aukeratzeko aginduak 156
8.1.- Betirako aukeraketa 156
8.1.1.- Aukeraketa baldintzatua.
"SELECT IF" agindua 156
8.1.2.- Aukeraketa sekuentziala. "N" agindua .. 157
8.2.- Aukeraketa denborala: 157
8.2.1.- Aukeraketa baldintzatu denborala.
PROCESS IF agindua 158
8.2.2.- Aukeraketa aleatorioa.
SAMPLE agindua 158

10.- ESTATISTIKA DESKRIBATZAILEA eta SPSS/PC+ 159

1.- FREQUENCIES azpiprograma 159

2.- DESCRIPTIVES azpiprograma *168*
3.- MEANS azpiprograma *171*
4.- PLOT azpiprograma *175*
5.- CORRELATION azpiprograma *178*

11.- INFERENTZIA ESTATISTIKOA: HIPOTESI-TESTAK ETA KONFIDANTZA-TARTEAK *183*

1.- Emaitzen orokorpena: Konfidantza-tarteak *184*
2.- Hipotesi-testa: I. eta II. Motako erroreak *186*
3.- Bi edo talde gehiagotan lorturiko emaitzen konparazioa *188*
4.- Sujetu batek lorturiko balioekiko erabaki-hartzea *191*

12.- LAGIN-TEORIA *193*

1.- Laginketa aleatorioak *193*
2.- Laginketa ez-aleatorioak *196*
3.- Laginaren tamaina erabakitzea *197*

13.- JI KARRATUAREN TESTA. CROSSTABS AZPIPROGRAMA *205*

1.- CROSSTABS azpiprogramaren estatistikoak *209*
2.- CROSSTABS azpiprogramaren opzioak *210*

14.- TEST PARAMETRIKOAK. T-TEST AZPIPROGRAMA ... *215*

1.- Bi talde independenteren batezbestekoen konparazioa . *216*
2.- Erlazionaturiko bi talderen batezbestekoen konparazioa *221*

15.- BARIANTZ ANALISIA: ONEWAY ETA ANOVA AZPIPROGRAMAK *225*

1.- Bariantz analisien kalkuluak *228*
2.- ONEWAY azpiprograma *233*
3.- Bariantzaren analisi multifaktoriala: ANOVA azpiprograma *240*

16.- ERREGRESIO ANIZKOITZA: REGRESSION AZPIPROGRAMA *245*

1.- Erregresio anizkoitzen analisien sarrera 246
2.- Bariantz analisiarekin duen erlazioa 251
3.- Erregresio sinplea SPSS/PC+ erabiliz 252
4.- Erregresio anizkoitza 254
5.- Erregresioa aplikatzeko baldintzak 260
6.- Erregresio anizkoitza SPSS/PC+ paketea erabiliz 261

17.- TEST EZ-PARAMETRIKOAK. NPAR-TEST 273

1.- Lagin bati dagokion inferentzia 276
1.1.- KOLMOGOROV-SMIRNOVen froga. Aldagai ordinalak 276
1.2.- CHISQUARE froga. Aldagai nominalak 278
1.3.- RUNS froga. Eskala nominalak 279
1.4.- BINOMIAL froga. Aldagai dikotomikoak 280
2.- Bi laginetarako inferentziak 281
2.1.- Bi lagin askerentzako froga ez-parametrikoak ... 282
2.1.1.- MEDIANAREN froga. Eskala ordinalak . 282
2.1.2.- MANN-WHITNEYen U froga. Eskala ordinalak 284
2.1.3.- KOLMOGOROV-SMIRNOVen froga. Eskala ordinalak 287
2.2.- Erlazionatutako bi laginentzako froga ez-parametrikoak 289
2.2.1.- MCNEMAREN froga. Eskala nominala edo ordinala 290
2.2.2.- ZEINUEN froga. Eskala ordinalak 292
2.2.3.- WILCOXONen froga. Aldagai ordinalak 293
3.- K laginetarako inferentziak 295
3.1.- K lagin askerentzako froga ez-parametrikoak .. 296
3.1.1.- MEDIANAREN froga. Eskala ordinalak. 296
3.1.2.- KRUSKAL-WALLISen froga. Eskala ordinalak 298
3.2.- K lagin erlazionaturentzako froga ez-parametrikoak 299
3.2.1.- COCHRANen Q froga. Eskala nominala. 299
3.2.2.- FRIEDMANen froga. Eskala ordinala.. 301
3.2.3.- KENDALLen froga. Eskala ordinala 303
4.- Opzioak eta estatistikoak 304

HITZAURREA

Azken aldi honetan eguneroko hizkeran sartu zaigu "Estatistika" hitza. Denok hitzegiten dugu estatistikez eta sarritan erabili ere egiten ditugu. Hauteskundeei buruzko aurreinkestak ohizkoak dira, bizimoduaren igoeren kostua kalkulatzeko hurbilketak maiz argitaratzen dira. Asko baliatzen gara Estatistikaz, gehiegi batzutan, eta ez beti helburu garbiekin.

Estatistika gaizki erabiltzeak behin baino gehiagotan honelako esaldiak entzunerazten dizkigu:

"Gezur txikiak, gezur handiak eta Estatistikak".
"Batezbeste 40 zentrimetroko sakonera zuen erreka batean ito zen".

Estatistika ondo erabiltzea, ostera, oso egokia izan daiteke eta batzuetan, bazterrezinezko tresna gerta daiteke.

Letretako karreratzat hartuak izan diren ikasketa unibertsitario askoren plangintzei Estatistika ikasgaia erantsi zaie. Honi buruzko arrazoi batzuk IÑAKI DENDALUZE irakasleak moldatu duen lehenengo gaian aurki daitezke. Karrera horietako ikasleei dago zuzendua liburu hau, hain zuzen ere.

Hurrengo gaietan, 2.tik 7.era, Estatistikaren oinarri teorikoak topatuko dituzu.

Alabainan, ezinezkoa zaigu ondorio estatistiko onak lortzea nahiz inferentzia aproposak burutzea, hartarako datu onik ez baldin badugu. B. MUNARRIZ eta M.B. ORDEÑANA irakasleek idatzitako 8. gaian aurki dezake irakurleak neurketa egoki baterako beharrezko ezaugarriak.

Estatistikaren oinarri teoriko-matematikoak aldatu ez arren, gauza nabaria da ordenagailuak agertzean -Hardware zein Softwarea- Estatistikan lan egiteko moduak bestelako bihurtu direla. Programa-paketeek arras erraztu dute ordenagailuaren nahiz datu-fitxategi handien erabilera.

Programa-pakete hauen artetik SPSS/PC(+) izenekoa aukeratu eta aztertu da liburu honetan, oso zabaldurik bait dago gaurregun Unibertsitateko lanetan. Pakete honi buruzko jakingarri garrantzitsuenen laburpen zabala egin du L. LIZASOAIN irakasleak eta liburu honetako 9. gaian aurkituko duzu.

Liburuaren azken zatiak, inferentzia estatistikoaren oinarriak garatu ondoren, aztertutako paketearen zenbait azpiprograma dakar, aplikazio-baldintza, aginduak eta berauen emaitzen interpretazioak arakatuz.

Testu liburu hau moldatzen lagundu didaten E.H.U.ko irakasleei ene eskerrik beroenak eman nahi dizkiet, hala nola, berau ulergarriagoa izan dadin lankidetza eskaini didaten Zorroagako Pedagogiako Egokitze-kurtsoko ikasleei.

Desiotxo hau agertu nahi nuke azkenik, hau da, liburu hau geuk estatistika gehiago eta hobeto erabiltzeko izan dadila eta bestek erabilitakoa zuhurtziaz gureganatzen jakin dezagula.

1. GAIA
*GIZA ZIENTZIAK ETA ESTATISTIKA **
(IÑAKI DENDALUZE)

1.- ISTORIO TXIKI BAT.

Istorio txiki batekin emango diot hasera kapitulu honi; izan ere beronen eduki edo mamia ez bait da oso atsegina. Gero ikusiko dugunez, Giza Zientziatan lan egiten dugunoi, zenbakiek gauza askotarako baliagarriak zaizkigula ikusteko denbora izango da. Oraingoz, besteak beste, bidaia interesgarriak egiteko balio dezaketela aurreratzen dizuet. Ni, New York-era (arlo honetako oinarrizko prestakuntzara eta, ondoren, azken berrien bilakuntzara), Stockholm-era (ikasle espainiarrek beste herri batzuetakoen aldean elkarbizitzarako duten heziketari buruzko ikerketa egitera), Paris-era (iturri bibliografiko frantsesak eta anglosaxoniarrak alderatzera), Frankfurt-era (Nazioarteko Hezkuntz Ikerketarako Erakunde Alemanera), eta zenbaitetan interesgarriagoak diren sona ez hain handiko toki batzuetara eraman naute.

Giza Zientzietako zenbakiei esker, kongresu, datu-bilketa, espezialistei egindako kontsulta, espezialista gisa egindako bidaia eta abarretan ispilatzen den nolabaiteko "turismo intelektuala" dagoela esan daiteke.

Nire istorio txiki hau gertatu zeneko bidaiak, nire Doktorego-Tesirako datuak biltzea zuen helburu. Tesian, maila sozio ekonomiko eta industrializazio-maila desberdineko zenbait herrialdetako 14 eta 16 urteko pertsona amerikarrak eta espainarrak konparatu nituen. Puntu konkretu batean, beraiengan bateratasunaren hipotesia betetzen ote zen ikusi nahi nuen. Bateratasunaren hipotesiak, munduan faktore ezberdin ugari daudela esaten du,(kultura, historia, hizkuntza,.....) eta, batez ere, faktore berdina den teknologiak eragindako uniformetasun-faktore nabari bat dagoela esaten du lehenik. Hipotesi honek gero esaten duena zera da: uniformetasun- edo berdina den teknologi faktorea, ezberdintasun-faktoreei nagusitzen ari zaiela, eta honen ondorioz zenbat eta teknologia aurreratuagoa eduki, orduan eta antz handiagoa edukiko dutela herriek.

Hipotesi hau kontutan hartzeak badu abantailarik egungo lasterketa teknologikoan atzean gelditu diren herriek, aurreratuago daudenen eredua ezagutzen dutelarik, egoera ho-

---

(*) "Estatistika/Analisi estatistikoak" eta "Datu-analisia" gauza ezberdinak dira. Datu-analisiak ez dituela Estatistika klasikoak dituen mugak esaten du Pilibossianek (1988). Ezberdintasun hau beste maila orokor batetara eraman genezake, hau "Neurketa" eta "Datu-bilketa"ren artekoa izango litzateke.
Datu-bilketa guztiek zenbakitara (Neurketa) eramaten ez duten bezala, datu enpirikoen analisiak ez dira analisi estatistikoak.
Proposamen hau ez doa Pilibossianen edo eskola frantsesaren ildotik eta honek eramaten gaitu eztabaida argitzaile batetara.

rri aurre egiteko baliabideak eskuratu ezean urte batzuk barru gertatuko zaienaren berri izateko aukera bait dute.

Lanaren munduak eskaintzen diguna baloratzean, balorazio guztiak bezala erabat pertsonala izanik, bateratasunaren hipotesia betetzen ote zen frogatzea zen nire tesiaren helbururik nagusiena.

Andalusiako ekialdean datuak bildu behar nituen. Trenez edota hegazkinez egindako bidaia luzeetan normala da oso era guztietako pertsonekin harremanetan jartzea. Bidaia hartan suertea eduki nuen eta errusiar eta amerikar unibertsitatetan bere ikasketak eginak zituen gazte japoniar bat ezagutu nuen eta bi sistemetako ikasketa-ohituren aldeak azaldu zizkidan.

Agure rabbi bat ere ezagutu nuen, Europako ekialdeko herriren batean sortua, eta ea zertan nenbilen galdetu zidan. Nire bidaiaren helburuen azalpena sinplifikatu eta arazo psikologiko bat argitzeko datu bila joana nintzela erantzun nion. "Zenbakiak eta arimaren azterketa: nahasketa zaila horixe" erantzun zidan.

2.- GIZA ZIENTZIETAN ESTATISTIKA EMATEAREN ARRAZOIAK

Asko dira, "Letratako" karreretan zenbakiez baliatzea eztabaidatzen dutenak. Karrerak "Letratan" eta "Zientziatan" mailakatzeak ea zentzurik duen erabakitzea litzateke lehen pausoa. Zientziatakoak al dira Ekonomi eta Enpresa-ikasketak?. Edota Letratakoak al dira Psikologia, Pedagogia eta Soziologia?. Egia esan, karrera horiek ikasten dituzten batzuek letratako ikasketak eginak dituzte aurrez, eta zientziatakoak besteak. Karrera horietako ikasketa-plangintzetan, pisu handia dute zenbakiek. Zientzi arloko gaiek presentzia handia dute. Ikerketarako eta Estatistika-alorreko prestakuntzari dagokionez, Letratako ikasleek, Zientziatako askok baino maila hobea izaten dute askotan. Begibistakoa da, gainera, Giza Zientziak nabarmenki "matematizatu" direla. Beharbada, zientziaren legezkotasuna baieztatzeko. Adibidez, neurketa eta datuak lortu ondoren hauek aztertzeko erabiltzen diren metodoak asko matematizatu dira. Honek ez du esan nahi, ordea, zenbat eta azterketa estatistiko sakonagoa egon, orduan eta "zientziatasun" handiagoa edo txikiagoa egongo denik. Ez da ahaztu behar beti ere Giza Zientzietan zenbakiek dituzten arazo eta mugak. Legeen azalpenean ere antzeko zerbait gertatzen da, espezialista batzuk era matematikoan legeak ematen saiatzen direlarik.

Matematizatzean, Giza Zientziek, argitasun eta zehaztasuna irabazi dute. "Zientzia" zer den eztabaidatzen hasi gabe, gauza jakina da metodo zientifiko tradizionala nagusitu dela betidanik Giza Zientzietan (batez ere, metodo hipotetiko-deduktiboa, analitiko--induktiboa atzerago geratuz). Honek Giza Zientziak, beste zientziekin parekatu ahal izatea ekarri du, Zientzia batzuetako eta besteetako espezialisten arteko elkarrizketa eta eritzi-ematea ekarri duelarik.

Aspalditxotik argudio baliagarriak erabiliz zalantzan hau jarri da: Giza Zientzietako metodo zientifiko bakarra metodo esperimentala den; baita kasu guztietan garrantzitsuena edota erabilgarriena ote den ere. Egungo joera nagusia Giza Zientzietan metodo anizkoitza edo plurala onartzea da, denak "zientifiko" izenaz deituak direlarik. Aniztasun honen barnean, batzuek metodo enpirikoak eta ez-enpirikoak banatzearen aldekoak gara, hau da, errealitatearekin kontrastea bilatzen dugunok eta metodologia espekulatibo historikoak etab. erabiltzen dituztenak. Aipatu beharra dago, halaber, ikuspegi enpirikoak egiaztatze--prozesuari besteko garrantzia teorizatze-prozesuari ematen diola.

Estatistikak Giza Zientzietan zuen tratamendua, ikuspegi enpiristan sartzen da bete--betean. Batetik, errealitateko datuek diotena "irakurtzeko" eta bestetik, lehenago Giza Zientzietan hain usu eta erraz egiten diren hitzezko azalpenetara jo beharrean zenbakiz adieraztea du helburu.

Kuantitatibo-kualitatibo eztabaida ere egun-egungoa dugu Giza Zientzietan. Debate bizia eta afektibitatez betetakoa berau. Ez da zertan datzan azaltzeko une egokia, baina bai ordea jarrera konkretu bat hartu aurretik zein mailatan eztabaidatzen ari den jakitekoa. Maila epistemologiko-filosofikoa (egiaren erizpidea, justifikazioaren logika), metodologikoa (ikerketaren planteamendua eta ondorioak ateratzeko jarraitu beharreko bideak) eta teknikoa (datuak bildu eta aztertzea) dira eztabaidagai. Argi dago, bestalde, ohizko azterketa estatistikoak metodo kuantitatiboetatik hurbilago daudela kualitatiboetatik baino; batez ere maila teknikoari lotzen bagatzaizkio.

Giza Zientzietan Estatistikak betetzen duen lekuan antzeko eragina duen beste joera bat Ikertze-ekintza prozesuaren hazkundea da. Formalagoa eta, sarritan, formalistagoa den ohizko ikerketaren aurrean, partaidetza handiagoa eskatuz, lantokian egindako ikerketa nagusitzen ari da.

Giza Zientzietan interesaturik egon edota Estatistikako ikasle izan daitekeenarentzat, ordea, zera da interesgarriena: bide bat edo beste hartu (Metodo kualitatiboei edota kuantitatiboen lantzeari lehentasuna ematea), beti somatzen da azterketa estatistikoak ezagutu, azterketak egin eta ulertu beharra, nahiz eta horretarako ordenadoreak erabili edo azterketa horietan arituak direnengana jo behar. Era berean, Giza Zientziatan metodo enpirikoekin lan egiten dugunok, garbi eduki behar dugu ezin ditugula baztertu estatistikoak ez diren azterketak.

3.- ESTATISTIKAREN ZENBAIT ERABILPEN GIZA ZIENTZIETAN

3.1.- Ikerketa

Lehenengo eta oinarrizko erabilpena ikerketan egiten da; Ikerketa da Giza Zientzien barnean burutzen den ekintza interesgarrienetako bat. Askoz ere erakargarriagoa da beti

gure ezagupenak eta interesatzen zaizkigun arazoei buruzko informazioa ekartzea, beste batzuk esan dutena ikastea edo aplikatzea baino. Prestigioa ezezik, laguntza ekonomikoa ere ekartzen du inoiz, lan batzuk argitaratu eta kongresu eta Biltzar profesionaletan parte hartzeko aukera emateaz gain.

Aztertu nahi dugun gaia edo azterketaren helburuaren arabera, ikertzeko era asko dago. Honen ondorioz, ikerketa batzuk zabalagoak dira, zenbait etapatan burutzen direnak, beste batzuk hasiera-hasieratik egituratuta daudelarik.

Argi eduki beharra dago beti zertarako ikertzen dugun. Normalean, interesatzen zaigun errealitatea egokiro deskribatzeko (zer nolakoa da errealitatea? eta zer ezagutu nahi dugu hobeto?), (zeintzuk dira errealitate hori ezagutzeko aldagairik egokienak?) edo errealitate hau esplikatzeko (zeintzuk dira X_1,X_2 aldagaiak zeren araberakoa da Y?) (Zeinekin du zerikusia Y-k?), edo aurrikusteko (Y, X_1 eta X_2-ren araberakoa dela jakinez gero eta aldagai hauei buruzko informazioa lortzen badugu, Y-k aurrez esana egin ahal izango dugu?), edota kontrolatzeko, bai nahi ez dugun Y bat aurrikusiz, bai nahi dugun Y bat bultzatuz (Y-rengan eragina duten X_1 eta X_2 eraldatuz), ikertzen dugu. Ikertzearen laugarren helburu honi buruz hitz egiterakoan idazle batzuek nahiago dute "kontrolatu" hitza baino "eraldatu" erabiltzea.

Giza Zientzietan gizakia dugu interesgunerik nagusiena, eta gizakiaren inguruan badago, bai, zer ikertu. Honen ideia bat emateko, aski litzateke, gaia aukera dezaten utziz gero, ikerketaren metodologia ikasten dihardutenek aukeratutako ikerketa-lanen zerrenda ematea. Duela hamar urte hasi nintzen gai hau irakasten eta, benetan, interesgarriak dira harez gero ikasleek aurkeztu dizkidaten lanak. Era askotan sailka daitezke Giza Zientzietan izan daitezkeen ikerketa- lanak. Alor bateko problematika nagusiaren ikuspuntu bat ematera zuzenduta egotea (adibidez, ezaugarri konkretu batzuk dituzten pertsonen edota anomalien tipo eta maiztasunak aztertzea, batzuk azterketa epidemiologikoa deitua), egoera bateko portaera edota egoera-motak diagnostikatu edo diagnostikatzeko erak hobetzea, edo interesatzen zaigun eta lan egiten duguneko errealitatean egiten ditugun eskusartzeak (heziketa-motak, tratamenduak, terapiak) hobetu eta ebaluatzea izan daitezke gaiak sailkatzeko erizpide batzuk.

Beste bat, ikerketaren helburu teoriko, tekniko edo praktikoaren araberakoa izan daiteke. Batzuetan, gure zientzia edo alor konkretuko teoria gehitu eta hobetzea nahi da, gure teoriak errealitatearekin kontrastatuz. Honela, legeak gero eta unibertsalagoak izateaz gain, kontzeptuak zabaldu eta hauen arteko loturak sistematizatu egingo dira, nolabait. Beste zenbaitetan, ordea, nahiago izaten da aldatu eta hobetu nahi dugun errealitatea lantzeko modu eta tekniken inguruko ikerketak egitea. Eta, azkenik, beste batzuk nahiago dute ikertzeko orduan ere praktiko-praktikora jotzea, nahiz eta "praktika onerako, hoberena teoria on bat" dela aitortu. Baina, egoera konkretu batzutan edota pertsona konkretu batzuekin egindako lana hobetu nahi duenak, nahiago izaten du teoriak nola teknikak ekar

diezaioketenaren konbinazio pertsonal eta espezifiko bat egitea.

Interesgune, gai, alor, eremu bakar baten era askotara plantea daitezke ikertu nahi dugunaren inguruko arazo eta galderak. Egia da, garrantzitsuena arazoa izaki, ikertze--metodologia arazoari egokitu behar zaiola. Baina egia da, halaber, ikertzaile bakoitzak aukeratzen duen metodologiaren arabera (kualitatiboa, kuantitatiboa...), arazo batzuetan edo besteetan jar daitekeela arreta, edota erantzuna ere aukeratutako metodologiaren araberakoa izan daitekeela.

Edozein ikerketatan, hasiera-hasieratik garbi eduki behar dugu zeintzuk diren helburuak, nahiz eta oso zabalak izan eta ikerketa aurreratu ahala osatuz joan. Kontutan eduki behar dugu, denborak, indarrak, eta baliabideak baldintzatzen gaituztela.

Argi eduki beharra dago, halaber, zer diren edo zertan datzan ikertu nahi ditugun gai edo aldagairik garrantzitsuenak. Bere burua ikertzaile enpirikotzat duenak gai/ aldagai bakoitzak zer esan nahi duen definitzeko konpromezua hartu behar du, kasuen arabera definizioen formalismo-maila aldatuko den arren.

Ikerketaren planteamendua (estrategia), eta nondik- norakoa ere hasiera-hasieratik egin beharreko lanak dira. Zenbait kasutan, ikerketaren "diseinua" deitzen zaio honi. Ohizko formalismoetatik at gelditu eta "diseinu" hitzak eduki ditzakeen konnotazio txarrak baztertzeko, komenigarria litzateke frantsesek erabiltzen duten "plangintza" hitza erabiltzea, anglosaxoniarrek erabili eta oso esanahi zehatzak hartu dituen "prozedura" hitza baino egokiagoa delako.

Ikerketaren "balio edo baliazkotasun" hitza edo mamia eztabaida daiteke. Ene ustez garrantzitsuena ez da hitza, edota beronen azalpena CAMPBELL eta STANLEY (1972) edo COOK eta CAMPBELL(1979)en adierazpena; ikerketaren ondorioetara iristeko bideak onak, sendoak eta zehatzak diren jakitea baizik. Ez da ahaztu behar ordea ikerketaren oinarri teorikoa. Ezta ikerketaren ondorioak beste pertsona, momentu edo egoera batzuei aplika dakizkiekeen jakitea interesgarria dela ere.

Har ditzagun kontutan orain aztergai diren sujetuak. Ikerketaren kalitateak ez du zerikusi handirik pertsona-kopuruarekin, maila batera iristen garenean behintzat. Dena den, ikertzaileak ondo baino hobeto ezagutu behar ditu sujetuen ezaugarriak, ea zein den hautaketa horren helburua eta aztertutako pertsona-kopuruak froga estatistikoen fidagarritasuna bermatzen duen ala ez. Laginketa egokia egitea da Estatistikaren erantzukizunetako bat Giza Zientzien arloan.

Azterketa estatistikoaren oinarri izango diren datuak kalitate onekoak izan eta ondo jasota egotea da ikertzaileen eta Estatistikaren beste lanetako bat. Askotan,

"baliagarritasunean" eta "fidagarritasunean" isladatzen da bildutako datuen kalitatea. Ikertzaile enpirikoak datuak agertu nahi denaren adierazgarri ote diren galdetu behar dio bere buruari. Datuak zenbaki bidez adierazi nahi ditugun eta zenbakien neurketa-maila finkatzea dira hartu beharreko lehen erabaki inportanteenak. Geroago, datu-bilketa egiteko era hautatu beharko dugu (inkesta, behaketa, testa, elkarrizketa, galdeketa, eskala, etab.).

Datuak estatistikoki aztertuko diren erabakitzea ere garrantzitsua da. Planteamendu eta datu-bilketa egoki batzuen ondorio izanez gero, ikertu nahi den arazoari buruzko berri interesgarri eta aipagarriak emango dizkigute datuek. Informazio hori atera, eta datuak esan diezaguketena dezifratzean datza gakoa. Galliez Pages-ek dioenaren arabera, fenomenoak aztertu ondoren hasieran begibistakoak ez ziren datuak deskubritu, aurkitu edota aurrez suposatzen ziren egiturak zehaztu behar dira. Hori egiteko teknika askotara jo dezakegu. Frogatua dago ondo erabilitako teknika estatistikoek balio handia dutela datuek dakarkiguten informazioa lortzeko orduan; bai zerrendatzeko, eta baita antolatu eta aztertzeko ere. Aipatu beharrik ez dagoen arrazoiak direla eta, kapitulu honetan ez ditugu aipatuko estatistikatik kanpo datuak aztertzeko dauden teknikak. Hauek ere baliagarriak dira datu enpirikoetatik informazioa ateratzeko, eta Giza Zientziek ezin ditzakete baztertu.

Azken bolada honetan, aldiz, garapen interesgarria eduki dute "meta analisiek". Hauekin lan egiteko moduetako bat, gai baten inguruko aldez aurretiko zenbait ikerketaren datuak aztertzetik abiatzen da, ikerketa bakoitzean jasotako datuak pilatuz.

Estatistika eta ikerketari buruzko atal honi amaiera eman baino lehen, onuragarria litzateke 1988. urtean ELMORE eta WOEHLKEk egin zutenaren antzeko azterketak egiten jarrai dezaten eskatzea. Hauek, Giza Zientzietako ikerketetan gehien erabiltzen diren metodoez eta ikerketa-mota bakoitzean erabiltzen den analisi estatistikoaz ihardun zuten. Aurrez, zer eta nola ikertzen den jakin beharko genuke, noski. Ikerketa-eremu desberdinetan (heziketa berezia, jarrera politikoak, portaera asozialak) azpimultzoak egin beharko lirateke batetik, eta arlo bakoitzean erabil daitezkeen metodologien sailkapena bestetik. Horrela argi geratuko litzateke nahikoa ikertuak ez dauden alorrak zeintzuk diren eta baliagarria izan litekeen sail batean zein metodologi mota ez den erabiltzen.

Jakina da azterketa estatistikoetako emaitzek ez dutela, berez, esanahi handirik, eta helburu, aurrez planteiatutako galdera edota hipotesi eta aldez aurreko teoria eta ikerketekin batera interpretatu behar direla. Interpretatzearena ezinbesteko lana da ikerketa guztietan, ikertzaileak bere gaitasun eta ahalmenak bete-betean garatu behar dituelarik. Hementxe datza, hain zuzen ere, ikuspegi hermeneutikoaren arrazoirik nagusiena.

3.2.- Neurketa

Aipatua dago Giza Zientzietan erabilitako datuekin Estatistikan duen zerikusia. Zati honetan, Estatistika eta Neurketaren arteko zerikusia aztertu nahi da, beti ere ikuspegi

orokor batetik abiatuz.

"Datuak esleitzea" eta "Neurketa" hitzen arteko aldea gogoratzetik hasi behar dugu. Idatziz, nik "Aspectos metodológicos de la investigación educativa" izeneko liburuari eginiko hitzaurrean aipatzen da lehenbiziko aldiz ezberdintasun hori. Ideiarik nagusiena zera da: nahiz eta zentzu orokorrean hitz egin, jadanik ez duela balio "neurketa" hitzak datu enpirikoak jasotzeko lana adierazteko.

Stevens-ek eman eta ohizko bihurtu den definizioak ederki adierazten du neurketaren esanahia: arau batzuei jarraituz objektuei zenbakiak* egokitzea. Zenbakien mundua eta zenbakiok agertzen(errepresentatzen) duten errealitatearen arteko nolabaiteko isomorfismoa suposatzen da (esan beharrik ez dago arauak zenbat eta hobeak izan orduan eta hobeak izango direla lortutako zenbakiak). Hala ere, neurketari buruz eman dira beste definizio batzuk ere. "Datuak esleitzeak", berriz, ezagutu nahi den errealitatearen ezaugarrien berri emango diguten datuen produkzioaz hitz egiten digu. Datua zenbaki bihurtzen ez bada, ez dago neurketarik.

Estatistikak datu-zenbakiarekin (zenbakiak dira Estatistikaren aztergai) eta datu--zenbakiari buruz (nola lortu zenbaki hobeak) lan egiten du. Hemendik uler daiteke neurketarekin duen zerikusia, neurketa maila nominala baino ez den arren. Estatistikak Giza Zientzietan neurketa hobetzeko zer egin dezakeen planteatu behar dugu atal honetan, teorien arteko aldeak aztertzen hasi gabe.

Honako hauek dira Giza Zientzietan neurketari dagokionez konpondu beharko liratekeen arazoak:

1.- Neurketa-tresnak eraikitzea (test, galdeketa, eskala, behaketa, inkestak, etab.)

2.- Beraien ezaugarri teknikoen azterketak.

3.- Banakako edo taldekako erabilpenerako prestakuntza.

4.- Neurketa-tresnen motak mailakatzea, eraikitze-teknika, erreferentzia edo edukiaren arabera.

(*) Gure "Hizkuntz Pedagogia eta Ikerketa-Metodoen" Saileko irakasle batek egindako tesi baliotsu eta interesgarri baten zuzendari izan nintzen aspaldi. Epaimahaiburuak, ospe handiko matematikoa bera, 32 doktoreen aurrean irakurri eta defendatu zuen bere tesia Moskun. Orduan beste istorio txiki bat kontatu nuen, bertako protagonista nire iloba bat zelarik. Zortzi bat urte zuenean, nik zer egiten nuen galdetu zidan. "Psikometria erakutsi" erantzun nion. Eta, hori zer da?. "Ba, gure gauzetan, 2 gehi 2, 3.99 edo 4.02 izan daitekeela erakustea; zeren, ziur asko, ez lehenengoa, ez bigarrena, ez dira benetako 2ak. Batuketa egitea, batzuetan bakarrik posible izango da, eta berdina ere ahalegin jator eta noble bat besterik ez da." "Jo, hau da erroiloa!", erantzun zidan.

5.- Eskuartean prest ditugun tresnak ezagutzea, erabiltzekotan izango lituzketen helburuak kontutan hartuz.

6.- Zenbait egoera profesionaletan (diagnostikatu edo ebaluatzerakoan, adibidez) erabilpen egokia.

Horietatik lehen hirurak dira Estatistikan eragin handienak dituztenak. Neurketa-tresnen eraikitzeari dagokionez, tresnaren arabera alda daiteke laguntza, eskala edo pareko metodoetan proba objetiboetan baino handiagoa delarik. Dena den, tresnen lehen bertsioen ezaugarri teknikoen analisiak azken bertsiora iristeko lagungarri izango dira.

Neurketa tresnen ezaugarri teknikoak finkatzeko analisiei dagokienez, faktore bat edo gehiago neurtzen ari garen (analisi faktoriala), ea neurketa-tresna fidagarria den, neurtu nahi duguna neurtzen ari ote garen, item-ek edo tresnen elementuek ondo funtzionatzen ote duten (ohizko item-en analisia edo item-ari erantzundakoaren inguruko teoria berria), item-ak edo neurketa-tresnak okerrak ote diren ikustea da arruntena.

Neurketa-tresnen bakarkako edo taldekako erabilpenerako prestakuntzari dagokionez berriz, analisi estatistikoak populazio eta forma desberdinak barematzeko orduan aplikatzen dira. Hemen ere, item-ari emandako erantzunaren inguruko teoriak gauza berri garrantzitsuak ekarri dizkigu. "Item-ari neurketa pertsonal librea" delako teoriari esker; item-ak neurtzen direnean, sujetuen portaera eskala batean kokatzen da, horrela item-en funtzionamenduak taldearekiko duen lotura estua apurtzen delarik. Guzti honek baremazio-lanak ezezik bezeroaren neurrira egindako testak eta erabiltzaileari egokitutako test konputerizatuak sortzea ere errazten du. Azkenik, adimen artifizialaren arloan egindako aurrerapenak erantzun irekiak masiboki erabili ahal izatea ekarri dutela gogoratzea komeni da.

3.3.- Diagnostikoa

Giza Zientzietako profesionalen lanetako bat, talde edo pertsona batzuen arazoa diagnostikatzea da, pertsona edo talde horiekin zein lan egin eta zertan hobetu daitekeen jakin ahal izateko. Estatistika, kasu honetan adibidez, erabilgarria da ezaugarri konkretu bat duten sujetuen portzentaiak atera, jarrera sintomatiko bat osatzen duten ezaugarrien arteko harremanak ikusi, tratamendu edo terapia zehatz bat behar duten talde edo pertsonak diagnostikatu, terapia horien ondorioak zeintzuk izan daitezkeen jakin eta gizabanakoen bilakaera kronologikoaren jarraipena egiteko. Erabilera guzti horiez gain, aurreko atalean aipatu diren neurketa-tresnen gisako diagnostikatzeko tresnak prestatu eta aztertzeko ere balio du. Pertsonari egokituriko neurketa-tresnak diseinatu eta egoki erabiltzeko ere ezinbestekoa da Estatistika. Neurketa-tresnek garrantzi handia dute diagnostikoaren ikuspegitik.

3.4.- Ebaluazioa

Gure lanaren emaitzek hasiera-hasieratik planteatutako helburuei erantzuten ote dieten jakitearen garrantziaz jabeturik, gero eta hobeto burutzen ari den lanetako bat da. Prozesuak zein emaitzak ebalua daitezke. Eta ebaluatu, erabakiak hartzeko ebaluatzen dugu. Batzuetan sujetuak, zenbaitetan erakundeak, eta beste batzuetan egindako programak ebaluatzen ditugu.

Sujetuak lortu behar dituzten erizpideak edo daudeneko taldeak kontutan hartuz ebaluatzen ditugu. Erakundeak ebaluatzean, sortzen direnean jarritako helburuak betetzen ote diren jakin nahi da, askotan antzeko helburuak dituzten erakundeen arteko konparaketak egin daitezkeen arren. Programak ebaluatzeak, berriz, beronen helburuak eta emaitzak konparatzea du xede. Bai erakunde, bai programen ebaluazioan kontestua ezezik antolaketa, baliabideak, azpiegitura etab. ere kontutan hartu beharrekoak dira.

Era askotako funtzioak betetzen ditu Estatistikak ebaluazio-prozesuan; bai maila indibidualean, bai taldekakoan, ikerketa, neurketa eta diagnostiko-lanak biltzen bait ditu, dimentsio ebolutibo-luzerakoak garrantzi aparta duelarik. Ebaluazioaren problematikak zenbait ezaugarri espezifiko duenez, Estatistikaren aplikazio batzuk ere espezifikoak direla esan behar, oraingoz gai hau alde batera utzi arren.

4.- ESTATISTIKA EGOKI ERABILTZEKO BETE BEHARREKO BALDINTZAK

Aurreko atalean ikusi dugunez, estatistikak Giza Zientzietan duen zereginik nagusiena zenbakiek dakarkiguten informazioa ateratzea da. Ez datu guztiak; jarritako helburuak betetzeko beharrezkoak ditugunak baizik. Metodo estatistikoak oker edo gehiegi erabil ez daitezen, zenbait baldintza kontutan hartu behar dira.

Asko dira espezialistak Giza Zientzien munduan erabil ditzakeen analisi estatistikoak. Horregatik, kasu bakoitzarentzat egokiena den froga aukeratu behar da lehen-lehenik.

Frogarik egokiena zein den erabakitzeko, berriz, dauzkagun helburuak hartu behar ditugu kontutan. Hauek, ikerketan egindako galdera eta hipotesietan isladatzen dira ongien. Besteak beste, helburu, galdera edo hipotesien planteamenduaren oinarria den eredu teorikoaren sinpletasun/zailtasunaren eta teknika estatistikoaren korrespondentziak egon behar du. Datu-bilketa planteatzeko erak ere balio dezake zein froga-mota erabili behar dugun argitzeko. Adibidez, zenbait ikerketatan test-aurreko eta test-ondoko neurketak egin ditzakegu, edota bi neurketa: bata talde esperimental batean eta bestea kontrolpeko talde batean. Kasu bietan Student-en testa erabil daiteke; baina lehenengo kasuan elkarrekin zerikusia duten taldeekin, eta bigarrenean, berriz, talde askeekin, hau da, inolako zerikusia ez dutenekin.

Aztertu beharreko datuen izaera litzateke beste mugatzaile bat. Zenbaki guztiek ez dute berdin balio eta bakoitzetik analisi desberdina egin daiteke. Neurketa-maila askoren (nominala, ordinala,...) ondorio dira zenbakiak eta zenbat eta maila altuagoa izan, orduan eta baliotsuagoak izango dira analisi estatistikoak. Hala ere baliteke maila batentzat egokiak diren azterketak, azpiko maila batentzat egokiak ez izatea. Datuen banaketa da kontsideratu beharreko beste faktore bat; egiten den banaketaren arabera egokiagoa izango bait da froga bat edo beste. Estatistikan espezialista'aditu direnek datuen banaketa-motak eta kasu bakoitza aztertzeko erarik zuzenenak ikertu dituzte. "Banaketa normala" da ezagunetariko bat (hain ezaguna den "Gauss-en kanpaia"ren itxura hartzen duena), banaketa honi erantzuten dioten datuen esanahia ulertzea erraz samarra delarik. Zorionez, gizakion ezaugarri askoren datuek kurba normalaren araberakoak dira.

Analisi estatistikoak baliozkoak izan dakizkigun, hauek ondo funtziona dezaten jarritako baldintzak errespetatu beharra dago. Adibidez, froga estatistiko batzuek datuak normalki banatzea eskatzen duten bitartean, beste batzuek ez dute horrelakorik eskatzen. Froga batzuk erabiltzeko, konparatzen ditugun taldeek aldagai berak izan behar dituzte. Matematikoek garrantzi handia ematen diote froga estatistikoen baldintzak errespetatzeari. Egia esan, askotan eskuartean darabilzkigun datuak ez dira aukeratutako frogei ongien dagozkienak.

Orain arte froga estatistikoen balio eta fidagarritasunari buruzko aipamen batzuk egin dira. Definizio tekniko eman gabe (hemen azalduko ez ditugun zenbait kontzeptu menperatu eta gero) zenbat eta froga sendoagoa egin, orduan eta informazio adierazgarriagoak aterako direla esango dugu. Adibidez, indar gutxiko analisi estatistiko bat ez litzateke nahikoa gure hipotesia onartzeko (ditugun talde bien arteko diferentziarik eza ezin frogatu). Gure frogak pisu handiagoa edukiko balu, berriz, alde adierazgarriren bat agertzeko arriskua edukiko genuke. Froga baten indarrak (fidagarritasunak) laginarekin zerikusi handia du. Gauza nabaria da, lagina zabala bada, diferentziarik txikienak ere nabarmenkiago azaltzen direla. Hau garbi ikusiko dugu, froga estatistikoak zatiketak direla kontutan hartuz gero: Zenbakitzaileak, ikusi nahi dugun faktorea adierazgarria ote den isladatzen du eta izendatzailea berriz, txikiagotu egiten da kasu-kopurua handiagotu ahala. Beraz, zenbat eta izendatzaile txikiagoa izan, orduan eta zatidura handiagoa edukiko dugu. Hori dela eta, signifikazio-maila handiagotzen den neurrian, are adierazgarriagoak izango dira emaitzak.

Nahikoa ahal da orduan, gure lagina handiagotzea, emaitzak fidagarriak izan daitezen?. Ba, askotan bai, baina ez beti. Honetaz aparte, ezin ahaztu honako bereizketa hau: gauza bat da estatistikoki adierazgarri diren emaitzak lortzea, eta beste bat emaitza horiek benetan garrantzitsuak izatea ala ez izatea. Honekin zera esan nahi da: aldagai batek sortutako diferentzien adierazgarritasunaz gain, aldagai bien kobariantzen magnitudea baloratu behar da.

5.- *HEZIKETA ESTATISTIKOA*

Atal honen hasieran, rabbi zaharrak Psikologian zenbakiek duten "rol"ari buruz egindako galdera erretorikoa jasotzen zen. Hemen, galdera hori beste modu batera egin daiteke, hau da, zein Estatistika eta zenbat estatistika jakin behar dugu Giza Zientzien alorrean lanean dihardugunok?. Behar bada, galdera zorrotzagoa ere egin liteke, gutxienez Estatistika apurren bat jakin beharko genukeen galdetuz. Erantzuna ez da betiko topikoak dioen "Jakintzak ez du lekurik hartzen", ez, gauza bat ikasteak beste batzuk ikasteko denbora kentzen duelako. Dena den aurreko ataletan azaldutakoaren ondoren, nekeza da Estatistika-apurren bat ez dela jakin behar argudiatzea. Nire eritzian, zenbat eta Estatistika gehiago jakin, hobe. Eritzi hau, jakina, zertxobait xehetu egin beharko litzateke, Giza Zientzietako espezializazio-arloen arabera. Eritzi honen itzulpen zehatza hau da: konpetentzia teoriko eta teknikoetan Estatistikaren menpe daudenek, horiexek, jakin beharko lukete Estatistika gehien (ikertzaile, neurketa-espezialista, diagnostikatzaile eta ebaluatzaileek).

Kuantifikazio edo kualifikazioaren aldeko joera haundiagoa izatea ez da erabakiorra, zenbat estatistika jakin behar denari dagokionean. Alde batetik, bakoitzak kuantifikazio edo kualifikazioaren aldeko aldiak edo faseak izaten ditu, eta hori aurrikusiz gero, ona da kuantifikazio-une edo aldietarako gertu egotea. Eztabaida liteke lanbidean aurrera egiteko kuantifikazio fasea ala kualifikazio fasean hastea hobe den. Nolanahi ere, etsaiaren argudioak eta ekarpenak ulertzea eta baloratzea komeni da. Horrela, guztiek jakin beharko lukete ahalik eta Estatistika gehien.

Estatistika deskribatzaileaz nahiz inferentzialaz jabetzea komeni da; bai aldagai bakarrekoez,bai aldagai bikoitzekoez, eta ahal den neurrian, aldagai aniztunez ere bai; gutxienez, aldagai aniztunen perspektiba eta abantailak ulertzeko adina. Batez ere Estatistika parametrikoa ikasi, baina ez- parametrikoak nondik nora doazen ere bai. Eta hasteko, teoria eta probabilitate-kalkuluari buruzko eta banaketa teorikoari buruzko ideiak eduki.

Ez dago inoiz soberan aldagai aniztunen analisi estatistikoaren garrantzia azpimarratzea. Egia da, batzuetan, beharrezkoak direla aldagai bakardunen analisiak (% zenbat pertsonek dauka Y ezaugarria?; zein da Y aldagaian, pertsonen batezbestekoa eta desbidazio tipikoa?, etab.). Halaber, aldagai bikoiztun analisiek oso informazio interesgarria ematen dute (zein da X eta Y aldagaien arteko koerlazioa?; zein da Y aldagaian dagoen batezbestekoen aldea, X ezaugarria ezberdina duten pertsonengan?), etab. Baina gauzak askoz argiago ikusten dira aldagai askoren arteko harremanak batera ikertzen badira. Horrela ikus daiteke, adibidez, X eta Y-ren arteko koerlazioa bi hauek batera hirugarren aldagai batekin duten harremanen menpe dagoela; Y ondoen aurresaten duena X dela; datu batzuek errepikatzen den informazioa ematen dutela, etab. Bidezkoa denez,

batzuek zera galdetuko dute: ea orduan zergatik ez den aldagai-anitzeko analisi gehiago egiten. Erantzuna hau da: aldagai-anitzeko analisi gehiago horiei buruz gehiago ez jakiteagatik ez da egiten. Beste era batera esanda, adibidez, ikerlariok estatistika gehiago bageneki, analisi hobeak egingo genituzke; datuetatik informazio gehiago eta hobea aterako bait genuke. Egia da ikertzaileek, arrazoi askorengatik, gero eta estatistika gehiago dakitela. Beraz, gero eta analisi estatistiko hobeak eta sakonagoak egiten dituzte. Eta egiten ez duena atzera geratzen da ikertze-lanetan.

Horiek horrela, arraroa irudi dakiguke duela urte batzuk parte hartu nueneko eztabaida bat. Orduan, gure taldekook, beste estatu batzuetakoekin alderatuz, Elkarbizitzarako heziketa-ikerlanei (pertzepzio, ezagupen, jarrerei) buruzko emaitzen aurkezpenera, entzute handiko ikertzaile ingeles bat gonbidatu genuen. Honek jendaurrean adierazi zuen Giza Zientzietan beharrezko diren analisiak ondoko hauek zirela: portzentaiatan azaltzen zirenak, batezbestekoak, desbidazio tipikoak, koerlazioak eta T-test X^2 edo antzerakoak; gainerakoak ez. Jakina, erantzun bila esan zuen esan zuena, Espainiako Heziketa-Ikerketetan aitzindari izandako ikerlan baten aurkezpenean hain zuzen ere. Ikerlari ingelesari emandako erantzuna hau izan zen: bereiztu egin behar dira emaitzak azaltzerakoan zein estatistika komeni den eta erabili behar den, analisiak egiteko orduan. Ikerketa baten ondorioak kaleratzerako orduan, emaitza horiek nori zuzenduak dauden kontutan hartzea da gauzarik garrantzitsuena. Izan ere, jende guztiari, politikoei, administrariei edo heziketa estatistiko handirik ez duten profesionalei zuzendua baldin badago, estatistika sinpleak erabili behar dira, erraz azaltzeko eta ulertzeko modukoak, hala nola, %ak, batezbestekoak, etab. Benetan trebetasun izugarria eduki behar da ideia handiak hitz sinple eta argiz azaltzeko.

Hala ere, datuek daramaten informazioa ateratzeko orduan, ezagutzen ditugun analisi bide onenak erabili behar dira, eta horretarako, batzuetan balio aniztuneko analisiak ezinbestekoak dira. Beste hau ere egia da: datuek egia garbi batzuk planteatzen dituzte, eta hauetaz jabetzeko ez dago entzungailu supersofistikatuen beharrik.

Estatistika ikasteko bideei buruz, hau esan behar da: Unibertsitatean zenbat eta oinarri handiago eskaini, orduan eta hobe. Dena den, galde liteke ikasle guztiei karreran zehar gutxienez zenbat estatistika ezarri behar zaien. Kontua ez da beharrik gabe "asignatura zaila" sortzea. Era berean, aitortu behar da, batez ere Letratako ikasleen artean zenbakizko munduaren inguruan aurrez sortzen diren tabuak gaindituta geratzen direla estatistikaren irakaskuntza ona dela eta gehiengoarentzat estatistika asignatura gaindiezina ez dela egiaztatzen denean. Halaber, "gehiengoez" ere zerbait esan behar da, hau da, zenbat estatistika eskainiko zaien (gutxienez aukerako asignatura bezala) ikasle adituenei edota giza zientzien zenbakizko arlorako zaletasuna dutenei, horrela analisi eta datu-prozesuan benetako espezialistak diren profesionalak egon daitezen.

Unibertsitateko karreran zenbat estatistika irakatsi behar denari buruzko arazoa edo

eztabaida, ikasketa-plangintza berrietan argituko da. Gerora, botere-joku eta nork bere izenaren hedapenerako joera behera joango da, oinarri oneko profesionalak sortzen diren neurrian. Jakingai batzuk, batez ere teknikoak, askoz hobeto jasotzen dira karreran zehar; beste batzuk, ordea, geroago lor daitezke autodidakta eran.

Profesional bakoitzak lanean aurkitzen dituen zalantzen aurrean ematen diren aholku praktiko eta ikastaroen bidez ikasi ohi da Estatistika. Ene eritziz, bakarkako lana eta talde txikientzat antolatutako tutoretza mugatuak ikastaro trinkoekin konbinatzean datza Estatistika ikasteko biderik onena. Ikastaroetan, gai espezifikoak irakasten zaizkie antzerako ezagupen-maila duten talde txikiei. Hasieran, nolabaiteko oinarria hartzeko ematen diete ikastaro bat, eta beste bat bukaeran, bakoitzak bere kabuz ikasi duena amankomunean jartzeko. Bakarkako lana ordenadorez eta tutoreak zuzendutako ariketez baliatuta burutzen da. Tutoretzak, mugatuak eta pertsonalak dira, ikaslearen martxa oztopa dezaketen arazo eta zalantzak argitzea dutelarik helburu.

Estatistika ikasteko orduan oso garbi eduki beharreko zenbait ideia aipatuko ditugu jarraian. Estatistikako ikasle orok lor dezake itxurazko maila. Horretarako ez da zenbakien "beldurrik" eduki behar, kalkulu numerikoaren erraztasuna, emaitzen interpretazioaren zailtasunik eza eta lortutako informazioaren balio eta erabilgarritasunaz jabetu behar dugularik.

Estatistikaren ikasketa-prozesuaren amaierarik ezak beldurra sorterazten du. Tunel beltz honek mehatxatzeko baino gehiago, erakartzeko balio behar du. Zenbat eta gehiago ikasi orduan eta hobeto ikusten dira kontzeptu eta kalkuluen arteko lotura estuak. Estatistikaren ikasketa, progresiboa denez, errepikatzen diren gauzek aurrez azaletik ikusitakoak sendotu eta sakontzeko balio dute. Besteak beste kalkuluaren mekanikan ez galtzea komeni da. Mekanikan sartu aurretik, kalkuluaren logikaren ardatz nagusiak eta nondik norakoak ezagutu behar dira.

Estatistika "eginez" ikasten dela da funtsezko ideiatako bat. Praktika nagusitu behar zaio beti teoriari. Ondoren, ikaslearentzat zentzua duten datuekin lan egiten hasi behar dugu. Oso motibagarria da nork bere mugak eta oztopoak gainditu eta gauza berrietan murgiltzea. Ondo egiten dituzten analisiak errepikatzea da batzuen joera. Onuragarriagoa da, ordea, menperatzen direnak atzean utzi eta, adituen laguntzarekin ezagutzen ez diren gaietan sartzea.

Zorionez erabat aldatu da Estatistikaren didaktika. Ordenagailuen erabilerak kalkuluaren mekanikatik askatu gaitu batetik, eta bestetik problema eta emaitzen arteko logika hobeto ikusteko aukera eman digu, denbora laburragoan etekin handiagoa ateratzen delarik. Estatistikako praktikek eskaintzen dizkiguten posibilitateak ez dira oraindik, bere osotasunean, aprobetxatzen. Klase teorikoetako baino talde txikiagoetan praktikak egiteko beharko litzatekeen ordu-kopurua eskatzeko nolabaiteko beldurra somatzen da, honek

lantaldeak zertan gelditukо diren ez dakitela irakasle-kopurua handitzea suposatuko lukeelako.

Estatistikaren didaktikako puntu konkretu bat matematizazio-mailari dagokiona da. Ona da Giza Zientzietako ikasketak egiten dituzten ikasleei Estatistika ematen dietenen artean matematiko batzuk egotea, batez ere departamentuko beste irakasleekin esperientzia eta ikuspuntuak komentatzen badituzte. Puntu honekiko betidanik esan dut irakasten denak eta ikas daitekeenak oinarri matematiko sendoa eduki behar dutela, sinbologia eta erakuspenak (demostrazioak) gehiegi matematizatu behar ez badira ere.

6.- INFORMATIKA ETA ESTATISTIKA

Esana dago, jadanik, informatikak zeregin handia duela Estatistika ikas/irakasteko orduan. Ordenadoreei esker, Estatistika gehiago eta hobeto ikasten da, prozesua erraztuz eta denbora irabaziz. Dena den, toki gutxitan atera zaio atera dakiokeen etekina.

Gutxienez hiru arlotan nagusitu behar da Informatika Estatistikaren ikaskuntza/irakaskuntzan: Kalkuluan, prestakuntzarako eragingarri izatean eta didaktikan. Ordenadoreek badituzte, ordea, beste erabilpen batzuk Giza Zientzien esparruan, hala nola, prozesuen simulazioan, adimen artifizialean, eta lehen aipaturiko sistema adituetan.

Kalkulu estatistikoaren sinplifikazioa da informatikaren emaitzik nabariena. Denbora eta lana aurreratzeaz gain, eskuz egitekotan planteatu ere egingo ez genituzkeen kalkuluak egiteko aukera ematen digu informatikak. Gainera berdintsuak edo konbergenteak diren metodo eta analisiak aplikatuz, lortuko ditugun emaitzak konparatzeko ahalmena edukiko dugu, honek teknika ezberdinak konparatu eta menperatzeko dakartzan abantaila eta guzti. Norberak eskatutakoari ordenadoreak eman diezaiokeen erantzuna zenbaterainokoa izan daitekeen ikustea da erabiltzailearen alderdi erakargarriena. Eskaintza, eskaria baino handiagoa dela esan genezake. Behar-beharrezkoa da erabiltzaileak prestakuntza ona edukitzea, esku artean daukanari etekinik handiena atera diezaion. Honekin batera, imitazio-norgehiagoka sanoa ere sortzen dela esan behar. Zenbaitek, beste batzuek ordenadoreak nola funtzionatzen duen dakitenez emaitza hobeak lortzen dituztela ikustean, ez dute normalean atzean gelditu nahi izaten.

Estatistikaren ikaskuntza/irakaskuntzan informatikak duen dimentsio didaktikoaz aski hitz egin da aurreko atalean. Nolanahi ere, ordenagailu eta ikus-entzungailuen arteko lotura sendotzeko lan handia egin behar da, bide batez "software"ari arreta handiagoa eskaini behar zaiolarik.

Aipatutako posibilitateak errealitate bihur daitezen, Estatistikarekiko azpiegitura informatikoa zaindu behar da. Zentzu honetan zaindu behar diren arloak lau dira: Euskarri

fisikoa (hardwarea), euskarri logikoa(softwarea), eragiketak eta aholku-ematea.

Euskarri fisikoa garaile ateratzen ari dela aitortu behar. Bazter guztiak ordenadorez josita daude. Handi eta txiki, era guztitakoak aurki ditzakegu. Erakunde ofizialek berek, pertsonak kontratatu beharrean, makinak erosten dituzte, erabilpen-egoera onean (mantenuak garrantzi handia dauka) egon eta etekina atera dakien. Zenbaitetan, ez dago politika koordinaturik material informatikoa erosteko orduan. Behin eta berriz azpimarratu behar da kalkulu-zentru handiek beharrezkoak izaten jarraitzen dutela, interesatuek, pantaila eta terminalez baliatuz, makroordenadoreak erabiltzeko posibilitatea izan dezaten. Erakunde publiko batzuek, toki irikietan -liburutegi publikotan- adibidez, jendearen esku ordenagailuak jartzea aurrerapausu nabarmena da.

Softwarea ere dagokion pisua eta tokia hartzen ari da. Administratzaileak, programa on eta egokirik gabeko makina baten balio ezaz jabetzen hasiak dira. Programa-paketeak (SPSS, BMDP, SPAD, SAS,etab....) ezinbestekoak dira Estatistikan lan egiten dugunontzat. Toki askotan salgai daude "bertsio ofizialak". Are zabalduago daude "bertsio piratak". Bertsio legal eta gaurkotuak nonnahi aurkitu ahal izatea lortu beharko litzateke. Hala ere, problema estatistiko konkretuak konponduko dituzten programa espezifikoak noiz agertuko zai egon beharko dugu erabiltzaileok. Askotan norberak dauzkan datuen behar espezifikoei erantzun egokia emateko programatxoak behar izaten direnez, bi irtenbide hauetatik bat aukera dezakegu: Diru askorik ordaindu gabe gure arazoari konponbidea emango dion pertsona bat kontratatu, edota maila apalagoko programatzaile bihurtu. Programatzaile ona izateko horretan aritu behar da buru-belarri. Boluntarismoa eta ondoko bulegoko lagunaren lepo bizitzea dira erro-errotik baztertu beharreko joerak.

Ordenadorez lan egitea, oro har, unibertsitate-mailan gaizki burutu ohi den prozesua da. Esana dago profesional berriak kontratatzeko beldurra nabaria dela. Hala ere erakunde publikoek badakite erabilgarri diren makinei probetxua ateratzeko ordenadoreen arteko loturak antolatu, periferikoekin konjuntzioa egin, ordenadoreen memoriaz eta erabiltzaileak ordenagailuarekin lan egiten hasteko momentu eta moduez arduratuko diren teknikoak behar direla. Bestalde, batzuen axolagabekeriak eta besteen urdurikeriak eragin dezaketen kalteen aurrean, ordenadoreen erabilpena kontrolatuko duten arduradunen premia azpimarratuko dugu.

Aholku ematearen gatazka gaizki planteaturik egoten da. Beharrezkoak diren zenbaiten artean (metodologikoak, datapen eta neurketari buruzkoak, etab.), prozesu informatikoen eta estatistikaren inguruko aholkuez hitz egin behar da bereziki. "Eginez ikasten diren" gauza orotan bezala. Problemak azaldu ahala, norberak baino gehiago dakien bati kontsultatu beharra egoten da; are gehiago zalantzak argitzeko eginda dauden eskuliburuak atsegin ez baditugu. Baina, askotan, arazo zailagoak konpontzeko gai direnengana jotzen dugu txikikeriatan aritzeko, giza errekurtsoen probetxamendua oker eginez. Prozesuan zehar gertatzen diren arazorik gehienak, makinen funtzionamenduaz

arduratzen direnak bidera ditzakete. Aholku-ematea bi mailatakoa izan daiteke: Lehena, hastapenei dagokiona da, eta, ordenadorea erabiliz egindako lehen prozesu estatistikoak direnez, makinak botatzen dituen zerrendak irakur eta interpretatzeko laguntza eman behar zaie hasiberriei.

Honetarako, makinen arduradun, karrerako azken kurtsoetako bekadun, edo gai esperimentaletan adituak direnez balia gaitezke. Aholku emate-mailarik gorena, ordea, esku artean daukagun arazoarentzat egokienak diren analisi estatistikoak planteatzean datza. Profesionalizatu beharreko lana berau ere. Estatistikako ezagupen handirik eduki ezean, kasurik gehienetan erabat beharrezkoa da tekniko bati aholkua eskatzea.

7.- ONDORIOAK

Kapitulu honetan azaldutako zenbait gauza, zabalago, hobeto edo beste era batean planteatzen dira liburu honetako beste kapitulu batzuetan. Errepikatze eta ebitaezinak diren aldeak aitortuta ere, ikuspegi orokorrean datza kapitulu honen interesa. Ikuspegi orokor hau, beste ikuspegi orokor edo partzialekin kontrastatuz hobetuko da.

Estatistika, gure lana merezi duen gauza baliagarria dela esanez amaitu nahi nuke. Giza Zientziatako edozein espezialistaren esku dago Estatistika menperatzea. Profesionala izateagatik, lanean azaltzen diren arazoak konpontzeko edota, askorentzat, oraindik ere berri samarra den ikerketaren mundu erakargarria ezagutzeko interesa dakiguke. Guzti honen azpian dagoen ideia zera da: nahiz eta planteamendu guztiz edo sasikualitatiboak eduki, Estatistika gehiago jakiteak profesional hobea izaten laguntzen du.

8. BIBLIOGRAFIA

- ALVARO, M.; DENDALUZE, I.; MUÑOZ-REPISO, M.; PEREZ SERRANO, G. (1982): *Los estudiantes españoles y los valores democráticos: Estudio empírico sobre la socialización política.* Ministerio de Educación y Ciencia. Madrid.

- CAMPBELL, D.T. + STANLEY, J.C. (1973): *Diseños experimentales y cuasi-experimentales en la investigación social.* Amorrortu. Buenos Aires.

- COOK, T.D. + CAMPBELL, D.T. (1979): *Quasi-experimentation: Design and analysis issues for field settings.* Rand McNally. Chicago.

- DENDALUZE, I. (1972): "La hipótesis de la convergencia de los países y la Psicología". *Revista de Ciencias de la Educación*, Oct-Dic. 519-533.

- DENDALUZE, I. (1980): "Psikometria".*Psikologia - Hiztegia.* Elkar. Donostia. 68-75.

- DENDALUZE, I. (1980): "Estatistika". *Psikologia - Hiztegia*. Elkar. Donostia. 234-–240.

- DENDALUZE, I. (Coord.) (1988a): *Aspectos metodológicos de la investigación educativa*. Narcea. Madrid.

- DENDALUZE, I. (1988b): *Una reflexión metodológica sobre la investigación educativa*. DENDALUCE,I (1988a).13-41.

- DENDALUZE, I. (1988c): *Contexto y criterios hacia una tipologización de las metodologías de la investigación en Pedagogía Diferencial*. Bordón, 40(4), 598-610.

- ELMORE, P.B. + WOEHLKE, P.L.(1988): *Statistical methods employed in "American Educational Research Journal", "Educational Researcher" and "Review o Educational Research"* from 1978 to 1987. Educational Researcher, 17(9), 19-20.

- PILIBOSSIAN, P.(1988): "Contribución de los métodos estadísticos al desarrollo de las Ciencias de la Educación". In DENDALUZE, I. (1988a). 216-227.

- TEJEDOR, J.(1988): "El soporte estadístico en la investigación educativa". DENDALUZE, I.(1988a). 228-244.

2. GAIA
ESTATISTIKA APLIKATUA. KONTZEPTU OROKORRAK. MAIZTASUNEN BANAKETA.

Ikertzaile batek giza errealitate bat edo/eta errealitate soziala ezagutu nahi duenean, ikuspegi askotatik jo dezake.Gizaki eta gizarteko jokabideen zailtasun logikoak ikusiz, bere "errealitatera" zehazpen batzuen bitartez hurbiltzen saiatzen da; hauek izango bait dira neurtu eta geroago aztertu nahiko dituenak. Estatistika, neurketa prozesuan lortutako zenbakiak edo zenbakizko asignazioak (egokierak) aztertzen saiatzen da.

Garbi utzi behar da fenomenoak ez ditugula neurtzen; baizik eta fenomenoen propietateen adierazleak edo agerpenak, hau da, benetan neurtuko ditugunak, ikertu nahi dugun fenomeno baten ezaugarriak izango dira.

Ikerlarientzat, Giza Zientzi mailan, Estatistika oso lagungarria izan daiteke, baina ez dugu ahaztu behar, estatistika ikertzailearen eskuetan dagoen tresna bat dela, eta ikerketaren zati garrantzitsu bat era dezakeela. Baina ziur asko edozein ikerketaren barruan badaude estatistika bezain garrantzitsuak izango diren beste alderdi asko, adibidez, planteamendu ona egitea, lanerako hipotesi onak formulatzea edo neurketa ona egitea...

Giza Zientzien barnean ikerketa aplikatu edo esperimental bat egiterakoan, ohizko araua da edozein eratako tresna psikometriko bat pertsona-multzo bati pasatzea. Pertsona--multzo baten errealitatea ezagutu nahi badugu, pertsona-multzo honi POPULAZIOA deituko diogu. Askotan ez da bidezkoa izango populazioko pertsona edo gauza guztien neurketak lortzea, arrazoi ezberdin askorengatik: denbora, dirua... Kasu hauetan populazioko zati bat aukeratu ondoren, azpimultzo honetako pertsonei diseinatutako froga psikometrikoa pasatuko diogu. Azpipopulazio honi LAGINA deituko diogu. Hemen arazo bat sortzeko arriskua dago, eta lagin honen aukeraketa da hain zuzen. Zenbat pertsonak osatu behar dute lagin hau?, eta ondoren zeintzuek izan behar dute lagin barruan egongo diren pertsonak?. Hau da, nola aukeratu lagineko pertsonak?. Azken batean, aurreko galdereei erantzun ona ematerakoan bilatzen dena zera da: aukeratutako lagina populazioen "errepresentagarriena" izatea. Honen arrazoia zera da: laginean lortutako emaitzak populazioari estrapolatzea nahi izatea. Orain arte esan ditugun arrazoiengatik oso garrantzi haundia edukiko du laginen hautaketak, zeren laginaren hautaketa txarra baldin bada lagin horretan lortutako emaitzek ez dute baliotasun estatistikorik izango, hau da, ez da posible izango gure datuen emaitzak populazio guztira estrapolatzerik. Laginen aukeraketa aztertzen duen zatia, Estatistika barruan LAGIN-TEORIA da eta 12. gaian ikusiko dugu.

Lagina aukeratu ondoren, lagin barruan dauden elementuak hartuz, eta neurketa--tresna egoki baten bidez, geroago aztertuko ditugun datuak lortuko ditugu.Estatistikako

arloetatik, gure ikerketarako interesatzen zaizkigun ezaugarrien neurketan zentratzen den zatia PSIKOMETRIA edo NEURKETAREN TEORIA da, eta 8. gaian teoria honen aurkezpen labur bat egiten da.

Tresna psikometriko hauen bitartez, elementu batzuetatik besteetara aldatzen diren ezaugarriak neurtuko ditugu. Ezaugarri hauei, orokorki hartuta ALDAGAI deituko diegu.

Neurtu nahi dugun ezaugarri-mota oso zabala eta era askotakoa izan daiteke. Neurtu nahi dugun ezaugarrien arabera edo/eta erabiliko ditugun tresnaren arabera, aldagai-mota ezberdinak izango dira. Aldagaien sailkapen posible bat orain ikusiko dugu.

1.- ALDAGAI KUALITATIBO EDO NOMINALAK

Ezaugarri kualitatibo bat neurtzearen emaitzari Atributu edo Aldagai kualitatibo deituko diogu. Pertsonak edo elementuak atributu batzuen arabera bereiz daitezke: ilearen kolorea, sexua, oinetako-mota.....

Batzuetan, erabilitako neurketa-tresnaren motagatik, edota beste batzuetan tresnaren faltagatik, hasiera batean kualitatiboak ematen ez duten aldagai batzuk (hain zuzen zenbakizko aldagaiak, pertsona baten pisua, altuera, adina edo ikasgai batean lortutako puntuazioak) aldagai kualitatibo eraldatzen dira. Adibidez, pisuaren arabera lagineko pertsonak lodi eta argaletan sailkatzen baditugu, edo/eta adinen arabera gazte eta zaharretan sailkatzen baditugu, aldagai hauek ere kualitatibo bezala kontsideratuak izango dira.

Aldagai kualitatibo batek, ondo definituta gelditzeko, ondorengo propietate hauek bete beharko ditu:

.- Klase ezberdinak ondo definiturik egon behar dute.

.- Klaseek, elkarren artean bereiztuak izan behar dute. Elementu edo pertsona bakoitza klase batekoa bakarrik izan daiteke.

.- Elementu guztiek klase batekoak izan behar dute. Hau da, elementu batek ezin du sailkatu gabe gelditu.

2.- ALDAGAI KUANTITATIBOAK

Neurtu nahi dugun ezaugarria zenbakizkoa (kuantitatiboa) denean, eta neurketa egiterakoan erabili dugun tresnak zenbakizko balioa esleitzen duenean, aldagai kuantitatibo bat edukiko dugu.

Jakina denez, datuak lortzeko erabili dugun neurketa-tresnaren arabera, ezaugarri

bakoitzari egiten zaion esleipen-mota ere aldatu egiten da.

Pertsona edo elementuen ezaugarri baten arabera ordenatzen direnean, sailkapen ordinala lortuko dugu. Hau ALDAGAI ORDINALA izango da.

Neurketa era ezberdin ugari daude eta sobera konplexua izango litzateke hemen deskribatzea.Hala ere, sailkapen bati buruz hitz egin beharra dago, gero izango dituen ondorioengatik, zeren eta analisi estatistikoak egiterakoan, datuak aztertzeko era aldatzen bait da.

Aldagai kuantitatiboak: DISKRETUAK-JARRAIAK

Aldagai kuantitatibo DISKRETUAK, zenbaki osoen bitartez beren balioak adieraz daitezkeenak dira. Beti balio batek bere hurrengoa edukiko du. Hau da, ondoz-ondoko bi balioren artean ezin du tarteko baliorik egon. Adibidez, familia bateko haur-kopuruaz hitz egiten ari garenean, bi eta hiru semeren artean ez da posible izango erdiko balio bat egoterik.

Aurretik azaldutakoari kontrajarriz, aldagai kuantitatibo JARRAIETAN, hitzaren adiera hertsian ezingo litzateke ondoz-ondoko balioez hitz egin, kaleko hizkeran hori erabiltzen bada ere. Aldagai jarraien adibide bezala pertsona baten pisua hartzen badugu, esan dezakegu 65 kg-ren hurrengoa 66 kg-ko pisua dela. Baina bien artean nahi dugun beste balio ezar dezakegu. Beti, bi edozein balioren artean tarteko infinitu balio daude. Hau da aldagai jarraien ezaugarria. Ohizko aldagaiak hauek dira: denbora, altuera, zabalera...

3.- *BANAKETAK*

Laginaren datuak jaso ondoren eta datuen deskribapena egiten lagunduko digun indize estatistiko posibleen kalkuluak egitera pasa baino lehen, datuak antolatu egingo ditugu. Askotan interesgarria izango da adierazpen grafikoren bat egitea ere. Adierazpen edo errepresentazio grafikoari buruz zera esan daiteke: lortutako emaitzak ondo adierazten dituzten guztiak onargarriak izango direla.Beraien artean garrantsitsuenak hauek direla esan daiteke: Histograma, poligonalak, sektorezko diagramak...

Datuen antolaketarekin, hauek ematen diguten informazioa laburtzea bilatzen da. Datuen antolaketa hiru era ezberdinetan egin dezakegu:

.- Maiztasun-banaketaz.
.- Maiztasun erlatiboen eta portzentaien banaketaz.
.- Maiztasun metatuen banaketaz.

Eman dezagun 1000 familiaren seme-alaba kopuruen ikerketa bat egiten ari garela. Bildu ondoren, datu hauek batabestearen atzetik aurkezten baditugu, ematen diguten informazioa ikustea nahikoa astuna izango da, zeren eta datu guztiak irakurtzera behartuta egongo bait ginateke. Horregatik datu hauek BANAKETA-taula baten bidez aurkezten dira.

Gure datuak, ondoko era honetan aurki ditzakegu. Gure adibidean X_i zutabea, seme-alaba kopurua adierazteko erabiliko dugu, eta M_i zutabea, seme-alaba kopuru bakoitza zenbat aldiz agertzen den adierazteko.

X_i	M_i
0	170
1	390
2	257
3	115
4	54
6	2

Datuekin egin dugun prozesua ondoko hau izan da. Kasu honetan aldagai diskretua denez, lehenengo agertzen diren balioak zeintzuk diren ikusi ditugu eta gero bakoitza zenbat aldiz agertzen den kontatu dugu.

Baina askotan hau ez da posible izango; esate baterako pisua bezelako aldagaietan, hau da, aldagai jarraietan. Kasu hauetan, lehenbizi gure aldagaiak hartzen dituen balioak tartekatu egin beharko ditugu. Hori da hurrengo adibidean egin duguna. Lehenengo, X_i aldagaia, pisua adierazten duena, tartekatu egin dugu.

X_i	M_i
40 kg edo gutx	12
40 - 50	115
50 - 60	170
60 - 70	257
70 - 80	390
80 - 90	54
90 kg edo gehi	2

Aldagai jarrai bat elkartzen edo tartekatzen denean, tarte-kopurua ikertzaileak erabaki behar du. Idazle batzuek, tarte-kopuruari buruz zera esaten dute: "Tarte-kopurua, elementu-kopuruen erro karratua baino txikiagoa izango da".

Banaketa portzentuala, maiztasun-banaketatik sortzen da. Askotan familietako %39k seme-alaba bakarra dutela jakiteak informazio gehiago ematen digu familia guztietatik 390 familiak seme-alaba bakarra dutela jakiteak baino. Batzuetan ehuneko hainbestekotan hitz

egin beharrean bateko hainbestekotan hitz egiten da. Beste era batera esanda, maiztasun portzentualaz hitz egin ordez, maiztasun erlatiboaz hitz egiten da. Tarte baten maiztasun erlatiboa kalkulatzeko, tarte edo klase honen kopurua lagineko pertsona edo elementu guztiez zatituta lortuko dugu. Hurrengo taulan lehen aipaturiko bi kasuak ikusiko ditugu. M_i, maiztasun absolutua da, P_i, maiztasun portzentuala eta p_i, maiztasun erlatiboa.

X_i	M_i	P_i (%)	p_i
0	170	17'0	0'170
1	390	39'0	0'390
2	257	25'7	0'257
3	115	11'5	0'115
4	54	5'4	0'054
5	12	1'2	0'012
6	2	0'2	0'002

X_i	M_i	P_i (%)	p_i
< 40 kg	12	1'2	0'012
40 - 50	115	11'5	0'115
50 - 60	170	17'0	0'170
60 - 70	257	25'7	0'257
70 - 80	390	39'0	0'390
80 - 90	54	5'4	0'054
< 90 kg	2	0'2	0'002

Azkenik zera esan behar da: badagoela oso erabilia den beste banaketa-mota bat, MAIZTASUN METATUEN BANAKETA. Banaketa-mota hau klase baten maiztasuna metatzean edo gehitzean datza. Beraz lehengo kasuko maiztasun metatuen banaketak ondoko hauek izango dira, MM_i-ren zutabea maiztasun metatuena izango delarik.

X_i	M_i	MM_i
0	170	170
1	390	560
2	257	817
3	115	932
4	54	986
5	12	998
6	2	1000

X_i	M_i	MM_i
< 40 kg	12	12
40 - 50	115	127
50 - 60	170	297
60 - 70	257	554
70 - 80	390	944
80 - 90	54	998
< 90 kg	2	1000

3. GAIA
ESTATISTIKA DESKRIBATZAILEA

Datuak jaso eta kodetu ondoren, hurrengo pausoa datuak deskribatzea izango litzateke eta era berean datuek ematen diguten informazioaren laburpena egitea. Beraz bi talde ezberdinetan lortutako emaitzak konparatu nahi baditugu, konparaketa hau emaitzaz emaitza egitea baino logikoagoa talde horien ezaugarri batzuk konparatzea izango da. Ezaugarri horiek taldeko balio adierazgarriak izatea komeniko da.

Datuen deskribapena egin ahal izateko lau eratako balioak kalkulatuko ditugu:

1.- **Zentralizazio-neurriak**: Lagin baten datu ordenatuen zentru inguruan kokatzen diren balio adierazgarri batzuk dira zentralizazio-neurriak.

2.- **Sakabanatze-neurriak**: Datu-multzo batean muturretako balioak erdiko balioetatik aldentzen direneko neurriaren zenbatekoa ematen duten estatistikoari, sakabanatze--neurria deritzogu.

3.- **Asimetria-neurriak**: Jasotako datuen konfigurazioa zenbateraino den asimetrikoa azaltzen diguten neurriak dira.

4.- **Zorroztasun-neurriak**: Jasotako datuen zorroztasuna azaltzen digutenak dira.

Neurri ezberdinen kalkuluak egiterakoan kontutan hartu beharko dugu datuak nola antolatu ditugun, hots, datu tartekatuak diren ala ez; eta datu tartekatuak izanez, datu hauek aldagaia diskretu edo jarraienak diren ere kontutan hartu beharko dugu.

1.-ZENTRALIZAZIO-NEURRIAK

Jaso ditugun datu guztien erdiko puntua zein den eta erdiko balioak zeintzuk diren esaten diguten neurriak, zentralizazio-neurriak izango dira. Normalean, datuen artean gehien errepikatzen direnak izaten dira. Bestalde zentralizazio-neurriak erdiko balioak direla esan daiteke, zeintzuek taldearen balio guztiak errepresentatzen bait dituzte eta era berean bi edo talde gehiagoren erdiko puntuen konparaketa egiteko baliagarriak izango bait dira.

Zentralizazio-neurririk garrantzitsuenak ondoko hauek dira:

A. Batezbesteko aritmetikoa.
B. Mediana.
C. Moda.

A. Batezbesteko aritmetikoa

Zentralizazio-neurrietan gehien erabiltzen den balioa da. Honela kalkulatuko dugu: aztertuko ditugun datuen batura zati kasu-kopurua.

Datu-multzo baten batezbesteko aritmetikoa kalkulatzeko ondoko formula hau erabiliko dugu:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_i + X_2 + \ldots\ldots + X_N}{N}$$

Hemendik aurrera sinbologia berezi bat erabiliko dugu makina bat formula idazterakoan, eta askotan ikusiko dugunetako bat "Σ" da. Σ sinboloak honen atzetik dagoen aldagaien balioen batuketa egin behar dugula esan nahi du.

Adibidez, sei mutikoren matematikazko notak ondoko hauek izango balira: 6, 7, 8, 7, 5, 7. Datu hauen batezbesteko aritmetikoa hau izango da:

$$\overline{X} = \frac{6+7+8+7+5+7}{6} = \frac{40}{6} = 6.66$$

Balio batzuk errepikatzen direnean, beste formula bat erabil daiteke. Orain aurkezten dugun formula erabiliz, operazio aritmetikoak asko laburtzen dira; batez ere lagin-kopurua handia denean.

$$\overline{X} = \frac{\sum_{i=1}^{N} M_i * X_i}{N} = \frac{M_1 * X_1 + M_2 * X_2 + \ldots\ldots + M_N * X_N}{N}$$

Eman dezagun familia batzuei inkesta bat pasatu ondoren beren seme-alaba kopuruari buruz taulan agertzen diren emaitza hauek dauzkagula. Taula honetan, X_i aldagaien balioak adierazteko erabiliko dugu, eta M_i maiztasunak adierazteko; hau da, seme gabeko familiak 8 dira, seme bat daukatenak 7 ... etab.

X_i	M_i	$M_i * X_i$
0	8	0
1	7	7
2	4	8
3	4	12
4	3	12
5	3	15
6	1	6
	30	60

$$\bar{X} = \frac{60}{30} = 2$$

Hirugarren kasua ere gerta daiteke, hau da, datuak tartekatuta egotea.

Askotan neurketa egiterakoan erabili dugun tresnagatik edo/eta beste arrazoi batzuengatik, datuak tartekatuta agertuko zaizkigu. Kasu hauetan orain arte erabilitako formulak ere baliagarriak izango dira, baina formula erabili baino lehen aldaketa txiki bat egin beharko dugu.

Eman dezagun lantegi bateko langileen soldatak honela batu ditugula:

X_i (tartea)	M_i
80000 baino gut.	2
80000 - 90000	4
90000 - 100000	6
100000 - 110000	4
110000 - 120000	2
120000 - 140000	1
140000 baino geh.	1
	20

Kasu honetan ikusten den bezala, datuak tartetan bilduta daudelako, X-en balioak ez dira ezagunak.

Arazo hau gainditzeko kontzeptu berri bat sortu behar dugu: Klase-ordezkaria.

Tarte baten ordezkaria, tarte hori adierazteko erabiliko dugun balio bat izango da. Kalkuluak egiterakoan klase bakoitzeko elementu guztiei klase ordezkarien balioa emango diegu.

Klase-ordezkari bezala gehienetan tartearen erdigunea onartuko dugu.Batzuetan izaten dira oztopo batzuk; lehenengo eta azkenengo tarteetan batez ere.Tarte hauetan klase--ordezkari bezala mugaren aldameneko balio bat hartzen da.

X_i (tartea)	X_i	M_i	$X_i * M_i$
80000 baino gut.	75000	2	150000
80000 - 90000	85000	4	340000
90000 - 100000	95000	6	570000
100000 - 110000	105000	4	420000
110000 - 120000	115000	2	230000
120000 - 140000	130000	1	130000
140000 baino geh.	150000	1	150000
		20	199000

$$\overline{X} = \frac{\sum_{i=1}^{N} M_i X_i}{N} = \frac{M_1X_1 + M_2X_2 + \ldots\ldots + M_NX_N}{N} = \frac{1990000}{20} = 99500$$

B. Mediana

Lehen aipatu dugun bezala, lagin bat deskribatzeko badira beste zentralizazio-neurri batzuk, eta horietako bat mediana da.

Mediana delako balioa honela definituko dugu: Txikienetik handienera ordenaturik dagoen multzo baten mediana, erdian dauden balioen batezbesteko aritmetikoa da.

Beste era honetara ere definitzen da balio hau: "Lagin batean, gainetik eta azpitik ehuneko berrogeitamarreko behaketa uzten duen puntuazioa edo balioa".

Adibide-pare bat ikusiko dugu:

X_i : 1 4 5 6 7 8 9 9 10 ..... mediana = 7

Y_i : 1 4 5 6 7 8 9 9 ..... mediana = $\frac{6+7}{2} = 6.5$

Datu elkartuak baditugu, mediana ondoko formula erabiliz kalkulatuko dugu:

$$Med = L_i + b_i \frac{N/2 - F_i}{n_i}$$

non i = medianadun klasea

N = laginaren maiztasuna

L_i = medianadun klasearen behemuga erreala

b_i = medianadun klasearen tarte-zabalera

n_i = medianadun klasearen maiztasun absolutua

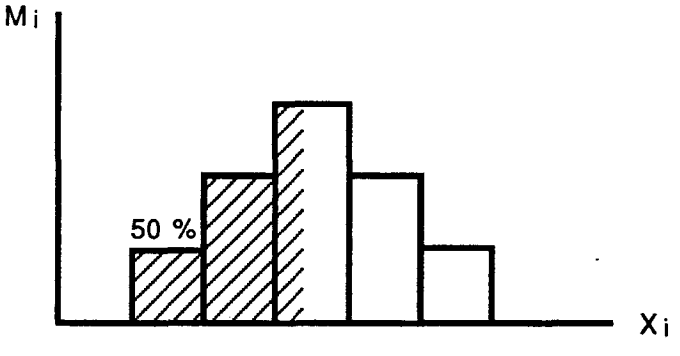
F_i = Medianadun klasearen behemuga errealari dagokion maiztasun metatua.

X_i (tartea)	M_i
80000 baino gut.	2
80000 - 90000	4
90000 - 100000	6
100000 - 110000	4
110000 - 120000	2
120000 - 140000	1
140000 baino geh.	1
	20

$$Med = 90000 + 10000 \frac{20/2 - 6}{6} =$$

$$Med = 90000 + 10000 * 4/6 = 96666.66$$

Medianaren adierazpen grafikoa ondoko hau izango da: Histogramaren azalera bi zatitan erdibitzen duen abzisa, mediana da. Beraz, ardatz horretatik eskuinetara nahiz ezkerretara gelditzen den azalera osoaren %50ekoa da. Beste era batera esanda, ardatz horretatik bi aldetara, maiztasun-kopuru berdinak daude.



C.- Moda

Hirugarren zentralizazio-neurria moda da.Modaren definizioa hau da: Zenbaki--multzo batean, gehien errepikatzen den balioa, moda da.

X_i : 1 4 5 6 7 8 9 9 10 ..... moda = 9 (modabakarra)

Y_i : 1 4 5 6 7 7 8 9 9 ..... moda = 7 eta 9 (modabikoa)

Datuak taldekatuta daudenean, Moda bezala, maiztasun handiena dagokion puntuazioa onartuko dugu.

X_i	M_i
0	8
1	7
2	5
3	4
4	2
5	3
6	1

Moda = 0

2.- *POSIZIO-INDIZEAK: KOANTILAK*

Posizio-indizeak, sakabanatze-neurriak eta zentralizazio-neurrien artean dauden adierazle batzuk dira. Batez ere froga edo psikometri tresnak neurtu edo barematu nahi ditugun kasuetan oso garrantzi handikoak izango dira.

Indize hauek helburu bikoitza betetzen dute:alde batetik emandako portzentaia bat bere ezkerretan uzten duen puntua adierazten digute; bestetik bere aldamenetan puntu batek uzten dituen portzentaiak kalkulatzeko baliotsuak zaizkigu.

Hots, talde baten soldataren maiztasun banaketa bat emanda, indize hauen bidez zera jakin genezake: alde batetik, 100.000 pta. baino gutxiago irabazten duten pertsonen portzentaia eta bestetik denetan gutxien irabazten duen %65a hartzen badugu, aukeratutakoetatik gehien irabazten duenak zenbat diru irabazten duen ere jakin genezake.

KOANTILAK: Koartil, Dezil eta Pertzentilak

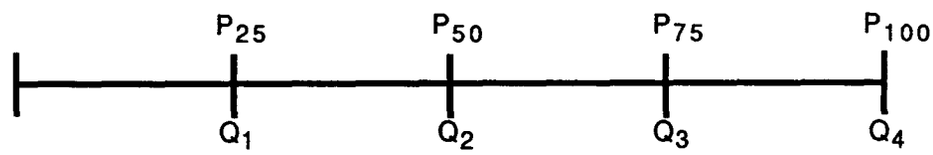
Datu-multzo bat dugunean, badakigu multzo hori bitan erdibitzen duen balioa mediana dela. Eta mediana honek naiz ezkerretara nahiz eskuinetara uzten duen maiztasuna N/2-koa da. Era berean multzo hau lau, hamar edo ehun zatitan zatitzen duten balioak kalkula ditzakegu. Balio hauek orokorki "Koantilak" ditugu. Lau zatitan zatitzen badugu Koartilak izango dira.Hamar zatitan banatuz gero, Dezilak eta ehun zatitan banatzen baditugu Pertzentilak.

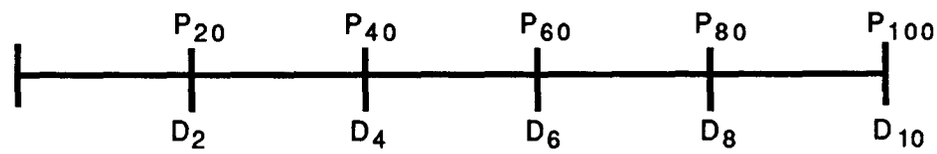
Pertzentilak adierazteko P letra erabiliko dugu.

Dezilak adierazteko D letra erabiliko dugu.

Koartilak adierazteko Q letra erabiliko dugu.

Koantilen artean dauden erlazioak aztertzeko, ondorengo eskemaz baliatuko gara.





Eskema horiek aztertuz, ondorengo identitateak azaltzen zaizkigu.

$Q_1 = P_{25}$

$Q_1 = P_{50} = Med = D_5$

$Q_3 = P_{75}$

Bestalde:

$D_1 = P_{10}$
$D_2 = P_{20}$
$D_3 = P_{30}$
.
$D_5 = P_{50}$
.
.
.
$D_{10} = P_{100}$

KALKULUAK

Lehen esan dugun bezala K-mailako pertzentila, bere azpian ehuneko K-ko banaketa uzten duen zenbakizko balioa da.

K mailako pertzentila kalkulatzeko, ondorengo formula hau erabiliko dugu:

$$P_k = L_i + b_i \frac{N*k/100 - F_i}{n_i}$$

non : i = medianadun klasea
N = laginaren maiztasuna
Li = medianadun klasearen behemuga erreala
b_i = medianadun klasearen tarte-zabalera
n_i = medianadun klasearen maiztasun absolutua
F_i = Medianadun klasearen behemuga errealari dagokion maiztasun metatua
k = Pertzentilaren maila

X_i (tartea)	M_i
80000 baino gut.	2
80000 - 90000	4
90000 - 100000	6
100000 - 110000	4
110000 - 120000	2
120000 - 140000	1
140000 baino geh.	1

40 mailako pertzentilak, kasu honetan bere ezkerretara 8 pertsona uzten ditu. Beraz hirugarren tartean egongo da. Hau kontutan harturik, honela kalkulatuko dugu:

$$P_{40} = 90000 + 10000 \frac{20*40/100 - 6}{6} =$$

$$P_{40} = 90000 + 10000 * \frac{8 - 6}{6} = 93333.33$$

Bigarren problema hau izango da: zenbat pertsonak irabazten duten 112.000 pta. baino gutxiago. Kasu honetan puntuen balioa ezaguna da, eta puntu honi dagokion zentila edo pertzentila jakin behar dugu, hau da, formulan agertzen den "k" balioa.

$$P_k = L_i + b_i \frac{N*k/100 - F_i}{n_i}$$

$$112.000 = 110.000 + 10.000 * \frac{20*k/100 - 16}{2}$$

eta hemendik, k balioa askatu behar dugu:

$$112.000 - 110.000 = 10.000 * \frac{20*k/100 - 16}{2}$$

$$\frac{2.000 * 2}{10.000} = 20*k/100 - 16 = 2'4$$

$$0'4 + 16 = 20*k/100 \;;\; 16'4 * 100/20 = k \qquad \underline{k = 82}$$

Honek zera esan nahi du: % 92 pertsonak, 112.000 pta. baino gutxiago irabazten dutela.

3.- SAKABANATZE-NEURRIAK

Orain arte, zentralizazio-neurriak erabiliz, lagin baten erdialdeko puntuak nola kalkulatzen diren ikusi dugu. Neurri hauen bitartez laginaren elementu gehienen posizioa jakingo dugu. Baina laginaren gehiengoa ez da lagin osoa. Eman dezagun ondoko bi lagin hauek dauzkagula:

X: 3,4,5,6,7
Y: 1,3,5,7,9

Zentralizazio-neurriak erabiliz garbi ikusten da bi laginen erdiko puntuak berdinak direla.Hau da:

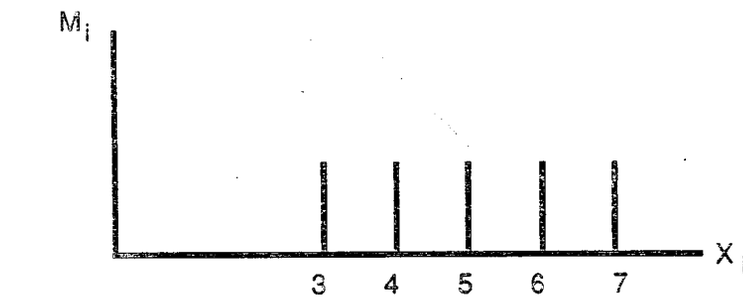
$\overline{X} = 5$; X-en mediana = 5

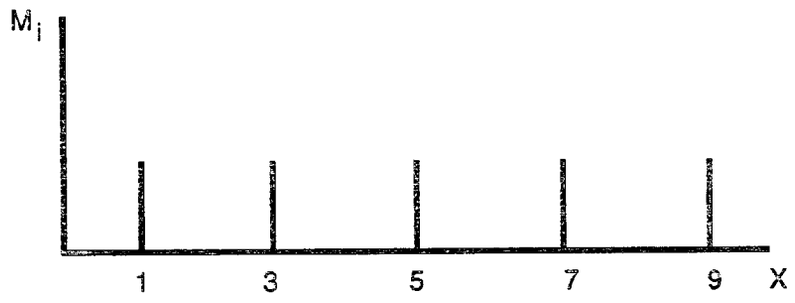
$\overline{Y} = 5$; Y-ren mediana = 5

Baina Y lagineko balioen arteko tarteak, X laginekoenak baino handiagoak dira. Hau da, Y lagineko balioak X-ekoak baino aldendu edo sakabanatuagoak daude.

Datuen arteko tarteak haundiak edo txikiak diren adierazteko, SAKABANATZE--NEURRIAK erabiliko ditugu.

Ondoko adierazpen grafikoetan ere garbi ikusten da Y lagineko balioak askoz ere sakabanatuagoak daudela. Beste era batera esanda, X lagineko balioak askoz ere batuagoak daude.





Ikusi dugu laginak zentralizazio-neurrien bidez definitzea ez dela nahikoa, zeren eta bestela aurreko bi kasuak berdinak direla esango bait genuke. Beraz garbi dago beste neurri batzuk beharrezkoak direla. Neurri berri hauek Sakabanatze-neurriak izango dira.

Definizio bezala zera esan dezakegu: datu-multzo batean muturreko balioak erdiko balioetatik aldentzen diren neurriaren zenbatekoa ematen duen estatistikoari Sakabanatzea deritzogula.

Sakabanatzea neurtzeko erabiliko ditugun neurriak hauek dira:

1.- Bariantza

2.- Desbidazio standarda

3.- Aldakuntz koefizientea

4.- Koartil arteko ibiltarterdia.

1.- BARIANTZA

Bariantza ondoko era honetan defini dezakegu: Puntu eta puntu hauen batezbestekoen arteko distantziaren karratuen batezbestekoa.

Kalkuluak egiterakoan, aurreko definizioak adierazten dituen pausoak emango

ditugu. Lehenik, puntu bakoitzetik batezbestera dagoen distantzia kalkulatuko dugu, ondoren diferentzia hauen berredura lortuko dugu, eta azkenik berredura hauen batezbestekoa kalkulatzeko, berreduren batuketa egingo dugu, eta azken pausuan batura hau N lagineko datu-kopuruaz zatituko dugu.

X aldagaien bariantza ondoko formularen bidez kalkula daiteke:

$$S^2x = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N} = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots\ldots + (X_n - \overline{X})^2}{N}$$

Eman dezagun bi datu-multzo ditugula, bakoitzaren batezbestekoen balioa 5 delarik. Bariantzaren kalkuluak ondoko hauek dira:

X : 4,5,6 $\qquad$ $Y = \overline{5}$

Y : 1,5,9 $\qquad$ $Y = \overline{5}$

$$S^2x = \frac{(4-5)^2 + (5-5)^2 + (6-5)^2}{3} = \frac{1+0+1}{3} = \frac{2}{3} = 0'66$$

$$S^2y = \frac{(1-5)^2 + (5-5)^2 + (9-5)^2}{3} = \frac{16+0+16}{3} = \frac{32}{3} = 10'66$$

X aldagaiaren bariantza, Y aldagaiaren bariantza baino txikiagoa dela ikus dezakegu, eta hori hain zuzen, X aldagiaren datuak euren artean Y aldagaiarenak baino hurbilago daudelako gertatzen da.

Bariantzako kalkuluak erraztuko dizkiguten beste zenbait formula badaude. Horien artean ondoko hau azpimarra daiteke:

$$S^2x = \frac{\sum_{i=1}^{n} X_i^2}{N} - \overline{X}^2 = \frac{X_1^2 + X_2^2 + \ldots\ldots\ldots + X_n^2}{N} - \overline{X}^2$$

Batezbestekoen kalkuluen kasuan bezala, bariantzaren kasuan ere, datuak elkarturik daudenean, formula berriak erabiliko ditugu, nahiz eta orain arte ikusi ditugun formula guziak baliagarriak izan.

Batezbestekoen kalkuluak egin ditugun adibideari berriz heltzen badiogu, ondoko maiztasun-banaketa ematen digu, bertan "X_i" aldagaien balioak adierazteko erabiliko dugu, eta M_i, maiztasunak adierazteko:

X_i	M_i	$M_i * X_i$
0	8	0
1	7	7
2	4	8
3	4	12
4	3	12
5	3	15
6	1	6
	30	60

$$\overline{X} = \frac{60}{30} = 2$$

Kasu honi dagokionez,ondoko bi formula erabiliko ditugu; hain zuzen bariantzaren kalkuluak egiterakoan sinpleenak direnak.

$$S^2x = \frac{\sum_{i=1}^{n} M_i*(X_i-\overline{X})^2}{N} = \frac{M_1*(X_1-\overline{X})^2+M_2*(X_2-\overline{X})^2+....+M_n*(X_n-\overline{X})^2}{N}$$

$$\text{non } N = \sum_{i=1}^{n} M_i$$

$$S^2x = \frac{\sum_{i=1}^{n} M_i*X_i^2}{N} - \overline{X}^2 = \frac{M_1*X_1^2 + M_2*X_2^2 + ... + M_n*X_n^2}{N} - \overline{X}^2$$

$$\text{non } N = \sum_{i=1}^{n} M_i$$

Bariantzaren edozein formulatako kalkuluan dagoen prozesua antzekoa da. Elkarren segidako zutabeak eraikitzen datza, zeinak zenbakitzaile eta izendatzaileen emaitzak emango bait dizkigu.

Hala nola, lehenengo formula erabili nahi izanez gero, adibidean aipatzen den ordenaren arabera zutabeak eraiki beharko ditugu, aurrez batezbestekoa kalkulaturik eduki beharko dugula kontutan harturik. Gure kasuko batezbestekoa 2 dela kontutan hartzen badugu:

X_i	M_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$M_i * (X_i - \overline{X})^2$
0	8	- 2	4	32
1	7	- 1	1	7
2	4	0	0	0
3	4	1	1	4
4	3	2	4	12
5	3	3	9	27
6	1	4	16	16
	30			98

Aldagaiaren bariantza hau izango da:

$$S^2x = \frac{\Sigma M_i * (X_i - \overline{X})^2}{N} = \frac{98}{30} = 3'2666$$

Bigarren formula erabiltzen badugu, kalkuluaren prozesua ondoko hau izango da, beti ere jadanik kalkulaturiko batezbestekoak 2 balio duela kontuan izanik:

X_i	M_i	$X_i * M_i$	$X_i^2 * M_i$
0	8	0	0
1	7	7	7
2	4	8	16
3	4	12	36
4	3	12	48
5	3	15	75
6	1	6	36
	30	60	218

$$S^2x = \frac{\Sigma M_i * X_i^2}{N} - \overline{X}^2 = \frac{218}{30} - 2^2 = 7'2666 - 4 = 3'2666$$

2.- DESBIDAZIO STANDARDA edo DESBIDAZIO TIPIKOA

Desbidazio standarda edo desbidazio tipikoa, bariantzaren erro karratu gisa defini dezakegu.

Sakabanatze-neurrietatik gehien erabiltzen dena desbiderapen standarda da. Hain zuzen, bere propietate estatistikoengatik eta sujetuen arteko distantziaren batezbestekoen errepresentazioari bariantzak baino askoz ere zuzenago erantzuten diolako. Gogora dezagun bariantza distantzien karratuen batezbestekoa dela. Aldiz erro karratutik sorturiko distantzia baino ez da desbidazio tipiko edo desbidazio standarda.

Kalkuluen prozesua beti berdina izango da: lehenengo datuen bariantza kalkulatuko dugu, zeinak ondoren desbidazio tipikoa kalkulatzeko balioko bait du.

Jakina da, bariantzaren kasuan bezala, datuak zenbat eta aldenduago egon desbidazio tipikoa handiagoa izango dela.

3.- ALDAKUNTZ KOEFIZIENTEA

Batzuetan gertatzen da bariantzaren balioa, eta ondorioz desbidazio tipikoaren emaitzak, distortsionaturik egotea; adibidez, datuak neurtzeko erabili dugun unitateagatik. Esate baterako, bi multzo ezberdinen barruko aldakortasuna parekatu nahi badugu, eta multzo hauen datuak bi unitate ezberdinetan neurtuta badaude, desbidazio standarda edo desbidazio tipikoa ez da baliagarria izango.

Eman dezagun ondoko bi datu-multzo hauek dauzkagula, baina bi eskala ezberdinetan neurtuta, hau da, lehenengo taldeko notak, 0-tik 10-era joango dira, eta bestekoak 0-tik 100-era joango dira.

A taldearen notak (emaitzak): 7, 6, 4, 5, 9, 5, 3, 8, 8, 5

B taldearen notak (emaitzak): 30,40,50,50,50,60,70,80,80,90

Talde bakoitzaren batezbestekoa eta bariantza kalkulatzen badugu, zera lortuko dugu:

$\overline{X}_a = 6$ $\qquad$ $\overline{X}_b = 60$

$S_a^2 = 3.4$ $\qquad$ $S_b^2 = 340$

$S_a = 1.84$ $\qquad$ $S_b = 18'4$

Argi ikusten da B taldeko emaitzaren desbidazio tipikoa, A taldeko elementuen desbidazioa baino hamar aldiz handiagoa dela. Bestalde bi taldeen puntuazioak berdinak direla argi ikusten da, baina ordena eta eskala ezberdinetan ezarrita.Talde bakoitzeko emaitzak ordenatuz gero eta A taldearen emaitzak bider 10 egin ondoren, B taldearen emaitzak edo puntuazioak lortuko ditugu. Logikoki B taldearen desbidazio tipikoa, A taldearen desbidazio tipikoa baino 10 aldiz handiagoa da. Horren esanahia garbia da: B taldeko datuen arteko tarteak edo distantziak A taldeko datuen artekoak baino 10 aldiz handiagoak direla .Baina bi taldeetan dauden tarte edo distantzia ezberdinak nahikoa fiktizio edo itxurazkoak dira. Honen arrazoia zera da: irakasleek notak jartzerakoan eskala ezberdinak erabiltzea.

Tarte edo distantzia hauen homogenotzea lortzeko ALDAKUNTZ KOEFIZIENTEA erabiltzen da:

$$A.K. = \frac{S}{\overline{X}} * 100$$

Talde bien A.K berdina dela erraz ikusi dezakegu:

$$\text{A.K. (A taldea)} = \frac{1'84}{6} * 100 = 30'6666$$

$$\text{A.K. (B taldea)} = \frac{18'4}{60} * 100 = 30'6666$$

A.K. erabiliz, bi aldagaien edo multzoen datuen aldakortasuna konpara dezakegu, nahiz eta eskala ezberdinetan egon.

4.- KOARTIL ARTEKO IBILTARTERDIA

Nahiz eta gehien erabiltzen diren sakabanatze-neurriak lehen aipatutakoak izan, batzuetan beste bi hauek ere erabiltzen dira: ibiltartea eta Koartil arteko ibiltarterdia.

Ibiltartearen definizio bezala ondoko hau har genezake "Ordenatutako datuetatik, handienaren eta txikienaren arteko kendura". Balio honek bere garrantzia du, dauzkagun datuak edozein arrazoi dela eta tartetako eskala batera pasatu nahi ditugunean.

Gehienbat erabiltzen den neurria Koartil arteko ibiltarterdia, Q, izango da. Batez ere maiztasun-banaketa asimetrikoa denean erabiltzen da.

Q, 3 ordenako koartilen eta 1 ordenako koartilen arteko distantziaren erdia bezala defini dezakegu.

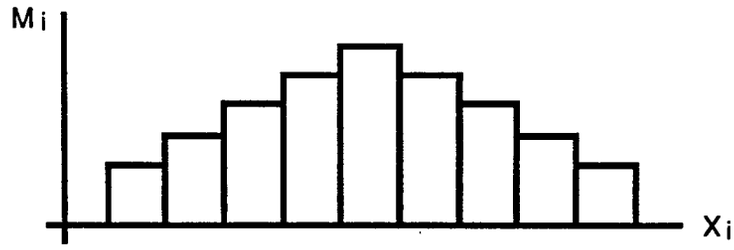
$$Q = \frac{Q_3 - Q_1}{2}$$

4.- ASIMETRI ETA KURTOSI-NEURRIAK

Nahiz eta Estatistika deskribatzailean zentralizazio- eta sakabanatze-neurriak garrantzitsuenak izan, batzuetan oso inportantea da gerta daitekeen asimetri- eta kurtosi-koefizienteen kalkulua. Honen arrazoia zera da: batzutan datuak egokitu egin behar zaizkio dira estatistikako edozein banaketa teorikori, eta egokitze hau egiteko beharrezkoa izango da gure banaketen asimetria eta zorroztasunak ezagunak izatea.

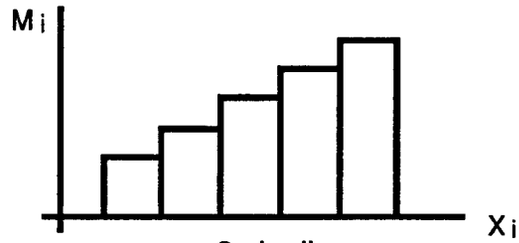
Asimetri indizeak erabiliko ditugu, guk dauzkagun datuak simetria orokor bati noraino hurbiltzen zaizkien jakiteko.

Hona hemen hiru adibide:



1. Irudia

2. Irudia



3. Irudia

Lehenengo errepresentazioa simetrikotzat har genezake; beste biak ez. Lehenengo kasuan argi ikusten dugu moda, mediana eta batezbestekoa berdinak direla.

Bigarren kasuan (2. irudia), asimetria positiboa dago, hirugarrenean (3. irudia) berriz datuen asimetria negatiboa da. Aurreko adierazpen grafikoak ikusi ondoren, asimetria positiboa era honetara defini dezakegu: puntuazio baxuak ugarienak direnean dagoena da. Kasu honetan batezbestekoa baino txikiagoak diren datuak asko dira, eta noski, batezbestekoa baino handiagoak direnak gutxiago. Honen ondorioz moda eta mediana, batezbestekoa baino txikiagoak izango dira. Asimetria negatiboan fenomeno berbera gertatzen da, baina aurkako zentzuan.

Asimetri koefizientea kalkulatzeko formulak ugari dira. Hauetako bat ikusiko dugu, eta formula honen bidez jakingo dugu gure datuak simetriko edo asimetrikoak zenbateraino diren:

$$A_3 = \frac{\dfrac{\sum_{i=1}^{n} m_i * (x_i - \bar{x})^3}{\sum_{i=1}^{n} m_i}}{S^3}$$

A_3 asimetri koefizienteen emaitza honela ulertzen da:

- A_3, 0 baino handiagoa bada, dauzkagun datuen asimetria positiboa izango da.

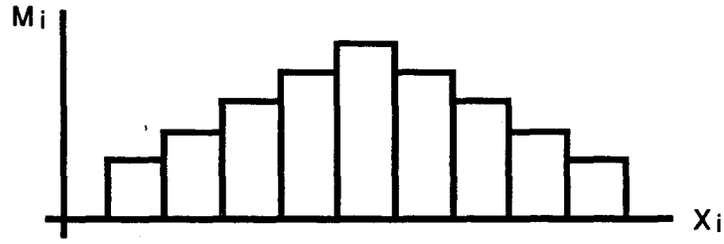
- A_3, 0 baino txikiagoa bada,dauzkagun datuen asimetria negatiboa izango da.

- A_3, 0 bada, datuen banaketa simetrikoa izango da.

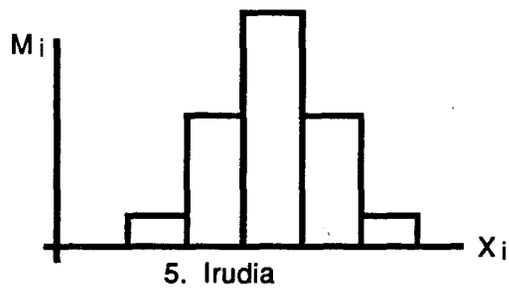
0-tik zenbat eta hurbilago egon Asimetri koefizientea, gure datuen banaketa simetrikoagoa izango da.

Gai honetan ikusiko dugun azken neurria, kurtosi edo zorroztasunena izango da. Estatistika teorikoan garrantzirik handiena daukan banaketa, banaketa Normala da. Banaketa hau hurrengo kapitulu batean aztertuko dugu. Indize honen helburua, gure datuen zorroztasuna banaketa normalen zorroztasunarekin konparatzea da.

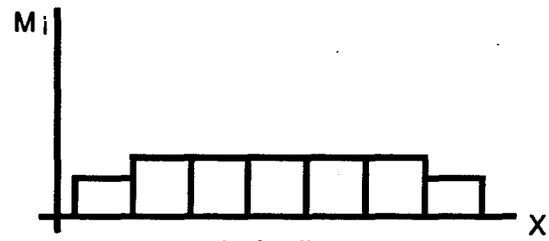
Ondoko irudiei begiratuz, garbi ikus dezakegu banaketen zorroztasunak ezberdinak direla.



4. Irudia



5. Irudia



6. Irudia

4. irudiko banaketen zorroztasuna maila ertainekoa dela ikus dezakegu, 5. irudiko banaketa aurrekoa baino zorrotzagoa da eta 6. irudikoa zorroztasun txikiagokoa.

4 irudiko banaketari, MESOKURTIKOA deritzogu, 5 irudikoa banaketa LEPTO-KURTIKOA eta 6 irudikoa PLATIKURTIKOA izango dira.

Asimetrian gertatzen den bezala, irudikapen grafikorik egin gabe kurba normalaren zorroztasun teorikoarekiko dituen desbidazioak jakiteko, formula ugari daude. Guk ikusiko dugun formula ondoko hau da:

$$A_4 = \frac{\dfrac{\sum_{i=1}^{n} m_i \ (x_i - \bar{x})^4}{\sum_{i=1}^{n} m_i}}{S^4} - 3$$

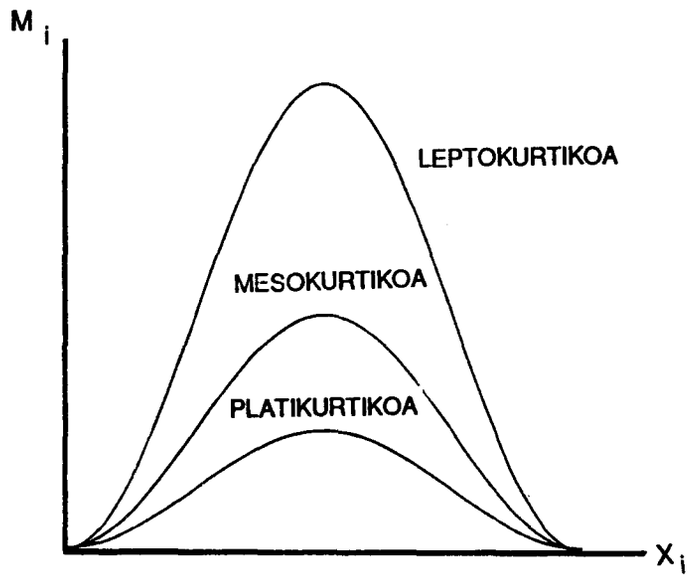
Estatistika teorikoak kurba edo banaketa normal batean $A_4 = 0$ dela frogatzen du.

Horregatik A_4 kurtosi koefizienteaaren balioen adierazpena hau izango da:

A_4 0 baino handiagoa bada, kurba leptokurtikoa da.

A_4 0 baino txikiagoa bada, kurba platikurtikoa da.

A_4 0 bada kurba mesokurtikoa da.



5.- *BANAKETA BIDIMENTSIONALAK*

Orain arte aldagai bakar bat kontsideratu dugu, eta beraren zentralizazio-, sakabanatze-, asimetri eta kurtosi-neurriak kalkulatu ditugu.

Hemendik aurrera, gai honetan bi aldagai batera kontsideratuko ditugu.Har ditzagun bi aldagai hauek: sexua eta abortuari buruzko eritzia.Bi aldagaiak batera analizatu nahi baditugu, bi sarrerako ondorengo taularen antzeko batekin hasiko gara.

	gizonak	emakumeak	
Abortuaren alde	75	80	155
Abortuaren kontra	45	30	75
	120	110	230

Kasu honetan aldagai biak batera azter ditzakegu, hau da, esate baterako, abortuaren alde dauden gizonezkoen portzentaia, abortuaren alde dauden emakumezkoen portzentaiarekin konpara dezakegu. Orokorki hartuta, abortuaren alde eta abortuaren kontra daudenak ere konpara ditzakegu.

Honela definitutako bi sarrerako taulari maiztasun bidimentsionaleko banaketen taula edo maiztasun bateratuen banaketa deituko diogu.

Kontsidera dezagun ondoko adibide hau: 1000 familiak dituzten seme-alaben kopurua eta beren etxeetan egunean zehar batezbeste telebista piztuta egoten den ordu-kopurua. Ondorengoaren antzeko maiztasunen taula bidimentsionala izango genuke.

Seme-alabak \ TB	Ordu 1	2 ordu	3 ordu	4 ordu	5 ordu	
0 seme-alaba	25	55	100	35	20	235
1 seme-alaba	40	65	125	80	40	350
2 seme-alaba	30	45	50	85	55	265
3 seme-alaba	5	10	15	15	25	70
4 seme-alaba	0	5	10	20	25	60
4 seme-alaba	0	0	5	5	10	20
	100	180	305	240	175	1000

Aurrekoa bezalako taula badugu, ordu-kopuruen batezbestekoa eta seme-alaben kopurua batezbestekoa kalkulatu ahal izango dugu, hau da, aldagai baten batezbestekoa kalkula dezakegu beste aldagaiarekin gertatzen denaz arduratu gabe.

Beraz, seme-alaben kopurua edo telebista-ordu aldagaiari buruz independenteki hitz egin genezake, eta ondorengo bi taula hauek bezalakoak atera ditzakegu:

Seme-alabak	Mi
0	235
1	350
2	265
3	70
4	60
5	20
	1000

Teleb.	Mi
1	100
2	180
3	305
4	240
5	175
	1000

Aurreko maiztasun-taula bakoitzari BAZTER-BANAKETAREN MAIZTASUN-TAULA deituko diogu. Banaketa marginalen taula ere deituko diogu. Bazter-banaketa hauetan batezbestekoa, bariantza eta oro har aldagai bakoitzarentzat aztertu ditugun estatistika deskribatzailearen indize guztiak kalkulatu ahal izango ditugu.

MAIZTASUN-BANAKETA BALDINTZATUAK.

Orain arte egindako kalkuluak maiztasun baldintzatuetara zabal ditzakegu. Adibidez seme-alaba bakarra duten familien kasuarentzat, indize desberdinen kalkulua plantea dezakegu, eta horrela banaketa bidimentsionalaren zati batekin gelditukо ginateke. Era berean, telebista ordubetean piztuta duten familien seme-alaben batezbestekoa kalkulatzea komenigarria izan daiteke. Kasu hauetan balioak eta famili kopuruak baldintzatu egiten ditugu. Familia guztien artean jarritako baldintzak betetzen dituztenak bakarrik konsideratzen ditugu.

Azaldu ditugun adibideetan, ondorengo maiztasun-banaketarekin lan egingo genuke:

Seme-alabak \ TB	Ordu 1	2 ordu	3 ordu	4 ordu	5 ordu	
1 seme-alaba	40	65	125	80	40	350

Telebista \ Se-al	0 se-al	1 se-al	2 se-al	3 se-al	4 se-al	5 se-al	
ordu 1	25	40	30	5	0	0	100

4. GAIA
KOERLAZIOA ETA ERREGRESIOA

Aurreko ikasgaian, aldagai aleatorio batek isolaturik hartuta zituen ezaugarriak aztertu ditugu. Ikasgai honetan bi aldagairen arteko harremanak aztertuko ditugu. Aldagaien arteko harreman edo erlazio hauek, KOERLAZIO-KOEFIZIENTE izendatuko ditugun indizeen bidez neurtuko ditugu.

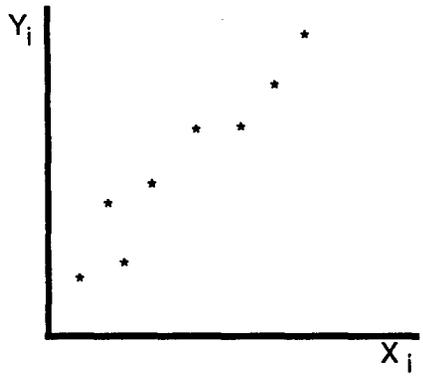
Argi dago, pertsonen altuera eta pisuaren artean lotura bat dagoela. Era berean, begibistakoa dirudi ikasleen adimen-koefizientea eta eskolako errendimenduaren arteko erlazioa. Zentzu honetan, itxuraz behintzat erlazioa duten aldagai-pare bat, edozeinek imajina dezake nahi duen arloan.

Aldagaien arteko harreman posibleak aztertzen saiatzerakoan, ondoko oinarrizko bi puntu bereiztu behar ditugu: Alde batetik aldagaien arteko harreman-maila aztertzea interesatzen zaigu, eta bestalde erlazio honen norantza. Lehenengo zati honetan ezin dezakegu kausa/efektu erlazioaz hitz egin. Horrela hitz egin ahal izateko, aldagai baten emaitzen erregresio edo aurresateari buruz hitz egin beharko genuke, beste aldagaiaren emaitzen arabera.

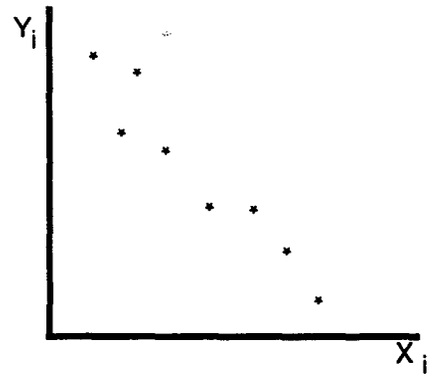
Bi aldagai erlazionaturik dauden bakoitzean ezin dugu esan bat bestearen kausa edo ondorio zuzena denik, kasu askotan erlazio hau eremu teorikoan parte hartzen duten beste aldagai batzuek ere baldintza dezaketelako.

Hasiera batean, erlazio zuzenak edo positiboak eta alderantzizkoak edo negatiboak bereiztu behar ditugu. Erlazio bat zuzena edo positiboa izango da aldagai bati zenbat eta balio handiagoak eman, baldin eta besteak ere gero eta balio handiagoak hartzen baditu. Adibidez, pisuaren eta altueraren arteko erlazioa positiboa izango litzateke "zenbat eta pisu gehiago altuera handiagoa" betetzen bada. Erlazioa alderantzizkoa edo negatiboa izango da, baldin eta aldagai batean zenbat eta balio handiago lortu bestean are eta txikiagoa lortzea espero bada. Adibidez, ikasle batek zenbat eta idazmen- eta irakurmen-arazo gehiago izan orduan eta errendimendu akademiko baxuagoa lortuko duela pentsa genezake.

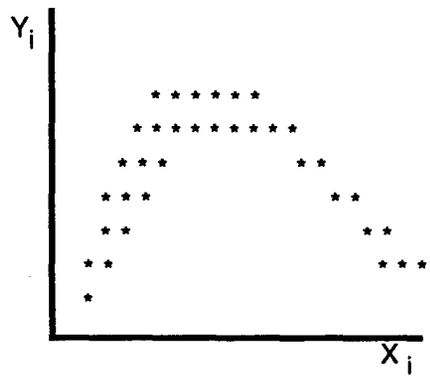
Ondoko adierazpen grafikoak aztertuz gero, lehenengoan erlazio positiboa eta bigarrengoan erlazio negatiboa ikusi ahal izango ditugu.



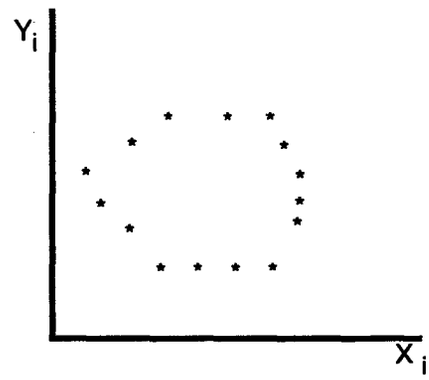
1. irudia



2. irudia



3. irudia



4. irudia

3. eta 4. irudietan ez dugu X eta Y aldagaien artean erlazio linealik ikusten. 3. irudian ordea, aldagaien artean nolabaiteko erlazio kurboa dagoenaz ohar gaitezke. 4. grafikak berriz, itxuraz behintzat, aldagaien artean erlaziorik ez dagoela pentsarazten digu.

Koerlazio-koefizienteek ere ikur positiboa edo negatiboa izango dute. Koerlazio-koefizientearen ikurra bat etorriko da erlazio-motarekin. Horrela, koerlazio-koefiziente positiboak erlazio positiboa adieraziko digu eta koefiziente negatiboak berriz, erlazio negatiboa. Koerlazio-koefizientea 0 edo 0-tik gertu badago, aldagaiak askeak direla adieraziko digu, hau da, aztertu ditugun datuek ez dutela, hasieran behintzat, batabestearekin zerikusirik.

Koerlazio-koefiziente batek har ditzakeen mutur-balioak +1 eta -1 dira. +1 balioak koerlazio-koefizienteak aldagaien arteko erabateko erlazioa adierazten digu, eta aldi berean positiboa dela. -1 balioak koerlazio-koefizienteak ere aldagaien artean erabateko erlazioa azaltzen du, baina negatiboa.

1. KOERLAZIO-KOEFIZIENTEAREN KALKULUA

Aldagaien artean dagoen harreman-maila kalkulatzea posible egiten duten formulak ugari dira. Erabilienak bakarrik aztertuko ditugu.

Bi aldagaien arteko koerlazio-koefizientea kalkulatzerakoan, kontutan izan behar dugu zein motatako aldagaiak diren. Mota hauen arabera 4 kasu ezberdin ikusiko ditugu.

1.1. Aldagaiak kuantitatiboak dira

Aldagaiak kuantitatiboak direnean, koerlazio-koefizientea kalkulatzeko ohizko formula "Pearson-en koerlazio-koefizientea" da. Formula hau, bere modu sinpleenean, era honetara azal daiteke:

$$R_{xy} = \frac{\frac{\Sigma XY}{N} - \bar{X}*\bar{Y}}{\sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}\ \sqrt{\frac{\Sigma Y^2}{N} - (\bar{Y})^2}}$$

Ondorengo adibide honen bidez nola kalkula dezakegun ikusiko dugu.

Eman dezagun bi aldagai hauek ditugula:

X_i : kurtso hasieran egindako test batetan, 10 ikaslek lortu dituzten emaitzak.
Y_i : Ikasleen kurtso-bukaerako notak.

X_i	Y_i	X^2	Y^2	$X*Y$
15	6	225	36	90
14	5	196	25	70
16	7	256	49	112
12	4	144	16	48
16	7	256	49	112
10	5	100	25	50
20	8	400	64	160
15	7	225	49	105
12	5	144	25	60
20	6	400	36	120
150	60	2346	374	927

$$\bar{X} = \frac{150}{10} = 15$$

$$\bar{Y} = \frac{60}{10} = 60$$

$$R_{xy} = \frac{\frac{\Sigma XY}{N} - X*Y}{\sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}\ \sqrt{\frac{\Sigma Y^2}{N} - (\bar{Y})^2}} = \frac{\frac{927}{10} - 15*6}{\sqrt{\frac{2346}{10} - 15^2}\ \sqrt{\frac{374}{10} - 6^2}} =$$

$$R_{xy} = \frac{92'7 - 90}{\sqrt{234'6 - 225}\ \sqrt{37'4 - 36}} = \frac{2'7}{\sqrt{9'6}\ \sqrt{1'4}} = \frac{2'7}{3'66} = 0'7364$$

Koerlazio-koefizientea lortu ondoren, gure kasuan 0,7364, emaitza honen esanahia interpretatu beharra dago. Koerlazio-koefiziente hau ikusi ondoren bi aldagai hauen artean harreman handia dagoela esan dezakegu, 1 baliotik oso hurbil bait dago.

Aldagaien arteko harremana positiboa da, hau da kurtso hasierako testean zenbat eta emaitza hobea lortu, hainbat eta nota hobea espero da kurtso bukaerako notetan.

Gai honen bukaeran mugatze-koefizientea ikusterakoan, aldagaien arteko harremanen interpretazioa hobeto ikusiko dugu.

1.2. Aldagai bat kuantitatiboa eta bestea dikotomikoa edo dikotomizatua direnean: koerlazio-koefiziente biserial-puntuala

Aldagai dikotomikoak edo dikotomizatuak giza zientziei egokitutako estatistikan asko erabiltzen dira: edozein item-i ondo edo gaizki erantzun, sexua.....

Nahiz eta aldagai dikotomiko edo kuantitatiboen harremana neurtzeko Pearson-en koerlazio-koefizientea egokia izan, erabiltzerakoan askoz ere errazagoak diren beste zenbait formula asmatu dira. Formula berrien emaitza interpretatzerakoan baliagarriak dira orain arte ikusitako pauso guztiak; formula berri hauek hasierako formulan oinarrituak bait daude.

Formula hauek aipatzerakoan gehien erabiltzen direnak koerlazio biserial-puntualaren koefizientea eta koerlazio biserialaren koefizienteak direla esan dezakegu.

Koerlazio-koefiziente biserial-puntuala kalkulatzeko ondoko formula hau erabiliko dugu, X aldagaia kuantitatiboa eta Y aldagaia dikotomikoa dela suposatuz:

$$R_{bp} = \frac{\bar{X}_p - \bar{X}_q}{S_t} \sqrt{p*q}$$

non:

X_p : Y aldagaiak 1 balioa hartzen duten elementuetan, X aldagaien batezbestekoa.
X_q : Y aldagaiak 0 balioa hartzen duten elementuetan, X aldagaien batezbestekoa.
p : 1 balioen maiztasun erlatiboa.
q : 0 balioen maiztasun erlatiboa.
St : X aldagaien desbiderapen standarda.

Adibide bat ikusiko dugu: Eman dezagun X eta Y bi aldagai, era honetan definituak dauzkagula.

X_i : Kurtso hasieran egindako test batean, 10 ikaslek lortutako emaitzak.
Y_i : 0 eta 1 balioak hartzen dituen aldagaia; kurtso bukaeran matematika-arloa gainditu duten ala ez adierazten duena.

X_i	M_i	X^2
15	1	225
14	0	196
16	1	256
12	0	144
16	1	256
10	0	100
20	1	400
15	1	225
12	0	144
20	1	400
150		2346

Formula aplikatzerakoan lehenengo pausoa X_p eta X_q kalkulatzea izango da:

$$\overline{X}_p = \frac{15 + 16 + 16 + 20 + 15 + 20}{6} = \frac{102}{6} = 17$$

$$\overline{X}_q = \frac{14 + 12 + 10 + 12}{4} = \frac{48}{4} = 12$$

$$\overline{X} = \frac{150}{10} = 15 \qquad p = 0'6 \qquad q = 0{,}4$$

$$S_t^2 = \frac{\Sigma X^2}{N} - (\overline{X})^2 = \frac{2346}{10} - 15^2 = 234'6 - 225 = 9'6$$

eta hemendik: $S_t = \sqrt{9'6} = 3'0986$

$$R_{bp} = \frac{\overline{X}_p - \overline{X}_q}{S_t} \sqrt{p*q} = \frac{17 - 12}{3'0986} \sqrt{0'6 * 0'4} =$$

$$R_{bp} = \frac{5}{3'0986} * 0'4898 = 0'7903$$

Koerlazio-koefiziente honen interpretazioa eta Pearson-en formulen emaitzen interpretazioa berdinak dira. Kasu honetan aldagaien arteko erlazioa positiboa dela esan dezakegu. Honek zera esan nahi du: pertsona batek X aldagaian zenbat eta puntuazio handiagoa hartu, orduan eta puntuazio handiagoak hartuko dituela Y aldagaian, hau da, kurtso hasieran pasatutako testen emaitza zenbat eta handiagoa izan, hainbat eta errazago izango da matematika aztergaia gainditzea, zeren eta Y aldagaiak errazago hartuko bait du 1 balioa.

1.3. Aldagai biak dikotomikoak dira: Φ Kerlazio-koefizientea edo koerlazio-laukoitza.

Ikusten dugun hirugarren kasu hau ere Pearson-en koerlazio-koefizienteen formularen ondorio edo aurreramendu zuzen bat da. Beraz lehen azalduriko interpretazio-arauak baliagarriak izango dira.

Φ-ren formula, aldagai biak dikotomikoak direnean erabiliko dugu. Eman dezagun 500 pertsonei test psikometriko bat pasatu diegula eta test honen lehenengo eta bigarren galderen emaitzen artean ea harremanik dagoen jakin nahi dugula. Horretarako, ondoren azaltzen dugun bezala, gure datuak sarrera bikoitza duen taula batean jarriko ditugu.

Taula honetan ikus dezakegunez, 100 pertsonak item biak ondo erantzun dituzte, 200 pertsonak lehenengoa gaizki eta bigarrena ondo, 80 pertsonak lehenengo itema ondo eta bigarrena gaizki, eta azkenik 120 pertsonen emaitzak gaizki daude bai lehenengo galderan eta bai bigarrenean.

	Item 1 Ondo 1	Item 1 Gaizki 0	
Item 2 Ondo 1	100 (a)	200 (b)	300 (a+b)
Item 2 Gaizki 0	80 (c)	120 (d)	200 (c+d)
	180 (a+c)	320 (b+d)	

Φ, koerlazio-koefizientea kalkulatzea posible egiten duen formula ondoko hau da:

$$\Phi = \frac{a*d - b*c}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \; ; \text{ Gure kasuan}$$

$$\Phi = \frac{100*120 - 200*80}{\sqrt{(300)(200)(180)(320)}} = \frac{12000 - 16000}{58787'75} = -0'0680$$

Interpretazioa aurreko bi kasuen berdina izango da. Item bien erantzunek ez dute zerikusi handirik; koerlazio-koefizientea 0 baliotik oso hurbil bait dago.

Ondoren beste adibide bat jarriko dugu eta kasu honetan ikusten denez itemen erantzunek zerikusirik badaukate, zeren eta item 3ari ondo erantzun dioten 220 pertsonak hartzen baditugu, hauetatik 200 pertsonak 4. itemari ere ondo erantzun diotela ikusten dugu. Aldi berean item 3ari gaizki erantzun dioten 260 pertsona kontutan hartzen baditugu, hauetatik 240 pertsonak 4. itemari ere gaizki erantzuten diotela ikusten dugu.

	Item 3 Ondo 1	Item 3 Gaizki 0	
Item 4 Ondo 1	200 (a)	40 (b)	240 (a+b)
Item 4 Gaizki 0	20 (c)	240 (d)	260 (c+d)
	220 (a+c)	280 (b+d)	

Kasu honetan kalkuluak hauek dira:

$$\Phi = \frac{200*240 - 40*20}{\sqrt{(240)(260)(220)(280)}} = \frac{48000 - 800}{61998'70} = 0'7613$$

Ikusten dugun Φ koefizienteak, aldagai bien arteko harremana egiaztatzen du.

4. Aldagai biak ordinalak dira: "ρ " Spearman-en koerlazio-koefizentea.

Batzuetan, gehienetan neurketa-tresna egokiak ez izateagatik, aldagaiari buruz daukagun informazio bakarra datuen ordenazioa da. Zenbait kasutan, datuek bete behar dituzten baldintza batzuk direla eta, nahiz eta informazioa galdu kuantitatiboak izan

daitezkeen datuak eskala ordinal batera aldatzea komeniko zaigu.

Hau da orain aztertuko dugun kasua: aldagai biak ordinalak direnekoa.

X_i	Y_i	d_i	d_i^2
1	8	7	49
2	7	5	25
3	6	3	9
4	9	5	25
5	4	1	1
6	3	3	9
7	10	3	9
8	5	3	9
9	2	8	64
10	1	8	64
			264

Kasu hauetan erabiliko dugun formula ondoko hau izango da:

$$\rho = 1 - \frac{\Sigma d^2}{N(N^2 - 1)}$$

non d_i = ordenen diferentzia den.

Formula hau ere, Pearson-en formularen ondorio da, eta noski, formula honen bidez lortutako emaitzen interpretazioa, aurrekoen berdina izango da.

Gure kasuan:

$$\rho = 1 - \frac{\Sigma d^2}{N(N^2 - 1)} = 1 - \frac{264}{10(100 - 1)} = 1 - \frac{264}{990} = 0{,}7333$$

Konklusio gisa, aldagaien edo ordenen artean badagoela harremanik, eta harreman hau positiboa dela esan dezakegu.

Orain aztertuko dugun adibidean, bietatik bat kuantitatiboa izango da eta bestea ordinala. Eman dezagun Xi, irakasle batek bere eritziaren arabera ordenatutako hamar ikasle direla, eta Y pertsona horiek matematika-arloan kurtso bukaeran izan zituzten puntuazioak.

1	8
2	7
3	6
4	8
5	5
6	5
7	5
8	7
9	4
10	6

Kasu honetan, koerlazio-koefizientearen kalkuluak egiten hasi baino lehen, aldagai kuantitatiboa ordenatu beharra daukagu. Adibidez, puntuazio handienetatik txikienetaraino ordena dezakegu.

Kontutan hartu behar dugu puntuazio batzuk berdinak direla. Kasu hauetan, beraien ordena bezala tokatzen zaizkion ordenen batezbestekoa ezarriko diogu. Gure kasuan 8ko bi puntuazio dauzkagu, eta kasu hauei lehenengoa eta bigarrena dagozkie. Beraz biei 1,5 emango diegu, eta hurrengo puntuazioari 3 zenbakia, ezarriko diogu.

Horrela gure datuak, Y aldagaia ordenatu eta gero ondoko taulan agertzen direnak izango dira.

X_i	Y_i	d_i	d_i^2
1	1'5	0'5	0'25
2	3'5	1'5	2'25
3	5'5	2'5	6'25
4	1'5	2'5	6'25
5	8	3	9
6	8	2	4
7	8	1	1
8	3'5	4'5	20'25
9	8	1	1
10	5'5	4'5	20'25
			70'50

$$R_o = 1 - \frac{6 * \Sigma d^2}{N (N^2 - 1)}$$

$$R_o = 1 - \frac{6 * 70'50,}{10 (100 - 1)} =$$

$$R_o = 1 - \frac{423}{990} = 1 - 0{,}4272 = 0{,}5728$$

Aurreko kasuetan bezala, emaitzaren interpretazioa beharrezkoa da. Kasu honetan ordenazioen arteko harremana positiboa dela egiazta dezakegu. Aldagai batean ordena edo postu ona izanda beste aldagaian ere postu ona espero da, eta alderantziz, lehenengo aldagaian pertsona batek azken postua badauka, bestean ere azken postua espero da.

2. *ERREGRESIO LINEALA*

Erregresioaren analisia, koerlazio-printzipioaren bidez, menpeko aldagai batean (aldagai dependente batean) beste aldagai aske edo independente batzuek eragiten dituen efektuak eta efektu hauen magnitudeak ikasteko edo ikertzeko metodo bat da. Koerlazio--koefizientea kalkulatzetik erregresio-ekuazioa kalkulatzera, oso salto kualitatibo handia dago. Koerlazioaren kasuan, aztertzen duguna zera da: aldagaien artean ia harremanik dagoen ala ez. Erregresioen kasuan, ostera, aldagai aske edo independenteen aldaketak menpeko (dependente) aldagaien aldaketak nola sortzen dituen aztertzen dugu. Eta azkenik, erregresioaren funtsezko helburua, ezaguna den aldagai independentearen bidez aldagai dependenteen aurresana egitea izango da.

Esan beharra dago erregresioaz hitz egiteko aldagaiek kuantitatiboak izan behar dutela.

Orain arte ez dugu aldagai independente (aske) eta dependenteez hitz egin. Aldagai dependentea, aztertu nahi dugun aldagaia izango da. Aldagai hau aztertzeko beste aldagai bat edo batzuk erabiliko ditugu, aldagai hauek aldagai aske edo independenteak izango dira. Aldagai askea bakarra denean ERREGRESIO SINPLEA deritzogu, eta aldagai askeak bat baino gehiago direnean ERREGRESIO ANIZKOITZA izango da.

Erregresioaren analisia ez da aldagaien arteko harremanak aztertzeko bakarrik erabiltzen; funtsezko beste hiru atal ere baditu:

.- Menpeko aldagai eta aldagai aske edo askeak hobeto erlazionatzen dituen funtzio matematikoa aztertzea.

.- Aurreko puntuan aipatutako funtzioa jakin ondoren, funtzioaren parametroak kalkulatu, eta egindako doikuntza edo egokitzearen hurbiltasuna kalkulatzea.

.- Aldagai askearen datuak kontutan harturik, menpeko aldagaien aurresanak egin eta aurresan hauen baliozkotasuna kalkulatu.

Aipatutako lehenengo puntua ez dugu aztertuko liburu honetan, jarritako helburuak gainditzen dituelako. Kasu guztietan, aldagaien arteko erlazio lineala dela onartuko dugu. Beraz, erregresio sinple lineala bakarrik ikusiko dugu gai honetan.

ERREGRESIO SINPLE LINEALA

Edozein zuzenen ekuazioa adierazteko modu bat hau da:

$$Y = a + bX$$

Gure arazoa hau izango da: gure datu multzoari (X_i, Y_i) gehien hurbiltzen zaion zuzenaren parametroak kalkulatzea. Hain zuzen "a" eta "b" parametroen balioak kalkulatzea. Hurbiltze hau "karratu txikien" zentzuan hartuko dugu, hau da, gure zuzena ondoko baldintza hau betetzen duena izango da: Puntuetatik zuzenera dauden distantzien karratuen batura minimoa egiten duena. Beste era batera esan dezakegu: aurresanen erroreak minimoak egiten dituena. Zuzen honi, ERREGRESIO-ZUZEN deituko diogu.

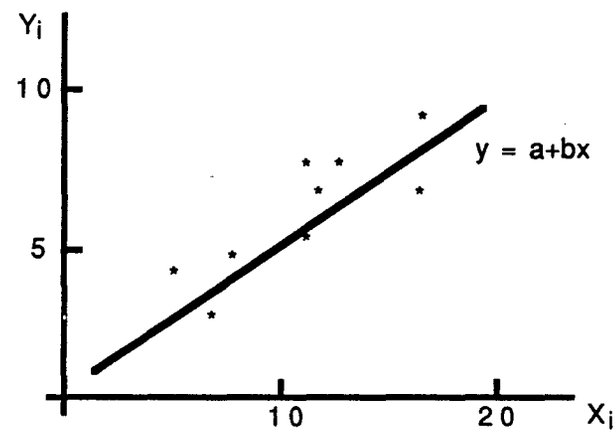
Kapitulu honetan aztertu dugun lehenengo adibidera bueltatuz, gogora dezagun X eta Y aldagaien arteko harremana aztertzen genuela. Aldagaiak hauek dira:

X_i : kurtso hasieran egindako test batean, 10 ikaslek lortutako emaitzak.
Y_i : Ikasle berdinen kurtso bukaerako notak.

Bagenekien aldagaien arteko erlazioa nahikoa handia zela, eta lortutako koerlazio-koefizientea R = 0.73649 zen.

Erregresioaren bitartez, kurtso bukaeran matematika-arloan ikasleak lortuko dituen notak aurresateko posibilitateak analizatzen saiatuko gara. Erregresio zuzenaren ekuazioak, aldagai dependentearen balioa aurresaten lagunduko digu; gure kasuan kurtso bukaerako matematikako notak aldagai independenteen arabera, hau da, testean lortutako puntuaziok kontutan harturik.

X_i	Y_i
15	6
14	5
16	7
12	4
16	7
10	5
20	8
15	7
12	5
20	6
150	60



Erregresio zuzenaren parametroak kalkulatzeko, ondoko formularen bidez egingo dugu:

$$b = \frac{N \sum X_i * Y_i - (\sum X_i)(\sum Y_i)}{N \sum X_i^2 - (\sum X_i)^2} \quad : \text{ERREGRESIO-KOEFIZIENTEA}$$

$$a = \overline{Y} - b\overline{X}$$

Azaltzen dugun adibidean, emaitzak hauek izango lirateke:

X_i	Y_i	X^2	X*Y
15	6	225	90
14	5	196	70
16	7	256	112
12	4	144	48
16	7	256	112
10	5	100	50
20	8	400	160
15	7	225	105
12	5	144	60
20	6	400	120
150	60	2346	927

$$b = \frac{N \sum X_i * Y_i - (\sum X_i)(\sum Y_i)}{N \sum X_i^2 - (\sum X_i)^2}$$

$N = 10$

$\sum X_i * Y_i = 927$

$(\sum X_i)(\sum Y_i) = 150 * 60 = 9000$

$\sum X_i^2 = 2346$

$(\sum X_i)^2 = 150^2 = 22500$

Beraz, gure kasuan erregresio-koefizientea hau izango da:

$$b = \frac{N \sum X_i * Y_i - (\sum X_i)(\sum Y_i)}{N \sum X_i^2 - (\sum X_i)^2} = \frac{10 * 927 - 9000}{10 * 2346 - 22500} = \frac{270}{960} = 0'28125$$

$$a = \overline{Y} - b\overline{X} = \frac{60}{10} - 0'28125 * \frac{150}{10} = 6 - 0'28125 * 15 = 1'78125$$

Aurreko kalkuluen bidez, datu-multzoei gehien hurbiltzen zaien zuzenaren ekuazioa,

hau da, erregresio-zuzenen ekuazioa hau da:

$$Y' = 1'78125 + 0'28125\,X$$

Erregresio-ekuazioa ezagutuz, menpeko aldagaien aurresanak erraz lortuko ditugu, hau da, pertsona batek X aldagaian 10 puntu lortu baditu, Y aldagairako egingo diogun aurresana, ekuaziotik zuzenean lortuko dut, X aldagaien tokian bere balioa jarriz.

Gure ekuazio honetan:

$$Y' = 1'78125 + 0'28125\,X \;;\; X = 10 \text{ bada } Y' = 1'78125 + 0'28125 * 10$$

Eta hemendik, gure aurresana:

$$Y' = 1'78125 + 2'8125 = \underline{4'59375}$$

Hurrengo galdera erraz agertzen da: noraino da baliagarria kasu honetan lortu dugun aurresana? Kontutan hartu behar dugu Estatistikan edozein balioren estimazio edo aurresana egiterakoan estimazioa bezain garrantzitsua zera dela: estimazioen baliozkotasuna, edo beste era batera esanda, aurresaterakoan egiten ari garen erroreari marjina jartzea. Azken batean edozein estimazio edo inferentziak, bere baliozkotasunen ezaugarriak batera eraman behar ditu.

Kasu honetan, hau da, erregresioa erabiltzen dugunean, aurresanaren baliotasuna kalkulatzea, zuzenen bidez puntuarekiko lortutako hurbiltasuna kalkulatzea izango da, hau da, zenbat eta hurbiltasun hobea lortu, errore txikiagoa edukiko dugu, eta gure aurresanen baliozkotasuna handiagoa izango da.

Lortutako hurbiltasuna kalkulatzeko, aztertuko ditugun distantziak edo diferentziak Y eta Y'-ren artekoak izango dira, Y aldagaiaren balio izanik, eta Y' pertsona horien aurresanak (beren X aldagaien balioak ikusiz lortutakoak). Distantzia hauek, erregresio--zuzenaren bidez egin ditzaketen aurresanak benetazko balioei asko ala gutxi hurbiltzen zaizkien ikusteko lagunduko digute. Gure kasuan, egingo nituen aurresanen eta lortuko nituen distantzien zutabeak egin ondoren, beste bi zutabe lortuko ditugu. Bat, aurrerago erabiltzeko, distantzien karratuekin osatuta, eta azkena, laugarrena, Y^2 aldagaien balioekin osatutakoa.

$$Y' = 1'78125 + 0'28125\,X$$

X_i	Y_i	Y'_i = 1'78125 + + 0'28125 X	di = Y_i - Y'_i	d_i^2	Y_i^2
15	6	6'00000	0'00000	0'00000	36
14	5	5'71875	0'71875	0'51660	25
16	7	6'28125	0'71875	0'51660	49
12	4	5'15625	1'15625	1'33691	16
16	7	6'28125	0'71875	0'51660	49
10	5	4'59375	0'40675	0'16544	25
20	8	7'40675	0'59375	0'35253	64
15	7	6'00000	1'00000	1'00000	49
12	5	5'15625	0'15625	0'02441	25
20	6	7'40675	1'40675	1'97894	36
150	60			6'40803	374

Konprobatu ahal izango dugunez Y_i' (aurresanak) puntuazioak ez dira bat etortzen behatutako Y-ren puntuazioekin. Bat etorriko dira aldagaien erlazioa orokorra denean bakarrik, hau da, koerlazio-koefizientea +1 edo -1 denean. Daukagun errorea, AURRESANEN ERRORE edo AKATS TIPIKOA erabiliz kalkulatuko dugu. Argi dago, aldagaien artean zenbat eta harreman handiagoa izan koerlazio-koefizientea hainbat eta handiagoa izango dela. Baina aldagaien artean harremana handiagoa bada, honek zera esan nahi du: (X_i, Y_i) puntuak, zuzen batetik egongo direla, hau da, puntuak zuzen batera gehiago hurbiltzen direla. Honek erregresio zuzenen bidez lortzea dudan hurbiltasuna handiagoa izango dela esan nahi du, hau da, errorea txikiagoa izango dela.

AURRESANEN ERRORE TIPIKOA, beraz honela definituko dugu:

$$S_{ee} = \sqrt{\frac{\Sigma d_i^2}{N}} \qquad d_i = Y_i - Y_i' \quad \text{izanik.}$$

Gure kasuan $$S_{ee} = \sqrt{\frac{\Sigma d_i^2}{N}} = \sqrt{\frac{6'40803}{10}} = 0'80050$$

Askotan beste formula hau erabiltzen da:

$$S_{ee} = \sqrt{\frac{\Sigma d_i^2}{N-2}}$$

Arrazioa zera da: bigarren formula honek propietate estatistiko hobeak dituela, baina arazo hau beste gai batean sakonduko dugu.

Lehen aipatu dugu koerlazio koefizientearen eta erregresio zuzena erabiliz egiten ditugun erroreen artean badagoela zerikusirik. Bien arteko erlazioa, ondoko formularen bidez adieraziko dugu:

$$\frac{S^2_{ee}}{S^2_y} = 1 - R^2_{xy} \qquad (***)$$

non: S^2_{ee} aurresanen errore tipikoen berredura

S^2_y Y aldagaien bariantza

R^2_{xy} koerlazio-koefizienteen berredura diren.

Interesantea da, formula honetan koerlazio-koefizienteen berredura azpimarratzea; baita MUGATZE-KOEFIZIENTE izendatua eta aurresanen errore tipikoen arteko harremana ere.

R = +1, edo R = -1 bada, badakigu aldagaien arteko erlazioa totala edo orokorra dela, eta balio hauek goiko formulan ordezkatuz, hau lortuko dugu:

$$\frac{S^2_{ee}}{S^2_y} = 1 - 1 = 0\,; \quad \text{hemendik zenbakitzailea: } S^2_{ee} = 0\,.$$

Honek, koerlazio-koefizientea ± 1 baldin bada, erregresio-zuzenen bidez lortzen dugun hurbiltasuna orokorra dela esaten digu, hau da, errorea 0 izango dela.

R = 0 bada, aldagaien artean ez dago inongo harremanik. Balio hau aurreko formulan ordezkatuz, lortuko duguna zera da:

$$\frac{S^2_{ee}}{S^2_y} = 1 - 0 = 1\,; \quad \text{hemendik zenbakitzailea: } S^2_{ee} = S^2_y$$

eta honek erregresio zuzenak puntuarekiko lortzen duen hurbiltasuna nulua dela esaten digu, zeren eta errore karratuen balioa eta Y aldagaien bariantzak berdinak bait dira.

Gure adibidean $S^2_y = \frac{374}{10} - 6^2 = 37'4 - 36 = 1'4$

$S^2_{ee} = 0'64080$

$R^2_{xy} = 0'54228$

Hemendik lortuko dugu:

$$\frac{S^2_{ee}}{S^2_y} = \frac{0'64080}{1'4} = 0'4577 \ ,$$

eta noski, balio hau eta $1 - R^2_{xy}$ balioak berdinak dira. Lehengo formulak esaten zuen bezala, gure kasuan: $1 - 0'54228 = 0'4577$.

(***) formulak adierazten duen harremana ikusirik, koerlazio-koefizienteen berredura, MUGATZE-KOEFIZIENTE deitzen dena, eta honen interpretazioa era honetara defini dezakegu: Menpeko (Y) aldagaien bariantzaren portzentai zatia, X aldagaiak esplikatzen duena da. Gure kasuan zera esan dezakegu: kurtso amaierako matematikazko noten aldakortasunen % 54 esplikatzen duela kurtso hasieran pasatu dugun testean lortu duten puntuazioak.

5. GAIA
PROBABILITATEA

Kaleko eguneroko hizkuntzan, askotan entzuten dira honelako esaerak: "Irabazteko probabilitateak oso handiak dira", "Bihar, mendira joateko probabilitatea dut".

Esaera hauek erabiltzen ditugunean, gure buruan daukagun ideia bat egiten da matematika- edo teoria-mailan probabilitateari buruz erabiltzen denarekin. Hala ere probabilitatearen kontzeptu estatistikoa sistematizatu beharra dago.

Probabilitate-kontzeptuaz hitz egin baino lehen, gertaera eta saiakuntza aleatorioari (zorizkoari) buruz hitz egin beharra dago.

Eman dezagun dato bat airera botatzen dugula. Guk ez dakigu begibistan agertuko den aurpegiak zenbat puntu edukiko dituen, baina badakigu sei posibilitate hauetatik bat agertuko dela: puntu bat, bi puntu, ..., edo sei puntu. Hau da saiakuntza aleatorio baten lehenengo garrantzizko propietatea. Bigarrena, aleatorioa edo zorizkoa izatea da. Honek ez du esan nahi alternatiba guztien probabilitateak berdinak izango direla, baizik eta zoriak parte hartzen duela emaitza bat lortzerakoan. Bi aldiz saiakuntza berdina egiten badugu, ez da derrigorrezkoa emaitza berdina lortzea. Hauek dira edozein esperimentu edo saiakuntza aleatoriok bete behar dituen baldintzak. Azpimarratu beharra dago aleatorio eta zorizko hitzak estatistikan esangura berdina daukatela.

Saiakuntza aleatorio bati lotuta GERTAERA ALEATORIOAK daude. Gertaera, edozein saiakuntzatan, edozein emaitza posible izango da.

Hurrengo kontzeptu berria LAGIN-ESPAZIOA izango da. Lagin-espazioa saiakuntza aleatorio baten emaitza guztiek osatzen duten multzoa da. Hau da, edozein gertaera, lagin-espazioen azpimultzo bezala kontsideratu ahal izango dugu. Dato bat airera botatzerakoan, gertaera posible bat zenbaki bikoitia ateratzea izango da, seikoa ateratzea beste gertaera bat izango da, eta noski biak lagin-espazioaren azpimultzoak izango dira, zeren eta kasu honetan lagin-espazioa ondoko hau da: $\{1,2,3,4,5,6\}$.

Gertaera aleatorioak definituta eduki ondoren, hauen sailkapena egingo dugu. Sailkapen honek probabilitatearen kalkuluarekin zerikusia dauka, eta honek bere garrantzia ematen dio.

.- GERTAERA SEGURUA: Gertaera bat segurua izango da, lagin-espazioarekin bat datorrenean. Hau da, gertaera hau burutzea segurua denean.

.- GERTAERA EZINEZKOA: Gertaera seguruen kontrakoa da; ez da posible gertaera hau burutzea.

.- GERTAERA BATERAEZINAK (INKONPATIBLEAK): Izenak esaten duen bezala, bi gertaera bateraezinak izango dira, biak batera gertatu ezin badira. Lehenengoa gertatzen bada, bigarrena ezin da gertatu. Adibidez datu bat airera botatzerakoan, 1 zenbakia lortzea eta 2 zenbakia lortzea bateraezinak izango dira.

.- GERTAERA BATERAGARRIAK: Bi gertaera bateragarriak izan daitezen bete behar duten baldintza bakarra, biak batera gertatzeko posibilitatea egotea da. Hau da, dato bat airera botatzerakoan, zenbaki bikoitia lortzea eta 4 baino zenbaki txikiagoa lortzea posible izango da. Baldintza hau betetzen dutelako bi gertaera hauek bateragarriak izango dira.

.- KONTRAKO GERTAERAK: Bi gertaera kontrakoak izango dira, bien artean gertaera segurua osatzen dutenean. Adibidez, dato bat airera botatzerakoan, zenbaki bikoitia lortzea eta zenbaki bakoitia lortzea kontrako gertaerak izango dira. Kontutan eduki behar dugu kontrako gertaerak, bateraezinak ere izango direla, baina gertaera bateraezinak ez direla beti kontrakoak izango.

.- GERTAERA ASKEAK (INDEPENDENTEAK): Bi gertaerak aske izateko bete behar duten baldintza hau da: bien artean ez dago inongo harremanik. Adibidez, bi dato airera botatzen baditugu, lehenengo eta bigarren datoko emaitzen artean ez dago erlaziorik.

.- MENPEKO GERTAERAK (DEPENDENTEAK): Bi gertaera menpekoak izango dira, lehenengoen gertaerak bigarrenena baldintzatzen badu. Hau da, dato bat airera botatzen badugu, eta lehenengo gertaera 3 edo 3 baino txikiagoa izatea eta bigarrena parea izatea baldin bada, lehenengo gertaera betetzen bada bigarrena baldintzatu egiten du.

1.- PROBABILITATEAREN DEFINIZIOAK

Probabilitate-kontzeptua definitzean, aplikazio-alorren arabera bi era ezberdinetan erabiliko ditugu.

Lehenengoak, probabilitatearen definizio klasiko deiturikoak, hau esaten du: gertaera baten probabilitatea, aldeko kasu eta kasu posibleen arteko proportzioan datza. Hau da A gertaeren probabilitatea honela definituko dugu:

$$P(A) = \frac{\text{Aldeko kasuak}}{\text{Kasu posibleak}} .$$

Adibidez karta-pilotik urre bat ateratzearen probabilitatea ondoko hau izango da:

$$P(A) = \frac{\text{Aldeko kasuak}}{\text{Kasu posibleak}} = \frac{10}{40} .$$

Batzuetan probabilitate-definizio hau, planteamendu batzuk argitzeko ez da nahikoa izango. Adibidez, hau galdetzen badigute: Zein da kalean topatzen dugun lehenengo pertsonak gabardina eramateko dagoen probabilitatea? Argi eta garbi dago kasu posibleen kopurua zenbatzea ez dela posible; ezta aldeko kasuen kopurua ezagutzea ere.

Bigarren definizioa edo probabilitate estatistikoaren definizioa, aurreko problemak argitzeko baligarria izango zaigu (non teorikoki, lagin-espazioa imajinaezina edota infinitua izan daitekeen). Probabilitatearen bigarren definizio honek, bere sorrera maiztasun erlatiboen kontzeptuan dago. Probabilitate hauek (probabilitate enpiriko ere deituak) momenturarte behatutako gertaerak etorkizunean iraungo duelakoan oinarritzen dira.

2.- PROBABILITATEAREN PROPIETATEAK

Probabilitatea, edozein kasutan 0 eta 1 balio tartean kokatzen da. Gertaera baten probabilitatea 0-koa izango da EZINEZKO gertaera denean, hots, aldeko kasuen kopurua 0 denean. Gertaera baten probabilitatea 1-ekoa izango da gertaera segurua denean, hots, aldeko kasuen kopurua eta kasu posibleena berdinak direnean.

GERTAEREN EBAKETA: PROPIETATEAK

A eta B gertaerak aldiberean agertzeko dagoen probabilitatea kalkulatzerakoan 3 kasu ezberdin aztertu beharko ditugu:

1) Gertaera bateraezinak dira: Kasu horretan eta aldiberean gerta ezin direnez, A eta B gertaeren ebakiduren probabilitatea 0 izango da.

2) Gertaera askeak badira, hau da, baten gertatzeak ez du eraginik bestearen gertatzearekin, gertaeren ebakiduren probabilitatea A eta B gertaeren probabilitateen biderkadura izango da.

$$P(A \text{ eta } B) = P(A \cap B) = P(A) * P(B)$$

Adibidez karta-pilotik bat atera eta irudidun urrea izatearen probabilitatea kalkulatzea nahi badugu, bi gertaeren ebakiduren probabilitatea kalkulatzea nahi dugu. Kasu horretan gertaera biak hauek izango dira; A: Urrea izatea eta B: Irudia izatea. Problema hau bi era ezberdinetan erabaki dezakegu:

a) $P\,(\text{irudidun urrea}) = \dfrac{3 \text{ aldeko kasu}}{40 \text{ kasu posible}} = \dfrac{3}{40}$

b) $P\,(\text{irudidun urrea}) = P\,(\text{irudia}) * P\,(\text{urrea}) = \dfrac{12}{40} * \dfrac{10}{40} = \dfrac{120}{1600} = \dfrac{3}{40}$

3) Menpeko gertaerak badira, hau da, baten gertatzeak bestearen gertatzearekin zerikusia dauka, ebakiduren probabilitatea ondoko hau izango da: Lehenengo gertaeraren probabilitatea bider bigarren gertaera gertatzeko dagoen *probabilitate baldintzatua*, lehenengo gertaera gertatua dela badakigularik.

$$P\,(A \text{ eta } B) = P\,(A \cap B) = P\,(A) * P\,(B/A)$$

P (B/A)-rekin zera ulertuko dugu: A gertaera gertatu dela jakinik, B gertaera gertatzeko dagoen probabilitatea.

Adibidez oraingo kasuan karta-pilotik bi karta elkarren segidan aterako ditugu, lehenengoa karta-pilora bueltatu gabe. Lehenengo karta urrea eta bigarrena kopa izatearen probabilitatea kalkulatu nahi badugu, honela jokatuko dugu:

$$P\,(\text{1.a urrea eta 2.a kopa}) = P\,(\text{1.a urrea}) * P\,(\text{2.a kopa/1.a urrea}) =$$
$$= \frac{10}{40} * \frac{10}{39} = \frac{100}{1560}$$

Atera dugun lehenengo karta urrea izan bada eta bueltatu ez dugunez, bigarrena ateratzerakoan 39 karta dauzkagu piloan, eta haietatik 10 kopa direla kontutan hartu behar dugu.

Aurreko esperimentu bera egin nahi badugu, baina bigarren karta ere urrea izanik, probabilitatea honela kalkulatuko dugu:

$$P\,(\text{1.a urrea eta 2.a urrea}) = P\,(\text{1.a urrea}) * P\,(\text{2.a urrea/1.a urrea}) =$$
$$= \frac{10}{40} * \frac{9}{39} = \frac{90}{1560}$$

Kasu honetan lehenengo karta urrea izanik, bigarren karta urrea izatearen probabilitatea 9/39-koa izango da; 39 karta gelditzen bait dira eta haietatik 9 urreak dira, lehenengo karta ere urrea izan delako.

GERTAEREN BILKETAREN PROBABILITATEA: A edo B gertaerak gertatzearen probabilitatea ondoko hau izango da: bi gertaeren probabilitatea ondoko hau izango da: bi gertaeren probabilitatearen batuketa ken gertaeren ebaketaren probabilitatea. Beraz, bi gertaeratik gutxienez bat gertatzearen probabilitatea hau izango da: gertaera bakoitzaren probabilitatearen batuketa ken bi gertaerak aldi berean gertatzeko probabilitatea.

$$P(A \text{ edo } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

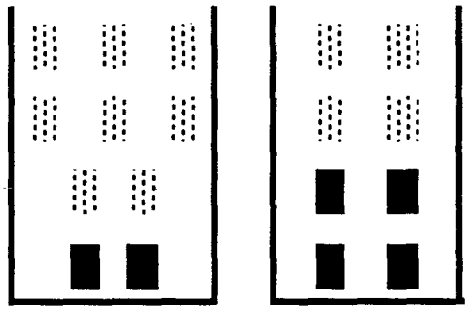
Adibidez, karten esperimentuan ateratzen dugun karta urrea edo irudidun karta izatearen probabilitatea kalkulatu nahi badugu:

$$P(\text{urrea edo irudia}) = P(\text{urrea}) + P(\text{irudia}) - P(\text{urrea eta irudia})$$

$$P(\text{urrea edo irudia}) = \frac{10}{40} + \frac{12}{40} - \frac{3}{40} = \frac{19}{40}$$

PROBABILITATE TOTALAREN TEOREMA

Demagun bi kutxa dauzkagula K_1 eta K_2. Lehenengoan, K_1-ean 8 bola zuri eta 2 beltz dauzkagu eta K_2-an 4 bola zuri eta 4 beltz dauzkagu.



Proposatzen den problema honako hau da: edozein kutxatatik ateratzen dugu bola bat, zein kutxatatik atera dugun jakin gabe eta bola hau zuria izateko dagoen probabilitatea kalkulatu nahi da. Era honetako probabilitateak kalkulatzeko egingo dugun planteamendua ondokoa izango da: bi alternatiba egon daitezke atera dugun bola zuria izan dadin.

a) K_1 kutxa izan da aukeratua. Kasu honetan probabilitatea 8/10-eko izango zen.

b) K_2 kutxa izan da aukeratua. Kasu honetan probabilitatea 4/8-eko izango zen probabilitatea.

Beraz hurrengo planteamendua egin dezakegu:

Bola zuria bi eratara lor dezakegu:

K_1 kutxa aukeratu eta bola zuria atera

edo

K_2 kutxa aukeratu eta bola zuria atera

$$P\,(\text{bola zuria}) = P\,(K_1 \text{ eta b. bola zuria}) + P\,(K_2 \text{ eta b. zuria})$$

$$P\,(K_1 \text{ eta b. zuria}) = P\,(K_1) * P\,(\text{b. zuria}/K_1) = \frac{1}{2} * \frac{8}{10} = \frac{8}{20}$$

$$P\,(K_2 \text{ eta b. zuria}) = P\,(K_2) * P\,(\text{b. zuria}/K_2) = \frac{1}{2} * \frac{4}{8} = \frac{4}{16}$$

Hemendik honako probabilitatea lortuko dugu:

$$P\,(\text{zuria}) = \frac{8}{20} + \frac{4}{16} = \frac{208}{320} = \frac{52}{80}$$

Orokorrean, probabilitate totalaren teoremak esaten diguna hauxe da: B_1, B_2, B_3, ... eta Bk, gertaerak badauzkagu, zeintzuen batura edo bildura gertaera segurua baldin bada, eta bi edozein gertaera hauen ebakidura hutsa baldin bada, hots, elkar baztertu egiten badute, A gertaera baten probabilitatea honela kalkula dezakegu:

$$P\,(A) = P\,(A \cap B_1) + P\,(A \cap B_2) + \ldots\ldots\ldots + P\,(A \cap B_k)$$

Eman dezagun tresna psikometriko bat daukagula eta % 90 kasuetan irakurmen/idazmen-arazoak detektatzen dituela ikasle batek arlo horretan benetako arazoak dituenean. Bestetik, test honek irakurmen/idazmen-arazoak ez eduki arren, frogaren ezaugarriengatik edo ikasleen erantzunen despisteagatik % 5 problematiko bezala eman dezake. Badakigu ari garen adinaren ikasleen artean % 20 irakurmen/idazmen-arazoa daukatela. Planteatzen ari garen arazoa hurrengoa da: ikasle bat zoriz aukeratu badugu, ikasle honek testean atera dituen ondorioetan irakumen/idazmen-arazoak dituela guri pentsarazteko dagoen probabilitatea kalkulatzea.

Kasu honetan bi kutxen kasuan bezala, bi ikasle-mota posibleei test baten bitartez arazoak detekta diezazkiekegu: benetan arazoak dituzten ikasleei, "problematikoak" deituko diegunei, eta despisteagatik edo beste gauzengatik teorikoki dagozkien baino emaitza okerragoak eduki dituztenei (hauei "arruntak" deituko diegu).

Beraz, testak arazoak detekta ditzan probabilitatea bi zatitan deskonposa dezakegu: ikaslea "problematikoa" edo "arrunta" denaren funtzioz.

P (arazoak detektatu) = P (problematikoa eta arazoak detektatu) +
P (arrunta eta arazoak detektatu)
(P problemat. eta arazoak det.) = P (problemat.) * P (araz.det/proble)
(P problemat. eta arazoak det.) = 0'20 * 0'90 = 0'18.
P (arrunta eta arazoak detekt.) = P (arrunta) * P (arz. det/arrunta)
P (arrunta eta arazoak detekt.) = 0'80 * 0'05 = 0'04

Eta hortik:

P (arazoak detektatu) = 0'18 + 0'04 = 0'22.

3. BAYES-EN TEOREMA

Analizatu dugun kasuan ikertzailearen arazo nagusia ez litzateke zoriz aukeratua izan den pertsona bati, testean arazoak detektatzeko dagoen probabilitatea zein den jakitea izango; baizik eta testuak arazoak detektatu ondoren ikasle hau benetan "problematikoa" izateko dagoen probabilitatea zein den ikertzea.

Arazo hau konpontzeko bidea Bayes-en teoremaren aplikazioa da. Lagin-espazioa bi gertaera osagarritan deskonposatzen dugunean, aipaturiko teorema formula honen bidez aplikatuko dugu:

$$P(B_1/A) = \frac{P(B_1) * P(A/B_1)}{P(B_1) * P(A/B_1) + P(B_2) * P(A/B_2)} \qquad (1)$$

Lagin-espazioa k gertaeraren B_1, B_2, B_3, ..., B_k, osagarritan deskonposatzen dugunean, eta gertaera hauek beren artean bateraezinak direnean, Bayes-en teorema beste formula honen bidez aplikatuko dugu:

$$P(B_1/A) = \frac{P(B_1) * P(A/B_1)}{P(B_1) * P(A/B_1) + P(B_2) * P(A/B_2) + \ldots\ldots + P(B_k) * (A/B_k)}$$

Hasiera batean planteatu dugun galdekizuna ondoko hau zen: pertsona batek test batean lortutako emaitzagatik problematiko bezala kalifikatu, edo arazo batzuk detektatu ondoren, pertsona hau benetan "problematikoa" izateko dagoen probabilitatea zein den jakitea.

Arazo hau honela eskematiza dezakegu:

P (problematiko/arazo.detektatu)

Gure problema, Bayes-en (1) formula erabiliz aska dezakegu.

$$P\,(PROB/araz.det) = \frac{P\,(PROBLEMAT) * P\,(araz.\ det/PROBLEMATIKOA)}{P(PROB) * P(araz.de/PROB) + P(ARRUN) * P(araz.de/ARRUN)}$$

$$P\,(PROB/araz.det) = \frac{0'20 * 0'90}{0'20*0'90 + 0'80*0'05} = \frac{0'18}{0'18 + 0'04} =$$

$$= \frac{0'18}{0'22} = 0'8181$$

Emaitza honekin jarritako galderari erantzuna ematen diogu: tresna honek ikaslearengan irakurmen/idazmenaren arazoak detektatu baditu, % 81'81 kasuetan pertsona hau benetako "problematikoa" izango da. Beste kasuetan, hau da % 18'19etan erabilitako tresnak detektatu dituen kasu hauek, ikaslearen despisteagatik edo erabilitako tresnaren erru posibleagatik gertatuko lirateke.

6. GAIA
ALDAGAI ALEATORIOAK. PROBABILITATE ETA BANAKETA DISKRETUAK

Esperimentu aleatorio bat definitzen dugunean bi gauzaz ari gara hitz egiten. Alde batetik gerta daitezkeen gertaerak zeintzuk direnaz eta bestalde gertaera hauen probabilitateaz.

Dato bat airera botatzen badugu, badakigu dauzkan sei aldeetatik bat irtengo dela eta alde bakoitzak irteteko duen probabilitatea 1/6 dela. Beste kasu bat ikusten badugu, adibidez, txanpon bat airera botatzen dugunean, badakigu aurpegia edo gurutzea irtengo zaigula eta kasu bakoitzaren probabilitatea 1/2 izango dela.

Aldagai aleatorioaren definizio klasikoa zera da: "Aldagai aleatorioa, esperimentu aleatorio baten emaitzen unibertso posiblean definituriko funtzio erreal bat da". Definizio hau, honela esplika dezakegu: esperimentu aleatorio bateko gertakizun posibleei balio bat esleitu behar diegu eta beraien probabilitatea ere adierazi beharko dugu.

Balio-multzoari eta beraien probabilitatei aldagaien dentsitate-funtzio edo probabilitate-funtzio deitzen zaio. Lehen aipatu ditugun bi adibideetan, bi aldagai aleatorio hauek defini ditzakegu:

Txanpon bat botatzen dugunean

GERTAERA	BALIOA	PROBABILITATEA
AURPEG.	1	p = 0'50
GURUTZ.	0	q = 0'50

Datoa botatzen dugunean

GERTAERA	BALIOA	PROBABILITATEA
1 atera	1	1 / 6
2 atera	2	1 / 6
3 atera	3	1 / 6
4 atera	4	1 / 6
5 atera	5	1 / 6
6 atera	6	1 / 6

Datoaren kasuan gertaera bakoitzari bere aldean azatzen den puntuen zenbakia jarri diegu. Txanponaren kasuan, aurpegia ateratzeari 1 balioa eta gurutzea ateratzeari 0 balioa eman diogu.

Aldagai aleatorioak DISKRETU eta JARRAITAN sailka daitezke. Aldagai aleatorio baten balioak zenbaki osoen bidez adierazi daitezkeenean DISKRETUA izango da, zeren eta ondoz ondoko balioren artean ezin da erdiko baliorik eduki. Lehen aipatu ditugun bi aldagaiak, diskretuak izango dira.

Aldagai aleatorio JARRAI batean ezin dezakegu elkarren segidako balioez hitz egin. Hau da, edozein balio hartzen badugu, ez da posible izango balio honen hurrengoez hitz egiterik. Adibidez gure aldagaia pertsonen pisua bada, ezin genezake jakin 50 kg-ren hurrengo balioa zein den, zeren eta 50en hurrengo bezala, edozein balio aukeratzen badugu ere, bien arteko tartean infinitu balio egongo bait dira.

1. ITXAROPEN MATEMATIKOA ETA ALDAGAI ALEATORIO BATEN BARIANTZA

Eman dezagun ondoko aukera egin behar dugula: 10 pezeta irabaztea %100 kasuetan, 100 pezeta irabaztea %50 kasuetan, edo, azkenik, 1000 pezeta irabaztea %10 kasutan. Denetatik bat bakarrik aukeratzekotan zein eta zergatik aukeratuko genuke? Planteamendu egokia edo zuzena hau izango litzateke: Luzarora zein aukerarekin lortuko nituzke etekin handienak?

Suposa dezagun esperimentua 100 aldiz errepikatzen dugula eta kalkula ditzagun kasu bakoitzean lortuko genituzkeen etekinak.

a) %100eko kasuetan 10 pezeta irabaziz = 10*100 = 1000 pzta.
b) %50eko kasuetan 100 pezeta irabaziz = 100*50 = 5000 pzta.
d) %10eko kasuetan 1000 pezeta irabaziz = 1000*10 = 10000 pzta.

Kalkulu hauek ikusi ondoren, bidezkoa da hirugarren aukeraren alde egotea.

Antzeko adibide bat hau izan daiteke. Pentsa dezagun lagun batekin datoka jolasean ari garela, aurrez erregela hauek jarri ondoren: 100 pezeta irabazten dituzu 6a ateratzen denean, 50 pezeta irabazten dituzu 5a irtetzen bada eta 30 pezeta galtzen dituzu ateratzen den balioa 5a baino txikiagoa bada .X aldagai aleatorioa nire ekonomiak joku bakoitzean jasaten dituen aldaketa bezala definitzen badugu, honela definiturik geratuko litzateke:

X_i	P_i
100	1/6
50	1/6
-30	4/6

Joku hau 300 aldiz egiten badugu, hau gertatzea espero dut: 100 pezeta irabaziko ditut 1/6 aldietan, 50 pezeta irabaziko ditut 1/6 aldietan eta 30 pezeta galduko ditut 4/6 aldietan. 300 joku hauetan nire ekonomiarekin gertatuko dena era honetara joango dela espero dut:

Totala = 100*1/6*300 + 50*1/6*300 - 30*4/6*300 = 5000 + 2500 - 6000 = 1500

300 joko hauek bukatu ondoren 1500 pezeta irabaztea espero dut. Baiezta dezakegu 300 joko hauetan batezbeste joko bakoitzean 1500/300 = 5 pezeta irabaziko ditudala ere.

Adibide honetan aldagai baten "ITXAROPEN MATEMATIKOA" izena duen parametroen kalkuluak egin ditugu. Ondorioz edozein aldagai kalkulatzeko posibilitatea eskainiko digun formulak definituko ditugu.

Aldagai diskretua bada:

$$E(X) = \sum x_i * f(x_i) = \sum x_i * p(x_i)$$

Aldagaia jarraia bada:

$$E(X) = \int x * f(x) * dx$$

Bigarren formula honetan aldagai jarraien kasuko formula idatzi dugu. Baina kalkuluak egiterakoan, integralekin kalkuluak egitea beharrezkoa izango da, eta gai honetako helburuak gainditzen dituenez, hemendik aurrera gure kalkuluetan aldagai diskretuetara mugatuko gara.

Beraz, honela definitutako X aldagaia badugu:

X :	0	1	2	3
f (X) :	.2	.3	.4	.1

Aldagai honen itxaropen matematikoa hau izango da:

$$E(X) = 0 * .2 + 1 * .3 + 2 * .4 + 3 * .1 = 1.4$$

Azken finean ITXAROPEN MATEMATIKOA eta lagineko batezbesteko matematikoak berdindu daitezke; batean eta bestean kalkuluak egiteko aukera ematen dizkiguten formulak baliokideak bait dira. Hau ikusteko aurreko adibidea beste era honetara jarriko dugu:

Pentsa dezagun 10 pertsonaz osaturiko lagin bat daukagula; 0, 1, 2, eta 3 balioak har ditzaketenak. Lortu dugun maiztasunen banaketa hau baldin bada:

X_i	m_i	$m_i X_i$	m_r
0	2	0	.2
1	3	3	.3
2	4	8	.4
3	1	3	.1
	10	14	1

$$\overline{X} = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{14}{10} = 1.4$$

Azaldu dugun maiztasunen banaketa eta lehen azaldutako probabilitate edo dentsitate-funtzioarekin bat datozela ohar gaitezke. Eskuineko zutabeari begiratzen badiogu, aldagai honen maiztasun erlatiboak eta aurreko adibideko probabilitateak berdinak direla konturatuko gara. Lehenengo kasuko itxaropen matematikoek eta bigarren kasuko batezbestekoek balio berdinak dauzkatela ere ikus dezakegu.

Estatistika deskribatzaileen kapituluan ikusten genuen bezala, aldagai aleatorioen kasuan ere zentralizazio- eta sakabanatze-neurriak ikusiko ditugu.

Bariantza eta desbidazio tipikoa definitu beharko ditugu. Desbidazio tipikoa, kasu honetan ere, bariantzaren erro karratu bezala definituko dugu.

Aldagai aleatorio baten bariantza era honetara definituko dugu:

$$\text{Bar}(X) = E(X^2) - E^2(X) = E[(X - E(X))^2]$$

Ikus dezagun aurreko formulan adierazitako kalkuluak nola egiten diren. Horretarako $E(X^2)$ kalkulatu beharra dugu, eta ondorengo formularen bidez egingo dugu:

$$E(X^2) = \Sigma x_i^2 * f(x_i) = \Sigma x_i^2 * p(x_i)$$

Gure kasuan, lehen era honetara definitu dugun aldagaiarekin jarraitzen badugu:

X :	0	1	2	3
f (X) :	.2	.3	.4	.1

$$E(X^2) = \sum x_i^2 * f(x_i) = 0^2 * .2 + 1^2 * .3 + 2^2 * .4 + 3^2 * .1 =$$
$$= 0 * .2 + 1 * .3 + 4 * .4 + 9 * .1 = 2.8$$

$E(X^2)$ balioa kalkulatu ondoren eta X aldagaien itxaropen matematikoa 1.4 dela kontutan harturik, bariantzaren formulan balio hauek ordezkatzen baditugu:

$$Bar(X) = E(X^2) - E^2(X) = 2.8 - 1.4^2 = 2.8 - 1.96 = 0.84$$

Desbidazio tipikoa edo standarda, bariantzaren erro karratua denez gero, gure kasuan:

Desbidazio tipikoa: $s = \sqrt{.84} = .9165$

Beraz, lanari ekiterakoan hurrengo pausoak kontutan hartu beharrekoak izango dira: lehenengo aldagaien probabilitatea edo dentsitate-funtzioa ezagutu, gero aldagaien batezbestekoa eta bariantza kalkulatu eta azkenik interesatzen zaigun probabilitatea kalkulatu.

Hala ere bizitza errealean planteatzen zaigun egoera edo arazo konkretu batzuei probabilitateen teoria aplikatzen diegunean, dentsitate- edo probabilitate-funtzio berri bat aurkitzea ez da beti beharrezkoa izango, zeren eta sarritan gure egoerak aurretik aztertuta daukagun beste egoera ezberdin bat edo batzuekin harreman handia edukiko bait dute.

Egoera edo alderdi asko errepikatzen direlako edo gutxienez hauen ezaugarriak era berdinekoak izango direlako, BANAKETA TEORIKO batzuk definituak izan dira. Banaketa teoriko hauek aztertu nahi diren egoera tipikoetara egokituta daude. Banaketa teoriko edo probabilitate-eredu batzuk aztertuko ditugu; giza zientziei aplikatuta dagoen Estatistikan garrantzi handia bait dute. Kasu askotan guri plantea dakizkigukeen arazoei banaketa teorikoek irtenbide bat emango diete.

Probabilitateen banaketa teorikoak bitan banatuko ditugu:

.- BANAKETA DISKRETUAK
.- BANAKETA JARRAIAK

Kapitulu honetan diskretuak aztertuko ditugu.

2. *BANAKETA DISKRETUAK*

2.1. *Bernouilli-ren banaketa*

Demagun txanpon bat airera botatzen dugula. Hau esperimentu estatistiko bezala kontsideratzen badugu, bi emaitza posible lortzeko posibilitatea daukagula ikus dezakegu, aurpegia edo gurutzea.

Edozein ikerlan egiterakoan, honelako egoera baten aurrean oso maiz aurkitzen gara, hau da, edozein aldagai dikotomikorekin: sexua, pertsona batek galdera bati erantzun egokia eman dion ala ez.....

Normalean aldagai dikotomiko hauei 1 eta 0 balioak emango dizkiegu, eta era honetako esperimentuei Bernouilliren esperimentua deritzogu. Esperimentu hau ondo definitua egoteko, bi emaitza edo alternatiba posibleen probabilitateek ezagunak izan behar dute.

Aurreko kasuarekin jarraituz, txanponaren aurpegia ateratzen denean arrakasta duen ekintza baten aurrean gaude. Normalean edozein aldagai dikotomikoren aurrean gaudenean, emaitza bati arrakastaren tratamendua emango diogu eta besteari, porrotarena. Aldagai dikotomiko bat era honetara definituko dugu:

$$X: \begin{cases} 0 & p \\ 1 & q = 1 - p \end{cases}$$

Aldagaiak 0 eta 1 balioak hartuko ditu, hauen probabilitateak p eta q izanik. Bi probabilitateen batura 1 izango da ($p+q = 1$).

Aldagai batetik bestera aldatuko dena probabilitateen balioa izango da, hau da, p eta q balioak, zeren Bernouilliren banaketako edozein esperimentutako eskema beti berdina izango bait da.

2.2. *Banaketa binomiala*

Eman dezagun txanpon bat hamar aldiz airera botatzen dugula, edo test bat pasatzen dugula, non ikasleak 20 galderari erantzun behar dien, eta galdera bakoitzak 2 alternatiba dituen. Aurreko bi esperimentuak Bernouilliren esperimentuen errepikapenak izango dira.

Eman dezagun galdera hauei erantzutea interesatzen zaigula: Hamar aldiz botatako txanponean 6 aurpegi lortzeko dagoen probabilitatea, edo ikasle batek 20 galderak zoriz

erantzuten baditu 12 galdera ondo erantzuteko dagoen probabilitatea.

Bernouilliren N esperimentu errepikatzen ditugunean eta k arrakasta lortzeko dagoen probabilitatea kalkulatu nahi badugu, lehengo bi galderetan bezala BANAKETA BINOMIALEKO eredu teoriko baten probabilitateak lortzen arituko gara.

Banaketa BINOMIAL bezala kontsideratua izateko, banaketa batek bi baldintza hauek bete beharko ditu:

.- Bernouilliren esperimentuko N errepikapen egiten dira.

.- Arrakasta lortzeko dagoen probabilitateak prozesu guztian konstante izan behar du.

Eman dezagun lehen planteatu dugun galdera erabakitzen ari garela. 10 aldiz bota dugu txanpona airera eta 6 aurpegi ateratzeko dagoen probabilitatea kalkulatu nahi dugu. Txanpon trukatu hau airera botatzen dugun bakoitzean aurpegia lortzeko dagoen probabilitatea, lehen esan dugunez, 0.6 da.

Kasu honetako probabilitatea jakiteko eman dezagun 6 aurpegiak jarraian atera zaigula lehenengo 6 kasuetan; eta gero lau gurutze, hau da, era honetara erori da gure txanpona: a a a a a a g g g g. Aurreko saila ateratzeko dagoen probabilitatea hau da:

$$0.6*0.6*0.6*0.6*0.6*0.6*0.4*0.4*0.4*0.4 = 0.00119439$$

Baina kalkulatu dugun probabilitatea aurreko seriearen probabilitatea izango da, eta ez guk eskatzen dugun probabilitatea, zeren eta 6 aurpegi beste edozein ordenatan lor ditzakegu. Orain erantzun behar dugun galdera zera da: gure 6 aurpegiak zenbat posizio ezberdinetan atera daitezkeen. Honen erantzuna, KONBINATORIAK era honetara ematen du: agertzeko era ezberdinak kalkulatzeko nahikoa dela 10 elementuen konbinazioak seinaka harturik kalkulatzea. Hemendik gure probabilitatea kalkulatzeko, lehen lortu dugun probabilitatea, konbinazio ezberdin posiblen kopuruaz biderkatzea nahikoa izango da, aurreko seriea edo beste edozein konbinazio lortzeko probabilitatea konstantea delako. Adibidez (a a a g g g a a g a) seriea lortzeko probabilitatea kalkulatu nahi badut, ondoko hau izango da:

$$0.6*0.6*0.6*0.4*0.4*0.4*0.6*0.6*0.4*0.6 = 0.00119439$$

hemendik, 6 aurpegi lortzeko probabilitatea:

$C_{10,6}$: 10 elementuen konbinaketak seinaka harturik, eta honen balioa :

$$C_{10,6} = \frac{10*9*8*7*6*5*4*3*2*1}{6*5*4*3*2*1 \ * \ 4*3*2*1} = 210$$

$$P \text{ (sei aurpegi)} = C_{10,6} * 0.00119439 = 0.2508$$

Baina kalkulatzeko prozesua nahikoa neketsua da, batez ere beste kasuan aurkezten dugun bezalako probabilitate nahaspilatuagoak kalkulatzea nahiko bagenu, zeren 20 galderei erantzuterakoan gutxienez horietako 12ri igartzeko probabilitatea kalkulatu nahi badugu, 12, 13,.... eta 20ri igartzeko probabilitateak kalkulatu beharko bait genituzke.

Hemen aipatu dugun adibidea edo antzerakoak askotan errepikatzen dira, eta gure probabilitateen kalkuluak errazteko taula batzuk eginda daude, non gure kasuko probabilitateak agertzen diren. Hurrengo orrialdetan ikus ditzakegu banaketa binomialen taulak. Taula hauek begiratzerakoan, hiru parametro hartu beharko ditugu kontutan: N, p, eta k. Parametro bakoitzaren esanahia orain ikusiko dugu:

N: Bernouilliren esperimentua errepikatu dugun aldi-kopurua adierazten du. Gure txanpona airera botatzen dugunean N = 10 izango da, eta testen kasuan N = 20.

p: Bernouilliren esperimentu bakoitzean arrakasta lortzeko dagoen probabilitatea. Gogoratu beharra dago esperimentu osoan arrakasta lortzeko dagoen probabilitateak konstantea izan behar duela.

k: Probabilitatea kalkulatu nahi dudaneko gertaeraren arrakasta-kopurua.

Banaketa binomialen itxaropen matematikoa edo batezbestekoen balioa N*p izango da, eta bariantza N *p* (1-p). Beti bezala desbidazio tipikoa bariantzaren erro karratua izango da.

Eman dezagun lehen aipatu dugun probabilitatea kalkulatu nahi dugula: Galdera bakoitzak bi aukera edukiz, 20 galderako test batean, zoriz erantzunda gutxienez 12 ondo igarrita egoteko dagoen probabilitatea.

Jakina da item bakoitzari igartzeko dagoen probabilitatea 0.5 dela. Bernouilliren esperimentua errepikatzen dugun aldi-kopurua 20 da, zeren eta gure testak 20 galdera bait ditu. Banaketa binomialen parametroak definitzeko, k-ren balioa ezagutu behar dugu.

Gure galdera 12 galderari igartzeko probabilitatea kalkulatzea izango balitz, kasu honetan k-ren balioa 12 izango litzateke, eta probabilitatea lortzeo modua hau izango litzateke: tauletan, N=20ri dagokion blokean, p=.50-ren zutabean eta K=12ren lerroan agertzen den balioa izango da, eta hemen aurkitzen dugun probabilitatea .1201 da.

Gure kasuan gutxienez 12 arrakasta lortzeko probabilitatea kalkulatzeko, baliagarriak diren k-ren balioak K=12,13,14... izango dira, eta hauei dagokien probabilitateen batura izango da kalkulatzen ari garen probabilitatea.

Hau da, gure kasuan batu behar ditugun probabilitateak hauek dira:

.1201 .0739 .0370 .0148 .0046 .0011 .0002 .0000 .0000

Eta emaitza hau izango da:

P (gutxienez 12 arrakasta) = 0.2517

Ondorengo orrietan azaltzen diren tauletan, agertzen diren probabilitateak p=0 .50 edo txikiagoari dagokionak dira, baina hasiera batean agertutako gertakizun edo antzerakoak sort daitezke, hau da, arrakasta lortzeko probabilitatea 0.50 baino handiagoa izatea (p > 0 . 50). Kasu hauetan hurrengo ariketetan bezala jokatuko dugu.

Eman dezagun gure txanpona airera botatzerakoan, kasu bakoitzean aurpegia lortzeko probabilitatea 0. 60 dela.

Probabilitate hau kalkulatzeko tauletara jotzen badugu, ez dugu p= .60ko zutaberik aurkituko, eta horregatik arrakasta gurutzea ateratzeari deituko diogu.

6 aurpegi ateratzearen probabilitatea kalkulatu nahi dugu, eta 6 aurpegi ateratzea eta 4 gurutze ateratzea, txanpona 10 aldiz airera botatzerakoan gertaera berdinak direla kontutan harturik, hauen probabilitateak berdinak izango dira, eta gurutzea ateratzeko probabilitatea, aldi bakoitzean .40 denez gero, gure gertaeren probabilitatea n = 10-en blokean, .4-ren zutabean eta k=4-ren lerroan aurkituko dugu, eta lortuko dugun emaitza .2508 da.

Egin dugun aldaketa hau da:

P(10 jaurtialditan 6 aurpegi ateratzea) = P(10 jaurtialditan 4 gurutze ateratzea)

Bukatzeko, 10 aldiz txanpona airera bota eta aurpegia lortzeko probabilitatea .60 dela jakinik, bi aurpegi baino gehiago lortzeko dagoen probabilitatea kalkulatuko dugu.

Kasu honetan lehengoaren antzera, gurutzeen probabilitatea kalkulatuko dugu. Berdintasuna honela lortuko dugu:

Bi aurpegi baino gehiago lortzea: 3,4,5,... 10 aurpegi lortzea da, eta aurpegi-kopuru hau lortzea eta 7,6,5,... edo 0 gurutze lortzea gauza berdinak dira. Beraz probabilitate hau

kalkulatzeko K-ren 7,6,5,... eta 0 balioei dagokion probabilitateak batu beharko ditugu, hau da, probabilitate hauen batura izango da gure emaitza:

.0425 .1115 .2007 .2508 .2150 .1209 .0403 .0060

Probabilitate hauek N=10en blokean, eta p=.40-ko zutabean bilakatuko ditugu.

Aurreko probabilitateen batura = 0.9877 da, hau da, gure galderaren emaitza .9877 izango da.

Batzuetan errazago izango da kontrako gertaerara pasatzea, eta 7,6,5,4,3,2,1, edo 0 gurutze ateratzeko kontrako gertaera, 8, 9 edo 10 gurutze ateratzea izango da. Eta kontrako gertaeren probabilitatea 1-gertaeren probabilitatea dela kontutan harturik, gure probabilitatea honela kalkula dezakegu:

P(7 gurutze edo gutxiago) = 1 - P(8 gurutze edo gehiago) =
P(7 gurutze edo gutxiago) = 1 - .0106 - .0016 - .0001 = 1 - .0123 = .9877

Bigarren era hau askotan askoz ere errazagoa izango da. Honelako probabilitateek propietate hau betetzen bait dute: probabilitate guztien baturak 1 izan behar du, eta ondorioz gertaera baten probabilitatea gehi kontrako gertaeren probabilitatea beti 1 izango da.

BANAKETA BINOMIALEN TAULA (1)

$$P(x = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

N	k	p: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.400	.4500	.5000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4450	.4800	.4950	.5000
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0040	.0911	.1250
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3840	.3456	.2995	.2500
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312
	1	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1562
	2	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0004	.0022	.0064	.0146	.0284	.488	.0768	.1128	.1562
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3826	.3847	.3355	.2760	.1977	.1373	.0896	.0548	.0312
	2	.0515	.1488	.2476	.2936	.3115	.2965	.2587	.2090	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0004	.0026	.0092	.0231	.0467	.808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039

BANAKETA BINOMIALEN TAULA (2)

$$P(x = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

N	k	p: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0277	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0016
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.2348	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0001	.0025	.0132	.0388	.0803	.1231	.1830	.2207	.2360	.2256
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005
12	0	.5504	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029
	2	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161
	3	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537
	4	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208
	5	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934
	6	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256
	7	.0000	.0000	.0006	.0033	.0115	.0291	.0591	.1009	.1489	.1934
	8	.0000	.0000	.0001	.0005	.0024	.0078	.0199	.0420	.0762	.2108
	9	.0000	.0000	.0000	.0001	.0004	.0015	.0048	.0125	.0277	.0537
	10	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0025	.0068	.0161
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0029
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002

BANAKETA BINOMIALEN TAULA (3)

$$P(x = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

N	k	p: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.3512	.3672	.2774	.1787	.1029	.0540	.0259	.0113	.0045	.0016
	2	.1109	.2448	.2937	.2680	.2059	.1388	.0836	.0453	.0220	.0095
	3	.0214	.0997	.1900	.2457	.2517	.2181	.1651	.1107	.0660	.0349
	4	.0028	.0277	.0838	.1535	.2097	.2337	.2222	.1845	.1350	.0873
	5	.0003	.0055	.0266	.0691	.1258	.1803	.2154	.2214	.1980	.1571
	6	.0000	.0008	.0063	.0230	.0559	.1030	.1546	.1968	.2169	.2095
	7	.0000	.0001	.0011	.0058	.0186	.0442	.0883	.1312	.1775	.2095
	8	.0000	.0000	.0001	.0011	.0047	.0142	.0336	.0656	.1089	.1571
	9	.0000	.0000	.0000	.0001	.0009	.0034	.0101	.0243	.0495	.0873
	10	.0000	.0000	.0000	.0000	.0001	.0006	.0022	.0065	.0162	.0349
	11	.0000	.0000	.0000	.0000	.0000	.0001	.0008	.0012	.0036	.0095
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.3593	.3559	.2539	.1539	.0832	.0407	.0181	.0073	.0027	.0009
	2	.1229	.2570	.2912	.2501	.1802	.1134	.0634	.0317	.0141	.0056
	3	.0259	.1142	.2056	.2501	.2402	.1943	.1366	.0845	.0462	.0222
	4	.0037	.0349	.0998	.1720	.2202	.2290	.2022	.1549	.1040	.0611
	5	.0004	.0078	.0352	.0860	.1468	.1963	.2178	.2066	.1701	.1222
	6	.0000	.0013	.0093	.0322	.0734	.1262	.1759	.2066	.2088	.1833
	7	.0000	.0002	.0019	.0092	.0280	.0618	.1082	.1574	.1952	.2095
	8	.0000	.0000	.0003	.0020	.0082	.0232	.0510	.0918	.1398	.1833
	9	.0000	.0000	.0000	.0003	.0018	.0066	.0183	.0408	.0762	.1222
	10	.0000	.0000	.0000	.0000	.0003	.0014	.0049	.0136	.0312	.0611
	11	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0033	.0093	.0222
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0019	.0056
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0009
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
	2	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
	3	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
	4	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
	5	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
	6	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
	7	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
	8	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
	9	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
	10	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
	11	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
	12	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0319
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

BANAKETA BINOMIALEN TAULA (4)

$$P(x = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

N	k	p: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.3706	.3294	.2097	.1126	.0535	.0228	.0087	.0030	0009	.0012
	2	.1463	.2745	.2775	.2111	.1336	.0732	.0353	.0150	.0056	.0018
	3	.3559	.1423	.2295	.2463	.2079	.1465	.0888	.0468	.0215	.0085
	4	.0061	.0514	.1311	.2001	.2252	.2040	.1553	.1014	.0572	.0278
	5	.0008	.0137	.555	.1201	.1802	.2099	.2008	.1623	.1123	.0667
	6	.0001	.0028	.0180	.0550	.1101	.1649	.1982	.1983	.1684	.1222
	7	.0000	.0004	.0045	.0917	.0524	.1010	.1524	.1889	.1969	.1746
	8	.0000	.0001	.0009	.0055	.0197	.0487	.0923	.1417	.1812	.1964
	9	.0000	.0000	.0001	.0012	.0058	.0185	.0442	.0840	.1318	.1746
	10	.0000	.0000	.0000	.0002	.0014	.0056	.0167	.0392	.0755	.1222
	11	.0000	.0000	.0000	.0000	.0002	.0013	.0049	.0142	.0337	.0667
	12	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0040	.0115	.0278
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0029	.0085
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.3741	.3150	.1893	.0957	.0426	.0169	.0060	.0019	.0005	.0001
	2	.1557	.2800	.2673	.1914	.1136	.0581	.0260	.0102	.0035	.0010
	3	.0415	.1556	.2359	.2393	.1893	.1245	.0701	.0341	.0144	.0052
	4	.0076	.0605	.1457	.2093	.2209	.1868	.1320	.0796	.0411	.0182
	5	.0010	.0175	.0668	.1361	.1914	.2081	.1849	.1479	.0875	.0472
	6	.0001	.0039	.0236	.0680	.1276	.1784	.1991	.1839	.1432	.0944
	7	.0000	.0007	.0065	.0267	.0668	.1201	.1685	.1927	.1841	.1484
	8	.0000	.0001	.0041	.0084	.0279	.0644	.1143	.1606	.1883	.1855
	9	.0000	.0000	.0003	.0021	.0093	.0276	.0611	.1070	.1540	.1855
	10	.0000	.0000	.0000	.0004	.0025	.0095	.0263	.0571	.1008	.1484
	11	.0000	.0000	.0000	.0001	.0005	.0026	.0090	.0242	.0525	.0944
	12	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0081	.0215	.0472
	13	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0021	.0068	.0182
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.3763	.3002	.1704	.0811	.0338	.0126	.0042	.0012	.0003	.0001
	2	.1683	.2835	.2556	.1723	.0958	.0458	.0190	.0069	.0022	.0006
	3	.0473	.1680	.2406	.2297	.1704	.1046	.0547	.0246	.0095	.0031
	4	.0093	.0700	.1592	.2153	.2130	.1681	.1104	.0614	.0291	.0117
	5	.0014	.0218	.0787	.1507	.1988	.2017	.1664	.1146	.0666	.0327
	6	.0002	.0052	.0310	.0816	.1436	.1873	.1941	.1655	.1181	.0708
	7	.0000	.0010	.0091	.0350	.0820	.1376	.1792	.1892	.1657	.1214
	8	.0000	.0002	.0022	.0120	.0376	.0811	.1327	.1734	.1864	.1669
	9	.0000	.0000	.0004	.0033	.0139	.0386	.0794	.1284	.1694	.1855
	10	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214

BANAKETA BINOMIALEN TAULA (5)

$$P(x = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

N	k	p: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
18 (JARRAIPENA)											
	12	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0045	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0658	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1491	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1789	.1771	.1442
	9	.0000	.0000	.0007	.0051	.0198	.0154	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0287	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.074+	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.2101
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0719
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

BESTE ZENBAIT BANAKETA DISKRETU: PASCAL, POISSON

Aldagai diskretoei dagokien beste bi banaketa berri aztertuko ditugu: Pascal edo binomial negatiboen banaketa eta Poissonen banaketa.

2.3. Pascal-en banaketa

Pascalen banaketa, banaketa binomialaren ondorio bezala kontsidera dezakegu; funtsean eskema teoriko berdinean oinarritzen bait da. Bernouilliren N esperimentu errepikatzen dira, baina banaketa binomialean bezala esperimentuen kopuruak aurretik jarria egon beharrean kasu honetan esperimentuen kopurua izango da, hain zuzen, aldagai aleatorioa. Hots, aldagai aleatorioa, k arrakasta lortzeko beharrezko diren N esperimentuen kopuru bezala definitzen da.

Nahiz eta Pascalen banaketaren berezitasunak aztertzera sartu ez, banaketa binomialarekin duen erlazioa azter dezakegu. Banaketa binomialean aldagai aleatorioa K arrakasta-kopurua zen, Bernouilliren esperimentua N aldiz errepikatzen genuenean. Pascalen kasuan K arrakasta lortzeko beharrezkoak diren N esperimentuen balioa izango da. Baina Ngarren saioan kgarren arrakasta lortzeko, N-1 esperimentutan k-1 arrakasta lortu beharko dira eta hurrengo saioan baita arrakasta lortu ere.

Beraz Pascal edo binomial negatiboen banaketa murriz dezakegu zera esanez: Bernouilliren esperimentua N-1 aldiz errepikatuz k-1 arrakasta izanik (hau banaketa binomial arrunt bat da), Ngarren esperimentuan ere arrakasta lortzen dela.

Probabilitateak kalkulatzeko identitate hauek idatz ditzakegu:

P (k-garren arrakasta n-garren saioan) =
P (k-1 arrakasta N-1 saiotan eta arrakasta N-garrenean) =
P (k-1 arrakasta N-1 saiotan) * P (arrakasta N-garrenean) =
Binomiala (N -1, K -1, p) * p

Beraz txanpon bat airera botatzerakoan aurpegia ateratzeko probabilitatea 0.50-koa bada, 5-garren aurpegia 8-garren aldian lortzeko probabilitatea kalkulatu nahi badugu, banaketa binomialaren taula honela erabiliko dugu:

Lehenengo arazoa birplanteatuko dugu: 8-garren saioan 5-garren arrakasta (aurpegia) lortzeko, lehenengo 7 saiotan 4 arrakasta lortu behar dira eta 8-garrenean berriro arrakasta edo aurpegia lortu behar dugu. Gure probabilitatea bi probabilitate hauek biderkatuz lortuko dugu.

P (5-garren arrakasta 8-garren saioan) =
P (4 arrakasta 7 saiotan eta arrakasta 8-garrenean) =
P (4 arrakasta 7 saiotan) * P (arrakasta 8-garrenean) =
Binomiala (7,4,0.50) * 0.50 = 0.2734 * 0.50 = 0.1367

Eta horrela Pascalen banaketari jarraitzen zaion edozein aldagairen probabilitateak kalkula ditzakegu.

2.4. Poisson-en banaketa

Batzuetan banaketa binomialen kasuan, gure esperimentua 20 aldiz baino gehiagotan errepikatuko dugu, hau da, txanpona 20 aldiz baino gehiagotan botako dugu airera, edo ikasle-talde bati pasatzen diogun testak 20 galdera baino gehiago edukiko ditu. Kasu hauetan, banaketa binomialari jarraitzen zaizkiola kontutan harturik, banaketa honen taulara jo beharko genuke, baina taula arruntetan, N=20 baino balio handiagoen probabilitaterik ez dugu edukiko. Honelakotan gure probabilitateen kalkulua Poissonen banaketaren bidez egingo dugu kasu batzuetan eta banaketa Normalaren bidez beste kasu batzuetan.

Aipatu dugun kasua, Poissonen banaketaren kasu tipikoena da, baina beste probabilitate asko kalkulatzeko ere erabiltzen da. Kasu nabarmenenak ondoren aipatzen ditugunen antzekoak izango dira:

.- Leihatila baten aurrean egongo den iladaren pertsona-kopuruen probabilitateak kalkulatzea.
.- Egun batean egongo den istripu-kopuruaren probabilitatea kalkulatzea.

Poissonen probabilitate-funtzioa ondoko formula honek adierazten du:

$$f(x) = P(x = k) = \frac{m^k}{k!} e^{-m}$$

Poissonen banaketaren probabilitatea kalkulatu baino lehen, banaketak daukan parametro bakarrak ezaguna izan beharko du. Parametro hau, batezbestekoa adierazten duena izango da eta "m" letraren bidez izendatuko dugu. Parametro hau kalkulatzeko, banaketa binomialen N eta p parametroak erabiliko ditugu.

$$m = N * p$$

m balioak adierazten duena zera da: arrakasten batesbestekoen kopurua.

Poisson banaketaren taulak erabiltzeko, m parametroen balioak ezaguna izan beharko du. Balio hau ezagutu ondoren, hurrengo orrialdeetan aurkezten dugun taulak erabiliko ditugu.

Gertaera baten probabilitatea kalkulatzeko, arrakasta-kopuruen balioa jakin ondoren m parametroen balioari dagokion blokean eta k arrakasta-kopuruen lerroan bilatuko dugu gure probabilitatea. Kontuan hartu behar dugu agertzen den balioa ez dela gure gertaeren probabilitatea; baizik eta probabilitae metatua dela. Adibidez, eman dezagun m = 1.6 izanda, K = 3 arrakasta lortzeko probabilitatea kalkulatu nahi dugula. m = 1.6 blokean K = 3 dagokion lerroan aurkitzen dugun balioa .21664 da. Baina balio hau 3 arrakasta edo gehiago lortzeko gertareren probabilitatea da. Guk 3 arrakasta lortzeko probabilitatea kalkulatu nahi dugunez, honela jokatuko dugu: 3 arrakasta lortzea era honetara idatziko dugu: 3 edo gehiago lortzea eta 4 edo gehiago EZ lortzea. Hortik gure probabilitatea:

P (3 arrakasta) = P (3 edo gehiago) - P (4 edo gehiago)
P (3 arrakasta) = .21664 - .07881 = .13783

Beste adibide bat egingo dugu.

Eman dezagun ikastola bateko ikasleetatik %5ek irakurmen-arazoak dauzkatela. Guk aztertu nahi dugun gelan 40 ikasle badaude, kalkulatu nahi ditugun probabilitateak ondoko bi gertaeren probabilitateak izango dira: Gure gelan 2 ikaslek arazoak edukitzea. Gure gelan 6 ikaslek edo gehiagok arazoak edukitzea.

Kasu bietan m parametroen balioa berdina izango da, eta honela kalkulatuko dugu:

$$m = N * p = 40 * 0.05 = 2.0$$

Bigarren gertaeren probabilitatea taulak zuzenean emango digu; lehen esan dugunez, taulak ematen dituen balioak probabilitate metatuenak bait dira. Hau da m = 2.0 blokean K = 6 balioari dagokion balioak adierazten duena, 6 arrakasta edo gehiago lortzeko daukagun probabilitatea izango da. Probabilitate hau kasu honetan: 0.1656 da.

Lehengo probabilitatea kalkulatzeko, k = 2 balioari dagokion probabilitatearen eta k = 3 balioari dagokionaren arteko kenketa egin beharko dugu. Hau da, 2 ikaslek arazoak edukitzeko probabilitatea, 2 edo gehiagok edukitzeko eta 3 edo gehiagok edukitzeko probabilitateen kendura izango da.

Gure kasuan hau izango da:

P (2 arrakasta) = P (2 edo gehiago) - P (3 edo gehiago)
P (3 arrakasta) = .59399 - .32332 = .27067

POISSONEN BANAKETA-TAULA. PROBABILITATE METATUAK

$$P(x \geq k) = \sum_{r=k}^{\infty} \frac{m^r}{r!} e^{-m}$$

k	m = 0.2	m = 0.3	m = 0.4	m = 0.5	m = 0.6
0	1.00000	1.00000	1.00000	1.00000	1.00000
1	.18127	.25918	.32968	.39347	.45119
2	.01752	.03694	.06155	.90020	.12190
3	.00115	.00360	.00793	.01439	.02311
4	.00006	.00027	.00078	.00175	.00336
5	.00000	.00002	.00006	.00017	.00039
6	.00000	.00000	.00000	.00001	.00004

k	m = 0.7	m = 0.8	m = 0.9	m = 1.0	m = 1.2
0	1.00000	1.00000	1.00000	1.00000	1.00000
1	.50341	.55067	.59343	.63212	.69880
2	.15580	.19120	.22751	.26424	.33737
3	.03414	.04742	.06285	.08030	.12051
4	.00575	.00908	.01345	.01898	.03377
5	.00078	00141	.00234	.00366	.00775
6	.00009	.00018	.00034	.00059	.00150
7	.00000	.00002	.00004	.00008	.000025
8	.00000	.00000	.00000	.00001	.00004
9	.00000	.00000	.00000	.00000	.00000

k	m = 1.4	m = 1.6	m = 1.8	m = 2.0	m = 2.5
0	1.00000	1.00000	1.00000	1.00000	1.00000
1	.75340	.79810	.83470	.86466	.91791
2	.40817	.47507	.53716	.59399	.71270
3	.16650	.21664	.26938	.32332	.45619
4	.05372	.07881	.10871	.14288	.24242
5	.01425	.02368	.03641	.05265	.10882
6	.00320	.00604	.01038	.01656	.04202
7	.00062	.00134	.00257	.00453	.01419
8	.00011	.00026	.00056	.00110	.00425
9	.00002	.00005	.00011	.00024	.00114
10	.00000	.00001	.00002	.00005	.00027

k	m = 3.0	m = 3.5	m = 4.0	m = 4.5	m = 5.0
0	1.00000	1.00000	1.00000	1.00000	1.00000
1	.95201	.96980	.98168	.98889	.99326
2	.80085	.86411	.90842	.93890	.95957
3	.57681	.67915	.76190	.82642	.87535
4	.35277	.46337	.56653	.65770	.73497
5	.18474	.27456	.37116	.46790	.55951
6	.08392	.14239	.21487	.29707	.38404
7	.03351	.06529	.11067	.16895	.23781
8	.01191	.02674	.05113	.08659	.13372
9	.00380	.00987	.02136	.04026	.06809
10	.00110	.00332	.00813	.01709	.03183
11	.00029	.00102	.00284	.00667	.01369
12	.00017	.00029	.00092	.00240	.00545
13	.00008	.00018	.00027	.00080	.00202
14	.00000	.00002	.00008	.000025	.00070

7. GAIA
BANAKETA JARRAIAK

Banaketa jarraiak aldagai jarraiei dagozkienak dira.

Probabilitateak kalkulatzeko era, banaketa diskretuetan erabilitakoen zeharo ezberdina izango da. Hau da bi banaketa-moten arteko ezberdintasunik garrantzitsuena. Lehenengo ezberdintasuna zera izango da: banaketa diskretuetan puntu edo balio konkretuen probabilitateak kalkulatzea posible izango dela, hau da, lehen egin ditugun adibideetan zenbait probabilitate honela kalkulatu ditugu: zein da 6 aurpegi lortzeko dagoen probabilitatea?,... Banaketa jarraietan galdera honek ez dauka zentzurik, eta beraz bere probabilitatea 0 izango da. Zentzurik ez daukala esan dugu, zeren eta edozein aldagai jarrai hartzen badugu, adibidez pertsona baten pisua, ez dago munduan 70 kilo dituen pertsonarik, hau da, guk askotan esaten dugu pertsona batek pisu hori daukala, baina matematikoki hartuta hau ez da, benetako pisurako hurbilketa bat besterik. Gertatzen dena zera da: pisua neurtzeko erabiltzen dugun tresnak askotan ez dituela gramoen zatitxoak bereizten, eta gure tresnaren bidez hurbilketa bat egiten dugula.

Banaketa jarraietan ez ditugu puntuen probabilitateak kalkulatuko tarteen probabilitateak baizik.

Lehen esan dugu banaketa jarraiak aldagai jarraiei dagozkienak direla, baina askotan (batez ere probabilitateen hurbilketak egiterakoan) aldagai diskretuen probabilitateak ere banaketa jarraiak erabiliz kalkulatuko ditugu.

Banaketa jarraien probabilitateak, tauletara joaterik nahi ez badugu, integralen bidez kalkulatuko ditugu. Guk gai honetan kalkulu guztiak taularen bidez egingo ditugu, gero SPSS/PC+ paketeak emango dituen balioak tauletako probabilitateekin konparatzea posible izango delako.

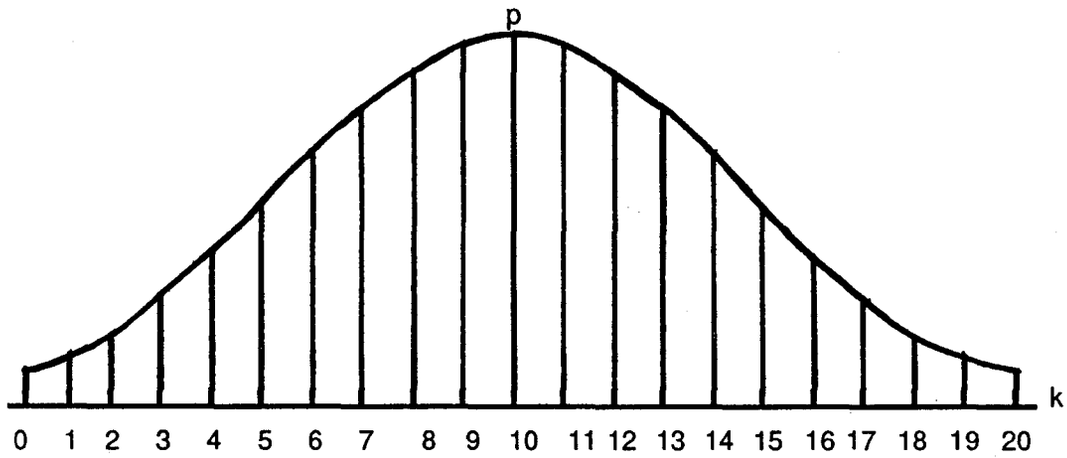
1. BANAKETA NORMALA

Banaketa jarraien deskribapena egiterakoan, Banaketa Normalarekin hasiko gara, denetan garrantzi handiena daukana bait da.

Banaketa normalaren funtzioa banaketa binomialen hurbiltze bezala sortu zen, eta lehenengo definitu zuen gizona De Moivre izan zen (1733. urtean). Baina benetan garrantzi handia hartu zuen XIX. mendean GAUSS eta LAPLACEk gaur ezagutzen diren formulak aurkeztu zituztenean, nahiz eta formula hauetako batzuk De Moivre-k argitu zituen formulen berdinak izan. Batzuetan banaketa normala GAUSS-LAPLACEren banaketa

izenaz ere ezagutzen da.

Lehen esan dugun bezala, banaketa normala banaketa binomialen probabilitateak kalkulatzeko sortu zen. N = 20 eta p = 0.50 parametro bezala dituen banaketa binomialen adierazpen grafiko bat egiten badugu, ondoren aurkezten dugun era lortuko dugu.



Gainean ezarri dugun lerro jarraiak banaketa normalen funtzioa adierazten duen kurbaren adierazpen grafikoa da, eta kurba honen propietateak aztertuko ditugu orain:

.- Edozein banaketa normal bi parametrorekin definitua geldituko da: Batezbestekoarekin eta desbidazio tipikoarekin. Banaketa normal bat izendatzerakoan, honela idatziko dugu N (μ,σ), μ batezbestekoa eta σ desbidazio tipikoa adierazten duten parametroak direlarik.

.- Banaketa normala adierazten duen kurba beti forma berdinekoa izango da. Aldatzen direnak parametroen balioak izango dira. Beraz parametroen arabera nahi dugun beste banaketa normal egongo dira.

.- Probabilitate-funtzioak, bere balio handiena batezbestekoan lortzen du.

.- Banaketa normalen kurba, batezbestearekiko simetrikoa da.

.- Batezbestekoa, moda eta medianaren balioak edozein banaketa normalen berdinak izango dira.

Lehen esan dugunez batezbesteko eta desbidazio standarden balioen arabera, kurba normal ezberdinen kopurua infinitua izango da.

Ezinezkoa izango litzateke banaketa normal guztien taulak edukitzea. honen ondorioz, probabilitateak kalkulatzerakoan beti BANAKETA NORMAL TIPIKOAri begiratuz lortuko ditugu. Banaketa normal tipikoa, batezbestekoa 0 eta desbidazio tipikoa 1

duenari deituko diogu. Hurrengo orrialdetan aurkezten ditugun taulak banaketa honi dagozkionak dira. Banaketa hau era honetara idatziko dugu N (0,1), eta hemendik aurrera Z letra erabiliko dugu banaketa normal TIPIKO hau adierazteko.

Kontutan hartu behar dugu banaketa normal tipikoen propietate garrantzitsuena kurba eta abzisa-ardatzen artean gelditzen diren zatien azalera 1 izatea dela. Orduan batezbestekotik ezkerretara gelditzen den azalera 0.50 izango da, eta batezbestekotik eskuinera gelditzen den zatiena ere 0.50koa izango da.

1.1. Aldagaien tipifikazioa

Lehen esan dugunez, banaketa normal guztiak baliokideak dira. Puntuen arteko baliokidetasun hauek lortzeko bi prozesu ezberdin ikusiko ditugu.

Edozein banaketa normalen balioei, banaketa normal tipikoan N (0,1) dagokion puntuazioa kalkulatzen saiatuko gara lehen kasu honetan.

Eman dezagun banaketa normal bati jarraitzen zaion X aldagai bat daukagula, eta gure aldagaien parametroak 100 eta 15 direla. Honek X-en batezbestekoa 100 eta desbidazio standarda 15 direla esan nahi du. Hau adierazteko honela idatziko dugu: $X \rightarrow N\ (100,15)$.

X aldagai honen balio batzuk badauzkagu, eta puntu hauei, banaketa normal arruntean, N (0,1)ean, dagozkion puntuazioak lortu nahi baditugu, ondoko bi pauso hauek eman beharko ditugu: lehengoan, oinarrizko balio aldagaien batezbestekoa kenduko diegu, eta bigarren pausoan lortutako emaitza desbidazio tipikoaz zatituko dugu.

Aipatutako pausoak formula honen bidez egingo ditugu:

$$Z = \frac{X - \mu}{\sigma}$$

N (100,15) banaketako puntu bat tipifikatu nahi badugu, egin behar dugun aldaketa ondoko hau izango da:

$$Z = \frac{X - 100}{15}$$

Prozesu honi aldagaiaren TIPIFIKAZIO-prozesu deituko diogu.

Batzuetan kontrako prozesua interesatuko zaigu, hau da, normal tipikoan emandako puntuazio bati edozein banaketa normalean dagokion balioa bilatzea. Prozesu honi DESTIPIFIKAZIO-prozesu deituko diogu. Honen prozesua tipifikazioaren kontrakoa da. Beraz, Z puntuazio tipiko baten baliokidea beste edozein banaketari N (μ,σ) jarraitzen

zaion X aldagai batean kalkulatu nahi badugu, prozesua hurrengoa izango da:

$$X = \sigma * Z + \mu$$

Z puntuazio tipiko bat, N (50,7) banaketari dagokion beste X batean aldatu nahi badugu, egin behar dugun aldaketa hurrengoa izango da:

$$X = 7 * Z + 50$$

1.2. Taulen erabilpenerako adibideak

Banaketa normalaren taulei bi ikuspuntu ezberdinetatik edota osagarritatik begira diezaiekegu.

Lehenengo kasuan, aldagai baten balioa emanda balio hau baino handiagoak edo txikiagoak direnen portzentaia edo probabilitatea kalkulatzeko erabil dezakegu. Kasu honetan emandako puntuen azpitik edo gainetik geratzen den zatien azalera kalkulatu beharko dugu. Puntua ezagutzen dugunez, badakigu gure probabilitatea zein zutabe eta zein lerrotan aurkituko dugun.

Bigarrena, alderantzizkoa izango litzateke. Probabilitatea edo azalera badakigu, eta kalkulatu behar duguna zera da: zein da bere ezkerrera edo eskuinera gelditzen den azalera emandakoa duen puntua. Kasu honetan badakigu zein den gure probabilitatea eta begiratu beharko duguna zein zutabe eta zein lerrotan dagoen da.

Banaketa normal tipikoen taulak erabiliz, adibide batzuk aztertuko ditugu.

1.- Zein da Z = 1.27 puntuazio tipikoak bere ezkerretara uzten duen probabilitatea? Galdera honekin eskatzen dena zera da: puntu hau baino txikiagoak izateko dagoen probabilitatea zein den.

Galdera honen erantzuna taulak zuzenki ematen digu, Z = 1.27 balioari dagokion balioa aurkituz. Balio hau aurkitzeko, 1.2ren lerroa hartu beharko dugu, .07ri dagokion zutabean, eta hor ikusiko dugun balioa .8979 izango da. Hau da guk egin dugun galderaren erantzuna. Honek 1.27 puntu baino gutxiago lortzeko probabilitatea .8979 dela esan nahi du.

2.- Zein da banaketa normal tipikoari jarraitzen zaion Z aldagaiak, -0.64 eta 1.53 balioen artean egoteko dagoen probabilitatea?

Kasu honetan balio hauen artean gelditzen den zatien azalera kalkulatu behar dugu.

Horretarako 1.53tik ezkerretara gelditzen den azalera kalkulatuko dugu eta probabilitate edo azalera honi -0.64tik ezkerretara gelditzen den azalera edo probabilitatea kenduko diogu.

Probabilitate edo azalera hauek tauletatik zuzenean lortuko ditugu, eta emaitzak hauek dira:

1.53tik ezkerretara : .9369
-0.64tik ezkerretara : .2611

Bi balioen arteko azalera edo probabilitatea:

.9369 - .2611 = .6758

3.- Zein da banaketa normal tipiko bateko puntua, zeinak bere gainetik %72 pertsona uzten duena?

Adibide hau, aurreko kasuen kontrakoa izango da, hau da, lehen kalkulatu dugun probabilitatea puntu batetik ezkerretara gelditzen zena zen. Kasu honetan puntu bat baino handiagoak zenbat diren kalkulatu behar dugu, hau da, eskuinetara gelditzen den portzentaia edo probabilitatea. Baina ezberdintasun hau garrantzizkoa izanik, ez da kasu honetan inportanteena; baizik eta orain komentatuko duguna. Orain arte egin ditugun bi adibideetan, puntua ezaguna zen eta kalkulatzen genuena probabilitatea zen. Ostera kasu honetan dakiguna probabilitatea da eta kalkulatu behar duguna puntua izango da.

Puntu batetik eskuinera gelditzen den probabilitatea .72 baldin bada, puntu honetatik ezkerretara gelditzen dena .28-koa da; bien arteko batura 1 bait da. Orain taulan begiratu beharko duguna zera da: .28 probabilitateei gehien hurbiltzen zaion probabilitatea zein den, eta probabilitate hau aurkitu ondoren zein lerro eta zein zutabetan dagoen. Gure kasuan gehien hurbiltzen den probabilitatea .2810 da, eta -0.5en lerroan eta 0.08n zutabean dago. Beraz bilatzen genuen puntua -0.58 da.

4.- Zenbat balio du koartil arteko ibiltarterdiak banaketa normal N(0,1) arrunt batean?

Gogora dezagun koartil arteko ibiltarterdia Q_3 eta Q_1-en arteko tarteen luzeraren erdia dela. Hau da, gure ibiltarterdien luzera kalkulatzeko, Q_3 eta Q_1 kalkulatu beharko ditugu lehen pausoan, eta hortik erraz lortuko dugu gure ibiltarterdia. Q_3-ren balioa kalkulatzeko, kontuan hartu beharra daukagu, puntu honek bere ezkerretara uzten duen probabilitatea .75 dela, eta Q_1-ek uzten duena .25 izango dela. Ezaguna da, Q_3-tik batezbestekora eta Q_1-etik batezbestekora dauden distantziak berdinak direla ere, zeren eta puntu hauek simetrikoak bait dira. Beraz gure kasuan nahikoa izango da bietatik bat kalkulatzea, zeren

eta besteen balioa berdina izango bait da, baina zeinua aldaturik.

Q_3 puntua zein den jakiteko gure tauletan begiratu behar duguna zera da: .7500-ri gehien hurbiltzen zaion probabilitatea zein lerro eta zutabetan dagoen. Gure kasuan gehien hurbiltzen den probabilitatea .7485 da eta 0.6-ren lerroan dago .07-ren zutabean. Beraz bilatzen dugun puntua 0.67 izango da, hau da, $Q_3 = 0.67$, eta honen ondorioz, $Q_1 = -0.67$ izango da.

Hemendik koartil arteko ibiltarterdia:

$$\frac{Q_3 - Q_1}{2} = \frac{0.67 - (-0.67)}{2} = 0.67$$

5.- Zein da N (15,6) banaketa normalari jarraitzen zaion aldagai batek 12.7 baino puntuazio txikiagoak hartzeko dagoen probabilitatea?

N (15,7) banaketa normalaren taularik ez daukagu. Beraz N (0'1) banaketaren taulak erabiliz kalkulatu beharko dugu gure probabilitatea, eta horretarako lehenengo pausoa N (15,7) banaketan emanda dagoen 12.7 punturi banaketa tipikoan dagokion puntuazioa kalkulatzea izango da. 12.7 punturi dagokion puntuazioa lortzeko, nahikoa da gure puntuazioa tipifikatzea. Beraz hau da egingo duguna:

$$Z = \frac{12.7 - 15}{6} = -0.38$$

Eskatzen genuen probabilitatea, hau da, 12.7 puntu baino gutxiago edukitzea, eta banaketa normal tipikoan -0.38 puntu baino gutxiago edukitzea berdinak izango dira, eta azken probabilitate hau tauletatik zuzenean lortuko dugu; -0.3ren lerroan eta .08ren zutabean agertzen dena bait da: 0.3520.

Hemendik 12.7 puntu baino gutxiago lortzeko dagoen probabilitatea 0.3520 izango dela esan dezakegu.

6.- Zein da N (15,6) banaketa normalari jarraitzen zaion aldagai baten puntuazioak 6.4 eta 19.5en tartean egoteko dagoen probabilitatea?

Aurreko kasuan bezala, probabilitatea kalkulatu nahi dugun tartearen bi muturrak tipifikatuko ditugu, beren baliagarriak banaketa normal tipikoan lortzeko. Gure kasuan:

$$Z = \frac{19.5 - 15}{6} = 0.75 \qquad Z = \frac{6.4 - 15}{6} = -1.43$$

Balio tipifikatuak ezagutu ondoren, tauletan begiratuz probabilitatea kalkula

dezakegu:

0.75tik ezkerretako probabilitatea : 0.7733
-1.43tik ezkerretako probabilitatea : 0.0764
Tartearen probabilitatea : 0.7733 - 0.0764 = 0.6969

Beraz, ezkatutako probabilitatea eta orain kalkulatu dugun tarte honen probabilitateak berdinak direnez geroz, eskatutako probabilitatea 0.6969 izango da.

7.- Zein da N (15,6) banaketa normalari jarraitzen zaion X aldagai baten puntua, zeinak bere eskuineko aldean %65 pertsona uzten duen?

Kasu honetan ere taularik ez daukagu, baina hala ere banaketa normal tipikoan jakin dezakegu probabilitate hau bere eskuinetara uzten duena zein puntu den. Tauletan agertzen diren probabilitateak, beti puntuetatik ezkerretara gelditzen direnak izango dira. Hala ere probabilitatearen batura orokorra 1 dela jakinik, ezkerreko probabilitatea 1-p izango da (ezkerreko aldera): 1 - 0.65 = 0.35.

Horrela bada N(0,1)-en tauletan ezkerretara %35 uzten duen puntua -0.39 dela ikus dezakegu, hau da, puntu honek bere eskuinetara %65eko portzentaia uzten du.

Baina puntu hau banaketa normal tipikoari dagokiona da, eta gure banaketako baliagarria lortzeko puntu hau destipifikatu beharrean aurkitzen gara. Lehen aipatutako formula erabiliko dugu:

$$X = \sigma * Z + \mu$$

Gure kasuan, N (15,6) banaketara pasatzeko:

$$X = 6 * (-0.39) + 15 = 12.66$$

Beraz, gure puntua 12.66 izango da.

8.- X aldagai N(15,6) banaketa normalari jarraitzen bazaio, zenbat balio du aldagai honen koartil arteko ibiltarterdiak?

4. 'ariketan bezala, 1. eta 3. mailako koartilak kalkulatu behar ditugu, beren arteko diferentzien erdia jakiteko.

Badakigu $Q_1 = -0.67$ eta $Q_3 = 0.67$ direla banaketa normal tipikoen kasuan, eta gure banaketa normaleko balioak kalkulatzeko nahikoa izango da balio hauek

destipifikatzea, eta lortutako emaitzak hauek dira:

$$Q_1 = 6 * (-0.67) + 15 = 10.98$$
$$Q_3 = 6 * (0.67) + 15 = 19.02$$

Hemendik koartil arteko ibiltarterdien balioa atera dezakegu, ondoko formula erabiliz:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{19.02 - 10.98}{2} = 4.02$$

Banaketa normal tipikoko ibiltarterdi eta gure banaketakoen arteko erlazioa eta desbidazio tipikoen arteko erlazioak berdinak direla ikus dezakegu, hau da, gure banaketako desbidazio tipikoa 6 denez, gure banaketa, banaketa tipikoa baino 6 aldiz sakabanatuagoa dela esan dezakegu, eta proportzio berdina gertatzen da koartil arteko ibiltarterdiekin. Banaketa tipikoen kasuan 0.67 zen eta kasu honetan 4.02, eta ikus dezakeguna zera da: bien arteko zatidura 6 dela.

1.3. Normalaren bidez, banaketa binomialen hurbilketa

Atal honen hasieran aipatu dugu banaketa binomialaren probabilitateen hurbilketa lortzea izan zela banaketa normalen jatorria. Ikus dezagun hurbilketa hau nola egiten den.

Demagun txanpon bat 100 aldiz airera botatzen dugula, eta kasu bakoitzean aurpegia lortzeko dagoen probabilitatea 0.50 dela. Demagun ondoko bi probabilitate hauek kalkulatu nahi dugula: 45 aurpegi baino gutxiago lortzeko dagoen probabilitatea eta bigarren kasuan zehazki 45 aurpegi lortzeko dagoen probabilitatea. Normalean probabilitate hauek banaketa binomialen taulak erabiliz kalkulatuko genituzke, baina taula egokirik ez dugu lortuko, zeren eta gehienbat N = 20-rainokoak izaten bait dira. Beraz, hurrengo pausoan Poissonen banaketaren bidez saiatuko gara probabilitate hau kalkulatzen, baina m = N*p kasu honetan, m = 100*0.50 = 50 izango da eta Poisson-en taularik ere ez dugu edukiko. Beraz, gelditzen zaigun posibilitate bakarra probabilitate hau banaketa normalaren bidez kalkulatzea izango da.

Baina hurbilketa hau banaketa normal baten bidez egiteko, batezbesteko eta desbidazio tipikoen balioak ezagutu beharko ditugu. Badakigu edozein banaketa binomialen batezbestekoa N*p dela, gure kasuan $\mu = 100 * 0.50 = 50$, eta bariantza $\sigma^2 = N * p * (1-p)$ izango da, gure adibidean $\sigma^2 = 100*0.50*0.50 = 25$. Hemendik, desbidazio tipikoa = 5; bariantzaren erro karratua bait da.

Banaketa normalen parametroak berdinak dira, hau da, gure probabilitateak kalkulatzeko erabiliko dugun banaketa normala N (50,5) izango da.

Eskatutako lehen probabilitatea 45 aurpegi baino gutxiago lortzekoena da, hau da, 44 edo gutxiago lortzea. Banaketa diskretutik, banaketa jarraira pasatzeagatik zuzenketa bat egin behar dugu; kontutan hartu behar bait dugu banaketa binomialean, hau da, txanpona airera botatzen dugunean, ezin dezakegula 44.5 aurpegi lortu. Ez du zentzu handirik era honetako hitzak erabiltzeak. Ostera banaketa jarraian bai. Hau dela eta, banaketa normalez hitz egiterakoan, 44 edo balio txikiagoa lortzeko probabilitatea, 44.5 baino txikiagoagatik aldatuko dugu, eta balio honekin kalkulatuko dugu gure probabilitatea.

Kalkulatzeko prozesua aurreko adibideetan erabilitakoaren berdina izango da. Lehenengo, balio hau tipifikatu beharko dugu:

$$Z = \frac{44.5 - 50}{5} = -1.1$$

Eta eskatutako probabilitatea, -1.1 banaketa tipikoko puntutik ezkerretara dagoen probabilitatea izango da. Hau tauletatik lortuko dugu eta 0.1357 da, eta aipatu dugun lehenengo gertaeraren probabilitatea izango da.

Bigarren probabilitatea kalkulatzeko gauza bera egingo dugu, baina 45 aurpegi lortzeko probabilitatea kalkulatu beharrean 44.5 aurpegi baino gehiago eta 45.5 baino gutxiago lortzeko probabilitatea kalkulatuko dugu. Bi balio hauek tipifikatu eta taularen bidez probabilitatea lortuko dugu:

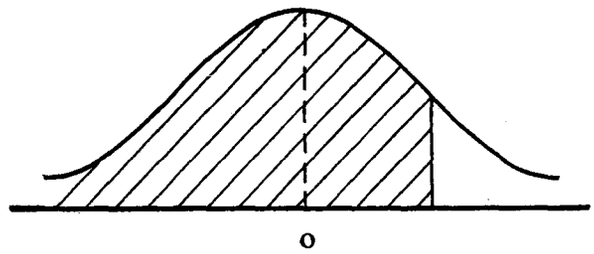
44.5 tipifikatzen badugu -1.1 lortuko dugu eta honi dagokion probabilitatea, taulan begiratuz, 0.1357 da.

45.5 tipifikatzen badugu -0.9 lortuko dugu eta honi dagokion probabilitatea, taulan begiratuz, 0.1841 da.

Eskatutako probabilitatea 0.1841 - 0.1357 = 0.0484 da.

Ondoren banaketa normalaren taulak azaltzen ditugu, bi zatitan banatuta: balio positiboentzat eta balio negatiboentzat. Balio positiboentzako taulak 0.5000 baino probabilitate handiagoak ematen dizkigu beti eta balio negatiboenak berriz 0.5000 baino txikiagoak.

BANAKETA NORMAL TIPIKOEN N(0,1) TAULA
(BALIO POSITIBOENTZAT)

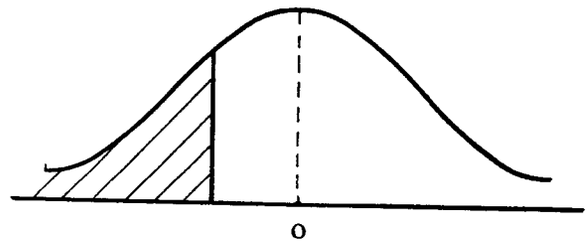


Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5039	.5079	.5119	.5159	.5199	.5239	.5279	.5318	.5358
0.1	.5398	.5438	.5476	.5517	.5556	.5596	.5635	.5674	.5714	.5753
0.2	.5792	.5831	.5870	.5909	.5948	.5987	.6025	.6064	.6102	.6140
0.3	.6179	.6217	.6255	.6293	.6330	.6368	.6405	.6443	.6480	.6517
0.4	.6555	.6591	.6627	.6664	.6700	.6736	.6772	.6808	.6843	.6879
0.5	.6914	.6949	.6984	.7019	.7045	.7088	.7122	.7156	.7190	.7224
0.6	.7257	.7290	.7323	.7356	.7389	.7421	.7453	.7485	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7733	.7763	.7793	.7823	.7852
0.8	.7881	.7910	.7938	.7967	.7995	.8023	.8051	.8078	.8107	.8132
0.9	.8159	.8185	.8212	.8238	.8263	.8289	.8314	.8339	.8364	.8389
1.0	.8413	.8437	.8461	.8485	.8508	.8531	.8554	.8576	.8599	.8621
1.1	.8643	.8665	.8686	.8707	.8728	.8749	.8769	.8790	.8810	.8829
1.2	.8849	.8868	.8887	.8906	.8925	.8943	.8961	.8979	.8997	.9014
1.3	.9032	.9049	.9065	.9082	.9098	.9114	.9130	.9146	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9250	.9264	.9278	.9292	.9305	.9318
1.5	.9331	.9344	.9357	.9369	.9382	.9394	.9406	.9417	.9429	.9440
1.6	.9452	.9463	.9473	.9484	.9495	.9505	.9515	.9525	.9535	.9544
1.7	.9554	.9563	.9572	.9581	.9590	.9599	.9608	.9616	.9624	.9632
1.8	.9640	.9648	.9656	.9663	.9671	.9678	.9685	.9692	.9699	.9706
1.9	.9712	.9719	.9725	.9732	.9738	.9744	.9750	.9755	.9761	.9767
2.0	.9772	.9777	.9783	.9788	.9793	.9798	.9803	.9807	.9812	.9816
2.1	.9821	.9825	.9830	.9834	.9838	.9842	.9846	.9850	.9853	.9857
2.2	.9861	.9864	.9867	.9871	.9874	.9877	.9880	.9884	.9887	.9889
2.3	.9892	.9895	.9898	.9901	.9903	.9906	.9908	.9911	.9913	.9915
2.4	.9918	.9920	.9922	.9924	.9926	.9928	.9930	.9932	.9934	.9936
2.5	.9937	.9939	.9941	.9943	.9944	.9946	.9947	.9949	.9950	.9952
2.6	.9953	.9954	.9956	.9957	.9958	.9959	.9960	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9972	.9973
2.8	.9974	.9975	.9976	.9976	.9977	.9978	.9978	.9979	.9980	.9980
2.9	.9981	.9981	.9982	.9983	.9983	.9984	.9984	.9985	.9985	.9986

Beste probabilitate batzuk:

3.0	.99865	3.5	.9997674	4.0	.9999683
4.5	.9999966	5.0	.9999997		

BANAKETA NORMAL TIPIKOEN N(0,1) TAULA
(BALIO NEGATIBOENTZAT)



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
- 0.0	.5000	.4961	.4921	.4881	.4841	.4801	.4761	.4721	.4682	.4642
- 0.1	.4602	.4562	.4524	.4483	.4444	.4404	.4365	.4326	.4286	.4247
- 0.2	.4208	.4169	.4130	.4091	.4052	.4013	.3975	.3936	.3898	.3860
- 0.3	.3821	.3783	.3745	.3707	.3670	.3632	.3595	.3557	.3520	.3483
- 0.4	.3445	.3409	.3373	.3336	.3300	.3264	.3228	.3192	.3157	.3121
- 0.5	.3086	.3051	.3016	.2981	.2955	.2912	.2878	.2844	.2810	.2776
- 0.6	.2743	.2710	.2677	.2644	.2611	.2579	.2547	.2515	.2483	.2451
- 0.7	.2420	.2389	.2358	.2327	.2297	.2267	.2237	.2207	.2177	.2148
- 0.8	.2119	.2090	.2062	.2033	.2005	.1977	.1949	.1922	.1893	.1868
- 0.9	.1841	.1815	.1788	.1762	.1737	.1711	.1686	.1661	.1636	.1611
- 1.0	.1587	.1563	.1539	.1515	.1492	.1469	.1446	.1424	.1404	.1379
- 1.1	.1357	.1335	.1314	.1293	.1272	.1251	.1231	.1210	.1190	.1171
- 1.2	.1151	.1132	.1113	.1094	.1075	.1057	.1039	.1021	.1003	.0986
- 1.3	.0986	.0951	.0935	.0918	.0902	.0886	.0870	.0854	.0838	.0823
- 1.4	.0808	.0793	.0778	.0764	.0750	.0736	.0722	.0708	.0695	.0682
- 1.5	.0669	.0656	.0643	.0631	.0618	.0606	.0594	.0583	.0571	.0560
- 1.6	.0548	.0537	.0527	.0516	.0505	.0495	.0485	.0475	.0465	.0456
- 1.7	.0446	.0437	.0428	.0419	.0410	.0401	.0392	.0384	.0376	.0368
- 1.8	.0360	.0352	.0344	.0337	.0329	.0322	.0315	.0308	.0301	.0294
- 1.9	.0288	.0281	.0275	.0268	.0262	.0256	.0250	.0245	.0239	.0233
- 2.0	.0228	.0223	.0217	.0212	.0207	.0202	.0197	.0193	.0188	.0184
- 2.1	.0179	.0175	.0170	.0166	.0162	.0158	.0154	.0150	.0147	.0143
- 2.2	.0139	.0136	.0133	.0129	.0126	.0123	.0120	.0116	.0113	.0111
- 2.3	.0108	.0105	.0102	.0099	.0097	.0094	.0092	.0089	.0087	.0085
- 2.4	.0082	.0080	.0078	.0076	.0074	.0072	.0070	.0068	.0066	.0064
- 2.5	.0066	.0061	.0059	.0057	.0056	.0054	.0053	.0051	.0050	.0048
- 2.6	.0047	.0046	.0044	.0043	.0042	.0041	.0040	.0038	.0037	.0036
- 2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0028	.0027
- 2.8	.0026	.0025	.0024	.0024	.0023	.0022	.0022	.0021	.0020	.0020
- 2.9	.0019	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014

2. *BANAKETA NORMALEAN OINARRITUTAKO BANAKETAK*

Ondoren ikusiko ditugun hiru banaketak, banaketa normalari jarraitzen zaien aldagaietan oinarrituak izango dira. Banaketa hauek garrantzi handikoak dira inferentzia estatistikoan ikusten diren hipotesi-testak erabiltzeko.

Hauetatik lehenengo ikusiko duguna Ji karratuaren (X^2) banaketa izango da.

2.1. *X^2-aren BANAKETA*

X^2-aren banaketaren iturburua N (0,1)-i jarraitzen zaizkien aldagai desberdinetan dago. Izan bedi $Z_1, Z_2, Z_3, \ldots Z_n$ beren artean aske (independente) diren aldagaiak, zeren eta aldagai bakoitza N(0,1) banaketari jarraitzen bait zaie. Aldagai aleatorio hauen karratuak batuta lortuko dugun aldagaiak X^2 banaketa, n askatasun graduekin, segituko duela esango dugu. Hau da, $X^2_n = Z^2_1 + Z^2_2 + Z^2_3 + \ldots + Z^2_n$ N ASKATASUN-GRADUKO JI KARRATUAREN BANAKETAri jarraitzen zaio.

Banaketa honen dentsitate-funtzioen formula ikertzen ez gara hasiko, zeren eta gure helburuak gainditzen bait dira. Kontutan hartu beharko duguna zera bakarrik da:

X^2_n-aren banaketari jarraitzen zaion aldagai baten batezbestekoa eta bariantza hauek direla:

$$E(X) = n, \qquad VAR(X) = 2n.$$

Ezagutzea komeni zaigun bigarren propietatea hau da:

X aldagaiak n_1 askatasun graduko X^2-ari jarraitzen bazaio eta Y aldagaiak n_2 askatasun graduko X^2-ari jarraitzen bazaio eta elkarrekiko aldagai askeak (independente) badira, X + Y aldagaiak $n_1 + n_2$ askatasun graduko X^2-ari jarraituko zaizkio.

Hau da $X \rightarrow$ Ji karratua n_1 askatasun-graduekin
$X \rightarrow$ Ji karratua n_2 askatasun-graduekin

Eta aldagai biak askeak badira:

$X+Y \rightarrow$ Ji karratua $n_1 + n_2$ askatasun-graduekin

Hurrengo orrialdean aurkezten dugun taulak, alde batetik askatasun-graduen zenbakia eta bestetik probabilitateak eskaintzen dizkigu, taularen barruan agertzen diren balioak bilatutako balioei dagozkienak izanez.

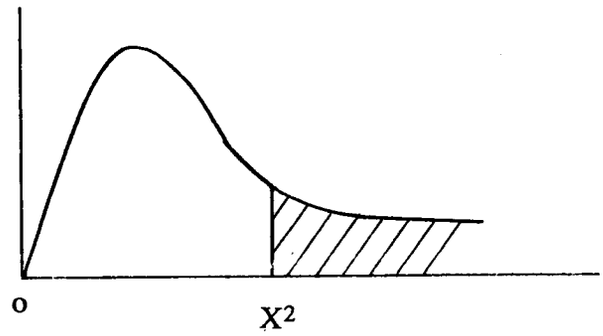
Goiko aldean agertzen diren probabilitateak puntutik eskuin aldera geratzen diren balioei dagozkie, edo berdina dena, aldagaiak puntu hori baino balio handiagoak lortzeko duen probabilitateari dagokio.

Horrela, 12 graduko askatasuna daukan eta Ji karratuari jarraitzen zaion aldagai batek 9.03 puntu baino balio handiagoak lortzeko dagoen probabilitatea jakin nahi badut, era honetara begiratu beharko dut. Lehenengo bilatu beharko dudana zera da: 12 graduri dagokion lerroa, eta lerro honetan bilatu beharko dut 9.03 balioa. Balio hau 0.70en zutabean dago. Beraz eskatutako probabilitatea 0.70 izango da.

15 graduko askatasuna daukan eta Ji karratuari jarraitzen zaion aldagai bateko puntu bat bilatzen badugu, zeinak bere eskuinera 0.95 probabilitatea uzten duen, era honetara begiratu beharko dugu: kasu honetan 15en lerroan eta 0.95eko zutabean aurkituko dugu gure balioa, eta hau 7.26 izango da.

Azkenik 10 graduko Ji karratuari jarraitzen zaion aldagai bat 4.86 eta 21.16 tartean egoteko dagoen probabilitatea kalkulatu nahi badugu, bi balio hauei dagozkien probabilitateak kalkulatuko ditugu lehen pausoan, eta balio hauek 0.90 eta 0.02 direla jakin ondoren, puntu bien arteko probabilitatea lortutako bi probabilitateen kendura izango da, hau da: 0.90 - 0.02 = 0.88.

JI KARRATUAREN TAULA



N	P: .95	.90	.50	.10	.05	.02	.01	.001
1	.0039	.016	.46	2.71	3.84	5.41	6.64	10.83
2	.10	.21	1.39	4.60	5.99	7.92	9.21	13.82
3	.35	.58	2.37	6.25	7.82	9.84	11.34	16.27
4	.71	1.06	3.36	7.78	9.49	11.67	13.28	18.46
5	1.14	1.61	4.35	9.24	11.07	13.39	15.09	20.52
6	1.64	2.20	5.35	10.64	12.59	15.03	16.81	22.46
7	2.17	2.83	6.35	12.02	14.07	16.62	18.48	24.32
8	2.73	3.49	7.34	13.36	15.51	18.17	20.09	26.12
9	3.32	4.17	8.34	14.68	16.92	19.68	21.67	27.88
10	3.94	4.86	9.34	15.99	18.31	21.16	23.21	29.59
11	4.58	5.58	10.34	17.28	19.68	22.62	24.72	31.26
12	5.23	6.30	11.34	18.55	21.03	24.05	26.22	32.91
13	5.89	7.04	12.34	19.81	22.36	25.47	27.69	34.53
14	6.57	7.79	13.34	21.06	23.68	26.87	29.14	36.12
15	7.26	8.55	14.34	22.31	25.00	28.26	30.58	37.70
16	7.96	9.31	15.34	23.54	26.30	29.63	32.00	39.29
17	8.67	10.08	16.34	25.77	27.59	31.00	33.41	40.75
18	9.39	10.86	17.34	25.99	28.87	32.35	34.80	42.31
19	10.12	11.65	18.34	27.20	30.14	33.69	36.19	43.82
20	10.85	12.44	19.34	28.41	31.41	35.02	37.57	45.32
21	11.59	13.24	20.34	29.62	32.67	36.34	38.93	46.80
22	12.34	14.04	21.34	30.81	33.92	37.66	40.29	48.27
23	13.09	14.85	22.34	32.01	35.17	39.97	41.64	49.73
24	13.85	15.66	23.34	33.20	36.42	40.27	42.98	51.18
25	14.61	16.47	24.34	34.38	37.65	41.57	44.31	52.62
26	15.38	17.29	25.34	35.56	38.88	42.86	45.64	54.05
27	16.15	18.11	26.34	36.64	40.11	44.14	46.96	55.48
28	16.93	18.94	27.34	37.92	41.34	45.42	48.28	56.85
29	17.71	19.77	28.34	39.09	44.56	46.69	49.59	58.30
30	18.49	20.60	29.34	40.26	43.77	47.96	50.89	59.70

2.2. Student-en T banaketa

Banaketa normalean oinarritutako bigarren banaketa hau bi aldagaien bidez lortuko dugu, batak N (0,1) banaketa normalari eta bestea X^2-aren banaketari jarraitzen zaiolarik.

Banaketa hau, bi laginen arteko batezbestekoen konparazioa egiteko erabiltzen da gehienbat.

Beraren formula hau litzateke:

$$T_n = \frac{Z}{\sqrt{\frac{X^2 n}{n}}}$$

non:

Z, N(0,1)ri jarraitzen zaion aldagaia den.
X^2, n askatasun-graduko X^2-aren banaketari jarraitzen zaion aldagaia den.

T banaketan ere, askatasun-graduaz hitz egin beharko dugu, eta hauek Ji karratuari jarraitzen zaizkion X aldagaiarekin bat datoz.

Probabilitate- edo dentsitate-formula aztertzen ez gara hasiko. Banaketa hauen bi propietate bakarrik ikusiko ditugu:

.- 0 balioa duen batezbestekoarekiko simetrikoa da.
.- Askatasun-graduak handitzen diren neurrian banaketa normalari gehiago hurbiltzen zaizkio, honela askatasun-graduak 120 direnean lortzen den hurbilketa oso haundia izanik. Askotan, askatasun-graduak 30 baino handiagoak direnean, kurba normalaren bidez egiten den hurbilketa erabiltzen da.

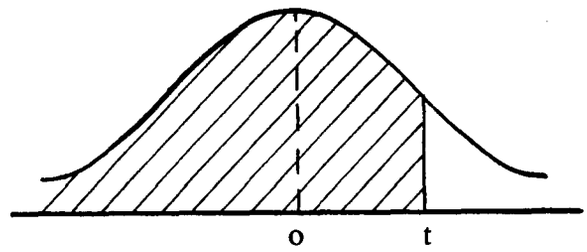
TAULAK ERABILTZEKO MODUA

Hurrengo orrian agertzen den T banaketaren taulak, ezkerreko zutabean askatasun--graduak zehazten ditu. Erabili nahi dugun bakoitzean, gure banaketaren askatasun-graduak ikusi beharko ditugu. Taulan oso probabilitate txikia duten balioak agertzen zaizkigu; hauek izango bait dira hipotesi-testetan erabiliko ditugun balioak.

Honela, adibidez, 12 askatasun-graduko Student-T bati jarraitzen zaion aldagai bat baldin badugu eta 0.01 probabilitatea bere eskuinera uzten duen puntua jakin nahi badugu, 12 askatasun-graduko lerroa aukeratu beharko dugu eta ondoren 0.01 zutabean begiratuz, behar dugun balioa lortuko dugu. Kasu honetan 2.68 izango da.

Aldiz, gure aldagaiaren balioa 1.33 baino txikiagoa izateko probabilitatea jakin nahi badugu, 20 askatasun-gradua duen aldai batean 20 graduari dagokion lerroan aurkituko dugu balio hau, eta 0.10 zutabean aurkitzen dela ikus dezakegu. Hau izango da aldagaiak 1.33 balioa baino handiagoak lortzeko probabilitatea. Beraz guk jakin nahi dugun probabilitatea 1 - 0.10 = 0.90 izango da.

STUDENT-EN TAULA



N \ P=	.75	.90	.95	.975	.990	.995
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7276	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9207
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3161	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6922	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3061	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4114	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0221	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896
60	0.6789	1.2957	1.6713	2.0002	2.3898	2.6602
80	0.6779	1.2918	1.6639	1.9901	2.3741	2.6387
100	0.6771	1.2903	1.6601	1.9838	2.3652	2.6261
200	0.6761	1.2857	1.6532	1.9721	2.3448	2.6012
500	0.6751	1.2831	1.6479	1.9650	2.3343	2.5858
∞	0.6742	1.2803	1.6451	1.9600	2.3258	2.5761

2.3. Snedeckor-en F banaketa

Banaketa normalean oinarrituriko hirugarrena Snedeckor-en banaketa da. Bariantzaren analisian erabiltzen da eta baita bi lagineko bariantzak konparatzeko ere.

F-ren banaketa honela definituko dugu: Ji karratuari jarraitzen zaizkion bi aldagaien zatiketa da. X, aldagai aleatorio bat izanik (n askatasun-graduko Ji karratu baten arabera banatuta) eta Y beste aldagai aleatorio bat izanik (hau m askatasun-graduko Ji karratuaren banaketari jarraitzen zaiona) ondoren definituriko F aldagaia n eta m askatasun-graduko Snedeckor-en banaketari jarraituko zaio.

$$F = \frac{X/n}{Y/m}$$

$F_{n,m}$-ri jarraitzen zaio.

Banaketa honekin lan egiterakoan, askatasun-graduak zehaztu beharko ditugu; bai zenbakitzailearenak, bai zatitzailearenak.

Noski, Snedeckor-en banaketari jarraitzen zaion aldagai batek balio negatiborik ezingo du hartu, X^2-aren arabera banaturiko aldagai batek balio negatiboak hartu ezin dituen bezala.

Banaketa honen probabilitate- edo dentsitate-funtzioa aztertzen ere ez gara hasiko. Taulak erabiltzeko era bakarrik aztertuko dugu.

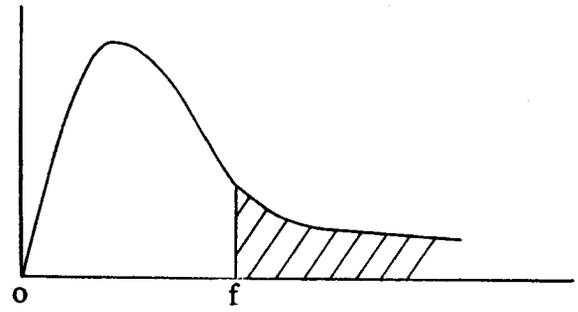
Ondoko orrialdean aurkezten dugun taulan aldagaien balioak bakarrik azaltzen ditugu, beren eskuinera 0.01 eta 0.05 probabilitateak dituztenenak; banaketa hau hipotesi-testetan eta konfidantzako tarteak eraikitzeko erabiltzen bait da.

Plantea daitekeen lehen arazoa zera da: 12 eta 7 askatasun-graduko Snedeckor-en F banaketari jarraitzen zaion aldagai baten puntua kalkulatzea, zeren eta puntu honek bere eskuinetara 0.05eko probabilitatea uzten bait du.

Puntu hau kalkulatzeko zenbakitzailean 12 eta izendatzaile edo zatitzailean 7 graduari dagokion zutabe eta lerroan aurkituko dugu. Hemen bi balio agertzen dira: bata, goikoa, 0.01 probabilitateari dagokiona da eta bestea 0.05 probabilitateari dagokiona. Beraz, gure puntua 3.5 izango da.

Bigarren arazoa hau izan daiteke: 24 eta 10 askatasun-graduko Snedeckor-en F banaketari jarraitzen zaion aldagai bateko 4.3 balioak bere eskuinetara uzten dituen probabilitatea kalkulatzea. Erantzuna lehengo kasuan bezala, 24ari dagokion zutabean eta 10 balioari dagokion lerroan aurkituko dugu, eta kasu honetan agertzen diren bi balioetatik gainekaldean gure balioa agertzen denez, eskatutako probabilitatea 0.1 izango da.

SNEDECKOR-EN F BANAKETA



A.G izen-dat.	p	zenbakitzaileen askatasun-graduak								
		1	2	3	4	5	6	12	24	
	0.01	4052	4999	5403	5625	5764	5859	6160	6234	6366
1	0.05	161.4	199.5	215.7	224.5	239.1	233.9	249.9	249.0	254.3
	0.10	39.8	49.5	53.5	55.8	57.2	58.2	59.2	62.0	63.3
	0.20	9.4	12.0	13.0	13.7	14.0	14.2	14.9	15.2	15.5
	0.01	98.4	99.0	99.1	99.2	99.3	99.3	99.4	99.4	
2	0.05	18.5	19.0	19.1	19.2	19.3	19.3	19.4	19.4	19.5
	0.10	8.5	9.0	9.1	9.2	9.2	9.3	9.4	9.4	9.4
	0.20	3.5	4.0	4.1	4.2	4.2	4.3	4.4	4.4	4.4
	0.01	34.1	30.8	29.4	28.7	28.2	27.9	27.0	26.6	26.1
3	0.05	10.1	9.5	9.2	9.1	9.0	8.9	8.7	8.6	8.5
	0.10	5.5	5.4	5.3	5.3	5.3	5.2	5.2	5.1	5.1
	0.20	2.6	2.8	2.9	2.9	2.9	2.9	2.9	2.9	2.9
	0.01	21.2	18.0	16.6	15.9	15.5	15.2	14.3	13.9	13.4
4	0.05	7.7	6.9	6.5	6.3	6.2	6.1	5.9	5.7	5.6
	0.10	4.5	4.3	4.1	4.1	4.0	4.0	3.9	3.8	3.7
	0.20	2.3	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4
	0.01	11.2	13.2	12.0	11.3	10.9	10.6	9.8	9.4	9.0
5	0.05	6.6	5.7	5.4	5.1	5.0	4.9	4.6	4.5	4.3
	0.10	4.0	3.7	3.6	3.5	3.4	3.4	3.2	3.1	3.1
	0.20	2.1	2.2	2.2	2.2	2.2	2.2	2.1	2.1	2.1
	0.01	13.7	10.9	9.7	9.1	8.7	8.4	7.7	7.3	6.8
6	0.05	5.9	5.1	4.7	4.5	4.3	4.2	4.0	3.8	3.6
	0.10	3.7	3.4	3.2	3.1	3.1	3.0	2.9	2.8	2.7
	0.20	2.0	2.1	2.1	2.0	2.0	2.0	2.0	1.9	1.9
	0.01	12.2	9.5	8.4	7.8	7.4	7.1	6.4	6.0	5.6
7	0.05	5.5	4.7	4.3	4.1	3.9	3.8	3.5	3.4	3.2
	0.10	3.5	3.2	3.0	2.9	2.8	2.8	2.6	2.5	2.4
	0.20	2.0	2.0	2.0	1.9	1.9	1.9	1.9	1.8	1.8
	0.01	11.2	8.6	7.5	7.0	6.6	6.3	5.6	5.2	4.8
8	0.05	5.3	4.4	4.0	3.8	3.6	3.5	3.2	3.1	2.9
	0.10	3.4	3.1	2.9	2.8	2.7	2.6	2.5	2.4	2.2
	0.20	1.9	1.9	1.9	1.9	1.9	1.8	1.8	1.7	1.7
	0.01	10.5	8.0	6.9	6.4	6.0	5.8	5.1	4.7	4.3
9	0.05	5.1	4.2	3.8	3.6	3.4	3.3	3.0	2.9	2.7
	0.10	3.3	3.0	2.8	2.6	2.6	2.5	2.3	2.2	2.1
	0.20	1.9	1.9	1.9	1.8	1.8	1.8	1.7	1.7	1.6
	0.01	10.0	7.5	6.5	5.9	5.7	5.3	4.7	4.3	3.9
10	0.05	4.9	4.1	3.7	3.4	3.3	3.2	2.9	2.7	2.5
	0.10	3.2	2.9	2.7	2.6	2.5	2.4	2.2	2.1	2.0
	0.20	1.8	1.9	1.8	1.8	1.8	1.7	1.7	1.6	1.6

A.G izen-dat.	p	zenbakitzaileen askatasun-graduak									
		1	2	3	4	5	6	8	12	24	
	0.01	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
11	0.05	4.84	3.98	3.59	3.36	3.29	3.09	2.95	2.79	2.61	2.40
	0.10	3.23	2.86	2.66	2.54	2.45	2.39	2.30	2.21	2.10	1.97
	0.20	1.86	1.87	1.83	1.80	1.77	1.75	1.72	1.68	1.63	1.57
	0.01	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
12	0.05	4.75	3.88	3.49	3.26	3.26	3.11	3.00	2.85	2.69	2.50
	0.10	3.18	2.81	2.61	2.48	2.39	2.33	2.24	2.15	2.04	1.90
	0.20	1.84	1.85	1.80	1.77	1.74	1.72	1.69	1.65	1.60	1.54
	0.01	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
13	0.05	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
	0.10	3.14	2.76	2.56	2.43	2.35	2.28	2.20	2.10	1.98	1.85
	0.20	1.82	1.88	1.78	1.75	1.72	1.69	1.66	1.62	1.57	1.51
	0.01	8.86	6.51	5.56	5.08	4.69	4.46	4.14	3.80	3.43	3.00
14	0.05	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	3.13
	0.10	3.10	2.73	2.52	2.39	2.31	2.24	2.15	2.05	1.94	1.80
	0.20	1.81	1.81	1.76	1.78	1.70	1.67	1.64	1.60	1.55	1.48
	0.01	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
15	0.05	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
	0.10	3.07	2.70	2.49	2.36	2.27	2.21	2.12	2.02	1.90	1.76
	0.20	1.80	1.79	1.75	1.71	1.68	1.66	1.62	1.58	1.53	1.46
	0.01	8.53	6.23	6.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
16	0.05	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
	0.10	3.05	2.67	2.46	2.33	2.24	2.18	2.09	1.99	1.87	1.72
	0.20	1.79	1.78	1.74	1.70	1.67	1.64	1.61	1.56	1.51	1.43
	0.01	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
17	0.05	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
	0.10	3.03	2.64	2.44	2.39	2.22	2.15	2.06	1.96	1.84	1.69
	0.20	1.78	1.77	1.72	1.68	1.65	1.63	1.59	1.55	1.49	1.42
	0.01	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
18	0.05	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
	0.10	3.01	3.62	3.42	2.29	2.20	2.13	2.04	1.93	1.81	1.66
	0.20	1.77	1.76	1.71	1.67	1.64	1.62	1.58	1.53	1.48	1.40
	0.01	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
19	0.05	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
	0.10	2.99	2.61	2.40	2.27	2.18	2.11	2.02	1.91	1.79	1.63
	0.20	1.76	1.70	1.66	1.63	1.61	1.57	1.52	1.46	1.39	1.31
	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
20	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
	0.10	2.97	2.38	2.25	2.16	2.09	2.00	1.89	1.77	1.61	1.54
	0.20	1.76	1.75	1.70	1.65	1.62	1.60	1.56	1.51	1.45	1.37

A.G izen-dat.	p	zenbakitzaileen askatasun-graduak 1	2	3	4	5	6	8	12	24	
	0.01	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
21	0.05	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
	0.10	2.96	2.57	2.36	2.23	2.14	2.08	1.98	1.88	1.75	1.59
	0.20	1.75	1.74	1.69	1.65	1.61	1.59	1.55	1.50	1.44	1.36
	0.01	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
23	0.05	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
	0.10	2.94	2.55	2.34	2.21	2.11	2.05	1.95	1.84	1.72	1.55
	0.20	1.74	1.73	1.68	1.63	1.60	1.75	1.53	1.49	1.42	1.34
	0.01	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
25	0.05	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
	0.10	2.90	2.53	2.32	2.18	2.09	2.02	1.93	1.82	1.69	1.52
	0.20	1.73	1.72	1.66	1.62	1.59	1.56	1.52	1.47	1.41	1.32
	0.01	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
27	0.05	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
	0.10	2.90	2.51	2.30	2.17	2.07	2.00	1.91	1.80	1.67	1.49
	0.20	1.73	1.71	1.66	1.61	1.58	1.55	1.51	1.46	1.40	1.30
	0.01	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
29	0.05	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
	0.10	2.89	2.50	2.28	2.15	2.06	1.99	1.89	1.78	1.65	1.47
	0.20	1.72	1.70	1.65	1.60	1.57	1.54	1.50	1.45	1.39	1.29
	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
30	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
	0.10	2.88	2.49	2.28	2.14	2.05	1.98	1.88	1.77	1.64	1.46
	0.20	1.72	1.70	1.64	1.60	1.57	1.54	1.50	1.45	1.38	1.28
	0.01	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
40	0.05	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
	0.10	2.84	2.44	2.23	2.09	2.00	1.93	1.83	1.71	1.57	1.38
	0.20	1.70	1.68	1.62	1.57	1.54	1.51	1.47	1.41	1.34	1.24
	0.01	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.97	1.60
60	0.05	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
	0.10	2.79	2.39	2.18	2.04	1.95	1.87	1.77	1.66	1.51	1.29
	0.20	1.68	1.65	1.59	1.55	1.51	1.48	1.44	1.38	1.31	1.18
	0.01	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
120	0.05	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
	0.10	2.75	2.35	2.13	1.99	1.90	1.82	1.72	1.60	1.45	1.19
	0.20	1.66	1.63	1.57	1.52	1.48	1.54	1.41	1.35	1.37	1.12
	0.01	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00
	0.05	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00
	0.10	2.71	2.30	2.08	1.94	1.85	1.77	1.67	1.55	1.38	1.00
	0.20	1.64	1.61	1.55	1.50	1.46	1.43	1.38	1.32	1.23	1.00

8. GAIA
NEURRIA, GIZA ZIENTZIEN BARNEAN.
(B. ORDEÑANA eta B. MUNARRIZ)

1. NEURRIA. SARRERA

Neurria, beharrezkoa da zientzia guztietan, zientzia esperimentalez aritzen garenean behintzat.

Zentzu honetan, psikometriak giza jokabideari buruzko neurrien metodo eta printzipioak aztertzen ditu, hau da, adimena edota pertsonalitatearen beste edozein jokabide.

Psikometriaren barnean, neurriari buruz hitz egiten bada, giza zientzietan eginiko neurketaren emaitzaz hitz egiten ari gara.

Neurri hauek, jokabidearen behaketatik sortzen dira, edota iharduera konkretu baten aurrean eduki daitekeen jokabidetik; bai definizio teoriko batetik abiatuz, edo askotan gertatzen ohi den bezala, neurtutako ezaugarrien definizio operatiboan geratuz.

Neurgailu psikometrikoak, pertsona ezagutzen laguntzen gaitu. Pertsona bat dimentsio guztietan ezagutzea ezinezkoa da ordea. Izan ere, osorik deskribatzea ezinezkoa bait da.

Hala ere, zenbait atributu ezagut ditzakegu; bai deskribapen kualitatibo zabal bat eginez edota ezagutu nahi dugun atributuari dagozkion jokabideak kuantifikatuz.

Kuantifikazio hauek, aukeraturiko atributuen neurketa bat egitea suposatzen dute.

Thorndike-k (1970) dioen bezala, edozein arlotan neurketa bat egiterakoan gutxienez ondoko hiru pauso hauek eman beharko ditugu:

1.- Neurtu behar dugun kualitate edo atributua definitu.

2.- Neurtu nahi dugun atributua, non ager daitekeen zehaztu, hau da, kualitatea errealitatean ikusteko eragiketak zehaztu.

3.- Adierazpenak edo eta behaketak, proposizio kuantitatibo bihurtzeko, prozedurak finkatu.

Pauso hauen garapena lagungarri gertatuko zaigu psikologian eta hezkuntz zientzietan neurketa burutzerakoan sortzen diren arazoak ulertzeko. Atributua definitzeak, zer neurtu nahi dugunaren arazoa sorterazten digu. Atributu fisikoak definitzea nahikoa erraza gertatzen da, baina atributu psikiko bat definitzerakoan arazoak sor daitezke.

Adibidez, "agresibitatea" neurtu nahi badugu, lehenbizi agresibitateaz zer ulertzen dugun argitu beharko dugu eta ondoren agresibitatearen definizio teoriko garbia ematen saiatu. Definizioaren arabera, aldagai honen ezaugarriak non eta noiz agertzen diren jakingo dugu. Bi pauso hauek elkarrekin erlazionatuta daude: ezaugarriaren definizio txarra emateak, planteaturiko eragiketek atributua oker neurtzea suposa lezake.

Atributu berdin bati definizio desberdinak ematen zaizkionez, neurgailu desberdinak aurki daitezke atributu bera neurtzeko.

Bonboir-ek (1974), "adimena" hain atributu arrunta izanik ere, definizioa ematerakoan zailtasunak daudela adierazten du. Horretarako, zenbait autore amerikarri galdetu egin zien: "Zer da adimena?"

Nahiz eta batzuek antzeko ideiak eskaini, guztietan agertzen dira ezberdintasunak, eta hauek adimena neurtzeko garaian, neurgailu ezberdinak eraikitzera garamatzate.

Binet-ek, adimenaren lehen eskala metrikoa egin zuenak, ere, "adimena" bere testak neurtzen zuena zela zioen. Honekin garbi ikus daiteke, denek ohartuko duten "adimenaren" definizio bakarra ematea oso zaila dela.

Hurrengo pausoa, planteaturiko eragiketen emaitza gai kuantitatibotan adieraztea litzateke.

Hemen, unitateen berdintasunak sorterazten duen arazoarekin topo egiten dugu. Normalean, buruturiko bi edozein lani puntuazio berdina ematen zaie. Badakigu ordea, eragiketa ezberdin horiek erabakitzeko zailtasunak desberdinak izan direla.

Horiek horrela, "ezagupena"-prozesuari dagokion item batek, "aplikazio"-prozesuari dagokion beste batek baino zailtasun txikiagoa izango du. Ia beti ordea item edo galdera biek puntuazio berdina jasotzen dute.

Guzti honek, psikologia eta hezkuntz zientzietan erabiltzen diren neurrien izakera erlatiboa garbi adierazten digu, eta ondorioz, datu hauek interpretatzerakoan izan behar dugun ardura.

2. *NEURRIAREN MUGAK GIZA ZIENTZIETAN*

Giza zientzietan, atributuaren definizioa eman beharra arazo larria izan da. Definizioen zehaztasun faltan oinarrituz, neurgailuei eginiko kritikak ugari izan dira; batez ere zientzia fisikoetan eginiko neurketen konparaketa egin ondoren.

Nahiz eta autore askok egin dituzten konparaketen aurka egon, kritika hauek neurri "psiko-hezitzaileen" zentzu eta erlatibotasuna argitzeko baliagarriak suertatu direla ere esan beharra dago.

Giza zientzietan egiten den neurketari ipinitako zenbait objekzio, ondorengo aurresuposamenduetan oinarritzen da:

1.- "Materiala" den guztia neur daiteke. Ez ordea "psikikoa" dena; hau, espirituala bait da.

Argi dago ezin dugula neurketa fisiko bat izango balitz bezala neurtu, baina psikikoan ere gradu edo kantitate desberdinak badaude (adib.: gehiago edo gutxiago egokituta egon) eta ezberdintasun hauek, eskala ordinala erabiltzea behintzat posible egiten dute.

2.- Edozein neurketak kuantifikazioa suposatzen du; atributu psikikoa kualitatiboa da ordea.

Gradu ezberdinak daudela eta hauek kuantifika ditzakegula onartzen badugu, ondorioz, nahiz eta errealitate osoa ezin neurtu, hauek neur ditzakegu. Neurketa fisikoetan bezala, alderdi amankomunak neurtzen ditugu, beste alderdietan ezberdintasunak egon arren.

3.- Psikikoaren aldaketa jarraiak, hauek errepikaezin bihurtzen ditu.

Ezin daiteke errepikatu, baina aldaketa horren erritmoa neur dezakegu. Sujetua garatzen doa. Heldutasun-froga batek gaur emaitza bat ematen badigu eta aurrerago beste bat, horrek ez du esan nahi neurketak ez duela zentzurik edota emandako emaitzaren artean ezberdintasun handiak daudenik. Zenbait arlotan aldaketa handiak egon daitezke (interesak, idealak, ezagupenak...) neurtzen dugun atributuaren aldakortasunean eta egoera pertsonalen menpe.

Neurketaren egoerak neurria alda dezake, baina ez du horregatik ezeztatzen. Hau, interpretatzeko garaian kontutan hartu behar da; ez bait da hain zehatza izango.

Neurketa psikologikoak, mugak barne, metodo esperimentala erabiltzen duten giza

zientzietan beharrezkoak izango dira; jokabidearen hipotesiak egiaztatzeko bide bakarra neurketa bait da.

Bestalde gizakia bat da, eta alderdi askoren jabe: alderdi desberdinetan aldaketa jarraia duenez, kontuan izan beharko dugu neurriak, sujetuaren nortasun, biografia eta egoeraren funtzioan interpretatu beharko direla.

Aipatutako kritikatik kanpo, egia da beste asko eta asko egiten direla; itemen alborapenetan oinarrituz adibidez. Baina egia da, alborapen hauek baztertzen joateko estatistikaren aurrerapenak asko laguntzen duela ere.

Froga guztiek beharrezko dituzten analisi zehatzak, bai kuantitatibo bai kualitatiboak, giza zientzietako neurketa hobeak lortzen lagunduko dute, nahiz eta zehaztasun osoa lortzea ezinezkoa izan.

3. ESTATISTIKA ETA PSIKOMETRIEN ARTEKO ERLAZIOA

Giza zientzien arloan ikasten duten ikasle gehienek, hasieran galdera bat planteatzen dute: Zertarako balioko digu estatistikak? Pedagogia eta psikologian zehazki, arlo hauen erlazioak argia dirudi; ebaluaketak, diagnostikoak, ikerketa esperimentalak... neurgailuak erabiltzen bait dituzte eta erabilitako tresnen emaitzak ezagutzea eta ulertzea beharrezkoa da.

Tresna psikometriko guztiek, baliagarriak izateko, bi propietate eduki behar dituzte: BALIOZKOTASUNA eta FIDAGARRITASUNA. Metodo estatistikoak erabiliz galdera askoren erantzunak lor ditzakegu. Adibidez: item-ak guk nahi duguna neurtzen duen ala ez, neurketa hau ondo egiten duten ala ez, sujetuak talde barruan duen lekua eta bere test-ean duen emaitza.

Ezinezkoa da, estatistikarik gabe, neurketa psikologikorik egitea. Psikometria datu psikologiko edo psikofisikoen eskala normalizatu edo estandarizatuak lortu nahi dituen neurketa-metodoa dela onartuz, esan beharra dago ez dela nahikoa izango pertsona baten puntuazio bat ematea, erreferentzi sistema baten barnean puntuazio hau kokatzea baizik. Beraz lortutako datuak eskala normalizatu batera bildu behar dira.

Sujetu-talde baten datuak eskala psikometrikoetan jartzeko, zenbait ezaugarri bete behar dituzte.

Neurtutako datuen berezitasunen arabera (aldagai kuantitatiboak, kualitatiboak) eskala bat edo bestean taldeka daitezke. Honek ondorio bezala analisi estatistiko ezberdinak erabiltzera behartuko gaitu. Bestalde, laginaren berezitasunen arabera

(errepresentagarria izan ala ez) estatistika deskribatzailearen teknikak edo estatistika inferentzialaren teknikak erabili beharko ditugu.

Neurketa psikometrikoei aplikatuz, teknika estatistikoak era honetan sailka daitezke:

a) Errepikapen-funtzioa.

Talde normatiboan parte hartzen duten pertsona guztiak baliokideak dira. Pertsona baten erantzunen probabilitatea iragartzeko, talde guztiko datuak erabiliko dira.

b) Fidagarritasun-printzipioa.

Datuak, baldintza berdinetan, denboran zehar ez dira aldatzen. Printzipio honek eskala psikometriko batekin eginiko esperimentua errepikatzeko dagoen posibilitatea adierazten digu.

d) Aleatorizazio-printzipioa:

Garrantzi handia izango du; batez ere, lagin errepresentagarrien aukeraketa egiterakoan.

e) Aldakortasun-printzipioa.

Alde batetik, laginaren tamainu eta heterogenotasunaren funtzioan, eta bestetik, erabilitako teknika psikometrikoen funtzioan, lortutako puntuazioen aldakortasuna kalkulatzea oso komenigarria izango da.

Estatistika, zientzia matematiko bezala ezagutu behar dugu; bai maila deskribatzailean bai maila inferentzialean.

Ikerketa- edo neurketa-arazo baten aurrean, zein metodo estatistiko erabili jakin behar dugu eta baita zein balio estatistiko aplikatu ere. Estatistikaren logika ezagutu beharko da, neurketa ezberdinen arabera metodo desberdinak noiz eta nola erabili behar diren jakiteko

Analisi estatistikoen interpretazioa beharrezkoa da; bai gure lanarentzat, psikometriaren arloan, eta baita giza zientzien arloan ere. Azkenik, argi esan behar da, gizazientzien arloan ikerketa ez dela estatistikara mugatzen. Estatistikak egiten duena zera da: helburuak lortzeko, gure hipotesiak baieztatzeko, gure eskuetan tresna batzuk jartzea.

4. TRESNA PSIKOMETRIKOAREN PLANGINTZA

Edozein neurgailu eraikitzerakoan etapa edo urrats ezberdinak bete beharko ditugu. Etapa edo lanprozesu hauen helburua zera da: lortuko dugun test psikometriko honek teknika edo/eta teoria aldetik edozein testek bete behar duen baldintzak betetzea.

Etapa edo lanprozesua, eskematikoki, ondoko hau izango da:

1.- Atributu psikologiko edo pedagogikoa, argitasun eta zehaztasun handienaz definitu.

2.- Tresnaren eraiketan planteatzen diren helburuak ondo zehaztu, (testa nori dagoen zuzenduta, puntuazioekin zein inferentzi mota egitea interesatzen zaigun).

3.- Eskema operatibo bat edo lan-plangintza (espezifikazioen koadroa) osatu. Plangintza honek ondoko helburuen antolaketa orokorra izan behar du: zein helburu eta prozesuk errepresentaturik egon beharko duten, zein adierazlek zuzenduko duten item--laborazioa, nola orekatu test-luzerari buruzko erabakiak, adina eta indibiduoen ezaugarriak, denbora-faktoreak, galderen formatua...

4.- Aldez aurreko fasean hartutako erabakien arabera itemen laborazioa, tresnak behin-betirako edukiko dituen itemak baino gehiago eginez. Hau da, item-banku bat antolatuko dugu.

Laboratuak diren item-multzoetariko laginketa aleatorio edo epaile adituen bidez aukeratutako lagina hartuko da ondoko saikuntzarako.

5.- Aplikatzeko arauak zehaztu, zuzentzeko gakoak eta frogako lehenengo protokoloa laboratu. (Adibidez, material inprimatua; arkatz eta paperezko testa izatekoan sujetuei aurkeztuko zaiena).

6.- Proba pilotua edo proben aplikazioa. Tresna pasatu, emaitzak aztertu eta froga hobetzeko erabakiak hartu.

Etapa honen barruan pauso desberdinak eman beharko dira:

- Neurgailua erabiliko deneko populazioaren laginaren hautaketa: testa zuzentzen deneko populazioa adierazkorrena izatea komeni da.

- Alderdi desberdinen ebaluazioa:
. Aurkeztu den protokoloa eta aplikatzeko baldintzen egokitasuna.
. Lortutako puntuazioen analisi estatistiko-deskribatzaileak egin eta emaitzak aztertu.

. Item-analisiak.
. Fidagarritasunen lehenengo hurbilketa egin.

- Lortutako emaitzen arabera frogaren berregituraketa.

7.- Tresnaren baliazkotasuna eta fidagarritasuna kalkulatzeko, behin-betirako testaren aplikazio orokorra. Puntu honetan, edozein pertsonak lortu duen puntuazioa interpretatzeko, testaren baremazioa ere egin beharko dugu.

8.- Test-eskuliburu edo manual teknikoaren laborazioa. Eskuliburu honek ondoko puntuen zehaztasunak eduki behar ditu: neurketaren tresna zertan datzan, zelan edo nola aplikatzen den, nola puntuatu, lortutako puntuazioa nola interpretatzen den testaren kalitatea zein oinarri teoriko egiaztatutan oinarritzen den.

Edozein test erabiltzeko ikertzaileak estatistika eta psikometria buruzko oinarrizko ezagumenduak izan beharko ditu. Ezagumendu hau ez da testa ondo aukeratzeko bakarrik; baizik eta testa ondo pasatu eta lortutako emaitzak ondo interpretatzeko, bestela testen kalitatea murrizteko arriskuan egongo bait gara.

5. LAGINAREN AUKERAKETA

Edozein test baremazeko eta test baten baliozkotasuna eta fidagarritasuna kalkulatzeko, populazio orokorraren zati edo lagin batekin lana egingo dugu. Argi dago lagin honen aukeraketak daukan garrantzia. Laginak bete behar duen baldintza zera da: laginean gure neurgailuarekin gertatzen dena, populazioan gertatu behar du. Hau da, gure laginak populazioaren "adierazgarria" izan behar du.

Populazioaren lagin adierazgarri bat aukeratzeko erabiltzen diren metodoak, 12. gaian ikusiko ditugu, eta hor esaten diren baldintza eta posibilitate guztiak, hemen ere baliagarriak dira.

6. ITEM-EN ANALISIAK

Giza zientzietan erabiltzen diren testetan oinarrizko unitateak itemak dira. Item gehienak desegokiak edo akastunak badira edo adierazgarriak ez badira, item hauek osatzen duten testa ez da egokia izango.

Test bat prestatzerakoan item-ekin frogak egiten dira beraien jokaera analizatuz, populazioaren lagin adierazgarri batean. Lehen aipatu dugunez, test psikometriko bat eraikitzerakoan, lehen pausoa item-banku bat egitea da. Gero, item guztiak probatu ondoren, hauen analisi kualitatibo eta kuantitatiboak egiten dira eta analisi hauen arabera

egokienak aukeratuko dira.

Analisi logiko eta kualitatiboen helburua, item bakoitzaren egokitasuna jakitea da. Edukia ondo idatzia, era egokian adierazia eta sujetuentzat ulergarria den ala ez ere kontrolatu behar da.

Analisi kualitatiboak osatzeko hainbat analisi kuantitatibo ere egiten dira. Hauen helburuak batez ere ondokoak dira: item bakoitzaren zailtasuna neurtu, neurtzen den aldagaia diskriminatzeko duen ahalmena kontrolatu eta azkenik itemen baliozkotasuna kalkulatu.

6.1. Zailtasun-indizea

Indize honek, aztertutako itemari erantzun zuzena eman dioten sujetuen proportzioa adierazten du. Bere kalkulua (p) proportzio batetik abiatzen da eta itemen batezbesteko aritmetiko bezala uler daiteke.

$$p = \frac{\Sigma X_i}{N}$$

non:

- X_i, i pertsonen emaitza adierazten duen aldagai baten balioa izango da. 1 balioa hartzen du itema ondo erantzunda baldin badago eta 0 balioa gaizki erantzunda dagoenean.
- N itemari erantzun dioten pertsona-kopurua.

"P" zailtasun-indizearen alderantzizkoari "q" deritzo eta aztertutako itema gaizki erantzun duten sujetuen proportzioa adierazten du.

$$q = 1 - p$$

P-ren balioa zenbat eta handiago izan, hainbat eta errazagoa izango da itema. Eta noski, q balioa zenbat eta handiagoa izan, hainbat eta zailagoa izango da item-a.

Test bat osatuko duten itemak aukeratzerakoan, alde batetik itemen zailtasuna kontutan hartu beharko dugu, baina baita testen helburuak ere. Helburu hauen arabera zailtasun ezberdineko itemak aukeratuko dira.

Errendimendu- eta gaitasun-testetan indize hauek baliagarriak dira, tresna eratzeko itemek duten zailtasuna oinarritzat harturik.

6.2. Diskriminazio-indizea

Indize honek zera adierazten du: item batek ze puntutaraino sujetuen artean aldagai bat zeintzuk betetzen duten gehiago edo gutxiago desberdin dezakeen. Neurri hau testaren egiteko eraren arabera adierazten da.

Interesatzen zaigun informazioa lortzeko prozedura desberdinak daude: gehien erabiltzen direnen artean homogenotasun-indizea deritzona dugu. Indize hauek koerlazio--indizeen bidez kalkulatzen dira.

Homogenotasun-indizeak, item konkretu batek item-multzoak neurtzen duen aldagai bera noraino neur lezakeen dio. Pertsona bakoitzak itemari erantzuterakoan lortutako puntuazioa eta sujetu guztiek testan lorturiko puntuazio osoaren arteko koerlazioa lortu beharra dakar honen kalkuluak.

Erabiliko den koerlazio-mota konkretua, aztertzen diren datuen izaeraren menpean dago. Itemak bitarrak direnean, kasurik orokorrena, (ondo 1, gaizki 0), koerlazio biserial puntuala (rbp) eta koerlazio biseriala (rb) erabiliko dira.

$$rbp = \frac{\overline{X}_p - \overline{X}_t}{S_t} \sqrt{pq}$$

Non:

- $\overline{X}_p$: Itema ondo erantzuten duten pertsonek testean lortu duten puntuazioaren batezbestekoa.
- $\overline{X}_t$: Test osoaren batezbestekoa.
- S_t : Testaren desbidazio tipikoa
- p : Zailtasun-indizea.
- q : p-ren alderantzizkoa (q = 1-p).

$$rb = \frac{\overline{X}_p - \overline{X}_t}{S_t} \sqrt{p/y}$$

Non:

- $\overline{X}_p$: Itema ondo erantzuten duten pertsonek testean lortu duten puntuazioaren batezbestekoa.
- $\overline{X}_t$: Test osoaren batezbestekoa.
- S_t : Testaren desbidazio tipikoa
- p : Zailtasun-indizea.
- y : Banaketa normal tipikoan, bere ezkerratara p probabilitatea uzten duen puntuko ordenatua.

Formula bat zein bestea erabiltzean lortutako emaitzak ezin dira beraien artean konparatu; rbp-ren balioak, item berberetan rb-ak kalkulatzerakoan lortutakoak baino baxuagoak dira. Teorikoki gutxienez, rbp hobea dela esan daiteke.

Indize hauen interpretazioa edozein koerlazio-koefizienteren berdina da. Normalean aukeratuko diren itemak hauexek izango dira: alde batetik homogenotasun edo diskriminazio-indizea positiboa dutenak eta bestetik diskriminazio-ahalmen haundiena erakusten dutenak (erlazioaren indarra).

6.3. Baliozkotasun-indizea

Item bat baliagarria dela diogu egokitu zaion helburua betetzeko aproposa den neurrian, hots, testarekin neurtu nahi diren psikologi edo hezkuntz esparruetako ezaugarriak neurtzeko balio duenean. Indize hau kalkulatzeko testaren kanpoko erizpide bat hartu eta honen eta itemaren puntuazioen arteko koerlazioa kalkulatuko dugu. Homogenotasun-indizean bezala, erabiliko den koerlazio-mota erlazionatu diren datuen izaeraren menpean egongo da.

Gehien erabilitako estatistikoak, koerlazio biserial-puntuala (rbp), koerlazio biseriala (rb), f_i koefizientea eta koerlazio tetrakorikoa (rt) izango dira.

Emaitzen interpretazioa edozein koerlazio-koefizienterena bezalakoa da, koerlazio positibo eta altuak dituzten itemak hobeagoak direlarik.

6.4. Erantzun ezuzenen (distraktoreen) analisia

Distraktoreak, aukera anitzeko itemetan opzio erakargarriak baina ezuzenak izaten dira.

Ikuspuntu teorikotik, aukera zuzena hautatzen ez duten sujetu guztientat, distraktoreek erakarpen berbera edukitzea komeniko da. Baina praktikan hau errealitate bat baino gehiago desio bat da, eta distraktoreen artean hautapen-proportzio desberdinak

agertzen dira.

Erabilitako analisi estatistikoak diskriminazio-indizean erabilitakoekin bat datoz, baina hemen aukera ezuzenak kontutan hartuko dira.

Gehienetan distraktore bakoitzarentzat rb edo rbp-a lortzen da; koerlazio negatiboak lortzen direnean distraktoreak efikazak dira, hau da, diskriminatzeko botere negatibo haundia dutela esan dezakegu.

7. TESTAREN ANALISI TEKNIKOA

Item onak eta planteatutako helburuentzat egokiak direnekin osaturiko testak ona izan behar duela esan dezakegu eta edozein testek bete behar duen baldintzak beteko dituela, baina, hala ere, test osoaren efikazia egiaztatu egin beharko da.

Edozein testek beharrezko bi baldintza bete beharko ditu: fidagarria eta baliagarria izatea. Giza zientziatan erabilitako neurgailuak akastunak dira; eta ez dira ez guztiz fidagarriak, ez guztiz baliagarriak, baina akats hauek txikiak edo zilegiak diren ala ez baloratu beharko dugu, eta honela neurgailua segurtasun osoaz erabil daiteke.

7.1. Testaren fidagarritasuna

Fidagarritasunaren arazoak, neurgailu psikometrikoak lortutako neurrien kalitatea eztabaidatzen du. Testa aplikatzerakoan lortutako puntuazioek zein finkotasun eta zehaztasun daukaten jakitea interesatuko zaigu.

Test baten fidagarritasuna bi ikuspuntu desberdinetatik azter daiteke, baina azken finean biak bat etorriko dira.

1.- Neurrien egonkortasuna edo finkotasuna: test bat fidagarria da honen puntuazioak sujetu berberean neurketa desberdinetan berregin daitezkeen neurrian. Bere egiaztapenerako koerlazio-koefizienteak erabiliko ditugu; hauei fidegarritasun-koefiziente deritze.

2.- Zehaztasun eta doitasuna: testa fidagarria da, neurketaren erroreak txikiak diren neurrian; hau da, ezaugarrian sujetuen egiazko puntuazio eta puntuazio enpirikoaren artean dagoen desberdintasuna txikia denean. Interesatzen zaigun kontzeptu estatistikoa neurketaren errore tipikoa da (S_e); eta sujetu-talde batek egindako neurketa erroreen desbidazio tipikoa bezala definituko dugu. Bere kalkulua fidagarritasun-koefizientearen bidez kalkulatzen da.

7.1.1. FIDAGARRITASUN-KOEFIZIENTEA

Fidagarritasun-koefizientea teorikoki honela definitzen da: behatutako bariantzatik egiazkoa den proportzioa.

Fidagarritasun-indizea kalkulatzeko metodo desberdinak daude eta gehien erabiltzen direnak hauek dira: test-retest metodoa, era paraleloena, bi erdien metodoa, Kuder--Richardson eta Cronbach-en alfa metodoa.

TEST-RETEST METODOA

Egonkortasun-koefiziente ere deitzen zaio. Koefiziente hau pertsona berberei bi aldiz test bera pasatu eta lortutako puntuazioen artean dagoen koerlazioa izango da.

ERA PARALELOEN METODOA

Baliokidetasun-koefiziente ere esaten zaio. Test bateko bi forma paralelo edo baliokide pasatzen zaie pertsona berberei, eta lortutako puntuazioen artean dagoen koerlazioa izango da.

Metodo honek daukan oztopo bakarra zera da: era paraleloak edo baliokideak ondo definitzea. Askotan bi forma hauek lortzea oso gauza zaila izango da; forma biek berdinak izan behar bait dute ondoko gauza guzti hauetan: helburu berdinak, item-kopuru berdinak, zailtasun-indize berdinak item guztietan, hau da, zailtasun-banaketa berdinak, elementuen formatuan berdina izan behar dute, denboraren muga berdinak...

BI ERDIEN METODOA

Metodo honek eta hurrengo biek, aurrekoen ondoan abantaila bat daukate: testa behin bakarrik pasatu behar izatea.

Azken batean metodo hauek item ezberdinetan ematen diren erantzunen finkotasuna neurtzen du.

Metodo hau aplikatzeko, lehenengo pausoa gure testa bi zatitan banatzea izango da. Eta test bakarra pasatuz pertsona bakoitzarentzat bi puntuazio emango ditu. Bi puntuazio hauen arteko koerlazio-koefizientea lortzea izango da metodo honen helburua. Lortutako koefizientea, gure testen item-kopuruaren erdia daukan test baten koefizienteak emango du. Beraz koefiziente hau nolabait zuzendu egin behar da eta horretarako formula berezi bat aplikatu beharko dugu.

Test osoko fidagarritasun-koefizientea kalkulatzeko formula ugari daude, eta formula

hauetatik, Spearman-Brownek proposatutako bat aurkeztuko dugu:

$$R_{xx} = \frac{2\,R_{ip}}{1 + R_{ip}}$$

Non:

R_{xx} : Test osoen fidagarritasun-koefizientea.

R_{ip} : Bi erdien metodoaz lortutako koerlazio-koefizientea den.

Metodo honek daukan problema, testa bi zatitan banatzea izango da. Egingo ditugun bi zatiek berdinak izan behar dute; ez item-kopuru aldetik bakarrik, baizik eta beste betebehar guztietan ere bai.

KUDER-RICHARDSON-EN METODOA

Aurreko kasuan bezala metodo honek item ezberdinetan ematen diren erantzunen tinkotasuna neurtzen du. Metodo honetan ez da testa erdibitzen; item bakoitzaren ikerketa bat egiten da baizik.

Kuder-Richardson-ek aurkeztutako formuletatik ezagunean KR20 da, eta formulak honela dio:

$$R_{xx} = \frac{n}{n-1}\left(1 - \frac{\Sigma pq}{S_t^2}\right)$$

non:

- R_{xx} : Testaren fidagarritasun-koefizientea.
- n : Testaren item-kopurua.
- R^2_t : Testean lortutako puntuazioen bariantza.
- Σpq : Item bakoitzean ondo erantzun dutenen proportzioa p bada, eta gaizki erantzuten dutenen proportzioa q baldin bada, Σpq, item bakoitzean egiten dugun p eta q biderkaduren batura izango da.

CONBACH-EN ALFA METODOA

Kuder-Richardson-en metodoen jeneralizazio bezala kontsidera dezakegu metodo hau. Batez ere testa osatzen duten itemak bitarrak ez direnean erabiltzen da. Erabiliko dugun formula ondoko hau izango da:

$$\alpha = \frac{n}{n-1} \quad (1 - \frac{\Sigma S_i^2}{S_t^2})$$

non:

- α : Testaren fidagarritasun-koefizientea.
- n : Testaren item-kopurua.
- S^2_i : Item bakoitzaren bariantza.
- S^2_t : Testaren bariantza.

NEURKETAREN ERRORE TIPIKOA

Edozein testen errore tipikoa, bere fidagarritasunaren zuzenean erlazionatuta egongo da. Zenbat eta fidagarritasun handiagoa izan, orduan eta errore tipiko txikiagoa izango du. Bi koefizienteen arteko harremana ondoko formula honen bidez adieraziko dugu:

$$S_e = S_x \sqrt{1 - r_{xx}}$$

non:

- S_e : Testaren neurketaren errore tipikoa.
- S_x : Testaren desbidazio tipikoa.
- r_{xx} : Testaren fidagarritasun-koefizientea diren.

Giza zientziatan erabiltzen diren neurgailuak, beste zientziatan erabiltzen direnak baino akastunagoak dira. Arrazoia neurtzen diren aldagaietan erraz aurkituko dugu. Giza zientziatan egiten diren neurketen zorizko erroreek ematen dituzten puntuazio enpiriko eta benetazko puntuazioen arteko ezberdintasunak. Neurketaren errore tipikoek, konfidantza-–tarteak eraikitzeko posibilitatea ematen digute testean lortutako puntuazioa, benetazko puntuazio bat besterik ez bait da.

a erroreko konfidantza tartea, era honetara idatz dezakegu:

$$X_e - Z_\alpha * S_e \leq X_b \leq X_e + Z_\alpha * S_e$$

non:

X_e : Testean lortu duen puntuazio zuzena edo enpirikoa

Z_α : Banaketa normalean bere eskuinera $\alpha/2$-ko probabilitatea uzten duen puntua.

S_e : Neurriaren errore tipikoa diren.

7.2. Testaren baliozkotasuna

Test baten baliozkotasunari buruz hitz egiten dugunean, neurtu nahi duena benetan zein puntutaraino neurtzen duen adierazi nahi da.

Hiru eratako baliozkotasuna erabiltzen dira: edukinezkoa, aurresan edo erizpideari dagokiona eta azkenik konstruktoarena. Kontzeptu hauek aztertuko ditugu orain.

EDUKINEN BALIOZKOTASUNA

Batez ere ERIZPIDE eta ERRENDIMENDUKO testetan erabiltzen da kontzeptu hau.

Testaren edukina aztertzen da, ea gure testarekin neurtzen diren ezaugarriak neurtu nahi dugun aldagaiaren adierazgarri diren ala ez ikusteko.

Test bat zentzu honetan aztertzeko, adituen eritziaren bidez analizatzen da, eta analisi kualitatibo baten bidez egiten da.

ERIZPIDEAREN BALIOZKOTASUNA

Edozein pertsonaren jokabidea aurresateko gure testak duen gaitasuna neurtzen du erizpidearen baliozkotasunak. Test batekin bilatzen dena, batzuetan gutxienez, zera da: pertsona baten jokabidea aurretik zenbateraino jakin dezakegun. Hau da kasu honetan neurtu nahi duguna. Enpirikoki neurtuko dugu balidezia hau, eta erabiliko dugun prozesua, kanpoko beste erizpide batekin koerlazio-koefizientea kalkulatzea izango da. Kanpoko erizpide hau ondo aukeratu behar dugu, eta aurresanen erizpidearekin bat etorri beharko du.

Baliozkotasun hau erabiltzen da, adibidez, pertsonen aukeraketa-prozesuetan, edota pertsona baten gaitasunak neurtzerakoan.

KONSTRUKTOAREN BALIOZKOTASUNA

Ikuspegi teoriko eta zientifiko aldetik baliozkotasuna motarik nagusiena da; psikologi aldetik testak neurtzen duenaren esanahia ulertzeko saio zientifikoa adierazten duelarik.

Konstruktoaren baliozkotasuna konplexua eta aberatsa da, bai konzeptualizazio--mailan, bai kalkulu-metodoetan. Bere baliozkotasunerako prozedurak tresnaren balioa teoria psiko-pedagogikoaren fidagarritasuna bezain ondo frogatzen du.

Azken finean, zertan datza? Laburtuaz, hipotesiak formulatu ondoren testean

lortutako puntuazioetan teoriak ezartzen duen erlazio berbera mantentzen den ala ez egiaztatu behar da.

Horretarako hipotesia egiaztatzen lagunduko gaituen diseinu esperimentala behar dugu. Baina kontutan hartu behar dugu diseinu esperimental desberdinetan erabil daitezkeen estatistikoak eta ezarriko ditugun hipotesien kopurua mugagabea dela.

Konstruktoaren baliozkotasuna ez da finkatzen esperientzia bakar batez. Egiaztapen honek informazioen metaketa bat behar du, testak neurtzen duen jokabideen zatia hobeto ulertzeko.

8. TESTAREN BAREMAZIOA

Testuaren baliozkotasuna eta fidagarritasuna frogatu ondoren neurgailu horrekin lortutako puntazioen esanahia mugatzea garrantzitsua da. Puntuazio zuzenak ia ez du adierazpen garbirik sujetuarengan. Beharrezkoa da taldera jotzea, dagokion populazioarekin egindako konparaketaren bidez interpretatzeko.

Froga bat barematzea edo tipifikatzea zera da: puntuazio zuzenak puntuazio normatibo bihurtzea alderdi horretan sujetu bati dagokion populazio edo talde normatiboan duen erlatibozko kokapena jakiteko.

Sujetu batek lortu duen puntuazio zuzenen interpretazioa errazteko abiapuntutzat zenbait arau-mota hartuko dugu. Nagusienak hauek dira: arau kronologikoak, kurtso edo graduko arauak, perzentilak eta arau tipikoak.

8.1. Arau kronologikoak

Oinarrizko prozedura, adin bakoitzarentzat baremazio berezi bat lortzea izango da. Honetarako adin bakoitzeko pertsonek lortu dituzten puntuazioen batezbestekoak eta baremazioak kalkulatzen dira.

Baremazio-mota hau haurtzaroko periodo eta adimen-testetan soilik erabili da; adinarekiko begibistako aldaketak nabaritzen dituzten seinaletara mugatzen den aplikazioa bait du.

8.2. Gradu-arauak

Aurrekoetatik erreferentzi taldea adimenean oinarritzen ez delako desberdintzen dira; sujetuen ikasturtean baizik.

Baremazio-mota hau ikasturte bakoitzeko pertsonek lortu dituzten puntuazio zuzenen

batazbestekoak kalkulatuz lortzen da.

8.3. Pertzentilak

Puntuazio zuzenak pertzentil bihurtzen dira, non pertzentilak puntuazio enpiriko mugatu baten azpitik dagoen baremazio baten proportzioa adierazten bait du.

Sujetu batek testean 70eko puntuazio zuzena lortzen badu eta honek 80. Pertzentila suposatzen badu, sujetuaren puntuazioa baino txikiagoak lortu dituztenenak % 80 direla adierazten digu.

8.4. Arau tipikoak

Puntuazio zuzenen tipifikazio bat egiten da, eta pertsonen puntuazioen arteko distantziak, desbidazio tipikoa unitate bezala hartuta adierazten du.

Aurreko kapituluan ikusi dugun bezala, puntuazio tipikoetara edozein puntuazio zuzen pasatzeko, ondoko aldaketa egiten da:

$$Z_i = \frac{X_i - \overline{X}}{S_t}$$

non:

- X_i : Puntuazio zuzena.
- $\overline{X}$: Talde honetako batezbestekoa testean.
- S_t : Talde honetako desbidazio tipikoa testean.

9. TESTAREN ESKULIBURU TEKNIKOA

Testaren eskuliburua testari buruz beharrezkoa den informazioa gordetzen duen liburu edo txosten bat izango da. Bere barnean testaren informe adierazkor bat eduki behar du, hala nola, neurketa horren kontzeptualizazioa, helburuak, oharrak, zuzenketa erak eta gaiaren ezaugarri psikometrikoak adierazten dituzten datu teknikoak.

Edozein eskuliburu teknikok honako datu psikimetrikoren informazioa gorde behar du:

- Datu teknikoak eman dizkiguten sujetuen laginaren deskribaketa.

- Item analisietan lorturiko ondorioen aurkezpena. (Zailtasun-indizeak sartzen dira

askotan eta baita ahalmen baztertzaile edo diskriminatzaileari buruzko datuak ere).

- Fidagarritasun-datuen emaitzak, hauek lortzeko erabili dugun metodoa edo metodoak azalduz eta zein taldetan lortu den esanez.

- Testearen baliozkotasunari buruzko datuen emaitzak, nola egiaztatu dugun deskribatuz: zein erizpide, nola neurtu den erizpidea, nork egin dituen analisi logikoak,...

- Baremazio-taulen aurkezpena, talde normatiboarekiko puntuazioen interpretazioa nolakoa izan den adieraziz.

Eskuliburu batean agertzen diren testaren datu tekniko guztiek, testa zuzentzen zaion populazioaren talde adierazgarri batean lortutakoak izan beharko dute.

Eredu-frogan edo froga pilotuan lortutako datuak ez dira eskuliburuetan agertzen; eredu-frogetan erabiltzen diren talde adierazgarriak baino talde zabalagoari pasatu diogun testean lortutako azken emaitzak baizik.

9. GAIA
ESTATISTIKA ETA INFORMATIKA: SPSS/PC+ programa-paketea
(LUIS LIZASOAIN HERNANDEZ)

Zientziako arlo askotan gertatu den bezala, Estatistikan informatikaren sorrerak garrantzi handia eduki du.

Nahiz eta atal honen helburua izan ez, informatikak, estatistikaren garapenean eduki dituen bi epoka garrantzitsuenak aipatuko ditugu.

Lehenengoan, analisi estatistiko bakoitza erabiltzeko FORTRAN, BASIC edo beste edozein programazio-lengoaiaz analisi horri zegokion programa berezi bat egiten zen. Horretarako lehenengo pausoan, erabiliko zen algoritmoen diseinua egin behar zen eta gero algoritmo hau erabiliko genuen "hizkuntzara" itzuli behar genuen. Prozesu guzti hau egitea nahikoa astuna, eta oso pertsona gutxiren eskuetan zegoen, zeren eta estatistikatik aparte informatikako lengoaiak menperatu behar ziren.

Bigarren epokan, hau da, gaur egun, estatistikako analisiak burutzeko gure eskuetan dauzkagu programa-pakete estatistikoak. Programa-pakete hauek estatistika eta informatika-gaietan pertsona adituak egindako azpiprograma-pilo batez osatuak daude. Pakete hauek metodo eta analisi estatistiko gehienak erabiltzeko baliagarriak izango dira.

Esan dugunez, lehenengo metodoaz gure analisi estatistikoak burutzeko informatika eta informatikaren lengoaien ezagupen handia eduki behar da. Beraz lehenengo epokatik bigarrenera pasatu garenenan, gauzak asko erraztu dira; batez ere edozein programa-pakete erabiltzerakoan, programak EGINDA bait daude. Programa-pakete bat erabiltzerakoan ikertzaileak egiten duena zera da: bere helburua lortzeko behar duen azpiprograma aukeratu eta hau erabiltzeko aginduak idatzi. Agindu hauek idaztea oso gauza erraza izango da. Arrazoia garbia da; zenbat eta pakete erabilgarriagoa izan, hainbat eta gehiago saltzen bait dute.

Beraz, pekete estatistiko hauek asko erraztu dute analisi estatistikoak egitea, eta hemendik ondo uler daiteke estatistika-munduan pakete hauek eduki duten arrakasta.

Aipatu ditugun pakete estatistikoetatik ezagun eta erabilienak ondoko hauek dira: SPSS, BMDP, SAS eta SPAD (datu-analisia eta korrespondentzi analisia).

Merkatuan agertu ziren lehenengo pakete estatistikoak erabiltzeko, ekipo informatiko haundiak behar ziren, baina ordenadore pertsonalak eduki duten hedapena alde batetik, eta ekipo txiki hauek momentu honetan daukaten ahalmena bestetik kontutan harturik, pakete

hauen bertsio berri batzuk agertu dira merkatuan; ordenadore txiki hauetan erabiltzeko pentsatuak. Guk hemendik aurrera aztertuko duguna hauetariko bat izango da: SPSS/PC+ izenekoa.

Pakete hau aztertzen hasi baino lehen pakete hau erabiltzerakoan ez dugula programaketarik egingo azpimarratu behar da, nahiz eta hitz hau erabili; baizik eta eginda dauden programak kargatu eta gure datuetara aplikatu. Honen ondorioz hemendik aurrera ikusiko ditugun gaien helburuak hauek izango dira: azpiprograma ezberdinak ezagutu, azpiprograma bakoitza erabiltzeko idatzi behar diren aginduak azaldu, eta azkenik gure datuetara komeni dena aplikatu ondoren lortuko ditugun emaitzak aztertu.

1. SPSS/PC+ PAKETERAKO SARRERA

Lehen aipatu dugunez, SPSS/PC+ paketea, ordenagailu pertsonaletan erabiltzeko pentsatua den SPSS paketearen bertsioa da.

SPSS/PC+ programa-pilo batez osatutako pakete bat da eta zatika eros eta instalatzen da. Paketea 7 zati ezberdinetan banatzen da:

BASIC : datuak, fitxategiak eta oinarrizko analisi estatistikoak erabiltzeko azpiprogramez osatua dago.

ADVANCED : Goi-mailako analisi estatistikoak erabiltzeko azpiprogramak. Zati honekin analisi estatistiko hauek egin daitezke: Faktoreen analisia, Diskriminatzailea, Cluster analisia, Log-lineala, eta bariantzaren analisi anizkoitza edo multibariantea.

TRENDS : Denborazko serieak egiteko azpiprogramak. Box-Jenkins eta analisi espektrala.

DATA ENTRY : Orri elektroniko eran datuak sartzeko erabiltzeko azpiprogramak.

GRAPFICS : Zati honek bi atal ezberdin dauzka, Microsoft enpresak egindako izen berdineko aginduak, eta bestalde SOSS-GRAPH. Grafikoak egiteko erabiltzen da.

TABLES : Taulak egiteko erabiltzen da.

SQC : Kalitate-kontrolari dagokion analisi estatistikoak erabiltzeko azpiprogramak.

SPSS/PC+ paketea, gutxienez 364 Kb. RAM, eta 10 Megako disko gogorra duen edozein ordenadore pertsonal bateragarritan erabil daiteke.

Eman dezagun paketea gure ordenadorean sartuta daukagula. Ikus dezagun martxan nola jartzen den. Giltza-disketa A unitatean sarturik, eta SPSS azpidirektorioan egonik SPSS/PC+ martxan jartzeko SPSSPC idatzi beharko dugu:

C :\SPSS>SPSSPC<Rt>

Honen ondorioz, SPSS/PC+ paketearen logotipoa agertuko zaigu, eta giltza-disketa A unitatean sartu gabe edo gaizki baldin badago, logotipoen azpian ingeles hizkuntzan idatzita agindu hau agertuko da: "Mesedez, giltza-disketa baliagarria sartu A unitatean eta gertu zaudenean OK idatzi, gelditzea nahi baduzu QUIT idatzi".

Paketea martxan dagoela adierazteko, logotipoa desagertu eta pantailan hau agertuko da:

SPSS/PC:

Batzuetan, agindu honen aurretik, beste agindu bat idatzi behar da, RAMKEY azpiprograma martxan jartzeko, eta hau izango da, SPSS azpidirektorioan egonik, eta giltza-disketa A unitatean sartuta edukirik:

C :\SPSS>A:RAMKEY

Hau idatzi ondoren lauki bat aterako da, eta ondoren SPSSPC idatzi beharko dugu lehen esan dugunez.

SPSS/PC+ paketearekin bukatzeko "FINISH." agindua idatzi beharko dugu. Kontutan hartu behar dugu, SPSS/PC+-eko agindu guztiak bukatzerakoan beti puntu bat ezarri beharko dugula, puntu hau bait da sistemari esaten diona agindu bat bukatu egin dela.

2. IDAZKERA-ARAU OROKORRAK

SPSS/PC+ ondoko era honetan lan egiten duen pakete informatiko iteraktibo bat da:

1.- SPSS/PC: agertzen da, aginduak ulertzeko prest dagoela adierazteko.

2.- Ascii kodean edozein agindu idatz dezakegu, BETI PUNTU BATEKIN BUKATUZ.

3.- Sistemak agindu hau beteta erantzuten du.

4.- Berriz lehenengo puntutik hasteko posibilitatea daukagu. Bestela "FINISH." idatziz buka dezakegu.

2.1. Aginduen sintaxia

Aginduen sintaxien arauen artean ondoko hauek aztertuko ditugu:

1.- Edozein agindu bi zatiz osatzen da. Gako-hitzak edo KEYWORDS eta hauen espezifikazioak. Lehenengoak, gako-hitzak agindu eta azpiaginduen izenak izango dira, eta bigarrenak, norberaren datuak aztertzeko eta agindu hauek osatzeko erabiltzen diren espezifikazioak.

2.- Agindu guztiak burutzeko, PUNTU bat idatzi behar da.

3.- Agindu bat idazterakoan edozein zutabetan hasteko posibilitatea daukagu. Arau honen salbuespena, datuak sartzea bukatzerakoan idazten den END DATA agindua. Datuak bukatzen diren hurrengo lerroko lehenengo zutabean idatzi behar da, kapitulu honen 7. atalean ikusiko dugun bezala.

4.- Edozein lerrotan gehienbat 80 karaktere (letra, zenbaki edo beste edozein ikur) idatz dezakegu agindua bukatzeko erabiliko dugun puntua ere kontutan harturik. Agindu bat idazteko 80 karaktere baino gehiago behar baditugu, bi lerrotan idatziko dugu, batetik bestera pasatzeko RETURN tekla erabiliz.

5.- Edozein agindu idazterakoan, letra maiuskulak edo minuskulak noiznahi erabil ditzakegu. Hala ere komenigarria izango da beti maiuskulak erabiltzea, zeren eta bestela paketeak berak maiuskulatara pasatu beharko ditu; horrela irakurtzen bait du.

6.- Kontutan eduki behar dugu, SPSS/PC+ paketea ingelesez idatzita dagoela. Beraz Ñ letra, eta azentua duten bokalak ez ditu ulertzen.

7.- Agindu guztien izenak beren GAKO HITZEN izenen lehenengo hiru letrak idatziz paketeak ulertzen ditu. Honen salbuespena, WUTH hitza izango da. Hala ere paketea erabiltzerakoan, lehenengo egunetan behintzat, komenigarria izango da izen osoak idaztea; ulergarriagoa izango bait da.

8.- Hitzak eta aginduak apartatzeko hutsune bat erabiltzen da. Batzuetan marra "/" ere erabiltzen da geroago ikusiko dugunez.

2.2. Agindu-era ezberdinak

SPSS/PC+ paketean hiru erako agindu ezberdinak bereiz ditzakegu:

- Eragiketa-aginduak.
- Datuak definitu eta aldatzeko aginduak.
- Prozedura-aginduak.

2.2.1. ERAGIKETA-AGINDUAK

Eragiketak zein eratara egitea nahi ditugun esateko erabiliko ditugu. Ez dira datuak aztertzeko baliagarriak eta sartu ahala betetzen dira.

Beren helburuak eta eskatzeko erak hauek dira:

- Erabiltzaileari laguntasuna emateko: HELP, SHOW, TUTOR eta DISPLAY.

- SPSS/PC+-en konfigurazoa aldatzeko erabiltzen dira. SET hitzarekin hasten dira denak.

- Paketeak dakarren testu-prozesadorea kargatzeko: REVIEW.

- Sistema eragilearen aginduak, pakete barrutik erabiltzeko: EXECUTE.

- SPSS/PC+-eko lanak bukatu eta sistema eragilera irteteko: FINISH.

2.2.2. DATUAK DEFINITU ETA ALDATZEKO AGINDUAK

Izenburuak esaten duen bezala, agindu hauek alde batetik aztertu nahi ditugun datuak nolakoak diren adierazteko izango dira eta bestetik datu hauek aldatzea nahi izanez gero ere agindu hauek erabili beharko ditugu.

Agindu hauek ez dira betetzen irakurri bezain agudo; baizik eta, irakurritakoan, paketeak konprobatzen du ea ondo idatzita dauden, eta geroago, PROZEDURA agindu batean beharrezkoak direnean atal honetan ikusten ditugun aginduak betetzen dira.

Beren helburuak eta eskatzeko erak hauek dira:

- Datuak sartu eta irakurtzeko: DATA LIST, BEGIN DATA, END DATA, IMPORT eta GET.

- Datuak aldatzeko: RECODE, COMPUTE, IF eta COUNT.

- Datu ausente edo galduak definitzeko: MISSING VALUE.

- Datu batzuk aukeratu edota ponderatzeko: SELECT IF, PROCESS IF, N,

SAMPLE, WEIGHT.

- Balio edo aldagai ezberdinei etiketak ezartzeko: TITLE, SIBTITLE, COMMENT, VARIABLE LABELS, VALUE LABELS eta FORMAT.

2.2.3. PROZEDURA-AGINDUAK

Agindu hauen bidez, datuen analisiak egiten dira. Era honetako agindu bat idazterakoan, sistemak, aurretik idatzita dauden eta hemen erabili behar den edozein definizio edo aldaketa-agindu betetzen du. Hau egiterakoan fitxategi "aktibo" bat eraikitzen du (Active file). Fitxategi honetan sartutako datuak, hauen aldaketak, eta behar duen informazio guztia egongo da, (aldagaien etiketak, izenak, datu galduen informazioa...).

Fitxategi aktibo hau eraiki ondoren, gure PROZEDURA aginduan eskatutako burutuko du. Erabil ditzakegun PROZEDURA-AGINDUAK hauek izango dira:

- Datuak azaltzeko erabiliko ditugunak: LIST, PLOT eta REPORT.

- Fitxategi ezberdinak erabiltzeko: WRITE, AGGREGATE, SORT CASES, JOIN, EXPORT eta SAVE.

- Analisi estatistiko ezberdinak egiteko erabiliko ditugunak:

  - Estatistika deskribatzailea: DESCRIPTIVES eta FREQUENCIES.
  - Kontingentzi taulak aztertzeko: CROSSTABS.
  - Batezbestekoak konparatzeko: T-TEST, ONEWAY, MEANS eta ANOVA.
  - Erregresio sinple eta anizkoitza: REGRESSION
  - Test ez-parametrikoak: NPAR TESTS.

Hauek dira BASIC modulua osatzen duten azpiprogramak. Kapitulu honetan garrantzi handiena daukaten aginduak ikusiko ditugu, analisi estatistikoak egiteko erabiliko ditugunak alde batera utzirik; hurrengo kapituluetan aztertuak izango bait dira.

2.3. Prozedura-era ezberdinak

SPSS/PC+ paketea erabiltzeko bi era edo metodo ezberdin daude: alde batetik metodo interaktiboa, hau da agindu bat idatzi eta jarraian makinak erantzuna ematen du, eta bestea, ZATI edo LOTEZ egitea. Bigarren metodo hau izango da erabilgarriena. Beraz guk hau bakarrik ikusiko dugu. Inork metodo interaktiboa ezagutu nahi badu, paketeak dakarren eskuliburuan begira dezake, baina esan beharra dago metodo bietan lorzen diren emaitzak berdinak izango direla, eta hemen aztertuko dugun metodoa askoz ere errazagoa dela.

2.3.1. LOTEZ ERABILTZEKO PROZEDURA

Metodo hau erabiltzerakoan, programa bat osatzen duten agindu guztiak testu-prozesadore baten bidez grabatuko ditugu, eta gero dena batera aztertzen ari garen paketeak irakurriko du.

Edozein testu-prozesadore izan daiteke baliagarria, baina beti ASCII kodean idatzi beharko ditugu, bai datuak eta bai aginduak. Adibidez, Word Start testu-prozesadorea erabiltzen badugu, N-erara idatzi beharko dugu. Hau da, dokumentua ez den fitxategi bat ireki beharko dugu.

Programaren fitxategia grabatu ondoren, SPSS/PC+ paketeak irakur dezala adierazteko, INCLUDE agindua erabili beharko dugu.

Eman dezagun SPSS azpidirektorioan grabatuta daukagula ondoko azpiprograma, ADIBIDE1.SPS izenarekin.

```
SET LIS 'ADIBIDE1.LIS'.
SET LOG 'ADIBIDE1.LOG'.
GET FILE 'ADIBIDE.SYS'.
LIST.
COMPUTE TELEBIS=TELEBIS1+TELEBIS2.
LIST TELEBIS.
SAVE OUTFILE 'ADIBIDE1.SYS'.
```

Gure paketeak programa hau irakurri eta bete dezan idatzi behar dugun agindua hau izango da:

```
SPSS/PC : INCLUDE  'ADIBIDE1.SPS'.
```

Programa hau irakurri eta burutu ondoren, inongo errorerik ez badago, gure programa burututa dagoela adierazteko, ondoko lerroa agertuko zaigu gure pantailan:

```
End of include file.

SPSS/PC:
```

Momentu honetan, "FINISH." idazten badugu, paketetik kanpora irtengo gara. Beste programa bat exekutatzea nahi izanez gero berriz, prozedura berdina egin beharko genuke, INCLUDE.....

Askotan, programa bat idazterakoan, azken lerroan "FINISH." agindua idazten da, eta orduan automatikoki, gure programa burututakoan, ordenadorea sistemara irtenda gelditukо da.

Gure programan erroreak baldin badaude, programa bukatzerakoan egin dugun errore-kopurua adieraziko du. Errore hauek zein lerrotan dauden kusteko, LOG atzizkia daukan fitxategian ikus dezakegu. Beraz kasu hauetan, testu-prozesadorea erabiliz, gure programa konpondu eta berria aipatutako pausu guztiak eman beharko ditugu.

3. LAGUNTASUN-AGINDUAK

Atal honetan hasiko gara aginduen deskribapena egiten. Horretarako lehen egin dugun sailkapen-ordena segituko dugu.

Atal honetan paketeari buruz, pantailan lor ditzakegun laguntzak nola eskatu behar ditugun ikusiko dugu.

Laguntza hau bi eratakoa izan daiteke:

1.- SPSS/PC+ paketeari dagokion sintaxien arauen informazio orokorra: HELP agindua.
2.- Lanean ari garen datu-matrizeen aldagaien izenak, etiketak, datu galdu edo ausenteen balioak...: DISPLAY agindua.

3.1. HELP agindua

Paketea nola erabiltzen den eta agindu ezberdinak idazteko arauak ematen dizkigu, agindu hau erabiltzen badugu. Maila ezberdineko laguntasuna eskaintzen digu agindu honek. Paketearen informazio orokorra, agindu batzuei buruzkoa, agindu konkretu batena, edo agindu bati dagokion azpiagindu baten informazioa eska diezaiokegu.

Funtzio hau erabiltzeko, HELP idatzi beharko dugu eta honen ondorioz nahi ditugun espezifikazio guztiak, nahi dugun zehaztasuneraino ailegatu arte. Beti bezala bukatzerakoan puntu bat eta Return teklari eman beharko diogu.

Eskatutako informazioa pantailan agertuko da. Ordenadoreak guk eskatu diogun informazioa bere memorian ez badauka, hau adierazi egingo du, ondoko lerroa idatziz: (NO HELP AVAILABLE).

3.1.1. LAGUNTZA OROKORRA

Ondoko edozein agindu idazten badugu:

SPSS/PC:HELP ALL.
SPSS/PC:HELP TOPICS.

Lortuko dugun erantzuna, ondoko hau izango da, non informazioa daukaten atal guztiak agertzen diren:

PROCEDURES.....
TRANSFORMATIONS.....
CASE SELECTION.....
DATA DEFINITION.....
SYNTAX RULES.....
COMMAND.....
FILES.....
DISPLAY.....
SET.....
SHOW.....
EXECUTE.....
INCLUDE.....
MATRIX DATA.....
NEWS.....
SPSS MANAGER.....
JOIN.....

3.1.2. GAI BATI BURUZKO INFORMAZIOA

Bigarren mailako informazioa lortzea nahi badugu, hau da, aurreko pantailan agertu den atal bateko informazioa, adibidez, fitxategiari buruzko informazioa lortzea nahi izanez gero, honela idatziko dugu:

SPSS/PC:HELP FILES.

Lortuko dugun erantzuna hau izango da:

DATA FILES.....
INCLUDE FILES.....
SYSTEM FILES.....
PORTABLE FILES.....
LISTING FILES.....

ACTIVE FILES.....
HELP FILES.....

3.1.3. AGINDU BATI BURUZKO INFORMAZIOA

Informazioaren hirugarren maila, honelakoa izan daiteke. Eman dezagun FREQUENCIES azpiprogramaren informazioa nahi dugula. Hau lortzeko era honetara idatz dezakegu:

SPSS/PC:HELP FREQUENCIES.

Ikusten dugunez, HELP aginduaren formatua beti berdina da, beti HELP idatzi behar dugu, eta gero nahi dugun mailako espezifikazioa. Eskatu dugun informazioaren erantzuna hau izango da:

FORMAT.....	HBAR.....
BARCHART.....	NTILES.....
HISTOGRAM.....	PERCENTILES.....
MISSING.....	STATISTICS.....

3.1.4. AZPIAGINDU BATEN INFORMAZIOA

Aurreko era berdinean eskatuko diogu. Eman dezagun Frequencies azpiprogramaren Statistics-en informazioa lortzea nahi dugula. Honela idatziko dugu:

SPSS/PC:HELP FREQUENCIES SATTISTICS.

Lortutako erantzuna hau izan da:

MEAN.....	KURTOSIS.....
SEMEAN.....	SEKURT.....
MEDIAN.....	RANGE.....
MODE.....	MINIMUM.....
STDDEV.....	MAXIMUM.....
VARIANCE.....	SUM.....
SKEWNESS.....	DEFAULTS.....
SESKEW.....	ALL.....

3.2. DISPLAY agindua

Fitxategi aktiboa osatzen duten aldagaien informazioa ematen digu. Bi era ezberdinetan eska dezakegu.

DISPLAY bakarrik idazten badugu, fitxategiko aldagai guztien zerrenda eta beren etiketak azaltzen ditu.

DISPLAY eta aldagaien zerrenda bat idazten badugu, aldagai hauen kodeak eta balio etiketak azaltzen ditu.

4. SISTEMAREN KONFIGURAZIO-AGINDUAK. SET AGINDUAK

Instrukzio-multzo hau, denak SET hitzarekin hasten direnak, SPSS/PC+ paketearen parametroak konfiguratzeko erabiliko ditugu.

26 agindu ezberdinak dira, baina guk 6 bakarrik aztertuko ditugu, garrantzitsuenak, hain zuzen: LISTING, LOG, LENGTH, EJECT, COLOR eta MORE.

Aztertzen ari garen paketeak, lortzen dituen emaitzak, alde batetik, pantailan agertzen ditu, eta bestetik, disko gogorrean grabatzen ditu, eta guk, emaitzen fitxategiari izen berezi bat ezartzen ez badiogu, SPSS.LIS izena ezartzen dio.

Lortuko ditugun emaitzei izen bat ezartzea nahi badugu, honela idatz dezakegu:

SET LISTING 'IZENA.LIS'.

Gure paketeak, beste fitxategi bat ere eraikitzen du, hau da, emaitzen fitxategiaren aurkibidea ematen duen fitxategi bat. Honek LOG atzizkia eramaten du, eta izen berezirik ezartzen ez badiogu, SPSS.LOG izenaz grabatzen du. Izen hau aldatzea nahi badugu, honela idatz dezakegu:

SET LOG 'IZENA.LOG'.

Aipatutako fitxategi hauek grabatzerik nahi ez badugu, honela idatziko dugu:

SET LISTING OFF.
SET LOG=OFF.

Batez ere lortu ditugun emaitzak paperean inprimitzea nahi dugunean, orrialde bakoitzean zenbat lerro idatzi behar duen esatea komenigarria izan daiteke. Horretarako erabiliko dugun agindua SET LENGTH izango da. Eman dezagun paper jarraien luzera arrunta den 72 lerro dituela gure paperak eta gure emaitzak formatu honetan gordetzea nahi ditugula. Hori adierazteko, agindu hau idatziko dugu:

SET LENGTH=72.

Gure datuak inprimatzerakoan orrialde batetik bestera salto txiki bat egotea ere nahi badugu, honela eskatuko diogu:

SET EJECT ON.

Aurreko agindua ezartzen ez badiogu, orrialde batetik bestera pasatzerakoan puntuz osatutako lerro bat idatziko du.

Batez ere kolorezko pantaila badaukagu, hurrengo agindua interesgarria izan daiteke. Pantailako koloreak aldatzeko erabil dezakegu. Gu oraindik ez gara maila honetara ailegatu. Beraz ez ditugu agindu honen kodeak probatu, baina inor interesatuta baldin badago, paketearen eskuliburuari begiratu eta han ikusiko ditu dauzkan posibilitate guztiak.

Atal honetan ikusiko ditugun aginduetatik, azkena hau izango da: SET MORE.

Eskatutako emaitzak bi eratara lor ditzakegu: alde batetik emaitza guztiak jarraian agertuko dira. Era honetara emaitzak lortzea nahi badugu, SET MORE OFF. idatzi beharko dugu. Orrialde bakoitza betetzen denean paketea gelditzea nahi badugu, SET MORE ON. idatzi beharko dugu. Azken era hau aukeratzen badugu, orrialde bakoitza agertzen denean, pantailaren goiko lerroan, eskuinaldean, MORE hitza agertuko da, eta hurrengo orrialdera pasatzeko Return tekla zanpatu beharko dugu.

Ikusi ditugun aginduak, denak jarraian ere idatz ditzakegu. Adibidez formatu labur hau ere erabil dezakegu:

SET LIS 'ADIBDE.LIS'/LOG 'ADIBIDE.LOG'.
SET LEN=72/EJECT ON/MORE OFF/COLOR OFF.

5. *DATUEN PRESTAKETA*

Datu-analisien prozesua hemen hasten da. Lehenengo pausoan liburu guztian aztertuko ditugun datuak aurkeztu eta datu hauen prestaketa egingo dugu. Kontutan eduki behar dugu hemen aurkezten ditugun datu guztiak asmatuak direla. Beraz datu eta datu hauek aztertzerakoan lortuko ditugun emaitzetatik ezingo dira ondorioak atera.

5.1. Adibidearen aurkezpena

Aurkeztuko ditugun datuak, hiru herri ezberdinetan bildutakoak direla esan dezakegu: EIBAR, GASTEIZ eta ATAUNen.

Asmatu dugun ikerketa honetarako, lagina osatzen duten ikasleei ondoko aldagaien neurketa egin genien.

Aldagaien izenak eta hauen definizioak ondokoak dira:

SEXUA : 1 emakumezkoa
2 gizonezkoa

HERRIA : 1 Ataun
2 Eibar
3 Gasteiz

INTELIGE : Lagineko pertsonen adimena neurtu dugu. Aldagai kuantitatiboa da.

ULERMENO : Ahozko ulerkuntza neurtu dugu. Aldagai kuantitatiboa da.

ESKOLAMO : 1 Ikastola
2 Publikoa
3 Pribatua

TELEBIS1 : Telebistaren aurrean astean zehar ematen duten ordu-kopurua. Lehenengo neurketa. Aldagai kuantitatiboa da.

FRUSTRAZ : Lagineko ikasleen porrot edo frustazio-maila neurtu dugu. Aldagai kuantitatiboa da.

ANTSIETA : Lagineko ikasleen antsia edo pirpira neurtu dugu. Aldagai kuantitatiboa da.

MAILASOZ : Lagineko ikasleen familiaren gizarte-maila.

1 Oso goikoa
2 Goikoa
3 Erdikoa
4 Apala
5 Oso apala

TELEBIS2 : Telebistaren aurrean astean zehar ematen duten ordu-kopurua. Bigarren neurketa. Aldagai kuantitatiboa da.

Lagin osoa 125 ikaslerekin osatuta dago.

Elementu bakoitzen datuak lerro batean idatzi ditugu. Datuak honela kodetu eta idatzi ditugu:

01	1	1	108	14	1	24	80	75	3	21
02	2	1	100	18	2	20	73	64	3	17
03	1	2	106	13	1	31	84	81	2	29
.....										
38	2	2	999	99	9	99	99	99	9	99
.....										
45	2	3	999	99	9	99	99	99	9	99
.....										
123	1	2	90	13	1	31	84	81	2	29
124	2	1	102	16	2	12	40	39	4	14
125	2	3	110	18	3	17	47	50	2	12

Puntu honetara ailegatzeko bi pauso eman behar izan ditugu. Lehenengoa aldagaiak izendatzea izan da eta bigarrena bildutako datuak kodetzea. Bi gauza hauek aztertuko ditugu hurrengo bi ataletan.

5.2. Aldagaien izenak

Badakigu, estatistika deskribatzailea edota inferentziala erabiltzeko datu batzuk bilduta eduki behar ditugula. Datu hauek aldagaien ezaugarriei dagozkienak izango dira. Datu hauek bildu ondoren aztertze-prozesura pasatuko gara.

Aztertze-prozesua hasi baino lehen datu hauek kodetu eta izendatu egin beharko ditugu.

Izendatze-prozesua, SPSS/PC+ paketeak oso erraz egiten du, eta bete behar diren baldintzak hauek dira:

- Aldagai bakoitzari nahi dugun izena ezar diezaiokegu. Izen honek gehienez 8 letra edo zenbaki eduki ditzake. Izenaren lehenengo karaktereak letra bat edo @ izan behar du. Hurrengoak letrak, zenbakiak edo ikurrak izan daitezke, ondoko hauek ezik: &, /, !, ... Bi letraren artean ezin da hutsunerik utzi.

- Aldagai ezberdinak izendatzeko izen ezberdinak erabili behar dira.

- Badira aldagaiak izendatzeko erabili ezin diren hitz batzuk. Hauek dira: ALL, AND, BY,

EQ, GE, GT, LE, LT, NE, NOT, OR, TO eta WITH.

- Azkenik, gure aldagaiak izendatzkeo, komenigarria izango da izen ulergarriak erabiltzea.

Guk erabili ditugun izenak hauek izan dira: ZENBAKIA, SEXUA, HERRIA, INTELIGE, ULERMENO, ESKOMALO, TELEBIS1, FRUSTAZ, ANTSIETA, MAILASOZ, TELEBIS2.

5.3. Balioen kodeketa

SPSS/PC+ paketearen bidez datu-matrize bat aztertzeko, bi eratako datuak izan daitezke: aldagai numerikoak eta alfanumerikoak. Lehenengoak zenbakiaren bidez kodetuta daudenak izango dira. Bigarrenak letra eta zenakien bidez kodetuta daudenak.

Egia esan, nahiz eta era bietako aldagaiak erabiltzeko posibilitatea eduki, errazena aldagai numerikoekin lana egitea izango da, eta gero VALUE LABEL aginduaren bidez, etiketak ezartzea.

Beraz, guk aztertuko ditugun datu guztiak kode numeriko baten bidez kodetuak izango dira, eta horregatik lehen aipatu ditugun aldagai kualitatiboko atributu ezberdinei zenbaki batzuk ezarri diegu:

SEXUA : 1 emakumezkoa
2 gizonezkoa

HERRIA : 1 Ataun
2 Eibar
3 Gasteiz

ESKOLAMO : 1 Ikastola
2 Publikoa
3 Pribatua

MAILASOZ : Lagineko ikasleen familiaren gizarte-maila.

1 Oso goikoa
2 Goikoa
3 Erdikoa
4 Apala
5 Oso apala

Beraz gure datuak ordenadorean grabatzerakoan herriaren izenak idatzi beharrean,

Ataungo ikasle bat baldin bada, herriari dagokion zutabean, 1 zenbakia idatzi dugu, eta era berean egin dugu beste aldagai kualitatibo guztiekin.

5.4. Datu galdu edo ausenteen kodeketa

Askotan gertatzen den gauza bat zera da: datuak batu ondoren elementu batzuen datu batzuk ez edukitzea, hau da, pertsona batek, neurtu ditugun aldagai batzuetan, adibidez erantzunik ez duela eman, edo pertsona baten maila sozioekonomikoa jakitea ezinezkoa izan dela, edo beste mila arrazoigatik datu batzuk ez edukitzea.

Elementu hauekin bi eratara joka dezakegu, lehenengoa datu datu hauek baztertzea izango zen, hau da aztertuko dugun matrizetik datu hauek kentzea. Bigarrena, eta hemen aztertuko duguna, pertsona hauen datu ausente edo galduei tratamendu berezi bat ematea izango da. Eta honela, hainbeste kostatu den matrizetik ez dugu elementurik baztertuko.

Eman dezagun pertsona baten adimenaren puntuazioa ez dugula lortu edo galdu egin zaigula. Pertsona honi, INTELIGE aldagaiari dagokion zutabean, adibidez, 999 puntuazioa ezarriko diogu. Pertsona honen sexua ezezaguna baldin bada, adibidez SEXUA aldagaiari dagokion zutabean 9 balioa ezarriko diogu.

Baina ezarri ditugun bi balio hauek, 9 eta 999, ez dira aldagaien balioak; baizik eta datu galdu edo ausenteak direla adierazteko erabili ditugun bi balio. Hori horrela dela adierazteko MISSING agindua erabiliko dugu. Agindu hau honela idatz dezakegu:

MISSING VALUE SEXUA (9)/INTELIGE (999).

Esan beharrik ez dago, datu galduari ezarriko diogun balioa, aldagaiak hartu ezin duen balio bat izan behar duela. Hau da, sexuaren balio galdu bat adierazteko ezingo dugu 1 eta 2 balioak erabili, zeren eta bestela gizon eta emakumeak adierazten dituenarekin nahastu egingo lirateke.

Gure datu-matrizea analizatzeko idatzi dugun MISSING agindua hau izan da:

MISSINS VALUE INTELIGE (999)/ESKOLAMO MAILASOZ(9)/
ULERMENO TELEBIS1 FRUSTRAZ ANTSIETA TELEBIS2 (99).

Agindu honen bitartez, INTELIGE aldagaian datu galduak adierazteko 999 idatzi dugula esaten diogu, ESKOLAMO eta MAILASOZ aldagaietan 9 balioa eta ULERMENO, TELEBIS1, FRUSTRAZ, ANTSIETA eta TELEBIS2 aldagaietan 99.

5.5. Datu-matrizeen elaborazioa

Aurreko ataletan ikusi ditugun arauak kontutan harturik, datuen kodeketa eta grabaketa egin beharko dugu azken momentuan. Datu hauek grabatzeko edozein testu-prozesadore erabil dezakegu, baina beti ASCII-kodearen erara. Lehen esan dugunez, Word Star testu-prozesadorea erabiltzen badugu, gure datuak N-erara sartu beharko ditugu.

Datuak era ezberdinetan grabatzeko posibilitate edukiko dugu, SPSS/PC+ paketea erabiltzen dugunean. Hala ere, guk arruntena ikusiko dugu. Hau da, formatu finko batean ezarriko ditugu gure datu guztiak. Formatu finko batean sartzen baditugu, bete behar den baldintza bakarra hau da: Lagineko elementu guztien aldagai berdineko balioek, zutabe berean egon behar dute. Beraz kontutan hartu beharko dugu aldagai bakoitzarentzat zenbat zutabe erreserbatzen ditugun, zeren eta posible izango da aldagai bateko emaitza pertsona batentzat bi zutabeko emaitza eta beste batentzat hirukoa izatea. Adibidez pertsonen pisua hartzen badugu, posible izango da pertsona baten pisua 67 kilo izatea, baina datuak sartzen hasi aurretik kontutan eduki beharko dugu posible dela pertsona baten pisua 103 kilo izatea. Beraz aldagai honen balioak sartzeko hiru zutabe gorde beharko ditugu.

Datu hamartarrak edo dezimalak sartzerakoan, ez dugu komatxorik erabiliko; baizik eta zati osoa eta zati hamartarra apartatzeko puntua erabiliko dugu.

Pertsona edo elementu bakoitzaren datuak 80 baino gehiago baldin badira, datu hauek ez dira lerro batean sartuko. Orduan elementu bakoitzarentzat bi lerro erabiliko ditugu, eta era berean, hemen ere sartzen ez badira hiru, edo lau...

Datuak sartzea bukatutakoan, kontuan eduki behar dugu, eta hau askotan gertatzen den errorea da, non bukatu behar dugun ere.

Azken pertsonaren datuak sartzen ditugunean, return teklari emango diogu, eta hurrengo lerroko lehengo zutabean kurtsorea ezarriz gure fitxategia grabatu beharko dugu. Bestela azken pertsonen datuak ez ditu irakurriko.

Kontutan eduki behar dugu datu-fitxategiari izen bat ezarri behar diogula. Normalean, datu-fitxategiak izendatzeko, DAT atzizkia erabiltzen da. Guk liburu honetan aztertuko ditugun datuak izendatzeko LIBURU.DAT erabili dugu.

Beraz gure datuak, LIBURU.DAT fitxategian grabatu ditugunak, zutabe hauetan idatziak izan dira:

ZENBAKIA	1-2-3
SEXUA	5
HERRIA	7
INTELIGE	9-10-11
ULERMENO	13-14
ESKOLAMO	16
TELEBIS1	18-19
FRUSTRAZ	21-22
ANTSIETA	24-25
MAILASOZ	27
TELEBIS2	29-30.

LIBURU.DAT

001	1	1	108	14	1	24	80	75	3	21
002	2	1	100	18	2	20	73	64	3	17
003	1	2	106	13	1	31	84	81	2	29
004	2	1	102	16	2	12	40	39	4	14
005	2	3	108	18	3	17	47	50	2	12

.....

6. DATUAK IRAKURTZEKO AGINDUAK

Aurreko atalean ikusi dugunez, gure datuak grabatuta dauzkagu, eta normalean, disko gogorreko SPSS azpidirektorioan sartuta.

Datu hauen informazioa SPSS/PC+ paketeak irakur dezan ondoko helburuak betetzen dituzten aginduak erabili beharko ditugu:

- Titulua ezarri.
- Datu-matrizea definitu:
  - Gure datuak grabatuta dauden fitxategiaren izena eman.
  - Aldagaiei izenak eman.
  - Aldagai ezberdinak zein zutabetan dauden.
- Datu galdu edo ausenteen balioen kodeak definitu.
- Aldagaien etiketak ezarri.
- Balioen etiketak ezarri.
- Informazio guzti hau beste fitxategi batean gorde, datu-matrize hau aztertu behar dugun bakoitzean berriz idatzi beharrik ez edukitzeko.

Aipatu ditugun helburuak betetzeko, agindu ezberdinak erabiliko ditugu. Hauek

izango dira orain aztertuko ditugunak.

6.1. Programaren izenburua. TITLE agindua

Normalean hau izango da edozein programaren lehenengo agindua.

Idazteko era oso erraza da, lehenengo TITLE hitza idatziko dugu, eta ondoren, komatxoen artean, ezarri nahi dugun titulua. Gehienez 58 karaktere eduki ditzake. Gure adibidean ezarri duguna hau izan da:

TITLE 'OINARRIZKO AGINDUEN ADIBIDEA'.

Konturatuko zineten agindu hau ere puntu batez bukatu dugula. Agindu guziak bezala.

Ezarri dugun titulua, emaitzetako orrialde guztietan azalduko da.

6.2. Datu-matrizeen definizioa. DATA LIST agindua

Beharrezkoa den agindu bat izango da. Agindu honen bidez paketeari zein datu-fitxategi aztertuko dugun esaten diogu.

Agindu honek bi atal edukiko ditu; biak garrantzi handikoak. Lehenengoan, aztertuko dugun datu fitxategiaren izena esan beharko diogu, eta bigarrenean, fitxategi honetako datuak nola irakurri behar dituen.

Lehen aipatu dugunez, gure datuak era ezberdinetan sartzeko aukera edukiko dugu. Guk, bat bakarrik ikusi dugu; arruntena, errazena eta gehien erabiltzen dena, hau da, formatu finkoa daukana. Orain datuak irakurtzerako aginduak ikusterakoan, formatu finkoa irakurtzeko idatzi behar den agindua ere ikusiko dugu.

Gure datuak irakurtzeko erabili dugun agindua hau izan da:

DATA LIST FILE='LIBURU.DAT'/ZENBAKIA 1-3 SEXUA 5 HERRIA 7
INTELIGE 9-11 ULERMENO 13-14 ESKOLAMO 16 TELEBIS1 18-19
FRUSTRAZ 21-22 ANTSIETA 24-25 MAILASOZ 27 TELEBIS2 29-30.

Ikusten dugunez, lehenengo zatian DATA LSIT FILE idatzi ondoren gure datu-fitxategiaren izena komatxoen artean idatzi dugu, eta berdinketa-zeinuaren ondoren, Bigarren zatian gure fitxategia den LIBURU.DAT nola antolatu dugun esaten diogu. Hau da, zein aldagai zein zutabetan dagoen.

Ikusiko dituen lehenengo datuak ZENBAKIA aldagaienak izango direla esaten diogu, eta datu hauek 1,2 eta 3 zutabean egongo direla. Horretarako ZENBAKIA 1-3 idatzi dugu. Hurrengo datua sexuari dagokiona izango dela, eta 5. zutabean dagoela. Beraz, 4. zutabeak hutsik egon beharko du fitxategi osoan. Hori esateko: SEXUA 5, idatzi dugu. Horrela idatzi dugu gure agindu guztia. Ikusi dugunez, agindu hau nahikoa luzea da eta gehienetan ez da lerro baten barruan kabituko. Orduan agindu hau idazteko bi, hiru... lerro erabiliko ditugu. Lerro batetik bestera pasatzeko ez daukagu ezer idatzi beharrik; Return teklari eman bakarrik eta aurrera.

Gure datu-fitxategiko elementu bakoitzaren datuak bi lerrotan daudenean, hau da, elementu bakoitzaren datuak 80 baino gehiago direnean, DATA LIST FILE aginduaren bidez hau paketari esan beharko diot. Esateko era oso erraza da. Marra bat '/' idaztea nahikoa izango da.

Eman dezagun elementu baten datuak hauek direla:

```
                ARRAROAK.DAT
001  55555555656666666666666666666666666666666666666666677777
565666566665656665667776
```

Hau adierazteko honela esango diot:

```
DATA LIST FILE='ARRAROAK.DAT'/ZENBAKIA 1-3 DATUBAT 5-60/
DATUBI 1-24
```

Hamartarra edo dezimala daukaten datuak edukiko bagenitu, alde batetik aldagai honek zenbat zutabe okupatzen dituen eta bestetik dezimalak zenbat diren esan beharko diogu, bigarren datu hau parentesien artean ezarriko dugularik.

Eman dezagun ondoko datu hauek ditugula:

```
DEZIMAL.DAT
1  1.12  124.1  23.45  1.1
2  2.65   23.5  19.33  0.9
```

Fitxategi hau irakurtzeko idatziko dugun agindua hau izango da:

```
DATA LIS FILE='DEZIMAL.DAT'/ZENBAKIA 1 ALDAGAI1 3-6(2)
ALDAGAI2 8-12(1)  ALDAGAI3 14-18(2)  ALDAGAI4 20-22(1).
```

Kontutan eduki behar dugu gure datu-fitxategian dauzkagun datu guztiak zenbakizkoak direla, hau da, letrarik ez daukagula, zeren eta azken kasu honetan

irakurtzeko agindua idazterakoan beste agindu berezi batzuk idatzi beharko bait genituzke.

6.3. Datu ausente edo galduen definizioa. MISSING agindua

Datu galduak edo ausenteak edukitzea oso gauza arrunta denez gero, eta datu galduak dauzkaten elementuak gure matrizetik ez kentzea erabakitzen badugu, tratamendu berezi bat emango diogu, 5.4. atalean esan dugunez. Han ikusten genuen datu hauek nola kodetu. Hemen ikusiko dugu datu hauek irakurtzeko idatzi behar dugun MISSING agindua nola erabiliko dugun.

Eman dezagun gure datu-matrizean, aldagai guztietan, sexua eta herria kendurik, datu galduak dauzkagula, eta datu hauek kodetzerakoan erabili ditugun zenbakiak hauek izan direla:

INTELIGE	999
ULERMENO	99
ESKOLAMO	9
TELEBIS1	99
FRUSTAZ	99
ANTSIETA	99
MAILASOZ	9
TELEBIS2	99

Gure programan hori adierazteko erabili dugun agindua hau izan da:

MISSING VALUE INTELIGE (999)/ESKOLAMO MAILASOZ (9)/ULERMENO TELEBIS1 FRUSTAZ ANTSIETA TELEBIS2 (99).

Agindu honetan lehenengo ezarri duguna, aginduaren izena: MISSING VALUE izan da, eta gero aldagai bakoitzean, datu galduak adierazteko erabili dugun zenbakia ezarri dugu.

Erabili dugun formatua formatu laburra izan da. Agindu bera beste era honetara ere idatz dezakegu:

MISSING VALUE INTELIGE (999)/ULERMENO (99)/ESKOLAMO (9)/
TELEBIS1 (99)/FRUSTRAZ (99)/ANTSIETA (99)/ MAILASOZ (9)/
TELEBIS2 (99).

Gure kasuan datu galduak adierazteko erabili ditugun balioak, 9, 99 eta 999 izan dira eta horregatik ezarri ditugun balioak horiek izan dira. Baina aldagai baten datu galduak

adierazteko 545 balioa erabili izan bagenu, era honetara idatzi beharko genuke:

MISSING VALUE ALDAGAIA (545).

6.4. Aldagai eta balioen etiketak. VARIABLE LABELS eta VALUE LABELS aginduak.

Esan dugunez, SPSS/PC+ paketeak aldagai eta balioei etiketak ezartzeko posibilitatea ematen digu. Askotan aldagai baten izena jartzerakoan, gogoratu gehienez 8 karaktere erabil dezakegula, komenigarria izango da izen honen esangura etiketa batean ezartzea. Honek bere garrantzia emaitzak interpretatzerakoan edukiko du, zeren askoz ere ulergarriagoak izango bait dira.

Aldagaien etiketa hauek, hutsuneak kontutan hartuta, gehienez 40 karaktere eduki dezakete.

Era berean, askoatan, aldagai kualitatibotan batez ere, komenigarria izango da aldagai hauen balioei ere etiketak eranstea. Eman dezagun SEXUA aldagaiarekin lanean ari garela. Emaitzak interpretatzerakoan, askotan ez dugu jakingo zein balio eman diogun gizonezkoari eta zein emakumezkoari. Hori gainditzeko ere, komenigarria izango da balioei etiketak eranstea.

Balioen etiketak, hutsuneak kontutan harturik, gehienez 20 karaktere eduki dezake.

Etiketak eransteko VARIABLE LABELS eta VALUE LABELS aginduak erabiliko ditugu.

Gure aldagai eta datuei ezarri dizkiogun etiketak hauek izan dira:

```
VARIABLE LABELS INTELIGE 'ADIMEN-KOEFIZIENTEA'
/ULERMENO 'AHOZKO ULERKUNTZA'
/TELEBIS1 'TELEBISTA AURREAN ORDUAN LEHENENGO NEURKE'
/MAILASOZ 'MAILA SOZIOEKONOMIKOA'
"TELEBIS2 'TELEBISTA AURREAN ORDUAK BIGARREN NEURKET'.

VALUE LABELS SEXUA 1 'EMAKUMEA' 2 'GIZONA'
/HERRIA 1 'ATAUN' 2 'EIBAR' 3 'GASTEIZ'
/ESKOLAMO 1 'IKASTOLA' 2 'PUBLIKOA' 3 'PRIBATUA'
/MAILASOZ 1 'OSO GOIKOA' 2 'GOIKOA' 3 'ERDIKOA' 4 'APAPALA'
5 'OSO APALA'.
```

Hemen ikusi dugunez, edozein etiketa idazterakoan, beti gauza bera egin beharko

dugu: lehenengo, aldagaien izena edo balio baten zenbakia ezarri eta ondoren komatxoen artean etiketa. Aldagai batetik bestera pasatzeko marra bat idatzi beharko dugu.

Azkenean, beti bezala, puntu batez bukatuko dugu.

6.5. Informazio honen grabaketa. SAVE agindua
Grabatutako informazioa berreskuratu. GET agindua
Grabatutako informazioa ikusi. LIST agindua

Ikusi ditugun bost aginduen bidez, gure datu-fitxategiko aldagaiak eta balioak definitu ditugu. Gure datu-fitxategia aztertzen dugun bakoitzean guzti hau idaztea beharrezkoa izango da eta suposatuko duzuenez nahikoa aspergarria izaten da.

SPSS/PC+ paketeak ematen digun posibilitatea hau da: guzti hau fitxategi batean grabatzea eta gero nahi dugunean berriz, paketearen bidez eskatzea. Beraz, informazio hau grabatzeko erabiliko dugun agindua SAVE izango da.

Kontutan eduki behar dugu aurreko agindu guztiak erabiltzen ditugunean gure paketeak sistemako fitxategi aktibo bat eraikitzen duela. Eta fitxategi aktibo hau izango da guk grabatuko duguna. Horregatik erabiliko dugun atzizkia SYS izango da. Esan beharra dago fitxategi aktibo hau datuak irakurri ondoren eraikitzen duela. Beraz, SAVE aginduak DATA LIST FILE aginduaren atzetik joan beharko du. Normalean aurreko bost aginduen atzetik idazten da, aurreko aginduko informazio guztia gordetzea interesatuko zaigulako.

Agindu hau idazteko era, hau da:

SAVE OUTFILE='LIBURU.SYS'.

Aurreko aginduarekin eskatzen duguna zera da: agindu honen aurretik dagoen informazio guztia LIBURU.SYS izeneko fitxategi batean gordetzea.

Grabatu dugun informazioa berreskuratzeko GET agindua erabiliko dugu, eta idazteko era hau izango da:

GET FILE='LIBURU.SYS'.

Agindu honen bidez, lehendik grabatuta daukagun LIBURU.SYS fitxategiarekin lana egiteko posibilitatea edukiko dugu.

Batez ere aurreko agindua erabiltzen dugunean, fitxategi honetan daukagun informazioa ikusteko interesgarria izan daiteke LIST agindua erabiltzea.

LIST agindu honen bidez, berreskuratu dugun fitxategiaren informazioa lortuko dugu, hau da, aldagaien izenak, eta zein eratara grabatuta dauden.

Agindu hau erabiltzeko era honetara idatz dezakegu:

LIST VARIABLES=aldagaien izenak.

Fitxategian daukagun informazio guztia ikustea nahi badugu, nahikoa izango da hau idaztea:

LIST.

Aurreko agindu guztiak ikusi ondoren, guk erabili dugun programa osoa hau izan da:

```
TITLE 'OINARRIZKO AGINDUEN ADIBIDEA'.
DATALA LIST FILE='LIBURU.DAT'/ZENBAKIA 1-3 SEXUA 5 HERRIA 7
INTELIGE 9-11 ULERMENO 13-14 ESKOLAMO 16 TELEBIS1 18-19
FRUSTRAZ 21-22 ANTSIETA 24-25 MAILASOZ 27 TELEBIS2 29-30.
VARIABLE LABELS INTELIGE 'ADIMEN-KOEFIZIENTEA'
/ULERMENO 'AHOZKO ULERKUNTZA'
/TELEBIS1 'TELEBISTA AURREAN ORDUAK LEHENENGO NEURKE'
/MAILASOZ 'MAILA SOZIOEKONOMIKOA'
/TELEBIS2 'TELEBISTA AURREAN ORDUAK BIGARREN NEURKET'.
VALUE LABELS SEXUA 1 'EMAKUMEA' 2 'GIZONA'
/HERRIA 1 'ATAUN' 2 'EIBAR' 3 'GASTEIZ'
/ESKOLAMO 1 'IKASTOLA' 2 'PUBLIKOA' 3 'PRIBATUA'
/MAILASOZ 1 'OSO GOIKOA' 2 'GOIKOA' 3 'ERDIKOA' 4 'APALA'
5 'OSO APALA'.
MISSING VALUE INTELIGE (999) /ESKOLAMO MAILASOZ (9) /ULERMENO
TELEBIS1 FRUSTRAZ ANTSIETA TELEBIS2 (99).
SAVE OUTFILE='LIBURU.SYS'.
FIN.
```

7. ALDAGAIEN BIRKODEKETA.
ALDAGAI BERRIAK SORTERAZTEKO ERAK

SPSS/PC+ paketeak daukan abantailetan, ez da txikiena atal honetan ikusiko duguna. Hau da, guk idatzi ditugun datu edo aldagaiak birkodetzea eta jatorrian daukagun aldagaietan oinarrituz aldagai berriak eraikitzeko ematen digun posibilitatea.

Egingo ditugun aldaketa guztiak, guk bereziki SAVE agindua erabiliz grabatzen ez baditugu, programa bukatzerakoan galdu egiten dira.

Nahiz eta aldaketak egin, lehendik daukagun informazioa ez da inoiz galduko. Beraz nahikoa lasai egin ditzakegu nahi ditugun froga guztiak.

Atal honetan ikusiko ditugun aginduak hauek izango dira: RECODE, COMPUTE eta IF.

Lehenengoak, aldagai baten balioak birdoketu egingo ditu. Bigarrena eta hirugarrena, aldagai berriak eraikitzeko erabiltzen dira.

7.1. Aldagaien balioen birkodeketa. RECODE agindua

Agindu hau erabiltzeko, idazkera oso erraza izango da. RECODE idatzi ondoren, birkodetu nahi ditugun aldagaien izena idatzi beharko dugu, eta ondoren, parentesi artean, lehengo balioa=balio berria. Berdintasun bakoitzerako parentesi bat ezarri beharko dugu.

Eman dezagun, gure adibidean daukagun MAILASOZ aldagaiaren kodeketa aldatu nahi dugula. Oinarrian, aldagai hau erabiliz elementu guztiak 5 sailetan banatu dira, 1:OSO GOIKOA, 2:GOIKOA, 3:ERDIKOA, 4:APALA eta 5:OSO APALA. Sailkapen hau aldatzea nahi badugu, eta pertsona guztiak hiru zatitan banatzea nahi badugu (OSO GOIKOA eta GOIKOA atal batean sartuz eta APALA eta OSO APALA beste batean) era honetan idatz dezakegu:

RECODE MAILASOZ (1,2=1) (3=2) (4,5=3).

Beraz hemendik aurrera, MAILASOZ aldagaiak hiru balio edukiko ditu, 1, 2 eta 3.

Beste adibide bat aipiniko dugu. Eman dezagun INTELIGE aldagaia birkodetzea nahi dugula. Aldagai hau, aldagai jarrai bat bezala kontsidera dezakegu. Gure lagineko elementuaren puntuazioak, 88 eta 112 punturen artean daude. Aldagai honetan lortu duten puntuazioen arabera, elementuen sailkapen bat egitea nahi badugu (adibidez hiru klasetan, lehenengoa 88-tik 95-erainoko puntuazioak lortu dituztenekin, bigarrena 96-tik 105-erainokoekin eta hirugarrena 106 puntu edo gehiago lortu dituztenekin) hau da idatzi behar dugun agindua:

RECODE INTELIGE (88,89,90,91,92,93,94,95=1) (96,97,98,99,100,101,102, 103,104,105=2) (106,107,108,109,110,111,112=3).

Edozein konturatzen da, idazteko era hau nahikoa astuna dela. Zorionez, pakete esta-

tistiko hau eratu zutenak ere konturatu ziren, eta aurreko agindua beste era labur batez idazteko posibilitatea ematen digute, hiru hitz hauek erabiliz: THRU, LO (lowest) eta HI (highest). Beraien esanahiak hauek dira: -RAINO, TXIKIENETATIK, HANDIENA-RAINO.

Beraz aurreko agindua beste era honetara idatz dezakegu:

RECODE INTELIGE (LO THRU 95=1) (96 THRU 105=2) (106 THRU HI=3).

7.2. Aldagai berrien sortzapena: COMPUTE agindua

Lehendik dauzkagun aldagaietan oinarrituz, agindu honen bidez aldagai berriak lortzeko posibilitatea edukiko dugu.

Agindu hau erabiltzeko, idatzi behar dugun instrukzioa era honetakoa izango da:

COMPUTE TELEBIS=TELEBIS1+TELEBIS2.

Ikusten dugunez, aurreko kasuetan bezala, lehenengo, aginduaren izena idatzi behar dugu, gero aldagai berrien izena eta azkenik, aldagai berri hau sortzeko, aldagai zeharrekin egin behar ditugun eragiketak idatzi beharko ditugu.

Aurreko adibidean, TELEBIS1 eta TELEBIS2 aldagaietan oinarrituz, TELEBIS aldagai berria eraiki dugu, oinarrizko bi aldagaien batura bezala.

Egin ditzakegun eragiketak eta hauek egiteko erabiliko ditugun ikurrak, hauek dira:

+ ..... Batuketa.
- ..... Kenketa.
* ..... Biderkaketa.
/ ..... Zatiketa.
** ..... Berredura.
ABS ... Balio absolutua.
SQRT.. Erro karratua.
EXP.... esponentziazioa.
LG10... 10 oinarriko logaritmoa (logaritmo arrunta)
LN ..... e oinarriko logaritmoa (logaritmo nepertarra)
SIN .... sinua
COS ... kosinua

Adibidez hau idazten badugu:

COMPUTE INTBERRI=10*INTELIGE + 4.

Aurreko instrukzioarekin INTBERRI aldagai berria sorterazi dugu. Aldagai berri hau INTELIGE bider hamar gehi lau eginda lortu dugu.

Datu galduen arazoa kontutan hartu beharko dugu. Beraz, INTELIGE aldagaian 999 balioa datu ausente bat adierazteko erabiltzen bagenuen, kasu honetan INTBERRI aldagaian edukiko dugun datu galdua 9994 izango da. Beraz aurreko COMPUTE egin ondoren, agindu hau ere idatzi beharko dugu:

MISSING VALUE INTBERRI (9994).

7.3. Aldaketa baldintzatuak. IF agindua

Agindu honen bidez, COMPUTE aginduaren erako aldaketak egingo ditugu, baina baldintza logiko batzuen bidez.

Idazteko era hau izango da: beti bezala lehenengo aginduaren izena, IF, gero espresio logikoa parentesien artean, eta azkenik aldagai berria eta kasu honetan hartuko duen balioa.

Espresio logikoak idazteko ondoko eragile edo operadoreak erabil ditzakegu. Ezkerrean dauden letrak eta sinboloak noiznahi erabil ditzakegu.

LE...	<=	.... txikiago edo berdina.
LT...	<	.... txikiago.
GE...	>=	.... handiago edo berdina.
GT...	>	.... handiago.
EQ...	=	.... berdin.
NE...	<>	.... ezberdin.

Bi espresio edo gehiago lotzeko, beste hiru eragile edo operadore ere erabil ditzakegu:

AND...	&	..... eta
OR...	:	..... edo
NOT...	~	..... ez

Eman dezagun gure lagineko elementuak sei taldetan banatzea nahi dugula, sexu eta herriari begiratuz. Ondoko sei instrukziok erabili beharko ditugu hau egiteko:

```
IF ( (HERRIA =1) AND (SEXUA =1) ) TALDE=1.
IF ( (HERRIA =1) AND (SEXUA =2) ) TALDE=2.
IF ( (HERRIA =2) AND (SEXUA =1) ) TALDE=3.
IF ( (HERRIA =2) AND (SEXUA =2) ) TALDE=4.
IF ( (HERRIA =3) AND (SEXUA =1) ) TALDE=5.
IF ( (HERRIA =3) AND (SEXUA =2) ) TALDE=6.
```

Eman dezagun telebistaren bi neurketak ikusi ondoren lehenengoa eta bigarrena konparatzea nahi dugula. Eta honetara beste aldagai bat ere lor dezakegu:

```
IF (TELEBIS1 GT TELEBIS2) ALDAKETA=1.
IF (TELEBIS1 LT TELEBIS2) ALDAKETA=2.
IF (TELEBIS1 EQ TELEBIS2) ALDAKETA=0.
```

Aurreko hiru aginduekin, aldagai berri bat eraiki dugu, ALDAKETA izena duena, eta bere balioak 1, 2 edo 0 izango dira, segun eta telebistaren aurrean ikasleek igarotzen dituzten orduen lehenengo neurketa bigarrena baino handiago, txikiago edo berdina den.

8. DATUAK AUKERATZEKO AGINDUAK

Nahiz eta gure fitxategian lagineko elementu guztiak eduki, batzuetan beharrezkoa izango da lagineko zati batekin lana egitea. Hau da, gure adibidean batzuetan Ataungo pertsonen analisi batzuk egitea bakarrik interesatuko zaigu, beste batzuetan, maila sozioekonomiko altua daukatenen analisiak,... Beraz kasu hauetan, lagineko elementuen aukeraketa bat egin beharko dugu. Aukeraketa hau programa guztirako izan daiteke ala hurrengo agindurako bakarrik. Lehenengo kasuan BETIRAKO AUKERAKETA izango da, eta bigarrenenan AUKERAKETA DENBORALA.

8.1. Betirako aukeraketa

Betirako hitz honekin esan nahi duguna zera da: programa guztirako dela, hau da, gure datu-fitxategia beti hor edukiko dugu, eta ez da inoiz galduko. Aukeraketa hau bi eratara egin dezakegu, baldintza logiko baten bidez, edo sekuentzia baten bidez.

8.1.1. AUKERAKETA BALDINTZATUA. "SELECT IF" AGINDUA

Agindu honen bidez, lagina osatzen duten elementuetan jartzen dugun baldintza betetzen dutenak bakarrik aukeratuko ditugu. Honela idatziko dugu:

```
SELECT IF (espresio logikoa).
```

Adibidez:
SELECT IF (HERRIA EQ 1).

Aurreko instrukzioa idatzi ondoren, egingo ditugun analisi guztiak Ataungo pertsonekin bakarrik egingo ditugu; hauek bait dira ezarri dugun baldintza betetzen duten bakarrak.

Bestetik, IF aginduan ikusi dugun bezala, bi espresio logiko instrukzio bakar batean batu ditzakegu, hau da, Ataungo emakumeen datuak aztertzea nahi badugu ondoko agindua erabil dezakegu:

SELECT IF ((HERRIA EQ 1) AND (SEXUA EQ 1)).

Aurreko agindua idatzi beharrean beste bi hauek idazten baditugu, gauza bera lortuko dugu. Beraz ondoko bi instrukzioekin, lehenengoan, Atauneko pertsona guztiekin geratzen gara, eta bigarrenarekin, emakumeak aukeratzen ditugu.

SELECT IF (HERRIA EQ 1).
SELECT IF (SEXUA EQ 1).

8.1.2. AUKERAKETA SEKUENTZIALA. "N" AGINDUA

Eman dezagun, oso lagin handia daukagula eta gure laginetik adibidez, lehenengo 30 elementuek osatzen duten azpilaginketa bat egitea nahi dugula, hemendik aurrera elementu hauekin analisi guztiak egiteko.

Agindu hau idaztea oso zaila izango da. Begira:
N 30.

Eta agindu hau erabiliz gure lagineko lehenengo 30 elementuak aukeratuak izango dira, eta agindu honen atzetik dauden analisi guztiak, 30 elementu hauekin egingo dira.

Kontuz ibili behar da, N agindua eta SELECT IF, edo orain ikusiko dugu PROCESS IF batera ezartzerakoan. Biak elkarren ondoan daudenean, beti lehenengo SELECT IF eta PROCESS IF aginduak beteko ditu.

8.2. Aukeraketa denborala

Orain ikusiko ditugun bi aginduak ere, gure fitxategiko elementuen aukeraketa bat egiteko baliagarriak izango dira, baina aukeraketa hauek, idatzita daukagun hurrengo agindurako bakarrik izango dira baliagarriak. Aurreko aukeraketaren kasuan bezala aukeraketa logikoa eta aleatorioa bereiztuko ditugu.

8.2.1. AUKERAKETA BALDINTZATU DENBORALA. "PROCESS IF" AGINDUA

Select if aginduaren berdina da. Bien artean dagoen ezberdintasuna hau da, SLECT IF aginduaren kasuan, AND, OR, eta NOT operadoreak erabiltzeko posibilitatea genuen. Kasu honetan hauek ezin dira erabili. Beraz hauek bakarrik erabil ditzakegu: EQ, NE, LT, LE, GT eta GE operadoreak. Beste era batera esanda, espresio logiko sinpleak bakarrik onartzen ditu.

Instrukzio baten aurretik bi PROCESS IF edo gehiago idazten baditugu, azkena bakarrik kontutan hartuko dugu. Beraz honetan ere SELECT IF aginduarekiko ezberdina da.

Idazteko era, hau da:

PROCESS IF (SEXUA EQ 1).

Aurreko aginduak eta orain ezarriko ditugun biek gauza BERA egiten dute, kasu bietan HURRENGO INSTRUKZIOKO ANALISIA lagineko emakumeekin bakarrik egingo da.

PROCESS IF (HERRIA EQ 1).
PROCESS IF (SEXUA EQ 1).

8.2.2. AUKERAKETA ALEATORIOA. "SAMPLE" AGINDUA

Agindu honen bidez lagineko azpilaginketa aleatorio bat egiteko posibilitatea ematen digu. Azpilaginketa hau egiteko bi era ezberdin erabil ditzakegu: bata lagin osotik zein zati aukeratu nahi dugun esanda, hau da, era honetara idazten badugu:

SAMPLE .3.

Agindu honen bitartez eskatzen duguna zera da: lagin osoko elementuetatik %30 zoriz aukera ditzala. Beraz hurrengo analisia elementu hauekin egingo du.

Bigarren era hau izango da. Era hau erabiltzeko lagineko elementu-kopuruak ezaguna izan behar du, hau da, adibidez guk badakigu liburu honetan adibideak jartzeko erabiliko dugun fitxategiak 125 elementu dituela. Beraz hauetatik 35 aukeratzea nahi izanez gero honela idatz dezakegu:

SAMPLE 35 FROM 125.

10. GAIA
ESTATISTIKA DESKRIBATZAILEA; FREQUENCIES, DESCRIPTIVES, MEANS, PLOT-PLOT *ETA* CORRELATION *AZPIPROGRAMAK*

3. eta 4. ataletan, Estatistika Deskribatzailea deritzon alderdiari dagozkion oinarri teorikoak aztertu ditugu. Zentralizazio eta Sakabanatze-neurriak, asimetria eta zorroztasuna (kurtosi), koerlazio-koefizienteak eta erregresio lineala, oinarri teorikoak eta kalkuluetarako prozedura ikusi ditugu. Honako atal honetan, bada, kalkulu horiek ikusiko ditugu, baina SPSS/PC+-ren azpiprogramak erabiliz.

Azalpen guztiak argiago eta errazago suerta daitezen, aurreko gaian arkeztutako zenbait datu erabiliko dugu. Hala ere, simulatutako datu-fitxategia dela azpimarratu beharra dago eta, beraz, ez da posible datu hauetan oinarriturik inolako ondoriorik ateratzea.

Horiek horrela, Estatistika Deskribatzailean maizen erabiltzen den azpiprograma hartuko dugu abiapuntutzat.

1. FREQUENCIES *AZPIPROGRAMA*

Aztertu gabeko datu-fitxategiari egin behar zaion lehenengo analisia honako hauxe dugu. Honela lorturiko emaitzak zenbait ikuspegi desberdinetatik erabil ditzakegu, lehena gaizki kodeturiko edota gaizki sarturiko datuak aurkitzeko delarik. Adibidez, 1 eta 2 balioaz kodeturiko SEXUA deritzon aldagaian 3 balioren bat agertuko balitz, erroreren bat dagoela argi dago; edo kodeketa-prozesuan edo datu-fitxategiaren eraiketan.

Egin dezakegun bigarren erabilera, zehatz-mehatz dagokiona da, hots, datu-multzoaren estatistiko deskribatzaileak kalkulatzea. Beraz 3. eta 4. ataletan ikusiriko neurri-klase desberdinen estatistikoak honen bitartez kalkula ditzakegu.

Erabilpenari dagokionez, azkenik, standarizatu nahi ditugun testen baremazioak ere egin ditzakegu. Aztertzen den aldagai jakin batekiko pertzentilen balioak kalkulatzeko aukera ere izango dugu.

Horiek horrela, gure datu-matrizeekin inongo analisirik egin baino lehen azpiprograma honen bitarteko analisiak egitea gomendagarria da.

SINTAXIA

Azpiprograma hau erabili ahal izateko ondorengo eskemari jarraitu behar zaio:

```
FREQUENCIES VARIABLES = aldagaien izenak  /Histogram
                                          /Statistics
                                          /Options.
```

Dena dela, aurrekoaz gain beste zenbait idazkera posible delarik, batzuk aztertuko ditugu eta ondorengo zerrendan deskribatu ere bai:

```
(1) FREQ VAR=HERRIA.

(2) FREQ VAR=INTELIGE/HISTOGRAM/NTILES=10/STATISTICS=ALL.

(3) FREQ VAR=SEXUA HERRIA INTELIGE TO TELEBIS2/BARCHART.
    RECODE MAILASOZ (1,2=1) (3=2) (4,5=3).
    VALUE LABELS MAILASOZ 1 'GOIKOA' 2 'ERDIKOA' 3 'BEHEKOA'.

(4) FREQ VAR=MAILASOZ/HISTOGRAM.
    COMPUTE TELEBIS=(TELEBIS1 + TELEBIS2)/2.
    RECODE TELEBIS (SYSMIS=99).

(5) FREQ VAR=TELEBIS/HISTOGRAM/STATISTICS=ALL.

  FIN.
```

Edonola ere aurretiko edozein agindu idatzi aurretik aldagaien formatua deskribatuta eduki beharko dugu edo, ez balego, aldez aurretik grabaturiko fitxategietatik berreskuratu.

Programa hasteko beraz, ondoren azaltzen ditugun bi era hauetara egin daiteke.

Lehenengo honetan eta aurreko atalean azaldu genuenez, programa osoa idatziko dugu, nahiz eta beharrezkoa ez izan.

```
SET MORE OFF.
SET LIST 'LIBURU.LIS'.
SET LOG 'LIBURU.LOG'.
DATA LIST FILE='LIBURU.DAT'/ZENBAKIA 1-3 SEXUA 5 HERRIA 7
INTELIGE 9-11 ULERMENO 13-14 ESKOLAMO 16 TELEBIS1 18-19
FRUSTRAZ 21-22 ANTSIETA 24-25 MAILASOZ 27 TELEBIS2 29-30.
VARIABLE LABELS INTELIGE 'ADIMEN-KOEFIZIENTEA'
/ULERMENO 'AHOZKO ULERKUNTZA'
/TELEBIS1 'TELEBISTA AURREAN ORDUAK LEHENENGO NEURKETA'
/MAILASOZ 'MAILA SOZIOEKONOMIKOA'
/TELEBIS1 'TELEBISTA AURREAN ORDUAK BIGARREN NEURKETA'.
```

```
VALUE LABELS SEXUA 1 'EMAKUMEA' 2 'GIZONA'
/HERRIA 1 'ATAUN' 2 'EIBAR' 3 'GASTEIZ'
/ESKOLAMO 1 'IKASTOLA' 2 'PUBLIKOA' 3 'PRIBATUA'
/MAILASOZ 1 'OSO GOIKOA' 2 'GOIKOA' 3 'ERDIKOA' 4 'APALA'
5 'OSO APALA'.
MISSING VALUE INTELIGE(999)/ESKOLAMO MAILASOZ(9)/ULERMENO
TELEBIS1 FRUSTRAZ ANTSIETA TELEBIS2(99).
```

Programa horren bitartez hasierako datu-irakurketa egin dugu eta fitxategi aktibo batean gordetzeko agindu hau idatzi beharko dugu:

```
SAVE OUTFILE='LIBURU.SYS'.
```

Aurreko aginduarekin gorde dugun fitxategi aktibo hau irakurtzeko, nahikoa izango da GET FILE baten bidez martxan jartzea. Honela idatz dezakegu.

```
SET LIS 'L.LIS'/LOG 'L.LOG'/MORE OFF.
GET FILE 'LIBURU.SYS'.
```

Gatozen bada ondoko instrukzioa erabiliz lorturiko emaitzak aztertzera. Honekin gure laginean herri bakoitzeko zenbat pertsona dagoen jakingo dugu.

```
FREQUENCIES VARIABLES=HERRIA.
```

Azpiprogramaren bitartez lortu ditugun emaitzak ondorengoak dira:

HERRIA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
ATAUN	1	46	36.8	36.8	36.8
EIBAR	2	31	24.8	24.8	61.6
GASTEIZ	3	48	38.4	38.4	100.0
	TOTAL	125	100.0	100.0	
Valid Cases	125	Missing Cases		0	

Lehenengo, ikertzen ari garen aldagaiaren izena azaltzen du, segidan aldagai-balioei dagozkien etiketak agertzen dizkigu baldin eta etiketa horiek ezarri baditugu, gure kasuan ATAUN, EIBAR eta GASTEIZ. Jarraian balioak eta dagozkien maiztasunak azaltzen dira. Gure kasuan, Atauneko 46, Eibarreko 31 eta Gasteizeko 48 pertsonak. Ondoko zutabean PERCENT izenburupean portzentaiak ematen dizkigu eta horren ondoan VALID

PERCENT izenburupean agertzen diren portzentaiak. Zutabe honek garrantzia izango du gure datu-fitxategian datu desagerturik dagoenean. Dena den hurrengo adibidean ikusiko dugu hori. Azkenengo zutabean portzentaia-metatuak azaltzen zaizkigu balio txi-kienetatik handienetaraino.

Ikus ditzagun, bada, azpiprogramaren bigarren idazkerapean lortzen ditugun emaitzak. Honetan FREQUENCIES azpiprogramaren bitartez INTELIGE aldagaiaren analisia egingo dugu. Aldagaiaren histograma eta 10 pertzentil kalkulatzea ere eskatzen diogu. Beraz, P10, P20, P30,....., P90 kalkulatuko dizkigu. Azkenik estatistiko guztiak kalkulatzea ere eskatzen diogu..

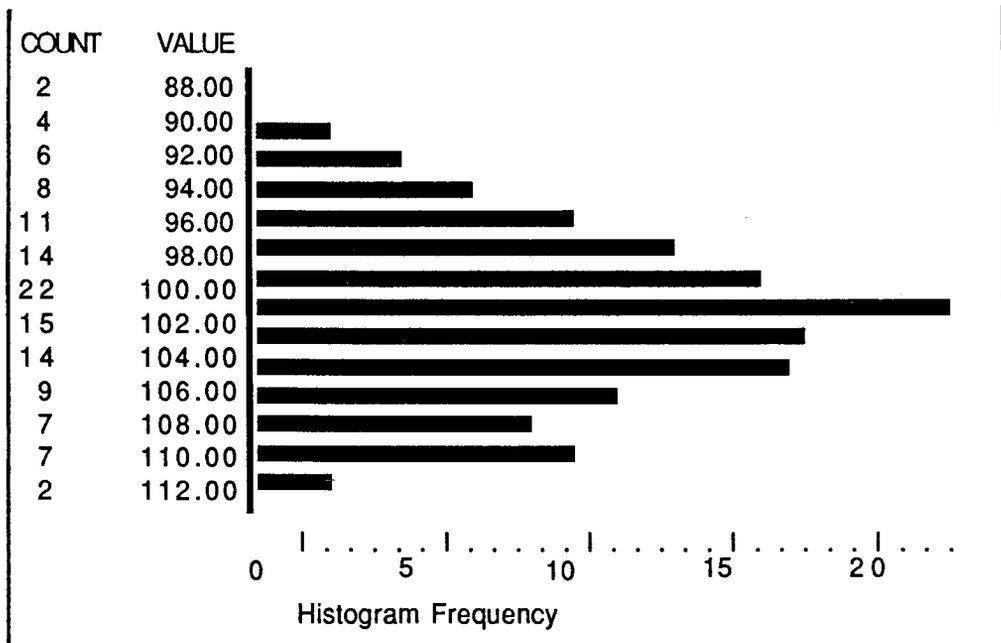
Idatzi dugun instrukzioa hau izan da:

FREQ VAR=INTELIGE/HISTOGRAM/NTILES=10/STATISTICS=ALL.

Lortu ditugun emaitzak hauek izan dira:

INTELIGE ADIMEN-KOEFIZIENTEA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	88	2	1.6	1.7	1.7
	90	4	3.2	3.3	5.0
	92	6	4.8	5.0	9.9
	94	8	6.4	6.6	16.5
	96	11	8.8	9.1	25.6
	98	14	11.2	11.6	37.2
	100	22	17.6	18.2	55.4
	102	15	12.0	12.4	67.8
	104	14	11.2	11.6	79.3
	106	9	7.2	7.4	86.8
	108	7	5.6	5.8	92.6
	110	7	5.6	5.8	98.3
	112	2	1.6	1.7	100.0
	999	4	3.2	MISSING	
	TOTAL	125	100.0	100.0	



Mean	100.479	Std Err	.498	Mediam	100.000
Mode	100.000	Std Dev	5.483	Variance	30.068
Kurtosis	-.436	S E Kurt	.437	Skewness	-.065
S E Skew	.220	Range	24.000	Minimum	88.000
Maximum					

Percentile	Value	Percentile	Value	Percentile	Value
10.00	94.400	20.00	96.00	30.00	98.00
40.00	100.000	50.00	100.000	60.00	102.000
70.00	104.000	80.00	106.000	90.00	108.000

Valid Cases 121 Missing Cases 4

Ikus dezakegunez, emaitza hauen lehenengo zatia eta lehen ikusi duguna formato berdinekoak dira. Hemen VALID PERCENT delako zutabeak duen garrantzia agerian dago. Izan ere datu desagerturik dagoenean eta soilik orduan bait da interesgarria. Datu horiek, desagertuak alegia, emaitza-irteeraren azpikaldean 999 baliopean azaltzen dira eta dagoz-kien maiztasuna 4 da. Horrek zera adierazten digu 125 sujeturen artean 121 sujeturen adimen-koefizienteari buruzko datuak ditugula. Gainontzeko 4 kasuetan ez daukagu daturik. PERCENT zutabean balio bakoitzari dagokion sujetu-portzentaia sujetu-–kopuruarekiko agertzen da; gure kasuan 125 pertsona. VALID PERCENT zutabean berriz, agertutako datuekiko portzentaiak kalkulaturik daude, hots, desagerturik daudenak bazterturik. Azken era honetara portzentaia metatuak ere kalkulatzen dira.

Ondoren aldagaiari dagokion histograma azaltzen digu.

Hurrengo zatian estatistikoak azaltzen dizkigu, gure kasuan STATISTICS=ALL idatzi dugunez posible diren guztiak azaltzen direlarik. Denak aztertuko ditugu, kasu bakoitzean azpiprograman erabili behar dugun hitza adieraziz.

FREQUENCIES azpiprogramaren estatistikoak

MEAN : aldagaiaren batezbestekoa kalkulatzen du. Gure kasuan 100.479 da.

MEDIAN : mediana. Gure kasuan MEDIAN izenburupean azaltzen den balioa 100 da.

MODE : moda. 100 gure adibidean.

VARIANCE : bariantza. (lagin-kuasibariantza). Kalkuluak eskuz egingo bagenitu balio hori lagin-kuasibariantzatik eginiko populazio-bariantzaren estimazioa dela ikus genezake. Gure adibidean 30.068 da.

STDEV : desbidazio tipikoa. Gure kasuan 5.483. Bariantzaren estimazioan bezala kalkulua kuasibariantzan oinarriturik dago, hots, desbidazio tipikoen kalkulua lagin-kuasibariantzan oinarritzen da.

SEMEAN : batezbestekoaren errore tipikoa;desbiderapen standarda eta lagin-tamainaren bigarren erroaren arteko zatiduraren ondorioa da. STD ERR izenburupean azaltzen da eta bere balioa .498. Irakurleak balio hau desbidazio tipikoa (5.483) eta sujetu-kopuruaren bigarren erroraren (11) arteko zatidura dela konproba dezake.

SKEWNESS: asimetri koefizientea. Bere balioa -0.065 da.

SESKEW : asimetri koefizientearen errore tipikoa. Bere balioa SESKEW izenburupean azaltzen da eta gure kasuan .220 da.

KURTOSIS : zorroztasun-koefiziente edo kurtosia. Gure kasuan -0.436.

SEKURT : zorroztasun-koefizientearen errore tipikoa. Gure kasuan .437.

RANGE : ibiltartea edo hedapena. (Baliorik handiena-Baliorik txikiena). Gure kasuan 24.

MINIMUM : baliorik txikiena. Adibidean 88.

MAXIMUM : baliorik handiena. Bere balioa 112 da.

SUM : aldagaiaren balio guztien batuketa. Gure kasuan 12158.

ALL : estatistiko guztiak.

Lorturiko emaitzak aztertzen jarraituz, ondorengo zatian, NTILES=10 idatzi dugunez 10 pertzentilen kalkuluak egingo dizkigu. Beraz, aldagaiaren ibiltartea 10 zatitan banatuko du, ebaki-puntuak, 9 guztira, dezilei dagozkienak direlarik. NTILES=4 idatzi izan bagenu, koartilak kalkulatuko zizkigun. Pertzentil jakin batzuk kalkulatzea nahi izango bagenu, esaterako 25, 33.3 eta 66.7, honela eska geniezaioke:

PERCENTILES=25 33.3 66.7

Aurrera joz lehen esaniko hirugarren instrukzioen aztertzeari ekingo diogu:

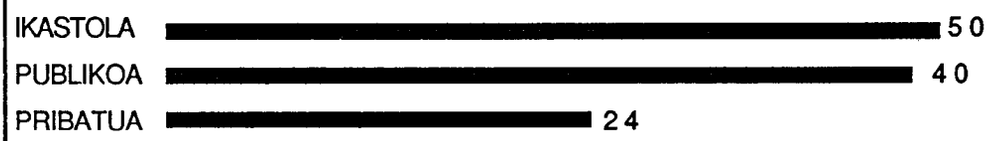
FREQ VAR=SEXUA HERRIA INTELIGE TO TELEBIS2/BARCHART.

Instrukzio honen bitartez zera eskatzen diogu: batetik, SEXUA eta HERRIA aldagaiak, eta bestetik INTELIGE eta TELEBIS2 aldagaien artean dauden aldagaien maiztasunak kalkulatzea, esandako biak barne. Aldagai bat baino gehiagoren maiztasunak kalkulatu nahi ditugunean, bi eratara egin dezakegu: lehenengoan VAR hitzen ostean aldagaien izenak banaka idatzi behar ditugu eta bigarrenean, DATA LIST FILE delakoan elkarren jarraian azaltzen direnean, TO hitza erabil dezakegu; nondik nora zehazten duen hitza alegia. Adibidez aztertzen dugun instrukzioan INTELIGE eta TELEBIS2 muturren arteko aldagaien analisia agintzen dugu, beti ere azken horiek barne direlarik.

Amaieran marra-diagrama egin diezagula eskatzen diegu (BARCHART). Ez ditugu ordenadore-irteera guztiak aztertuko. Adibide gisa ESKOLAMO delakoen emaitzak aurkezten ditugu; INTELIGE eta TELEBIS2 aldagaien artean dagoen aldagai batenak hain zuzen.

ESKOLAMO

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
IKASTOLA	1	50	40.0	40.7	40.7
PUBLIKOA	2	49	39.2	39.8	80.5
PRIBATUA	3	24	19.2	19.5	100.0
	9	2	1.6	MISSING	
	TOTAL	125	100.0	100.0	



Valid Cases 123 Missing Cases 2

Aurreko atalean aldagaiak aldatzeko bi posibilitate ikusi ditugu. Alde batetik, aldagaiak birkodetzeko posibilitatea agertu dugu, RECODE agindua erabiliz, eta bestetik aldagai berriak sortzearen posibilitatea ere aztertu dugu. Orain kasu bakoitzean erabil daitezkeen adibideak plazaratuko ditugu. Lehenengo kasuan MAILASOZ aldagaia muturretako taldeak multzokatuz RECODE instrukzioaren bitartez birkodetu dugu eta, logikoa denez, balioen etiketak birdefinitu behar ditugu. Ondorengo instrukzio-sortan aldagai berri bat sortzen dugu. Honako hau: sujetu bakoitzak telebistaren aurrean egiten duen orduen batezbestekoen kopurua. Horretarako TELEBIS aldagaia sortzen dugu eta 99 balioa (datu desagertuei dagokiena) kontutan hartu behar dugu. Beraz, MISSING instrukzioa sartu beharra edukiko dugu.

```
RECODE MAILASOZ (1,2=1) (3=2) (4,5=3).
VALUE LABELS MAILASOZ 1 'GOIKOA' 2 'ERDIKOA' 3 'BEHEKOA'.
FREQ VAR=MAILASOZ/HISTOGRAM.
COMPUTE TELEBIS=(TELEBIS1 + TELEBIS2)/2.
MISSING VALUE TELEBIS(99).
FREQ VAR=TELEBIS/HISTOGRAM/STATISTICS=ALL.
```

ADIERAZPEN GRAFIKOAK

Orain arte adierazpen grafikoetarako bi forma erabili ditugu: HISTOGRAM eta BARCHART,erabilienak alegia. Dena dela SPSS/PC+ paketeak badauka beste zenbait posibilitate. Beste bat bakarrik aipatuko dugu, interesgarria suerta daitekeelakoan:

HISTOGRAM=NORMAL.

Hemen aldagaiaren histograma ikus dezakegu eta gainetik, gertuen dagoen banaketa normal teorikoen adierazpena. Jakina denez kurba normal honen parametroak banaketarenak dira, hots, aztertzen ari garen aldagaiaren batezbestekoa eta desbidazio tipikoa.

EMAITZEN FORMATOA

Berehala jabetuko gara SPSS/PC+ erabiliz egindako analisien emaitzak inprimatzeko behar den paperen kantitateaz. Bestetik, dugun paperaren formatua dela eta, baliteke guri inprimatze-formatu hori aldatzea interesatzea. Horretarako FORMAT agindua erabiliko dugu. Agindu honek ematen dizkigun posibilitate garrantzitsuenak ondorengoak dira:

CONDENSE : Emaitzak hiru zutabetan azaltzen dira, portzentaien balioak gertuen duten zenbaki osoetara biribilduz.

NEWPAGE : Taula bakoitza orrialde berrian idatziko du.

NOTABLE : Maiztasun-taulak ez ditu azaltzen.

AFREQ : Aldagaiaren balioak maiztasun-gorakorreko ordenetan azaltzen ditu.

DEFREQ : Emaitzak maiztasun-beherakorreko ordenetan azaltzen ditu.

Adibide gisa, ondorengo azpiprogramaren emaitzak aurkezten ditugu. Horietan formatu kondentsatua erabiltzea eta banaketa normalaren kurba, aldagai horri dagokion histogramaren gainean ager dezala eskatzen diogu.

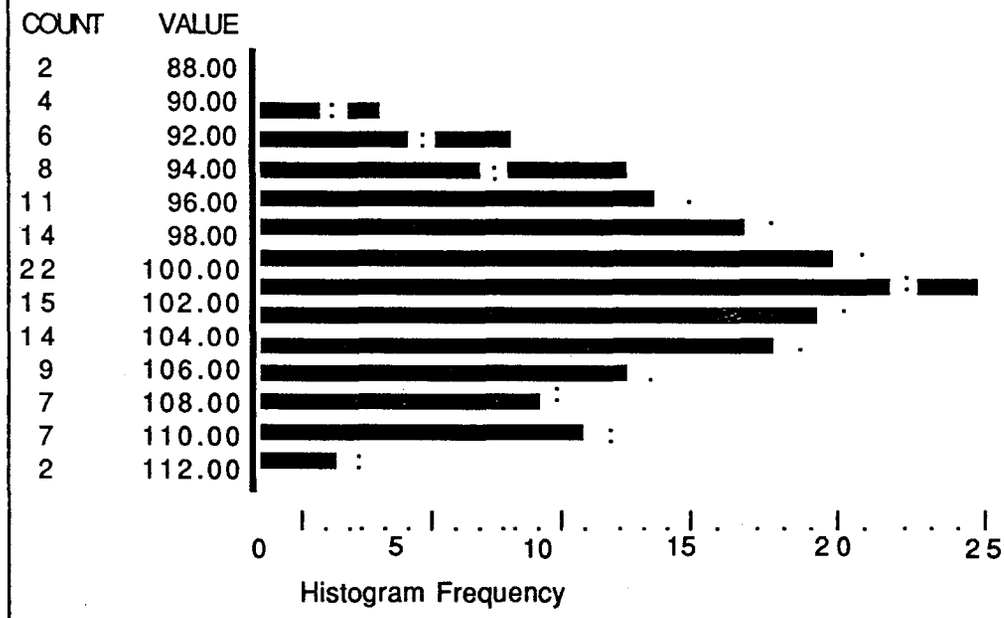
FREQ VAR=INTELIGE/FORMAT=CONDENSE/HISTOGRAM=NORMAL.

Hona hemen lorturiko emaitzak:

VALUE	FREQ	PCT	CUM PCT	VALUE	FREQ	PCT	CUM PCT	VALUE	FREQ	PCT	CUM PCT
88	2	2	2	98	14	12	37	108	7	6	93
90	4	3	5	100	22	18	55	110	7	6	98
92	6	5	10	102	15	12	68	112	2	2	100
94	8	7	17	104	14	12	79				
96	11	9	26	106	9	7	87				

MISSING DATA

VALUE	FREQ
999	4



2. DESCRIPTIVES *AZPIPROGRAMA.*

FREQUENCIESen antzerakoa da. FREQUENCIESen STATISTICS delako azpiaginduaren estatistiko berdinak erabiltzen ditu, baina ez du irudi-adierazpenik egiteko posibilitaterik.

Azpiprograma honen garrantzia aldagaien tipifikazioan datza eta gomendagarria izan daiteke, lantzen ditugun aldagaiak tipifikatzeko eta "aktibo" deritzon fitxategian gordetzeko. Horrela azpiprograma berean aldagai berri horiekin, tipifikatuekin hain zuzen, zenbait analisi garrantzitsu eta interesgarri egin dezakegu aurreragoko ataletan ikusiko dugunez.

Instrukzio honi dagokion idazkera honako hauxe da:

DESCRIPTIVES VAR=aldagai-zerrenda.

Gutxieneko agindua idatzirik batezbestekoaren balioa, desbidazio tipikoa, balio txikiena eta handiena, eta, baita baliagarriak diren kasu-kopurua ere.

OPTIONS azpiaginduak, FREQUENCIESen azaltzen ez zenak, datu desagertuen trataera arautu egiten du. Baita emaitzen aurkezpena ere ..... Dagokion sintaxia honako hau da:

/OPTIONS=aukerei dagozkien zenbakiak.

Hautabideak edo opzioak hauek izango lirateke:

OPTION 1 : erabiltzaileak esaniko datu desagertuak sartuko ditu.

OPTION 2 : balioen etiketak baztertzen ditu.

OPTION 3 : puntuazio tipikoak fitxategi aktiboan gordetzen ditu (ikus adibidea). Fitxategi aktibo hau dela bide, sorturiko aldagaiak gure fitxategiko beste edozein aldagairen antzera erabili ahal izango ditugu. Programa hau amaitzearekin batera fitxategi aktiboa desagertu egingo da SAVE azpiaginduaz salbatzen ez badugu.

OPTION 5 : aldagai-definiziotan izendaturiko aldagaietan balioren bat desagerturik duten kasuak ez ditu kontutan hartzen.

OPTION 6 : estatistikoak imprimatzerakoan, lerro bakoitzean bat imprimatzen du.

OPTION 7 : estatistikoak iladan inprimatzen dira.

OPTION 8 : aldagaien balioak baztertzen ditu.

Aldagaiak tipifikatu, eta fitxategi aktibo baten salbatzea nahiko bagenu, era honetako instrukzio bat idatzi beharko dugu. Ikusten dugunez, oinarrizko aldagai bakoitzen atzekaldean, parentesi artean, izen berri bat idatzi dugu. Izen hauek izango dira aldagai tipifikatuak gordetzeko erabiliko dituenak. Beraz, hemendik aurrera, adibidez, inteligentzi koefizienteen bi puntuazio edukiko ditugu: INTELIGE, eta ZINTELIG; aurreko aldagien puntuazio tipifikatuekin osatutakoa. Gauza bera gertatuko da TELEBIS1 eta TELEBIS2 aldagaiekin.

DESCRIPTIVES VAR=INTELIGE (ZINTELIG) TELEBIS1 (ZTELEBIS)
TELEBIS2 (ZTELEBI2)/OPTIONS=3.

Lortu ditugun emaitzak hauek dira:

Nurber of Valid Observations (Listwise) = 121.00

Variable	Mean	Std Dev	Minimum	Maximum	N
INTELIGE	100.48	5.48	88	112.00	121
TELEBIS1	19.11	7.24	9	32	123
TELEBIS2	16.48	5.85	9	29	123

The following z-score variables have been saved on your active file:

From Variable	To Z-score	Weighted Valid N
INTELIGE	ZINTELIG	121
TELEBIS1	ZTELEBI1	123
TELEBIS2	ZTELEBI2	123

Ikus dezakegunez emaitzetan bi zati diferente daude: alde batetik azterturiko aldagai bakoitzaren emaitzak eta bestetik azpikaldean OPTIONS=3 dela medio sorturiko aldagai berriak. Aldagai berri horiek guk agindutako izenekin ikus ditzakegu hurrenez hurren: ZINTELIG, ZTELEBI1 eta ZTELEBI2.

FREQUENCIES azpiprogramapean egin dezakegun moduan hemen ere estatistiko zehatzak kalkulatzea eska diezaiokegu. Honako zenbakiak erabiliko ditugu hain zuzen:

/STATISTICS=estatistikoei dagozkien zenbakiak.

zenbakia		estatistikoa
1	:	batezbestekoa
2	:	batezbestekoaren estimazio-errore tipikoa.
5	:	desbidazio tipikoa.
6	:	bariantza
7	:	kurtosi-koefizientea
8	:	asimetri koefizientea
9	:	ibiltartea edo hedapen osoa
10	:	minimoa
11	:	maximoa

12 : kasu guztien balio-batuketa
ALL : guztiak

3. MEANS *AZPIPROGRAMA*

Aldagai kualitatibo bat erabiliz, badakigu gure lagineko elementu guztien sailkapen bat egiten dugula. Adibidez, SEXUA aldagaia erabiltzen badugu, badakigu, gure lagina bi zatitan banatzen dugula: gizonezkoak eta emakumezkoak. Atal honetan ikusiko dugun azpiprogama, aldagai kualitatibo bat erabiltzen dugunean lortzen ditugun azpilagin bakoitzean edozein aldagai KUANTITATIBOren deskribapena egiteko baliagarria izango da. Bi aldagai kualitatiboen artean egon daitekeen elkarrekiko erlazioaren deskribapena CROSSTABS azpiprograma azaltzean plazaratuko dugu.

Lagin osoko balizko azpilaginak eratzerakoan nolabaiteko sailkapen-erizpide bati jarraituko diogu, hots, gure laginean dagoen aldagai kualitatiboren batean oinarriturikoa edota IF zein RECODE instrukzioz egoki sortu edo transformaturiko aldagai kuantitatibotan oinarritutakoa.

Oro har MEANS azpiprogramari dagokion sintaxia honako hauxe izango litzateke:

MEANS TABLES=ULERMENO BY HERRIA.

TABLES hitza ez dela beharrezkoa esan behar dugu, hots, instrukzioak gauza bera sortuko luke honela idatziko bagenu:

MEANS ULERMENO BY HERRIA.

Ikus ditzagun, bada, sortzen dituen emaitzak:

Summaries of ULERMENO AHOZKO ULERKUNTZA
By levels of HERRIA

Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			13.6016	3.3651	123
HERRIA	1	ATAUN	15.9565	2.7322	46
HERRIA	2	EIBAR	12.5667	2.8763	30
HERRIA	3	GASTEIZ	11.9574	2.9189	47

Total Cases = 125
Missing Cases = 2 OR 1.6 PCT.

Goiko laukian ikusten dugunez oinarrizko instrukzioak, batezbestekoak, desbidazio tipikoak eta klase bakoitzari dagozkion sujetu-kopurua azaltzen dizkigu. Sorturiko hiru klase horiek herriaren arabera egin dira. Beraz, agertzen zaiguna zera da: herri bakoitzari dagozkion batezbestekoa, desbidazio tipikoa eta sujetu-kopurua. Era berean, lehenengo lerroan aldagai kuantitatiboaren deskribapena, lagin osoarekiko egiten digu.

Lehen aipaturiko oinarrizko instrukzio horri ondoren azaltzen diren estatistikoak eta aukerak erantsi diezazkiokegu:

MEANS azpiprogramaren estatistikoak:

1. Bariantz analisi sinplea egiten du.

2. Linealtasun-testa egiten du. Karratuen batuketa, batezbesteko koadratikoak, eta osagai lineal zein ez-linealei loturiko askatasun-graduak kalkulatzen ditu. F arrazoia ere azaltzen du.

MEANS azpiprogramaren opzio edo hautabide garrantzitsuenak:

1. Desagerturiko datuak kontutan hartzen ditu.

2. Menpeko aldagaian datu desagertuak dauzkaten elementuak baztertu egiten ditu.

3. Aldagai eta balioen etiketak ez dizkigu azaltzen.

6. Talde bakoitzaren balio-batuketak azaltzen ditu.

7. Ez ditu talde bakoitzeko desbidazio tipikoak azaltzen.

8. Ez ditu balioen etiketak azaltzen.

10. Ez ditu aldagai aske edo independenteen etiketak azaltzen.

11. Ez ditu taldeen batezbestekoak azaltzen.

13. Taldeen bariantzak azaltzen ditu.

Lehen erabilitako instrukzioari estatistikoak eransten badizkiogu, honela idatziko dugu:

MEANS ULERMENO BY HERRIA/STATISTICS=2.

Lorturiko emaitzak jarraian aurkezten ditugunak dira eta haien artean lehenengo zatikoak eta arestian ikusitakoak berdinak dira.

Summaries of ULERMENO AHOZKO ULERKUNTZA
By levels of HERRIA

Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			13.6016	3.3651	123
HERRIA	1	ATAUN	15.9565	2.7322	46
HERRIA	2	EIBAR	12.5667	2.8763	30
HERRIA	3	GASTEIZ	11.9574	2.9189	47

Total Cases = 125
Missing Cases = 2 OR 1.6 PCT.

Summaries of ULERMENO AHOZKO ULERKUNTZA
By levels of HERRIA

Value	Label	Mean	Std Dev	Um of Sq	Cases
1	ATAUN	15.9565	2.7322	335.9130	46
2	EIBAR	12.5667	2.8763	239.6767	30
3	GASTEIZ	11.9574	2.9189	391.9149	47
Within Groups Total		13.6016	2.8390	967.1946	123

Analysis of Variance

Source	Sum of Squares	D.F.	Mean Square	F	Sig.
Between Groups	414.2851	2	207.1425	25.7002	.0000
Linearity	370.4406	1	370.4406	45.9606	.0000
Dev. Linearity	43.8445	1	43.8445	5.4398	.0213

R = -.5178 R Squared = .2681

Within Groups	967.1946	120	8.0600

Eta = .5476 Eta Squared = .2999

Emaitzen lehenengo zatian, talde ezberdinen deskribapena egiten du. Aurreko adibideko formatu berdina erabili du. Bigarren zatian, eskatutako bariantz analisien emaitzak agertzen ditu. Analisi honen emaitzak eta ONEWAY azpiprograman agertuko direnak berdinak direnez geroz, hemen ez ditugu aztertuko, zeren eta ONEWAY azpiprogramari dagokion atalean patxadaz eta zehatz aztertuko bait ditugu.

Deskribapen bat baino gehiago egin nahi dugunean, MEANSen instrukzioa laburrago idatz dezakegu:

MEANS ULERMENO INTELIGE BY HERRIA.

Aurreko horrek ULERMENO eta INTELIGE aldagaien deskribapena eskaintzen digu, baina HERRIA aldagai-balioen eraginez sorturiko taldeen arabera.

Bi edo hiru sailkapen-erizpideen arabera, sujetuen deskribapena egitea nahi izanez gero, ondorengo instrukzioa erabil dezakegu:

MEANS ULERMENO BY HERRIA BY SEXUA.

Hona hemen horrela lorturiko emaitzak:

Summaries of	ULERMENO	AHOZKO ULERKUNTZA			
By levels of	HERRIA				
	SEXUA				
Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			13.6016	3.3651	123
HERRIA	1	ATAUN	15.9565	2.7322	46
SEXUA	1	EMAKUMEA	15.2857	3.4065	14
SEXUA	2	GIZONA	16.2500	2.3827	13
HERRIA	1	EIBAR	12.5667	2.8730	30
SEXUA	1	EMAKUMEA	13.6316	2.4991	19
SEXUA	2	GIZONA	10.7273	2.6112	11
HERRIA	1	GASTEIZ	11.9574	2.9189	47
SEXUA	1	EMAKUMEA	11.0857	2.0346	35
SEXUA	2	GIZONA	14.5000	3.6556	12
Total Cases =	125				
Missing Cases =	2 OR 1.6 PCT.				

Hor, HERRIA aldagaiaren arabera, batezbestekoa, desbidazio tipikoak eta sujetu-kopurua, taldeka agertzen zaigu, hiru talde direlarik. Talde bakoitzean sexu bakoitzari dagozkion estatistikoak ere azaltzen zaizkigu. Azkenik, azpikaldean darabiltzagun hiru aldagaietatik, gutxienez batean datu desagerturik dutenen kasu-kopurua azaltzen digu.

4. PLOT *AZPIPROGRAMA*

Estatistika-eskuliburutan aldagai koantitatiboen arteko erlazioa azaltzean, sakabanatze-diagrama edo hodei-puntua abaiapuntutzat hartzea, ohizkoa da. Diagrama horretan aldagai independenteari ardatz horizontala eta menpeko aldagaiari, ardatz bertikala lotzen zaie.

Gogoan izan, bestetik, maiz suertatzen den errorea: koerlazio linealaren koefizientea ikusita, aldagaien arteko erlazio eza ebaztea eta hori beste erlazio motak izan daitezkeela pentsatu gabe. Irudi-adierazpenetan honelako auziak agerian gera daitezke.

SPSS/PC+-en inplementaturiko PLOT azpiprogramak bi dimentsioko irudiak sor ditzake eta aldagai biko sakabanatze-diagramak adierazteko erabiltzen da.

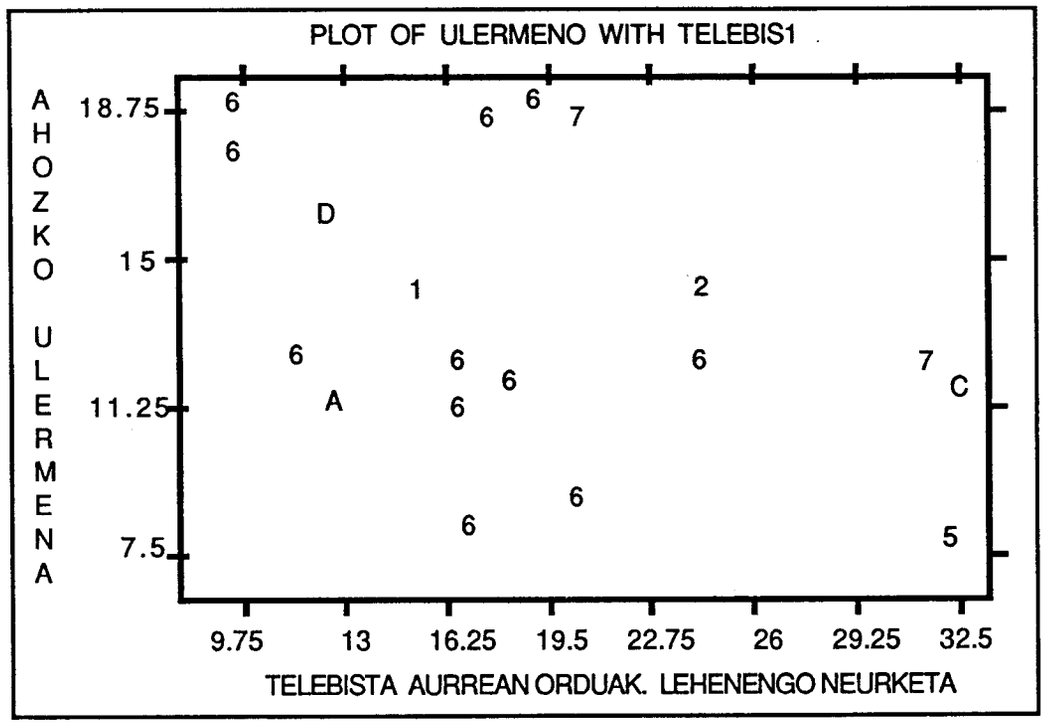
Oro har PLOT azpiprogramari dagokion sintaxia honako hau izango da:

PLOT PLOT= ardatz bertikalari lotutako aldagaiak WITH ardatz horizontalari
lotutako aldagaiak.

Demagun honako programa hau:

PLOT PLOT=ULERMENO WITH TELEBIS1.

Horrek ondoko irudia sortzen digu:



Puntu bakoitzari aldagai biko banaketaren maiztasun absolutua eransten dio. Maiztasuna 10 edo handiagoa denean automatikoki letrak eransten dizkie. (10=A, 11=B, 12=C,....).

Adierazpen anitzezko adibide bat:

PLOT PLOT=Y_1 Y_2 WITH X_1 X_2 X_3.

Honek ondorengo irudiak sortuko lituzke:

$Y_1 X_1$	Y_1X_2	Y_1X_3
$Y_2 X_1$	Y_2X_2	Y_2X_3

Beste era batera PAIR kode-hitzaz egin genezake:

PLOT PLOT=Y_1 Y_2 WITH X_1 X_2 (PAIR).

zeinen arauera $Y_1 X_1$ eta $Y_2 X_2$ "bikoteak" adierazten diren.

Amaitzeko ikus dezagun honako hau:

PLOT PLOT=Y_1 WITH X_1 X_2 ; Y_2 WITH X_3 X_4.

zeinek $Y_1 X_1$, $Y_1 X_2$, $Y_2 X_3$ ETA $Y_2 X_4$ irudiak sortuko bait ditu.

KONTROL-ALDAGAIAREN ERABILERA

Kontrol-aldagaia hirugarren aldagaia da; koantitatiboa zein koalitatiboa izan daitekeena alegia. Aldagai honen balioak bi dimentsioko espazioaren puntuak definitzeko erabiltzen dira. Ez da pentsatu behar, ordea, hori erabiltzerakoan hiru dimentsioko irudirik lortuko dugunik.

Oro har honelakoa izango litzateke instrukzioa:

PLOT PLOT=aldagaiak ardatz bertikala WITH aldagaiak ardatz horizontala BY kontrol-aldagaia.

Jakina denez PLOT azpiprogramaren bakoitzean kontrol-aldagai bakar bat erabil dezakegu.

Ikus ditzagun, bada, kontrol-aldagaia instrukzioan sartzeak dakartzan ondorioak zeintzuk diren:

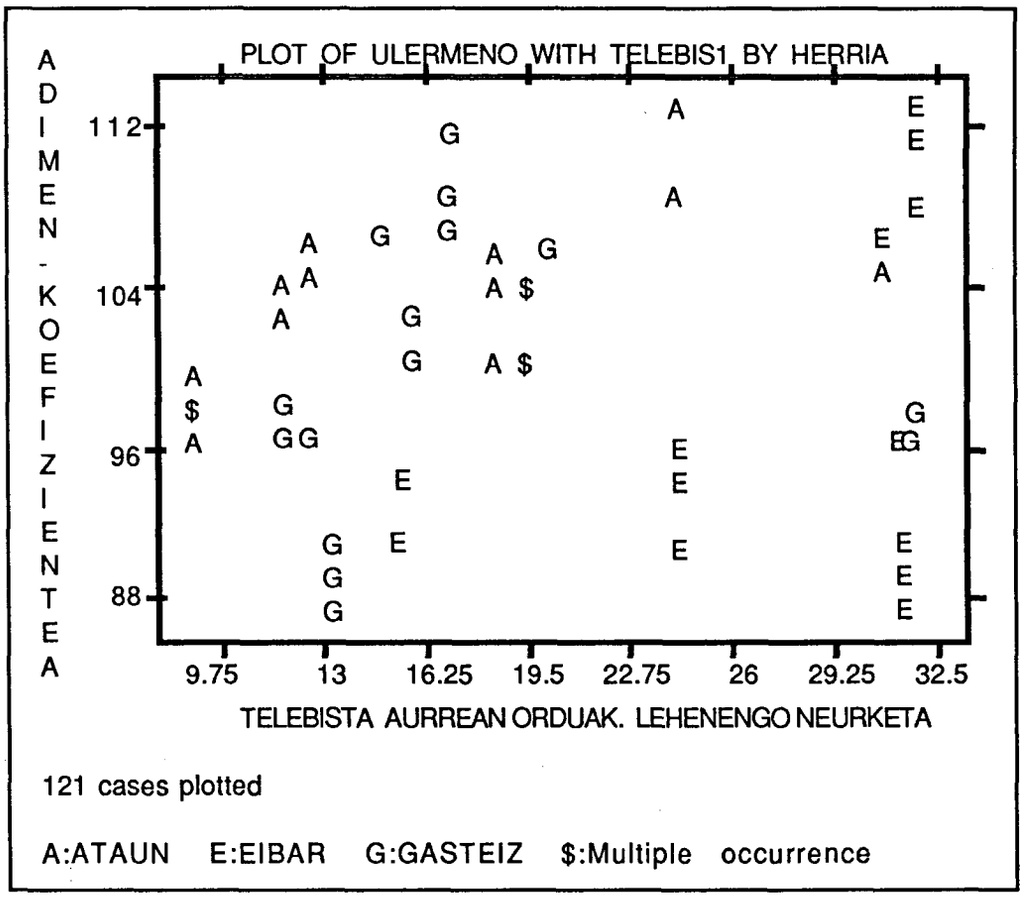
- kontrol-aldagaiaren balioak etiketa gabekoak badira, puntuei dagokion baliotatik lehenengo ikurra eransten zaie.

- balioak etiketadunak badira, puntuei eransten zaien ikurra etiketen lehenengoa izango da.

Hona hemen azpiprograma:

PLOT PLOT=INTELIGE WITH TELEBIS1 BY HERRIA.

horretan HERRIA kontrol-aldagaia delarik; beraren modalitateak G (GASTEIZ) E (EIBAR) eta A (ATAUN) dira eta honako irudi hau sortzen du:



Hots, puntu jakin batean sujetu bat baino gehiago dagoenean $ ikurra agertuko da (izan ere, puntu berean gizonezko eta emakumezkoak batera suerta bait daitezke). Bestetik A, E eta G eransteak, etiketak ATAUN, GASTEIZ eta EIBAR izatearen ondorio dira. Balio horiek etiketa gabekoak izan balira kontrol-aldagaiarekiko irudiko ikurrak aldagaiaren balioak izango zitezkeen; gure kasuan 1, 2 eta 3 alegia.

Gure helburuetatik at egonagatik azpiprograma honen azaleko aurkezpena hemen plazaratzea interesgarria iruditu zaigu. Dena den, baten batek honetan sakontzea nahi izango balu, paketearen eskuliburura jo lezake.

5. CORRELATION *AZPIPROGRAMA*

Baliteke aldagai kuantitatiboen artean, diskretoak edo jarraiak zein puntutaraino erlazionatuta dauden aztertzea interesgarria izatea, eta, baita, erlazio hau kuantifikatzea ere.

Hasieran, abiapuntu bezala eta sinpleena izanik, erlazio lineala planteatzen da. Bestetik metodo estatistiko sinpleen sorta bat, tartean anitzezkoak eta aldagai anitzezkoak direla, aldagai zein aldagai-multzoen arteko erlazio linealetan oinarritzen da.

Esan beharra dago, zerotik gertu dagoen Pearson-en koerlazio linealaren koefizien-

teak erlazio lineal eza baino ez duela adierazten. Beraz ez da baztertzen erlazio lerromakurraren posibilitatea.

Hauxe da gomendioa beraz: bi aldagai kuantitatiboen artean egon daitekeen erlazioa aztertu nahi dugunean Pearson-en koerlazio linealaren koefizientea erabiltzea hain zuzen eta hori CORRELATION delako azpiprogramaren bitartez egitea.

Azpiprograma honi dagokion instrukzioa idazterakoan VARIABLES deritzon azpiagindua ere idatzi behar dugu:

```
CORRELATION VAR=aldagaien zerrenda 1
                /aldagaien zerrenda 2
                /...
```

Aldagai guztien arteko koerlazioak lortu nahi baditugu, aldagaien zerrendak agertu behar duen lekuán ALL hitza idaztea nahikoa izango litzateke. Era berean bi aldagaien arteko aldagai guztiak erabili nahi baditugu, nahikoa da hasierakoa eta amaierakoa idatzi eta tartean TO hitza idaztea.

Horrela ondorengo instrukzioa idaztean:

```
CORRELATION VARIABLES=INTELIGE ULERMENO TELEBIS1 TO ANTSIETA.
```

Emaitzak ondoko hauek izango dira:

Correlations:

	INTELIGE	ULERMENO	TELEBIS1	FRUSTRAZ	ANTSIETA
INTELIGE	1.0000	.0564	.0576	-.1120	-.1024
ULERMENO	.0564	1.0000	-.4157**	-.3405**	-.3901**
FRUSTRAZ	-.1120	-.3405**	.7575**	1.0000	.9617**
ANTSIETA	-.1024	-.3901**	.8094**	.9617**	1.0000

N of cases: 121 1-tailed Signif: * - .01 ** - .001
" . " is printed if a coefficiente cannot be computed

Kalkuluetan erabilitako kasu-kopurua 121 da. Izan ere horrekiko beste aukerarik gabe (ikus 2.aukera), VARIABLES azpiaginduak izendaturiko aldagaien artean balioren bat desagerturik dutenak baztertu bait dira.

Koerlazio-koefizienteari dagokion signifikazio-testa aurrerantzean ikusiko dugu; hipotesien testa aztertu eta gero alegia.

Sintaxi aldetik beste idazkera:

CORRELATION=1. aldagaien zerrenda WITH
2. aldagaien zerrenda.

Sintaxi honen pean zerrendetan dauden aldagaien arteko koerlazioak kalkulatzen dira, hots, 1. zerrendakoek 2. zerrendakoekiko dituzten koerlazioak hurrenez hurren.

Demagun, esaterako, ondorengo programa:

CORRELATION VARIABLES=INTELIGE ULERMENO TELEBIS1 TO
ANTSIETA WITH TELEBIS2.

Emaitzak hauek dira:

Correlations:	TELEBIS2
INTELIGE	.0326
ULERMENO	-.3791
TELEBIS1	.8980**
FRUSTRAZ	.6842**
ANTSIETA	.7130**

N of cases: 121 1-tailed Signif: * - .01 ** - .001
" . " is printed if a coefficiente cannot be computed

OPZIOAK EDO HAUTABIDEAK

Dagokion azpiagindua hau da:

/OPTIONS=aukerei dagozkien zenbakiak

zenbakiak

1 : balio desagertuak kalkuluetan sartzen dira.

2 : koefizientea kalkulatzen den aldagaietan ez ditu balio desagertudunen kasuak kontutan hartzen.

3 : koerlazio-koefizienteen signifikazio-maila. Ez da zenbakiz agertuko; baizik eta esanguratsu direnei izartxo bat edo bi erantsiko dizkio.

4 : koerlazio-matrizea agertarazten du, koefiziente bakoitzaren kalkuluetan baliagarria izan den kasu-kopurua adieraziz.

5 : koerlazio koefiziente bakoitzari dagozkion kasu baliagarriak eta signifikazio maila azaltzen ditu.

ESTATISTIKOAK

Ondorengo eran eskatzen dira:

/STATISTICS=estatistikoei dagozkien zenbakiak

Estatistiko 1 : VARIABLES azpiagindupeko aldagai bakoitzaren batezbestekoa, desbidazio tipikoa eta kasu-kopurua.

Estatistiko 2 : bariantza-kobariantza matrizea.

Estatistiko biak nahi izango bagenitu, ALL hitza erabiliko genuke.

11. GAIA
INFERENTZIA ESTATISTIKOA: HIPOTESI-TESTAK ETA KONFIDANTZA-TARTEAK

Hemendik aurrera aztertuko ditugun gaiak, Inferentzia Estatistikoa deiturikoari dagozkionak izango dira, atal honetan aurkezpen orokor bat egingo dugularik.

Inferentzia hitzaz zera uler dezakegu: "gauza batetik ondorio bat atera". Guk gure aldetik Inferentzia Estatistikoen definizio bat ematerakoan zera esan dezakegu: "Estatistikari dagokion atal bat da, zeinak populazioaren lagin adierazgarri baten datuez baliaturik populazio guztirako ondorioak edota zenbait erabaki ateratzeko arauak ematen bait ditu".

Datu batzuen analisiari ekin baino lehen, zera kontutan hartu behar dugu: aztertu behar ditugun datuak ea populazio guztikoak diren edo lagin deritzogun populazio horren zati bati dagozkionak. Eta azken kasu honetan, horrela ateratako emaitzak ea populazioaren ezaugarritzat joko ditugun ala ez, hots, orokortu nahi ditugun ala ez. Demagun edozein hiritako biztanlegoaren altueren batezbestekoa aztertu nahi dugula. Horretarako momentu jakin batean biztanle guztiak neurtzea metodo bat izan liteke. Jakina denez horrela lorturiko neurrien analisiak populazio osoari dagozkionak dira eta, ondorioz, kalkulaturiko batezbestekoa eta biztanlegoaren batezbesteko erreala bat etorriko dira. Hala ere biztanlego osoa une jakin batean neurtzeak duen zailtasuna ez luke inork ukatuko. Horregatik, bada, aurreko populazioko zati bat erabili ohi da, eta azpipopulazio honetatik lorturiko emaitzetan oinarrituz populazio osoaren emaitzak "inferitu" egiten dira. Arestian esan dugunez, azpipopolazio hori "lagina" izendatzen dugu. Lagin honetako elementuen hautaketak garrantzi handia edukiko du, zeren eta aukeratutako elementuak izango bait dira gure neurgailuen bidez neurtuko ditugunak. Lagina zein izango den erabakitzerakoan, hots, populazioarekiko lagin adierazgarria izatea oinarrizko helburutzat jotzen dugunean (beste era batera esanda, populazioari dagozkion emaitzak eta gure laginarenak bat etor daitezen), hurrengo gaian aztertuko ditugun funtsezko bi erizpideri jarraitu behar gatzaizkie. Honako hauei hain zuzen:

- Neurtu behar dugun sujetu-kopuru minimoari.

- Sujetuen hautaketari.

Edonola ere, gai honetan laginaren hautaketa zuzena dela joko dugu. Beraz, darabiltzagun datuak, aztertzen den populazioarekiko adierazgarriak izango dira.

Datu-bilketa egin ondoren,atera nahi ditugun ondorioak honako hiru alderdi hauetan sailka ditzakegu:

1.- Laginean lorturiko emaitzak populazio osoari orokortzea.

2.- Bi taldetan edo gehiagotan lorturiko emaitzen arteko konparazioa egitea.

3.- Sujetu jakin bati buruz lorturiko emaitzen arabera erabakiak hartzea.

1. EMAITZEN OROKORPENA: KONFIDANTZA-TARTEAK

Lagin baten emaitzak populazio osoarekiko orokortu nahi ditugunean, errore-maila delakoaren arriskuaz jabetu behar dugu. Errore-maila hori, ikertzen dugun populazioaren azpipopulazio bateko datuekin ibiltzean datza eta, behaturiko emaitzak populazio-emaitzak estimatzeko baliagarriak badira ere ezin dezakegu ziurtasun osoz populazio-emaitzak berberak direnik esan. Errore-mailak probabilitate-eremura garamatza, hots, probabilitate-hitzak erabili behar ditugu. Adibidez, lagineko sujetuen batezbesteko altuera 155 zentimetro bada, ondorengo bi eratara mintza gaitezke. Lehena 155 cm populazioari dagokion batezbestekoa dela esatea, eta hau arriskutsuagoa izateaz aparte "estatistikoegia" ez litzateke izango. Prozedura horri Puntu-Estimazioa esaten diogu eta egiten dugun hurbilketa edo estimazioa, puntu baten balioa ematea izango da.

Emaitzak plazaratzeko bigarren forma, berriz, zera izan liteke: ikertzen dugun populazioaren altueren batezbestekoa 153 eta 157 zentimetroren artean dagoela eta hori %99ko probabilitateaz dagoela esatea. Honela eginez gero errealitatetik gertuago egoteaz gain estatistiko-hitzetan mintzatzen garen neurrian "zintzoagoa" ere izango da.

Azterketa-objektu den populazioaren emaitza-hurbilketaren prozedura horri Konfidantza Tartezko Estimazioa esaten zaio.

Hurbilketa konfidantza-tartearen bitartez egiten dugunean, hots, konfidantza-tartezko estimazioari ekiten diogunean, ondorengo lau alderdi, zeintzuek hurbilketan eragina duten, kontutan hartu behar ditugu:

- tartearen zabalera.
- signifikazio edo errore-maila.
- desbidazio tipiko edo standarda.
- laginaren tamaina.

Segituan azalduko dugunez, lau alderdi horiek elkarri loturik egongo dira.

Populazioaren errealitatera hurbildu nahi dugunean, hurbilketa hori ahalik eta zehatzena izan dadila nahi dugu; hau da, ez dugu estatistikoki ahalegin handirik egin behar populazioaren altueren batezbestekoa 110 eta 200 zentimetroren artekoa dela ziurtatzeko. Gure asmoa, gure helburua, konfidantza-tartearen zabalera ahal den txikiena izatean datza,

hots, emaitzen aurkezpena ahalik eta finena izan dadila. Baina jakina, hori eta ziurtasuna kontrajarriak dira emaitzen aurkezpena egitean: Altueren batezbestekoa 110 eta 200 zentimetroren artekoa dela %100eko ziurtasunez esan dezakegu; gure errore-maila %0 dela alegia. 153 eta 157 zentimetroren artean dagoela esango bagenu berriz, arrisku handiagoa edukiko genuke eta errore edukitzeko probabilitatea ez litzateke zero izango. Tartearen zabalera txikiagotuko bagenu eta batezbestekoa 154-156 tarteari dagokiola esango bagenu ere, erratzeko maila edo probabilitatea handiagoa izango litzateke. Edozein kasutan konfidantza-tarteen gakoa zera da: zabalera txikieneko tartea izan eta errore-maila ere txikiena izatea. Errore-maila minimoa izateak, egoki asmatzeko probabilitaterik handiena du.

Zabalera txikieneko tarte eta errore-maila minimoa lortzean, darabilgun laginaren tamainak eta desbidazio standardak eragina dute.

Argi dago 100000 pertsonaz osaturiko populaziotik 100 sujetuko laginarekin lana egiten dugunean, lortuko ditugun emaitzen ziurtasuna txikiagoa izango dela 10000 pertsonako lagin batekin lan egiten dugunean baino. Arrazoia garbia da: teorikoki behintzat 10000 pertsonekin lortzen dugun populazioen adierazgarritasuna askoz ere handiagoa izango delako. Tarte-estimazioan eragina duen azken faktorea, aldagaien desbiderapen estandarda izango da. Herri bateko urtebeteko umeen altueren batezbestekoen inferentzia egiterakoan lortuko ditugun emaitzak, herri horretako 18 urte dituzten gazteen altueren inferentziak egiterakoan baino zehatzagoak izango dira. Honen zergatia zera da: 18 urteko pertsonen altuerak, urtebeteko umeenak baino askoz ere sakabanatuagoak izatea. Hau kontutan harturik, desbiderapen standarda izango da konfidantza-tartearen zehaztasunean eragina edukiko duen azken faktorea.

Aipaturiko lau alderdien artean suertatzen diren erlazio orokorrak hauek dira:

.- Zenbat eta errore-maila handiago, orduan eta tarte-zabalera txikiago.

.- Zenbat eta lagin tamaina handiago, orduan eta tarte-zabalera txikiago.

.- Zenbat eta desbidazio standard handiago, orduan eta tarte-zabalera handiago.

Aurreko erlazio horiek beheko formulan, zeina konfidantza-tarteen eraikuntzan klasikoenetariko bat den, agertzen ditugu:

Konfidantza-tartea:

$$\overline{X} \pm T\alpha \frac{\sigma}{\sqrt{n}}$$

Non:

$\overline{X}$ = Laginaren batezbestekoa

$T\alpha$ = α balio jakin batekiko, taulatan aurkituriko balioa.

σ = Populazioaren desbidazio standarda.

m = Laginaren tamaina.

Tartearen zabalera honako hau izango da: $2 * T\alpha \frac{\sigma}{\sqrt{n}}$

2. HIPOTESI-TESTA: I. MOTAKO ERROREA ETA II. MOTAKO ERROREA

Argigarria izan daitekeen eguneroko bizimoduaren adibide bati loturik hipotesi-test baten planteamendu orokorra egingo dugu, geroxeago lagin bat edota lagin batzuen kasuetara zabalduko badugu ere.

Demagun pertsona bat medikuarengana doala eta, honek behaturiko sintomak bide, gaixoaren odol-analisia egitearen beharra adierazi diola, emaitzen arabera ebakuntza edo operazio bat behar den ala ez erabakitzeko.

Odol-analisiak burutu ondoren, medikuak dituen datu horiekin (demagun daukan informazio bakarra dela) erabaki bat hartzera behartuta dago: pertsona hau operatu ala ez operatu.

Horiek horrela, erabakia hartzetakoan ondorengo baliozko bi errore egin litzake:

Lehenengoa, analisiaren emaitzak bide, medikuak ebakuntza ez egitea erabakitzea (beharrezkoa ez delakoan) baina benetan, errealitatean, ebakuntza behar duenean. Horri I. MOTAKO ERRORE esango diogu.

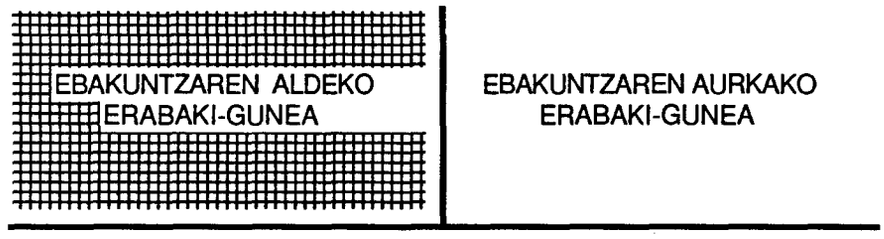
Bigarren errore-klasea kontrakoa izango litzateke,hots, bere gaixoak ebakuntza jasan dezala erabakitzea eta gaixotasuna "gezurrezkoa" izatea. Hau da, analisien emaitzak eta gaitzaren larritasunak bat ez etortzea. Errore honi II. MOTAKO ERRORE esango diogu.

Horrelako egoeran gure inongo irakurlerik ez egotea nahi dugun arren (ez mediku gisa ez eta gaixo gisa ere) planteamendu horren aurrean beheko sarrera biko taula bezalakoa egin dezakegula aitortu behar dugu:

	ERABAKIA	
	ebakuntza BAI	ebakuntza EZ
Bezeroak EZ du ebakuntza behar	II. MOTAKO ERROREA	ERRORERIK EZ
Bezeroak (BAI) ebakuntza behar du	ERRORERIK EZ	I. MOTAKO ERROREA

Dena den, jakina denez medikuak erabakitzerakoan informazio osagarriak izan ohi ditu eta horrela, I. MOTAKO ERROREA zein II. MOTAKO ERROREA ahalik eta txikienak izan daitezen saiatuko da. Inferentzi alorrean Estatistikaren erabiltzailea ordea, sarritan aurkituko da honelako egoeratan.

Estatistika, erroreak (bata zein bestea) txikiagotzen saiatuko da. Azken finean gure "galenoak" nahi lukeena zera izango litzateke: operatu behar diren eta behar ez diren gaixoen arteko muga bereizgarria aurkitzea. Beste era batera esanda, analisi-emaitzen ondorioz gaixoak bereiztu segun mugaren ezker aldean edo eskuin aldean geratzen diren. Grafikoki honela adierazi genezake:



Medikuak odol-analisien arabera erabakiko balu, erabakiak hartzean ondorengo bi kasuetan erratuko litzateke: batetik ebakuntzaren aldeko odol-analisiak izan, benetan operatu behar ez denean (irudian ezkerraldetik kokaturik leudekeenak) eta bestetik ebakuntzaren kontrako odol-analisiak izanagatik, benetan operatu behar denean (irudian eskuin aldetik kokaturik leudekeenak.

Arestian bezala errore-maila eta muga-kokapenaren arteko elkar-erlazioa (menpekotasuna) agerian dago. **Benetan** operatu behar diren pertsona guztiak opera daitezen nahi baldin badugu, erabakitze-muga askoz eskuinerago kokatuko genuke, baina, kasu horretan, operatu behar ez diren pertsona asko operatuko litzateke. Horren arrisku handiagoa izango lukete bederen, **benetan** operatu behar ez direnean. Aitzitik erabakitze-muga ezkerrantzago kokatuko bagenu, pertsona gutxi operatzea erabakiko genuke eta, beraz, benetan behar ez duten pertsona gutxi operatuko genuke, hots, horiekin gutxitan erratuko ginateke. Hala ere, beste alderditik, **benetan** ebakuntza behar dutenak ez operatzearen

arriskua handiagotu egiten da. Izan ere, erabakitze-muga dela eta, odol-analisiak probabilitate handiz ebakuntzaren kontrakoak izango bait lirateke.

Logikoa denez, erabakitze-muga tokirik egokienean kokatzeak Estatistikaren eginkizuna izan behar du, hots, bi errore horiek minimoak izan daitezen, errore horiek (I.a eta II.a) minimotzat jo ditzakegun lekua aurkitzea alegia. Hala eta guztiz ere erroreen elkarrekiko menpekotasuna (bat txikiagotzean bestea handiagotzen da) dela medio estatistikak hipotesi-testetan darabilen planteamendu orokorra bi erroreetariko bat txikieneratzean datza, beti ere, beste erroreak aldez aurretik ezarritako mugarik gainditzen ez duelarik.

Estatistikak hipotesi-testei ematen dien planteamendu orokorra honako hauxe da:

Bi errore-motei tratamendu berezia emango zaie. Bat oinarrizko edo garrantzitsuentzat joko dugu eta I. MOTAKO ERRORE esango diogu. Besteari berriz, II. MOTAKO ERROREA deritzogu. Tajuzko lana egin ahal izateko I. MOTAKO ERROREA kontrolatuz II. MOTAKO ERROREA txikieneratuko dugu. Planteamendu jakin bat egitean, kasu baterako, I. MOTAKO ERROREAk ez dezala % 5 edo % 1 gainditu. Balio horiek baino txikiagoa dela ziurtaturik, II. MOTAKO ERROREA txikieneratzen saiatuko gara.

Gure adibidean, planteamendua ondorengoa izan zitekeen: ebakuntza behar zuen pertsona operatu gabe geratzea larriagotzat jo dugu, beste errorea (zeina ebakuntza behar ez duen pertsona operatzean datzan) baino. Hori horrela bada,ebakuntza behar dutenen eta ez dutenen artean bereiz dezakeen muga non jarri erabakitzekotan, ebakuntza behar dute pertsonak operatu gabe gelditzeko daukaten arriskua % 1 baino txikiagoa izatea abiapuntutzat hartuko dugu. Errore hori esaniko portzentaia baino txikiagoa dela ziurtaturik, ebakuntza behar ez duten pertsonen operatze-arriskuaren portzentaia txikieneragotzen ahaleginduko gara.

Ondorengo kapituluetan aurreko planteamenduak nola gauzatzen diren aztertuko dugu, populazioen ezaugarriak ezagutu nahiz populazio horretatik ateratako lagin-datuez baliaturik inferentziak egiten ditugunean.

3. BI EDO TALDE GEHIAGOTAN LORTURIKO EMAITZEN KONPARAZIOA

Edozein ikertze-diseinuren formulazioan populazio ezberdineko sujetuz osaturiko lagin diferenteetako emaitzen arteko konparazioa ohizko eginkizuna izaten da.

Aurrera jo baino lehen bi kasu bereiztu behar ditugu:

1.- Datuak populaziokoak direnean. Hortaz ez dugu ezer inferitu behar. Izan ere, datuek neurtzen duten ezaugarria errealitate hartan adierazten bait dute. Neurtzean erratzen ez bagara argi dago ez dugula inolako errorerik egiten.

2.- Gure datuak populaziotik aukeratu dugun lagin batekoak dira. Populazioko ezaugarri jakin bat ezagutu ahal izateko lagin bat hautatu dugu. Hautateka- eta neurketa-prozedura egokia bide, aztertu nahi ditugun populazio-ezaugarriekiko lagin adierazgarria lortuko dugu. Adierazgarria ez balitz, ordea, azterketa-objektua den populazioarekiko ondorioak ateratzea zail samarra izango litzateke. Era berean, gaizki eginiko neurketak, analisi estatistiko egokiak izanagatik, alferrikako edo baliogabeko emaitzak ekarriko lizkigukete neurketa-errorea dela kausa. Oinarrizko erroreak badira ere, sarri askotan egiten direnak izaki, irakurlegoa honen garrantziaz jabe dadin nahi genuke. Aurrekoa kontutan harturik ere, emaitzak populazio osora orokortu nahi ditugunean arrisku jakin batzuen menpe gaudelarik ezinbesteko dugu arrisku horiek kontrolatzea.

Demagun, bada, behin esandakoak kontutan harturik, baserri-giroko eta kale-giroko pertsonek telebistaren aurrean pasatzen duten ordu-kopuruen arteko diferentzia posibleak aztertu nahi ditugula. Laginak aukeratu eta populazio diferenteetako sujetuak telebista aurrean egiten duten ordu-kopuruari buruzko datuak jasota, demagun ondoko emaitzak atera ditugula:

$$\overline{X}_b = 12.38 \qquad \overline{X}_k = 14.81$$

"Baserriko" (X_b) eta "kaleko" (X_k) taldeetako laginen bitartez lorturiko emaitzei behaturik, zera galde genezake: ea zoriak erangina diren edo, aitzitik, errealitatean baserri-giroko pertsonek telebistaren aurrean kale-giroko pertsonek baino ordu gehiago pasatzen duten, datuek agertzen duten moduan.

Egingo dugun planteamendua oro har, hipotesi-testetan egin ohi duguna izango da. Hasierako hipotesia ezarriz, zeinari OINARRIZKO HIPOTESIA edo NULUA esango diogun, bi populazioen arteko diferentziarik eza suposatuko dugu, hots, ikusitako diferentziak zoriak eragindakoak direla. Hipotesi nuluaren aurrez aurre ORDEZKO HIPOTESIA edo HIPOTESI ALTERNATIBOA jarriko dugu; gure kasuan kale-giroko pertsonek telebistaren aurrean baserri-girokoek baino ordu gutxiago egiten dutela, eta, ondorioz, ikusitako diferentziak zoriak eragindakoak ez eta benetan gertatzen den diferentzien isladapena izango lirateke.

Planteamendu orokorra, bada,:

OINARRIZKO HIPOTESIA edo NULUA H_o: $X_b = X_k$

HIPOTESI ALTERNATIBOA H_1: $X_b < X_b$

Ordezko hipotesia formulatzerakoan hautabide bi izango ditugu. Lehenengoa baserri-giroko ordu-kopurua kale-girokoa baino handiagoa dela esatea, baina, baserri-giroko eta kale-giroko ordu-kopuruak diferenteak direla ordezko hipotesitzat ere jo genezake. Bi planteamendu horiek kontraste-klase bi dakarkigute eta, ondorioz, errore-klase bi: ALDE BATEKOAK, adibidean guk plazaraturikoa, eta ALDE BIKOAK, hots, telebistaren aurrean pasatako ordu-kopuruen batezbestekoak diferenteak izatea ordezko hipotesitzat joko lukeena.

Arestian bi errore-motei buruz hitz egin dugu. Orain, Estatistikak adierazten duen bezala, ikus dezagun nola erlazionatzen diren Oinarrizko hipotesia edo Nulua eta I. eta II. motako erroreak.

I. MOTAKO ERROREA: Hipotesi nulua ukatzea, hipotesi hau egia denean. Gure kasuan baserri-giroko eta kale-giroko pertsonek telebista aurrean pasatako ordu-kopuruak diferenteak direla erabakitzea, eta, benetan, ikusitako diferentziak zoriak eragindakoak direnean. Hori I. motako errorea izatearen arrisku edo probabilitatea (α izendaturikoa) izango litzateke.

II. MOTAKO ERROREA: Hipotesi nulua onartzea faltsua edo gezurrezkoa denean. Gure kasuan talde batek zein besteak telebistaren aurrean pasatako ordu-kopuruaren arteko diferentziak zoriak eraginak direla erabakitzea baserri-girokoek, errealitatean, telebistaren aurrean ordu gehiago pasatzen dutenean. II. motako errorea izatearen arrisku edo probabilitatea (a izendaturikoa) izango litzateke.

Aldez aurretik ikertzaileak, eta ez beste inork, erabaki behar du zein den onar lezakeen I. motako errorerik handiena. I. motako erroreak ezarritako muga ezin dezakeela inolaz ere gainditu onarturik, II. motako errorea txikieneratzen saiatuko gara. Horretarako Estatistika teorikoak zenbait prozedura garatzen du; I. motako errorearen probabilitate-balio jakin bati eutsiz II. motako errorearen probabilitatea txikiagotzeko alegia.

SPSS/PC+-k jarraitzen duen prozedura beti bera izango da: emaitzetan I. motako errorea izatearen probabilitatea agertzen digu; hipotesi nulua egia izanda, ukatzeko probabilitatea alegia. Erabiltzaileak emaitza horretan oinarrituz, zeinerako II. motako errorea txikieneratzen prozedura estatistikoa erabilia izan bait da, erabakia zentzu batean edo bestean hartuko du. Arazo hau test edo analisi estatistiko bakoitzean aztertuko dugu.

Edozein kasutan ere Giza Zientziei Egokitutako Estatistikan sarrien erabiltzen diren errore-mailak % 5, %1 eta milako 1-a izaten dira. Lehen esan dugunez, gorengo errore maila ezartzen duena ikertzailea izango da.

4. SUJETU BATEK LORTURIKO BALIOEKIKO ERABAKI-HARTZEA.

Sarritan lagin edo populazioren datu edo ezaugarriekin baino sujetu jakin baten datuekin lan egitea interesgarriagoa suerta dakiguke. Horrelako kasuetan ondorengo bi zentzu diferenteetan sor dakizkiguke buruhausteak:

1. Sujetu bakar baten ezaugarriak direla eta, ezaugarri jakin batekiko sujetu "normalen" taldeko kide bezala jo al dezakegu ?. Hau da, sujetu jakin batek aldagai batean emaniko balioa, dagokion populazioaren balio arrunt edo normal bezala har al dezakegu ?

2. Posible al daiteke sujetu bakar baten ezaugarri jakin baten balioa aurresatea sujetu berak beste ezaugarrietan emaniko emaitzetan oinarrituz?. Horrelako arazoei aurre egiteko eta irtenbidea emateko erregresio-metodoa erabili ohi da.

3. Eman dezagun populazioko sujetu guztiak hainbat klase edo kategoriatan sailkatuta dauzkagula. Gure sujetua zeini dagokion ez dakigunean, sujetuak dituen ezaugarriak bide, probabilitate handienaz zein kategoriari egokituko litzaiokeen ezagutzea izango da hirugarren arazoa.

Aurreko arazo horiei era desberdinez erantzun dakieke, baina haien baitan konfidantza-tartea eta hipotesi-testen kontzeptuak daude (lehenengo kasua), konfidantza tarteak (bigarren kasua) eta hipotesi-testak eta sailkapen-metodoak (hirugarren kasua).

12. GAIA
LAGIN-TEORIA

Estatistikaren aplikazioetan, garrantzirik handienetakoa populazio bati dagozkion inferentziak egitea izango da. Inferentzi hauek, lagin batean bildutako informazioaren arabera egiten dira, eta beraien helburua populazioko parametro batzuk ezagutzea izango da.

Aipatutako helburuak lortzeko, garbi dago populazioa aztertzeko aukeratu dugun laginak, eta populazioak berak, karakteristika berdinekoak izan beharko dutela. Lagin bat aukeratzerakoan, beraz, kontutan hartu beharko dugu lagin honek populazioaren "ADIERAZGARRI" izan behar duela.

Adierazgarritasun hau lortzeko bi gauza zaindu behar dira bereziki: bata, laginen tamaina, hau da, aukeratutako pertsona- edo elementu-kopurua; eta hau erabaki ondoren, bigarrena,lagina osatzen duten elementuen aukeraketa, zeren eta azkenik populazioko emaitzen inferentziak elementu hauetan neurtuko ditugun ezaugarrien arabera egingo bait ditugu.

Gai honetan bi zati hauek aztertuko ditugu, baina ordena aldatuz, zeren eta lagineko pertsonak hautatzeko erabiliko ditugun metodoen arabera lagin-kopurua erabakiko bait dugu.

1. LAGINKETA ALEATORIOAK

Esan beharra dago laginak aukeratzeko dauden metodoetan aleatorioak populazioaren adierazgarritasuna garantizatzen duten bakarrak direla. Populaziorako inferentziak egiten ditugunean, errore batzuk edukitzeko arriskua ere badaukagu eta laginketa-metodo hauek erabiliz errore hauek edukitzeko probabilitateak kalkula ditzakegu.

Metodo aleatorioetan, garrantzitsuenak aztertuko ditugu.

LAGINKETA ALEATORIO SINPLEAK

Metodo ezagunena eta seguru asko gehien erabiltzen denetako bat hau izango da; populazioko elementu-kopurua oso handia ez denean batez ere.

N elementu dauzkan populazio batetik, n elementuz osatutako lagin bat aukeratu behar dugu, eta metodo hau erabiliz betetzen den propietatea zera da: populazioko edozein pertsonak aukeratua izateko daukan probabilitatea eta beste edozeinena berdinak izango dira.

Praktikan, laginketa aleatorioa elementuz elementu egiten da. Aukeraketa egiteko erabiltzen den prozedura, askotan behintzak, ondoko hau izango da: populazioko elementu bakoitzari zenbaki bat eransten zaio 1-etik N-raino, eta gero n zenbaki aleatorioa aukeratzen da, zenbaki aleatoriozko taulak erabiliz edota ordenadoreen bidez, zenbakien serie aleatorioak ateratzen dituen programa egoki bat erabiliz. Kontutan eduki behar da prozesuen edozein momentutan aukeratu gabe dauden elementuek aukeratuak izateko daukaten probabilitateak berdina izan behar duela.

Metodo honek daukan problema, batez ere populazioko elementuen kopurua handia denean, elementuen aukeraketa egiten denbora asko pasa behar izatea da, eta badakigu estatistika erabiltzen denean denborak eta diruak garrantzi handia dutela.... Horregatik askotan orain aipatuko ditugun metodoak erabiltzen dira, batez ere azkarragoak direlako.

LAGINKETA SISTEMATIKOA

Lagineko elementuen aukeraketa sinplifikatzen duen metodo bat da. Eman dezagun 5000 pertsonak osatzen duten populazio bat daukagula, eta 100 pertsonen lagin bat aukeratzea nahi dugula. Garbi dago populazioko 50 pertsonetatik bat aukeratua izango dela. Populazioko pertsona guztiei zenbaki bat erantsi ondoren, 1 eta 50-en arteko zenbaki bat aukeratuko dugu, eta zenbaki bakar hau izango da, ondoko era honetan, laginketa osoa erabakiko duena. Eman dezagun lortu dugun zenbakia 32 dela. Orduan populaziotik aukeratuko ditugun elementuak, 32, 82, 132,182.... zenbakizkoak izango dira. Aukeraketa-prozesu hau erabiltzeko lehenengo kalkulatu behar dena GORAPEN-KOEFIZIENTEA izango da. Koefiziente honek adierazten duena zera da: populazioko elementu-kopuru eta lagineko elementu-kopuruen arteko zatidura. Bere kalkulua, noski, bi kopuruen arteko zatiketen emaitza izango da.

Prozesu honek, arrisku bat dauka, eta populazioko elementuei zenbakiak ezartzeko erabili dugun metodoa da. Adibide gisa, eman dezagun gure populazioa osatzen duten 5000 pertsonak, 2500 bikote (andre-gizonak) direla, eta beti lehenengo zenbakia emakumeari ezarri diogula. Hau da, zenbaki bakoitiak dauzkaten elementuak emakumeak eta zenbaki bikoitiak dauzkatenak gizonak izango dira. Aukeratu dugun zenbakia, 32, parea denez gero, lagina osatzen duten elementu guztiak gizonezkoak izango dira, eta honen ondorioz gure laginean populazioko emakumeak ez dira ordezkatuak egongo, noski, gure helburuen kontra. Benetan, arrisku handia izango da eta kontutan eduki beharko dugu era honetako serieak populazioko elementuei zenbakiak eransterakoan ager ez daitezen.

LAGINKETA GERUZATUA (ESTRATIFIKATUA)

Metodo hau erabiltzeko lehenengo pausoa, populazio guztiko elementuak zati ezberdinetan banatzea izango da. Banaketa hau egiterakoan bilatzen dena zera da: zati bakoitzeko elementuak oso homogenoak izatea, eta horretarako erabiliko dugu populazioari buruz

aurretik daukagun informazioa.

Laginketa honela egiten badugu, lehenengo pausoan elementu guztiak, zati homogeno ezberdinetan zatituko ditugu, eta ondoren, zati bakoitzeko azpilagin aleatorio bat aukeratuko dugu, azpilaginketa hau egiterakoan edozein metodo aleatorio erabil dezakegula kontutan harturik.

Populazioko zatiketa egiterakoan, beharrezkoa izango da erizpide argi eta garbi batzuk edukitzea, eta hauen bidez azpipopulazio bakoitzeko homogenotasunak populazio guztikoen aldean handia izan beharko du, zeren eta bestela ez bait du pena mereziko honelako lana hartzeak.

Eman dezagun estatu-mailako inkesta elektoral bat egin nahi dugula. Alderdi edo koalizio batzuk probintzia batzuetan bakarrik aurkezten direla kontutan harturik, nahikoa logikoa dirudi populazioa den estatu guztia probintzietan banatzeak. Eta gure populazioko laginketa bakarra egin beharrean, probintzia bakoitzeko azpilagin bat aukeratuko dugu; horrela populazio guztiko lagin adierazgarri bat lortuko bait dugu. Kasu hauetan, populazio orokorren tratamendua emango diogu probintzia bakoitzeko populazioari.

Prozedura honek ere, bere akats txikia badu, eta hau da: gehienetan lagineko elementu-kopurua metodo hau erabiliz beste metodoak erabilita baino askoz ere handiagoa izango dela.

Aurretik lagineko elementu-kopurua mugatuta badaukagu, batzuetan arrazoi ekonomiko eta denboral batzuengatik azpipopulazio bakoitzeko azpilaginketen kopurua erabakitzerakoan bi erizpide ezberdin erabiltzen dira:

.- Neurtzen ditugun ezaugarrien puntuazioen emaitzen bariantzak azpipopulazio bakoitzean berdinak badira, azpilagin bakoitzeko elementu-kopurua, azpipopulazio bakoitzeko kopuruekiko proportzionala izango da.

.- Zati bakoitzeko bariantzak ezberdinak badira, azpilagineko elementu-kopurua kalkulatzerakoan kontutan eduki beharko dugu, alde batetik,zati bakoitzeko elementu-kopurua. Populazioan zenbat eta pertsona gehiago egon, hainbat eta pertsona gehiago egongo dira laginean. Bigarrena bariantza izango da eta orain ere zenbat eta bariantza handiagoa hainbat eta pertsona gehiago laginean.

KONGLOMERATUEN LAGINKETA

Gehienetan lagina aukeratzeko metodoetan erabiltzeko errazena hau izango da, baina esan dezakegu arriskutsuena izango dela ere, zeren eta askotan nahiko zaila izango da metodo hau erabiliz lagin adierazgarri bat lortzea.

Kasu honetan laginen aukeraketa ez dugu populaziotik zuzenean egingo; baizik eta lehenengo pausoan, populazioa azpipopulazio ezberdinetan zatituko dugu, baina kasu honetan, azpipopulazio bakoitza populazio guztiko adierazgarritzat kontsideratuko dugu. Populazioan egin dugun zatiketa honen zati bakoitza konglomeratu bat izango da eta azken batean lagin bezala konglomeratu bat hartuko dugu.

Ez dugu metodo hau laginketa geruzatuarekin nahastu behar. Kasu hartan zati bakoitzetik azpilaginketa bat aukeratzen genuen. Ostera konglomeratuen laginketan lagin bezala zati oso bat kontsideratuko dugu.

Askotan erabiltzen den adibide bat, hau da. Eman dezagun Donostian bizi diren pertsonen altueren batezbestekoa kalkulatu nahi dugula. Kasu honetan, kale bakoitzean era guztietako altuerako pertsonak bizi direla kontutan harturik, konglomeratu bezala kaleak kontsidera ditzakegu. Beraz populazio guztiko lagin adierazgarritzat Donostiako kale batean bizi diren pertsonak izango dira. Kalea, noski, zoriz aukeratuko dugu.

Laginketa-era honek daukan arriskua zera da: batzuetan konglomeratu bateko ezaugarriak eta populazio osokoak berdinak ez izatea.

Konglomeratu-laginketaren bariante bat ETAPAZKO KONGLOMERATUEN LAGINKETA izango da. Egiten dena zera da: lehenengo pausoan populazio guztia konglomeratu ezberdinetan zatitu, eta hauetatik, zoriz, bat aukeratu. Lehenengo konglomeratua aukeratu ondoren, eta honen beste zatiketa bat egin ondoren, lehenengo konglomeratuen adierazgarri bezala honen zati bat hartuko dugu, eta horrela behar den beste pauso eman dezakegu.

Eman dezagun Euskadiko pertsonen altueren batezbestekoa kalkulatu nahi dugula. Horretarako konglomeratu bezala probintziak kontsidera ditzakegu. Eman dezagun zoriz, Gipuzkoa izan dela aukeratua, Gipuzkoa zati ezberdinetan zatituko dugu, herritan adibidez, eta hauetatik Donostia izan da aukeratua eta berdin jarrai gintezke, Donostiako lehengo kalera ailegatu arte.

Argi ikusten dira metodo honek dauzkan arriskuak, baina aurrekoek ere berenak dituzte. Azken batean ikertzaileak izan beharko du, bere ahalmenak ikusirik, metodo bat edo bestea aukeratuko duena.

2. LAGINKETA EZ-ALEATORIOAK

Lehen aipatu dugu populazioko lagin adierazgarri bat lortzeko metodo seguru bakarrak aleatorioak direla. Baina makina bat arrazoigatik, askotan laginketa ez-aleatoriotara jotzen dute ikertzaileek. Hauetako bi bakarrik aipatuko ditugu.

LAGINKETA AKZIDENTALA : Giza zientzietan asko erabiltzen den metodo bat da. Datuak lortzeko erraztasun handiena ematen dituzten elementuak aukeratzen dira. Kontutan hartu behar dugu, erraztasun hitza zentzu zabalean kontsideratu behar dugula. Askotan, irakaskuntz metodo berri bat martxan jartzerakoan, eskola edo zentru pilotoak erabiltzen dira, eta zentru hauetako emaitzak konparatzeko antzerako edo ezaugarri berdineko zentru batzuekin konparatu beharko ditugu. Beste askotan, adibidez test edo neurketa-tresna bat estatu-mailarako barematu nahi badugu, diru-problemagatik edota beste arrazoi batzuengatik ezinezkoa izango da Madriletik buelta ibiltzea. Kasu hauetan laginketa aleatoriak ez du zentzurik. Kontutan hartu beharko dugu honela aukeratu dugun lagin batetik populazio osora inferentziak egitea nahi badugu, lagin hauek ez direla populazioaren adierazgarriak.

ERITZIZKO LAGINKETA : Lagineko elementuen aukeraketa, populazioari buruz aurretik dauzkagun informazioen arabera egingo dugu. Adibidez inkesta elektoral bat egiterakoan, badakigu aurreko hauteskundeetan nolako emaitzak egon diren herri ezberdinetan, eta hurrengo hauteskundeetako botuen joera jakiteko, laginen aukeratzea aurreko informazioa kontutan harturik egingo dugu.

3. LAGINAREN TAMAINA ERABAKITZEA

Populazioko lagin adierazgarri bat aukeratzerakoan, hartu behar den lehenengo erabakia, tamainarena izango da. Erabaki hau hartzerakoan, kontutan eduki beharko dugu zein laginketa-metodo erabiliko dugun. Ez da berdina izango laginketa aleatorio sinplea edo laginketa geruzatua erabiltzea.

Laginen tamaina erabakitzerakoan bereiztu beharra dago populazioko elementu-kopurua finitu ala infinitua den. Zentzu hertsian, Giza Zientziei egokitutako Estatistikan erabiltzen diren populazioko elementu-kopuruak, ia beti finituak izango dira, baina hala ere, tamaina handikoak direnean, infinituzkoak bezala kontsideratuko ditugu. Badira zenbait idazle, populazio eta laginen elementu-kopuruen arteko proportzioa 20-tik 1-era baino handiagoa denean populazio-tamaina infinituzkoa bezala kontsidera dezaketela esaten dutenak. Eritzi honen arrazoia gero ikusiko dugu.

Hemen aipatuko ditugun kontsiderazio eta formula orokorrak, laginketa aleatorio sinpleari dagozkionak izango dira.

Laginen tamaina erabakitzerakoan, batez ere populazioko hiru ezaugarri edo karakteristika hartu beharko ditugu kontutan:

Lehenengoa, laginetik populaziora egingo ditugun inferentzietan eduki dezakegun errore maximoa zein izango den jakin beharko dugu. Inferentzi zehatzagoak egiteko, lagin–kopurua handiagotu egin beharko dugu.

Bigarrena, populazioen tamaina izango da. Ez da berdin izango 1000 pertsonak osatzen duten populaziorako ondorioak ateratzea, edota 1000000 bat pertsona daukan populazio baterako ateratzea.

Hirugarrena, populazioan neurtu nahi dugun aldagaien ezaugarrien bariantza izango da. Hau da, aurreko kapituluan esaten genuen bezala, Gipuzkoako 4 urteko umeen altuerari buruz ondorioak atera nahi badugu, behar dugun lagin-kopurua askoz ere txikiagoa izango da, 20 urteko neska-mutilen altuerari buruzko ondorioak ateratzeko erabili beharko genukeena baino. Honen zergatia erraza da: 20 urteko pertsonen altuerak 4 urtekoenak baino askoz ere sakabanatuagoak direlako, eta, ondorioz, lagin adierazgarri bat lortzea askoz ere zailagoa izango da.

POPULAZIO INFINITUZKO LAGIN-TAMAINA

Lehen aipatu dugun bezala, populazioaren adierazgarri izan behar duen lagin baten tamaina zuzenki erlazionatuta egongo da lagin honetatik inferentziak edota konfidantza-tarteak egiten ditugunean onartuko dugun gehienezko errorearekin.

Aipatuko ditugun kasuak gehien erabiltzen direnak izango dira. Hau da, populazio bateko batezbestekoen konfidantza-tartea atera nahi dugunean, bariantza ezaguna edo ezezaguna denean, eta proportziotarako konfidantza-tarteak egiteko behar ditugun lagin-kopuruak ikusiko ditugu, beti gehieneko errorea mugatuz.

Hemendik aurrera, populazioan neurtu ditugun aldagaien ezaugarriak banaketa normalari jarraitzen zaizkiola suposatuko dugu, bere parametroak μ eta σ izanik.

BARIANTZA EZAGUNA DEN POPULAZIO INFINITUZKO BATEZBESTEKOEN KONFIDANTZA-TARTEA ETA LAGIN-KOPURUA

Eman dezagun banaketa normalari jarraitzen dion aldagai bat daukagula. Badakigu laginketa aleatorio bat egin ondoren eta lagin horretan X aldagaia neurtu badugu, lortutako emaitzen batezbestekoa banaketa normalari jarraitzen zaiola. Lagin bateko batezbestekoen parametroak, hauek izango dira: Batez-bestekoa: μ, eta desbidazio tipikoa $\sigma/\sqrt{n}$, n-k lagin-kopurua adierazten duelarik.

Esan duguna kontutan harturik, eta α errorea gehieneko bezala harturik, populazioko batezbestekoaren konfidantza-tartea, honela defini dezakegu:

$$[\overline{X} - T\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + t\alpha \frac{\sigma}{\sqrt{n}}]$$

Hau da, α errorea mugatuz, lortu dugun konfidantza-tarteen zabalera ondoko hau izango da:

$$\text{ZABALERA} = 2 * t\alpha \frac{\sigma}{\sqrt{n}}$$

Konfidantza-maila ezarri ondoren, eman dezagun kasu honetan %95ekoa dela. Banaketa normal tipikoko tauletan aurkituko dugu $t\alpha$-ren balioa: kasu honetan 1.96. Bariantza edo desbidazio tipikoa ezaguna baldin bada, eman dezagun gure kasuan σ=10 dela. Tarteen zabalera hau izango da:

$$\text{ZABALERA} : 2 * 1.96 \frac{10}{\sqrt{n}}$$

Lagin-kopurua 100ekoa bada, tarteen zabalera 3.92 izango dela ikus dezakegu. Lagin-kopurua haundiagotu egiten badugu, eman dezagun bigarren kasuan 10000koa dela, lortuko dugun tarteen zabalera 0.392 izango da. Planteamendu honetan oinarrituz, laginko-purua erabakitzerakoan azken hitza beti ikertzaileak edukiko du; beti noski, onartu duen errore handiena kontutan edukirik.

Hala ere, ohizkoa da estimazioaren errore edota zabaleren emaitzak zehaztasun-hitzetan aurkeztea. Esan beharra dago "zehaztasuna" eta "zabalera" kontzeptu berdinak direla. Edozein estimaziotako zehaztasuna "ϵ" ikurra erabiliz adieraziko dugu. Gehienetan batezbestekoen balioarekiko erabiltzen da. Era honetako esaldiak askotan ikusten dira: Batezbestekoen estimazio bat egin dugu, batezbestekoen %1eko zehaztasunaz.

Zehaztasun-hitzetan emaitzak aurkeztea nahi badugu, egin behar dugun prozesua ondokoa izango da. Lehenengo pausoan lagin pilotu baten bidez, neurtu nahi dugun ezaugarrien lehenengo ikerketa egingo dugu, populazioko batezbestekoen lehenengo hurbilketa lortzeko. Pauso honetan lortu dugun emaitzetan oinarrituz, azken laginketaren tamaina erabakiko dugu.

Eman dezagun batezbestekoen lehenengo estimazioa 100 puntukoa dela, eta populazioko batezbestekoa estimatzerakoan eduki dezakegun gehieneko errorea %5ekoa dela. Lortu behar dugun emaitzen zehaztasuna gure kasuan batezbestekoen %1 dela suposatuko dugu. Hau da, gure kasuan batezbestekoen lehenengo estimazioa 100 dela kontutan harturik, emaitzen zehaztasuna puntu batekoa izango da. Era honetako ikerketa egiteko behar dugun lagin-kopurua ondoko formularen bidez kalkulatuko dugu:

$$\text{ZABALERA} = 2 * t\alpha \frac{\sigma}{\sqrt{n}} = 1 = 2 * 1.96 \frac{10}{\sqrt{n}}$$

eta hemendik $\sqrt{n} = 39.2$. Honek zera esan nahi du: gure laginen tamainak $n = 1536.64$ izan behar duela.

Aipatu beharra dago "zehaztasun" kontzeptua bi eratara erabiltzen dela. Batzuetan "zehaztasuna" eta zabalera berdin erabiltzen dira, eta beste kasu batzuetan zehaztasun bezala zabaleraren erdia hartzen da. Idazleak ez dira honetan ados jartzen. Kontutan eduki behar dugu era batera edo bestera onartu lagin-tamaina 4 aldiz handiago edo txikiagoa izango dela. Emaitzak aurkezterakoan kontzeptu hau zein zentzutan onartu dugun adieraztea beharrezkoa izango da.

BARIANTZA EZEZAGUNA DEN POPULAZIO INFINITUZKO BATEZBESTEKOEN KONFIDANTZA-TARTEA ETA LAGIN-KOPURUA

Aurreko kasuan egin ditugun planteamendu guztiak kasu honetarako ere baliagarriak izango dira. Hau da, bariantza ezaguna edo ezezaguna izan, aldatzen den gauza bakarra erabiliko dugun formula izango da. Bariantza ezezaguna denean, populazioko batezbestekoentzako eraikiko dugun konfidantza-tartea hau izango da:

$$[\overline{X} - t\alpha \frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t\alpha \frac{s}{\sqrt{n}}]$$

Non s lagineko quasibariantzaren erro karratua den. Gogoratu beharra dago SPSS/PC+ paketeak bere emaitzetan quasibariantzaren erro karratuen balioa STD DEV izenaz azaltzen duela.

Kalkuluak eskuz egiten baditugu, lagineko quasibariantzaren emaitza lortzeko bariantzaren balioan (S^2), oinarrituz, honela kalkulatuko dugu:

$$s^2 = \frac{n * S^2}{n - 1}$$

Aurkeztu ditugun bi kasu hauek, luza ditzakegu beste kasu batzuetara ere, hau da, bi populazioko batezbestekoak konparatu nahi dugunean, desbidazio tipikoarentzat konfidantza-tarteak eraiki nahi ditugunean...; kasu hauetan erabiltzen diren planteamenduak berdinak izango bait dira.

Aipatuko dugun bakarra, proportzioko estimazioen kasua izango da.

PROPORTZIOENTZAKO KONFIDANTZA-TARTEAK ETA LAGIN-KOPURUA

Kasu honetan ere gure lagina laginketa aleatorio sinple baten bidez aukeratu dugula

suposatuko dugu. Orduan laginetik lortu dugun p proportzioen estimazioa, banaketa normalari jarraitzen zaio, beraren parametroak P eta $\sqrt{P*Q/n}$ direlarik, hau da :

$$p \rightarrow N(P, \sqrt{P*Q/n})$$

Aurreko banaketa kontutan harturik α mailako errore bat gehienekoa izanik, proportzioarentzako konfidantza-tartea ondoko hau izango da:

$$[\, p - t\alpha \frac{\sqrt{pq}}{\sqrt{n}} \leq \mu \leq p + t\alpha \frac{\sqrt{pq}}{\sqrt{n}} \,]$$

non p, P parametroen estimazioa den, eta q = 1 - p.

Kasu honetan tarteen zabalera :

$$\text{ZABALERA} = 2 * T\alpha \frac{\sqrt{pq}}{\sqrt{n}}$$

%95eko konfidantza-mailarako, $t\alpha = 1.96$ balioa tauletatik lortuko dugu, eta p*q-ren gehieneko balioa, p=.5 denean,0.25 izango dela kontutan harturik, tarteen zabalera ondoko hau izango da:

$$\text{ZABALERA} = \frac{1.96}{\sqrt{n}}$$

Tarteen zabalera mugatu ondoren lagineko tamaina lor dezakegu aurreko formula erabiliz. Aurreko kasuan bezala emaitzak zehaztasun-hitzetan aurkeztu nahi baditugu, lehe nengo pausoan, lagin pilotu batekin lana eginda, proportzioaren lehenengo estimazioa egingo dugu, eta hemen lortutako emaitzetan oinarrituz azken laginaren tamaina erabakiko dugu.

LAGIN-TAMAINA POPULAZIO FINITUETAN

Populazio finitu bezala, pertsona- edo elementu-kopurua finitua denean eta laginketa itzulera gabekoa egiten dugunean kontsideratuko ditugu, zeren eta nahiz eta populazio finitua izan, laginketa itzuleraduna egiten badugu, populazio honi infinituzko tratamendua emango bait diogu.

Hemendik aurrera populazioko pertsona- edo elementu-kopurua N letraz adieraziko dugu eta lagin-kopurua n-ren bidez.

Populazio finitu eta infinitukoen artean dagoen ezberdintasun garrantzitsuena, bariantzaren estimazioan oinarrituko da. Populazio infinituetan lagineko quasibariantza populazioko bariantzaren estimatzaile alboragabea da. Aldiz, populazio finituetan, bariantzaren estimatzaile alboragabea ondorengo hau da:

$$\frac{N - n}{N} * \frac{S^2}{n}$$

non S^2 lagineko quasibariantza den. Esan beharra dago SPSS/PC+ paketeak quasibariantza VARIANCE izenburuaz aurkezten duela.

Aurreko formulan ikusten dugunez populazioaren tamaina nahiz eta finitua izan, haundia denean, formulan egiten den zuzenketa (N - n)/N, 1 baliotik oso hurbil dago. Horregatik makina bat idazlek lagin-tamaina aukeratzerakoan populazioaren tamaina haundia denean infinituzkoa balitz bezala kontsideratzen dute.

Batezbestekoen tartea eraikitzerakoan, bariantza ezaguna denean, aurreko kasuko berdina izango da, zeren eta kasu honetan bariantzaren estimatzailea erabiltzeak ez bait dauka zentzurik, ezaguna delako.

BARIANTZA EZEZAGUNA DENEAN, aurrean aipatu dugun zuzenketa ezarri beharko dugu bariantzaren estimatzailean, eta ondorioz, batezbestekoen konfidantza-tartea ondoko hau izango da:

$$[\,\overline{X} - t\alpha \sqrt{\frac{N - n}{N}}\ \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t\alpha \sqrt{\frac{N - n}{N}}\ \frac{S}{\sqrt{n}}\,]$$

Populazio infinituetarako egin ditugun kalkulu berdinak errepikatuz, α gehieneko errore bezala onartzen badugu, eraikitako tarteen luzera ondoko formulak adieraziko du:

$$\text{ZABALERA} = 2 * t\alpha \sqrt{\frac{N - n}{N}}\ \frac{S}{\sqrt{n}}$$

Aurreko formulan aldaketa batzuk eginik, eta tarteko zabalera A izendatuz, ondoko formularen bidez laginen tamaina lor dezakegu:

$$n = \frac{4 * N * S^2 * t^2\alpha}{A^2 * N + 4 * S^2 * t^2\alpha}$$

10000 pertsona edo elementuz osatutako populazio bateko batezbestekoen konfidantza-tarte bat egin nahi badugu, bariantzaren estimazioa 100 baldin bada, eta tarteen zabalera 2 puntu baino txikiagoa izatea nahi badugu, α=.05 gehiengo errorea izanik, aurreko formula erabiliz aukeratu behar dugun laginaren tamaina erabaki dezakegu:

$$n = \frac{4 * 10000 * 100 * 1.96^2}{2^2 * 10000 + 4 * 100 * 1.96^2} = 396.94$$

Hau da, gure kasuan 370 elementuz osatutako lagin bat nahikoa izango litzateke, elementuen aukeraketa laginketa aleatorio baten bidez egin behar dugula kontutan harturik.

Lehen aipatu dugun bezala, batzuetan ZEHAZTASUN-hitzetan mintzatzen da. Kasu honetan, $\in$ zehaztasuna, zabaleren erdia bezala onartzen badugu:

$$n = \frac{N * S^2 * t^2\alpha}{\in^2 * N + S^2 * t^2\alpha}$$

PROPORTZIOEN ESTIMAZIOA POPULAZIO FINITUETAN

Batezbestekoen tarteak eraikitzerakoan egiten genuen bezala, proportzioen tarteak eraikitzerakoan, populazioko bariantzaren estimatzailea alboragabea izan dadin, zuzenketa-faktore bat ezarri beharko dugu. Kasu honetan bariantzaren estimatzaile alboragabea ondoko hau izango da:

$$\frac{(N - n) * p * q}{(N - 1) * n}$$

Hemendik aurrera, aurreko kasuan egin dugun prozesu bera erabiliz, α errorea eta A zabalera finkatuz, eta populazioko tamaina N izanik, ondoko formularen bidez, laginen n tamaina lortuko dugu:

$$n = \frac{4 * N * p*q * t^2\alpha}{A^2 * (N - 1) + 4 * p*q * t^2\alpha}$$

Zabaleraz hitz egin beharrean, $\in$ zehaztasun-hitzetan ari bagara, (zehaztasuna, zabaleraren erdia bezala definituz) laginen tamaina emango duen formula hau izango da:

$$n = \frac{N * p*q * t^2\alpha}{A^2 * (N - 1) + p*q * t^2\alpha}$$

13. GAIA
JI KARRATUAREN TESTA. CROSSTABS AZPIPROGRAMA

12. gaian Estatistika Inferentzialaren zentzu zabaleko hurbilpen orokorra plazaratu dugu. Hemendik aurrera ohizko hipotesi-testak, erabilienak beraz, aurkeztuko ditugu, SPSS/PC+-ren lankera zein den azalduz.

Test-sorta honi hasiera emateko, Ji karratuarena, erabilienetariko bat hain zuzen, dakarkizuegu.

Edozein ikerketatan hasierako suposizio edo hipotesian fenomeno edo gertaera baten emaitza zein izan daitekeen adierazten dugu. Azterketa enpirikoa egitean, gertatutako edo ikusitako emaitzak eta gure oinarrizko hipotesia betetzen denean espero ditugunak bat ez etortzea suerta daiteke. Desakordio hori garrantzitsua den ezagutzeko edo, gertatutakoak guk itxarondakotik nabarmenki aldentzen den jakiteko Ji karratu delako testa erabiliko dugu.

Funtsean Ji karratua honelako eginbeharretan erabiliko dugu:

1. Bi aldagai kualitatiboen arteko erlazioen analisian. Demagun horrelako aldagai bi ditugula: SEXUA (gizonezko edo emakumezkoa) eta HERRIA (Ataun, Gasteiz edo Eibar). Kasu honetan aztertuko duguna zera da: ea herri horietan aztertu ditugun sexuarekiko pertsonen portzentaiak berdinak diren.

2. Behaturiko datuak banaketa teoriko bati egokitzerakoan. Aldagai bat, aldez aurretik jarritako banaketa bati egokitzen zaion ala ez jakin nahi dugu hain zuzen.

Gai honetan lehenengo puntukoari bakarrik erantzungo diogu. Bigarrenengoa NPAR-TESTSen CHISQUARE delako azpiprograma ikustean aztertuko dugu.

Abiapuntutzat har dezagun 1. puntuan eginiko planteamenduaren modukoa. Jarraian azaltzen dugun sarrera biko taulan bi aldagai kualitatibo aztertzen ditugu, posible diren kategoriatan gertatutako maiztasunak agerteraziz. Taula horri MAIZTASUN-taula edo kontingentzi taula esango diogu.

HERRIA		ATAUN	EIBAR	GASTEIZ	
EMAKUMEA	1	14	19	35	68
GIZONA	2	32	12	13	57
		46	31	48	125

Ji karratuaren testa aplikatzerakoan hipotesi nulu bezala aldagaiak elkarrekiko independenteak direla hartuko dugu. Sujetu jakin bat aldagai baten kategoria bati atxekirik izateak ez du bigarren aldagaiaren kategoria zehatz batean atxekirik izatea baldintzatzen. Beste era batera esanda, kategoria ezberdineko portzentaiak aldatu gabe mantentzen dira. Sujetuak zoriz banatzen dira laukitxotan.

Gure kasuan hipotesi nulua honela esango genuke: aipaturiko hiru herrietan emakumezkoen portzentaiak eta gizonezkoen portzentaiak berdinak direla.

Oinarrizko hipotesia betetzen denean espero zitezkeen maiztasunak kalkulu erraz baten bitartez jakin ditzakegu. Adibidez pertsona guztien artean (125), 68 emakumezkoak badira, Atauneko emakumezkoen kopurua honek izan beharko luke:

$$\frac{68 * 46}{125} = 25.02 \text{ pertsona}$$

Dena dela itxarondako maiztasunak honako formularen bitartez kalkulatuko ditugu:

$$\text{Itxar. maiz.} = \frac{\text{(lerroko elem. kopurua)} * \text{(zutabeko elem. kopurua)}}{\text{elementu guztien kopurua}}$$

Itxarondako maiztasunen taula (parentesi artean) lortu dugu. Beheko irudian azaltzen denez, itxarondako eta ikusitako edo jasotakoen arteko diferentziak horien azpian idatzi ditugu.

HERRIA		ATAUN	EIBAR	GASTEIZ	
EMAKUMEA	1	14 (25) -11	19 (16.9) 2.1	35 (26.1) 8.9	68
GIZONA	2	32 (21) 11	12 (14.1) -2.1	13 (21.9) -8.9	57
		46	31	48	125

Horiek horrela Ji karratuaren testaren planteamendu orokorra honako hau izango litzateke: behaturiko maiztasunen eta itxarondako maiztasunen artean badagoela egon diferentziarik. Diferentziak bi arrazoigatik egon daitezke: zoriak eragindakoak izan daitezke edo abiapuntutzat jo dugun hipotesi nulua faltsua delako. Bigarrena gertatu balitz, gure aldagaiak elkarrekin erlazionaturik daude eta bizilekuak (herria) baldintzatzen du sexuen portzentaia edo, bederen, elkarrekin erlazionaturik daude. Alderantziz ere uler dezakegu. Izan ere analisi honek ez bait du erlazioaren norabidea adierazten.

Estatistika teorikoak, beste aldetik, X_1^2 estatistikoa K askatasun graduko Ji karratuaren banaketari jarraitzen zaiola esaten digu, non:

K = (lerro-kopurua - 1) * (zutabe-kopurua - 1)

$$X_1^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Ji karratuaren analisien kalkuluak eskuz egiten baditugu, segitu behar dugun prozedura hau izango da: laginaren X_1^2 kalkulaturik tauletara joko dugu diferentziak signifikatiboak diren ala ez jakin ahal izateko.

Gure kasuan kalkulu hauen bitartez ezagutuko dugu X_1^2-ren balioa:

$$= \frac{(14-25)^2}{25} + \frac{(19-16.9)^2}{16.9} + \frac{(35-26.1)^2}{26.1} + \frac{(32-21)^2}{21} + \frac{(12-14.1)^2}{14.1} + \frac{(13-21.9)^2}{21.9}$$

$$= \frac{(-11)^2}{25} + \frac{(-2.1)^2}{16.9} + \frac{8.9^2}{26.1} + \frac{11^2}{21} + \frac{(-2.1)^2}{14.1} + \frac{(-8.9)^2}{21.9}$$

$$= \frac{121}{25} + \frac{4.41}{16.9} + \frac{79.21}{26.1} + \frac{121}{21} + \frac{4.41}{14.1} + \frac{79.21}{21.9}$$

$$= 4.84 + 0.26 + 3.)3 + 5.76 + 0.31 + 3.62 = 17.82$$

Balio hau (17.82) tauletan azaltzen denarekin konparatuko dugu. Gure kasuan K=2 askatasun-graduaz, K-ren kalkulua gogoraraziz:

K=(lerro-kopurua - 1) * (zutabe-kopurua - 1)=(2-1)*(3-1) = 2

Errore-maila diferenteen arauera taulatan azaltzen diren balioak hauek dira:

Errore-maila $\alpha = 0.05$, $X^2_t = 5.99$
Errore-maila $\alpha = 0.01$, $X^2_t = 9.21$
Errore-maila $\alpha = 0.001$, $X^2_t = 13.82$

Darabilgun edozein errore-maila izanda, lagineko balioa (X_t^2) tauletakoak baino handiagoa denez gure ondorioa hipotesi nulua, zeinak diferentziak zoriak eragindakoak zirela adierazten zuen, baztertzea izango da. Hau da, gure erabakia hau izango litzateke: ikusitako eta itxarondako maiztasunen arteko ezberdintasunak ez dira zoriak eragindakoak izango, oker erabakitzeko arriskua milako bat baino gutxiago delarik.

Jarraian ikus dezagun,bada, emaitza horiek SPSS/PC+ paketea erabiliz nola lor genitzakeen. Ondorengoa da hain zuzen, X^2-aren analisiak egiteko idatzi behar dugun oinarrizko instrukzioa:

CROSSTABS TABLES= 1. aldagai kualitatiboa BY
2. aldagai kualitatiboa.

Funtsezko azpiprograma horren bitartez emaitza deskribatzaileak soilik lortuko ditugu, hipotesi-testa egin gabe eta balizko erlazioaren signifikazio-testa ere egin gabe.

Idatzi dugun azpiprograma hau da:

CROSSTABS TABLES=SEXUA BY HERRIA.

Lortu ditugun emaitzak hauek izan dira:

Crosstabulation : SEXUA
By HERRIA

HERRIA=> SEXUA	Count	ATAUN 1	EIBAR 2	GASTEIZ 3	Row Total
EMAKUMEA	1	14	19	35	68 54.4%
GIZONA	2	32	12	13	57 45.6%
	Column Total	46 36.8%	31 24.8%	48 38.4%	125 100.0%

Number of Missing Observations = 0

Erlazioaren adierazgarritasuna jakin nahiko bagenu, ondoren ikusiko dugunez 1 estatistikoa eskatu behar izango genukeen.

1. CROSSTABS AZPIPROGRAMAREN ESTATISTIKOAK

Estatistiko guztiak ez ditugu hona ekarriko. Funtsezkoak bakarrik; erabilienak alegia.

Zb. Kalkuluak

1. X^2-aren testa egiten du. 2 X 2 taulan 20 elementuk egon behar dute gutxienez. Laukitxoren batean itxarondako maiztasuna 5 baino txikiagoa balego automatikoki Yates-en zuzenketa egingo luke.

2. 2 X 2 tauletarako f_i koefizientearen kalkulua. Taula handiagoen kasuetan Cramer-en V kalkulatzen du.

3. Kontingentzi koefizientearen kalkulua.

4. Lambda, simetriko eta asimetrikoa.

6. Kendall-en Tau-b. "STRING" aldagaiak direnean ez du kalkulatzen.

10. "Eta" koefizientearen kalkulua. Zenbakizko aldagaiak direnean soilik.

11. Pearson-en r kalkulatzen du. Zenbakizko aldagaiak direnean soilik.

ALL Estatistiko guztiak kalkulatzen ditu.

2. *CROSSTABS AZPIPROGRAMAREN OPZIOAK*

Lehen bezala, hemen ere garrantzitsuenak baino ez ditugu agertuko:

1. Desagerturiko edo galdutako datuak sartzen ditu. Datu baliagarri bezala jotzen ditu eta kontutan hartzen ditu estatistikoen eta portzentaien kalkuluetan.

2. Aldagai eta balioen etiketak baztertzen ditu.

3. Portzentaiak lerroka azaltzen ditu.

4. Portzentaiak zutabeka azaltzen ditu.

5. Portzentaia orokorrak azaltzen ditu.

14. Itxarondako maiztasunak azaltzen ditu.

15. Itxarondako eta behaturiko maiztasunen arteko diferentziak edo hondarrak azaltzen ditu.

16. Hondarrak edo diferentziak tipifikaturik azaltzen ditu.

18. Laukitxo bakoitzean ahal den informazio guztia azaltzen du.

Aurreko kasua berrikusiko dugu, baina 14. eta 15. aukera eta 1 eta 2 estatistikoekin osaturik. Kasu honetan sintaxia honelakoa da:

```
CROSSTABS TABLES=SEXUA BY HERRIA/OPTIONS=14 15
/STATISTICS=1 3.
```

Eta emaitzak:

***** Given WORKSPACE allows for 9475 Cells with
2 Dimensions for CROSSTAB problem *****

Crosstabulation : SEXUA
By HERRIA

HERRIA=> SEXUA	Count / Exp Val / Residual	ATAUN 1	EIBAR 2	GASTEIZ 3	Row Total
EMAKUMEA	1	14 25.0 -11.0	19 16.9 2.1	35 26.1 8.9	68 54.4%
GIZONA	2	32 25.0 11.0	12 25.0 -2.1	13 21.9 -11.0	57 45.6%
	Column Total	46 36.8%	31 24.8%	48 38.4%	125 100.0%

Chi-Square	D.F.	Significance	Min E.F.	Cels with E.F. < 5
17.87790	2	.0001	14.136	None

Number of Missing Observations = 0
Contingency Coefficient .35373

Aurreko emaitzak aztertzen baditugu, erdiko taulan izan den aldaketaz jabetuko gara. Hor, 14 eta 15 aukerak dira bitarte; laukitxo bakoitzean itxarondako eta jasotako maiztasunen arteko diferentziak azaltzen bait zaizkigu.

1 eta 3 estatistikoak eskatu ditugunez Chi-square balioa, lagineko Ji karratuarena eta D.F. balioa, zeinak askatasun-graduak adierazten dizkigun, agertzen zaizkigu.

X^2-aren hipotesi-testa egitean funtsezko balioa Significance izenpean agertzen dena da. Balio horrek itxarondako eta jasotako maiztasunen arteko diferentziak zoriak eraginda izatearen probabilitatea adierazten digu. Gure kasuan Significanceren balioa .0001 izanik eduki ditugun diferentziak zoriaren ondorioa izateko dagoen probabilitatea: 1/10000koa izango litzateke. Edozein erabaki hartzekotan, onar dezakegun goreneko erroreari begiratu behar diogu. Gure kasuan % 1 edo % 5 izango balitz, aldagaiak elkarrekin erlazionaturik daudela esan beharko genuke.

SPSS/PC+-ren irteeran ikusi ahal izan dugunez, zoriz suertatzearen probabilitatea zuzenean azaltzen zaigu, hots, ez dugu tauletan begiratu behar.

Jarraian eta Yates-en zuzenketa (zeina 5 baino txikiagoko itxarondako maiztasunen kasuetarako ezinbestekoa den) beharrezkoa izan daitekeelakoan itxarondako gutxieneko

maiztasuna azaltzen digu eta, ostean, bost elementu baino gutxiago duten laukitxo-kopurua. Gure kasuan "None" jartzen du, zeren eta 5 baino gutxiago itxarondako laukitxorik ez bait dago.

Azkenik kontingentzi koefizientea azaltzen digu. Ildo honetan koefiziente hori aldagai kualitatiboen edota bi ezaugarri-multzoen arteko erlazio-neurri bat dela gogorarazi behar dizuegu. Bere kalkuluetarako formula hau erabiliko dugu:

$$C = \sqrt{\frac{X^2}{X^2 + N}}$$

Azpiprograma hau erabiltzeko beste idazkera posible bat hau izango da:

CROSSTABS TABLES=SEXUA MAILASOZ ESKOLAMO BY HERRIA/
OPTIONS=14 15/STATISTICS=ALL.

Kasu honetan instrukzio bakar batean 3 analisi egin ditzala eskatu diogu, hots, SEXUA, MAILASOZ eta ESKOLAMO aldagaiak HERRIA aldagaiarekiko gurutza ditzala.

Era berean beheko formatu hori, zeinetan bi analisi desberdin eskatzen bait dizkiogu, erabil genezake. Instrukzio honen bidez SEXU eta HERRIAren arteko erlazioa eta ESKOLAMO eta MAILASOZen arteko balizko erlazioa ikus genezake.

CROSSTABS TABLES=SEXUA BY HERRIA/ESKOLAMO BY MAILASOZ/
OPTIONS=3 4 5 14 15/STATISTICS=1 3.

Hurrengo kasuan X^2-aren hiru analisi egin ditzala eskatzen diogu. Bat SEXUA eta HERRIA, ESKOLAMOren LEHENENGO BALIOAREKIKO erlazionatzeko. Bi, gauza bera ESKOLAMOren BIGARREN BALIOAREKIKO eta, azkenik, ESKOLAMOren HIRUGARREN BALIOAREKIKO gauza bera egin dezala. Beste era batera esanda, sexu eta herrien proportzioak eskola-mota bakoitzean ea mantentzen diren ala ez. Analisi horiek egin nahi baditugu, honela idatzi beharko dugu:

CROSSTABS TABLES=SEXUA BY HERRIA BY ESKOLAMO/
OPTIONS=14/STATISTICS=ALL.

ESKOLAMOren lehenengo balioari dagozkion emaitzak, PUBLIKOA deritzon balio horri dagozkionak alegia, ondoren aurkezten dizkizuegu:

***** Given WORKSPACE allows for 7480 Cells with
3 Dimensions for CROSSTAB problem *****

Crosstabulation : SEXUA
By HERRIA
Controlling for ESKOLAMO = 2 PUBLIKOA

HERRIA=> SEXUA	Count / Exp Val / Row Pct / Col Pct / Tot Pct	ATAUN 1	EIBAR 2	GASTEIZ 3	Row Total
EMAKUMEA	1	6 12.2 26.1% 23.1% 12.2%	6 2.8 26.1% 100.0% 12.2%	11 8.0 47.8% 64.7% 22.4%	23 46.9%
GIZONA	2	20 13.8 76.9% 76.9% 40.8%	6 3.2 0.0% 0.0% 0.0%	6 9.0 23.1% 35.3% 12.2%	26 53.1%
	Column Total	26 53.1%	6 12.2%	17 34.7%	49 100.0%

Chi-Square	D.F.	Significance	Min E.F.	Cells with E.F. < 5
14.88116	2	.0006	2.816	2 OF 6 (33.3%)

Contingency Coefficient .48265

Taulako goi-ezkerraldeko angeluan adierazi eta ikus daitekeenez laukitxoetan azaltzen zaigun informazioa lehenengoarekin konparaturik aldatu egin da. Horrela azaltzen diren balioak (beheranzko ordenean) hauek dira: itxarondako maiztasunak, lerroko portzentaiak, zutabeko portzentaiak eta lagin-elementu guztiekiko portzentaia.

Zutabe-eta lerro-kopuruak aldatu ez direnen askatasun-graduak lehengokoak dira.

Kasu honetan ere % 1eko errore maila ezartzen badugu, hipotesi nulua ukatuko dugu, hots, eskola PUBLIKOAri atxekitako sujetuen artean SEXUA eta HERRIA direlakoen balioen portzentaiak mantentzen direnaren hipotesia ukatu egingo dugu, erabaki hau hartzerakoan gure arriskua 10000tik seikoa izango dela kontutan harturik.

Beraz jasotako eta itxarondako maiztasunen arteko diferentziak EZIN diezazkiokegu zoriaren eraginari egotzi.

Gure adibidean sarritan ikus dezakegunez eta, batez ere, lagin txikiak darabiltzagunean, 5 sujetu baino gutxiagoko den itxarondako maiztasuneko laukitxo batzuk aurki-tzen ditugu. Hori gertatzen denean SPSS/PC+ paketeak egoeraren berri ematen digu; 5 sujetu baino gutxiagoko laukitxo bi daudela, guztiarekiko portzentaia ere agertaraziz. Aipatzen ditugun maiztasunak ITXARONDAKOAK dira, gure kasuan 2 hain zuzen. Jasotako maiztasunak izango balira, laukitxo bakar bat izango litzateke.

Kasu hauetan, arestian esan dugunez, hipotesi-testaren emaitzak YATESen ZUZENKETAren bitartez zuzendurik azaltzen zaizkigu.

14. GAIA
TEST PARAMETRIKOAK. T-TEST AZPIPROGRAMA

Test parametriko guztien artean agian ezagunenetariko bat orain aztertuko duguna izan daiteke; batezbestekoen konparazio-testa alegia. SPSS/PC+ paketea erabiliz, test honen kalkuluak egiteko T-TEST azpiprogramaren bidez egin ditzakegu.

Horrekin talde biko diseinupean lan egin ahal izateko, talde horien batezbestekoen berdintasuna ala desberdintasuna aztertzea helburutzat hartuko dugu. Hori horrela bada ere azpiprograma honekin talde bien bariantzen berdintasuna ere azter dezakegu.

Esan dugunez bi datu-multzo izango ditugu. Orain problema zera da: batezbestekoen artean jasotako diferentziak zoriak eragindakoak diren, batezbesteko berdineko populaziokoak direlarik, edo, bestetik, gure laginak batezbesteko ezberdinak dituzten populaziokotatik aukeratuak izan diren erabakitzea.

Behaturiko diferentziak errealak edo zoriak eraginikoak izatearen probabilitateei buruz erabakitzerakoan, bi taldeen batezbestekoez gain, laginez desbidazio standardak eta haien tamainak kontutan hartu beharko ditugu.

Konparazioa egin aurretik, konparatu nahi ditugun bi taldeak ia askeak edo erlazionatuak (parekatuak) diren erabaki behar dugu, zeren eta aipatutako bi kasu hauetan erabiliko dugun prozedura ezberdina izango bait da.

Lagin erlazionatuak edo parekatuak kontsideratuko ditugu alde batetik pertsona berdinez osatutakoak, hau da, pertsona berberei bi momentu ezberdinetan neurketak egin dizkiegunean. Hau da kasu arruntena, eta ziur asko era honetako frogen abiapuntua. Hala ere, lagin erlazionatuak ere kontsideratuko ditugu; ondoko era honetan lortuko ditugunak: Eman dezagun populazio bateko pertsonak binaka ezarri ditugula, pare bakoitza ezaugarri berdinak dituzten pertsonez osatu dugularik. Parekatze hau egin ondoren, laginketa egiterakoan ez ditugu pertsonak aukeratuko, baizik eta bikoteak, eta ondorengo pausoa bikote bakoitzeko pertsona bat talde batean ezarriko dugu eta bestea bigarren taldean. Horrela aukeratutako taldeak ere erlazionatutzat kontsideratuko ditugu. Parekatze honen zergatia hau da: talde parekatuak aukeratu ondoren, adibidez bi tratamendu ezberdinen eragina konparatzeko, ezaugarri berdinak dituzten taldeak aukeratu ondoren, talde bakoitzeko elementuei tratamendu bat ezartzea, talde bakoitzean lortu ditugun emaitzak konpara ditzagun.

Beraz, talde erlazionatuak edo parekatuak ez badira, askeak edo independenteak izango dira.

Analisi honekin hasi aurretik betebeharreko baldintzak betetzen direla ziurtatzea ezinbestekoa da. Baldintza hauek honakoak dira:

.- Behaketak elkarrekiko independenteak izatea.

.- Zorizko lagina izan behar du, hots, prozedura aleatorioaren bitartez lorturikoa.

.- Lortutako laginei dagokien populazioan, neurtzen dugun aldagaia banaketa normalari jarraitu behar zaio.

.- Populazioko bariantzek berdinak izan behar dute. Baldintza hori, dena den, zuzenketa baten bitartez konpon daiteke. SPSS/PC+-k, hain zuzen, horrela egiten du.

.- Aldagaiek tarte-eskalan neurturik egon behar dute, hots, aldagai kuantitatibo izan behar dute.

SPSS/PC+ erabiliz analisi honekin zuzen jokatu ahal izateko, datuak independienteak ala binakaturikoak diren aldez aurretik bereiztu behar dugu.

Datuak elkarrekiko independenteak izango balira, ondorengo erako instrukzioa erabiliko dugu:

T-TEST GROUPS=aldagai dikotomizatua (v_1,v_2)/VARIABLES=aldagaiak.

Aldagai dikotomikoa (edo dikotomizatua) izango da bi datu-multzoen sortzailea. Gero "VARIABLES"en ostean aztertu nahi ditugun aldagai kuantitatiboen zerrenda, zeinetan aldagai horien batezbestekoak konparatu nahi ditugun, idatzi beharko ditugu.

Binakaturiko datuak direnean, azpiprogramaren eskema honelakoa izango litzateke:

T-TEST PAIRS=DENBORA1 DENBORA2.

Kasu honetan aldagaiak, zeintzuen emaitzak konparatzen ditugun, DENBORA1 eta DENBORA2 dira.

1. BI TALDE INDEPENDENTEREN BATEZBESTEKOEN KONPARAZIOA

Batezbestekoen konparazioari dagokion adibide bat ikusiko dugu; talde independenteen kasurako hain zuzen ere. Erabili dugun azpiprograma:

T-TEST GROUPS=SEXUA(1,2)/VARIABLES=TELEBIS1 INTELIGE ULERMENO.

Aurreko instrukzioan, lagineko sujetuak sexuarekiko bi taldetan banatu ditugu, eta

Aurreko instrukzioan, lagineko sujetuak sexuarekiko bi taldetan banatu ditugu, eta TELEBIS1, INTELIGE eta ULERMENO aldagaien batezbestekoen konparazioa egin dezala eskatu diogu programari.

Batezbestekoen konparazioa egiterakoan, arestian esan dugunez batezbestekoak funtsezkoak dira, baina darabiltzagun laginen bariantzak eta haien tamainak ere kontutan hartu beharrekoak dira.

Era berean hipotesi-testaren kalkulua egiterakoan, laginak bariantza berdineko populazioei atxekirik izateak ala populazioak bariantza desberdinekoak izateak, kalkulu diferenteak egitera behartzen gaituela jakin behar dugu.

SPSS/PC+-k aurreko paragrafoetan esandako guztia kontutan hartzen du. Jarraian agertuko ditugun emaitzek, SEXUA aldagai dikotomizatuaren arabera bi taldetan bereiztuko dute eta ULERMENO aldagaiaren batezbestekoen konparazioa egiten dute.

Independet samples of SEXUA
Group 1: SEXUA EQ 1 Group 2: SEXUA EQ 2

t-test for: ULERMENO AHOZKO ULERKUNTZA

	Number of Cases	Mean	Standard Desviation	Standard Error
Group 1	68	12.6618	3.006	.364
Group 2	55	14.7636	3.448	.465

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.32	.285	-3.61	121	.000	-3.56	107.92	.001

Hasieran SEXUA aldagaiaren bitartez bereizturiko bi taldeen deskribapena agertzen digu. Hipotesi-testean erabiliko diren parametroak hortik datozkigu, hots, talde bakoitzaren elementu-kopurua, desbidazio standarda eta errore standarda. Errore standarda, konfidantza-tarteak eraikitzeko ezinbestekoa dena, laginaren desbiderapen standarda eta lagin–tamainaren bigarren erroaren arteko zatidura da.

Ondoren batezbestekoen konparazio-analisien emaitzak agertzen dira.

Dena dela, batezbestekoen konparazioari ekin baino lehen Snedeckor-en F testaren bitartez bariantzen konparazioa egiten da. Izan ere, Student-en T estatistikoa ezarri baino lehen, laginak bariantza berdineko edo bariantza desberdineko populazioei atxekirik dauden jakin behar bait dugu. Kasu batean, bariantza berdineko populazioak direnean, eta bestean, bariantza desberdineko populazioak direnean, formula desberdinak erabili behar ditugu.

Bariantza berdineko populaziaren kasuan, talde bakoitzeko lagin-kopuruak n_1 eta n_2, eta lagineko bariantzak S_1^2, S_2^2 izanik, testaren kalkuluetarako erabiltzen den formula ondoko hau izango da:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S_t^2 / n_1 + S_t^2 / n_2}} \quad \text{non} \quad S_t^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

Bariantza diferenteko populazioko laginen kasuan:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{S_1^2 / n_1 + S_2^2 / n_2}}$$

Horrela lorturiko t balioak, kalkuluak eskuz egin izan bagenitu, Student-en T taulatan behatutakoekin konparatu beharko genituzkeen. Beti ere bariantza berdineko populazioen kasuan (zeina maiz agertzen den) n_1 eta n_2 lagin-tamainak izanik ($n_1 + n_2 - 2$) askatasun-gradu direla kontutan hartu behar dugu.

SPSS/PC+ paketeak bariantzen konparazio-analisiak agertzen dituenean helburua zera da: eskuinaldean dauden emaitzen artean zein aztertu behar dugun erabakitzea.

Prozedurari dagokionez, lehenik bariantzak berdintzat jo ditzakegun ala ez aztertu behar dugu. Horretarako ezkerraldean agertzen den lehenengo atalari begiratuz, F (1.32) balioa eta zoriak eraginikoa izatearen probabilitatea (.265) agertzen zaigu. Kasu honetan % 5eko signifikazio- maila aldez aurretik ezarririk balego, ateratako probabilitatea % 5 baino handiagoa da; onarturiko errore maximoa baino handiagoa alegia. Ezin dezakegu, beraz, bariantzen berdintasuna (hipotesi nulua) baztertu eta horiek horrela, dagokigun Student-en t analisi zuzena POOLED VARIANCE ESTIMATE izenburupean agertzen zaiguna izango da; erdiko atalean dagoena hain zuzen ere.

Aipaturiko atal horretan zera agertzen zaigu:

Laginetik lortu dugun t balioa: -3.61, askatasun-graduak: 121 (degrees of freedom) eta t balio horrek zoriak eragindakoa izatearen probabilitatea .000.

Lehen esan dugunez % 5 goreneko errorepean lan egiten badugu, batezbestekoak berdinak izatearen hipotesi nulua baztertuko dugu ($p < .05$).

Gure kasuan laginak batezbesteko berdineko populazioei atxekirik izatea edo, beste era batera, abiapuntutzat hartu dugun hipotesi nuluaren aurrean batezbesteko diferenteetako populazioei atxekirik izatearena, hots, ordezko hipotesia, onartuko dugu. Test honi alde biko test esango diogu (two-tailed test).

Hipotesi nulua eta ordezko hipotesia honela ere idatz genitzake:

$H_0: \mu_1 = \mu_2$ eta $H_1: \mu_1 < \mu_2$

Kasu honetan lagineko t balioari atxekiriko probabilitatea jakin ahal izateko, azpiprograman agertzen zaigunaren erdia izango da.

Horiek horrela, Student-en T emaitzak aztertzerakoan ondorengo eskema orokorrari jarraituko gatzaizkio:

Bi taldeen bariantzen berdintasuna aztertzen dugu. Hori ordenadore-irteeran ezker aldeko blokean aurki dezakegu.

Bariantzak berdinak direnean, bi taldeen batezbesteko berdintasun-azterketari ekingo diogu; ordenadore-irteeran POOLED VARIANCE ESTIMATE izenburupean, erdiko blokean dagoena aztertzeari hain zuzen.

Bariantzei buruzko erabakia bestelakoa denean, ezberdinak direnean alegia, gure hipotesi nulua baztertu egingo dugu. Taldeen batezbestekoen konparazio-emaitzak eskuin aldeko blokean daude; SEPARATE VARIANCE ESTIMATE izenburupean hain zuzen.

Ondorengo kasuan, % 5eko errore-mailapean lan egiten badugu (ez ordea % 1 denean) bariantza-berdintasunari buruzko erabakia diferenteak direnekoa izango da. Beraz, batezbestekoen konparazio-analisi egokia eskuin aldeko blokean SEPARATE VARIANCE ESTIMATE izenburupean agertzen da. Horren ondorioz taldeak batezbesteko berdineko populazioetatik datozkigula esango dugu. Izan ere t balioa 0.35 eta zoriak eraginikoa izatearen probabilitatea (2-tail. prob.) 0.726 bait dira. Signifikazio-maila, adibidez, 0.05 bada, hipotesi nulua onartuko dugu $p=.05$ baino handiagoa delako.

Independet samples of SEXUA
Group 1: SEXUA EQ 1 Group 2: SEXUA EQ 2

t-test for: TELEBIS2 TELEBISTA AURREAN ORDUAK BIGARREN NEURKETA

	Number of Cases	Mean	Standard Desviation	Standard Error
Group 1	68	16.6471	6.238	.757
Group 2	55	16.2727	4.690	.633

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.94	.038	.35	121	.726	.36	120.45	.722

Aurreko emaitzak ematen dizkigun instrukzioa honako hau izan da:

T-TEST GROUPS=SEXUA (1,2)/VARIABLES=TELEBIS2.

Beste idazkera batzuk ere badaude egon. Aldagai dikotomizatuaren balioak 1 eta 2 direnean, hauek ken daitezke eta instrukzioa honela idatz dezakegu:

T-TEST GROUPS=SEXUA/VARIABLES=TELEBIS2.

Bi taldeak eratzerakoan dikotomizatua ez den aldagaia erabil dezakegu. Horretarako aldagai kuantitatiboko elementuek hartzen duten balioen arabera sortuko genituzke lagineko bi taldeak. Esaterako:

T-TEST GROUPS=INTELIGE (100)/VARIABLES=ULERMENO.

Hor ULERMENO aldagaiaren emaitzen diferentziak aztertzen ditugu. Taldeak honela eratu ditugu: neurketa-tresnean 100 puntu edo gehiago lortu dituzten sujetuak alde batetik, eta bestetik, 100 puntu baino gutxiago lortu dituztenak.

OPZIOAK

Erabil ditzakegun hautabide edo aukerak:

2. Darabiltzagun aldagaietan gure datu-matrizean zenbait datu desagerturik duten elementuak analisi honetatik baztertu egiten ditu.

3. Aldagaien etiketak baztertzen ditu.

2. ERLAZIONATURIKO BI TALDEREN BATEZBESTEKOEN KONPARAZIOA

Kasu honetan elementu berdinek osatzen dituzten bi taldeko emaitzak konparatuko ditugu. Elementu berberei bi momentutan bi neurri desberdin hartu dizkiegu; egin diegu, adibidez, tratamendu baten aurreko eta osteko neurriak.

Horiek horrela, bi aldagai partehartzaileak, zeintzuen batezbestekoak konparatu nahi ditugun, azpiprograman idatzi behar ditugu.

Erabili dugun instrukzioa:

T-TEST PAIRS=TELEBIS1 TELEBIS2.

Paired samples t-test:
TELEBIS 1 TELEBISTA AURREAN ORDUAK LEHENENGO NEURKETA
TELEBIS 2 TELEBISTA AURREAN ORDUAK LEHENENGO NEURKETA

Variable	Number of Cases	Mean	Standard Desviation	Standard Error
TELEBIS1	123	19.1057	7.243	.653
TELEBIS2	123	16.4797	5.854	.528

(Differ.) Mean	Std Dev.	Std. Error	Corr.	2-Tail Prob.	t Value	Degrees Freedom	2-Tail Prob.
2.6260	3.233	.291	.900	.000	9.01	122	.000

Lagin independenteen kasuan bezala, hemen ere lehenengo pausoan aldagai bakoitzaren deskribapena eta analisian parte hartuko duten parametroak agertzen dizkigu. Parametro horiek elementu-kopurua, batezbestekoa, desbiderapen standarda eta errore standarda dira.

Erlazionaturiko laginen kasurako, Student-en T analisia t estatistikoaren kalkuluetan oinarritzen da; t horri dagokion formula honakoa izanik:

$$t = \frac{\bar{d}}{Sd / \sqrt{n}}$$

non: d : aldagaien arteko diferentzien batezbestekoa
S_d : diferentzia horien desbiderapen standarda
n : laginaren tamaina.

Azpikaldeko blokean bi aldagai horien diferentzi analisiaren emaitzak agertzen dizkigu. Ezkerretik hasita:

(DIFFERENCE MEAN: 2.6260),

diferentzien desbiderapen standarda:

(STANDARD DEVIATION: 3.233),

eta errore standarda:

(STANDARD ERROR : .291).

Erdiko blokean Student-en T analisiari dagozkion emaitzak agertzen zaizkigu. Lehenengo lagineko t, bigarrena askatasun-graduei dagokiena: n - 1, eta, azkenik, ikusitako t-ri dagokion signifikazio-maila alde biko kasurako. Alde bakarreko testari legokiokeena jakin nahi izan bagenu, agertzen den probabilitatea erdibitu egin beharko genuke.

Aurkezturiko adibidean eta datuak aztertu ondoren zera esan genezake: lehenengo momentuan eta bigarren momentuan telebista aurrean egiten duten ordu-kopuruen batezbestekoak signifikatiboki desberdinak direla.

Azpiprograma hau, bestetik, laginaren populazio-batezbestekoa eta emaniko balio jakin bat berdinak diren ala ez aztertzeko ere erabil dezakegu.

Gure adibidean argiago ikus dezakegulakoan telebista aurrean emaniko ordu-kopuruen diferentzia, lehenengo eta bigarren momentuan, 3 ordukoa izan daitekeela uste dugu. Hori da, hain zuzen, aztertu nahi duguna eta horretarako ondorengo azpiprograma idatzi dugu:

```
COMPUTE DIFEREN=TELEBIS1-TELEBIS2.
COMPUTE KONSTANT=3.
T-TEST PAIRS=KONSTANT DIFEREN.
```

Hasierako COMPUTE-ren bitartez DIFEREN delako aldagai berria sortu dugu. Bigarren aldagaia bigarren COMPUTEk sorturiko KONSTANT izango da, zeinak lagineko elementu guztietan 3 balioa hartzen duen. Hirugarren lerroan T-TEST instrukzioa bide, aldagaien konparazioari ekitea eskatzen diogu, edo, hobeto esanda, aldagaia eta konstantearen arteko konparazioari.

Aldagaien gehikuntzen artean dauden diferentziak aztertu nahi ditugu. Hori da, hain zuzen, gure planteamendua. DIFEREN eta 3 konstantearen arteko diferentziak zoriak eraginda sor ote daitezkeen (hipotesi nulua) ala signifikatiboak diren (ordezko hipotesia).

Emaitzak jarraian plazaratzen ditugu eta, jakina, interpretatze-eskema, lehengo bera izango da; eralazionaturiko laginen adibidean ikusi duguna alegia.

Paired samples t-test: KONSTANT
DIFEREN

Variable	Number of Cases	Mean	Standard Desviation	Standard Error
KONSTANT	123	3.0000	.000	.000
DIFERENT	123	2.6260	3.233	.291

(Differ.) Mean	Std Dev.	Std. Error	Corr.	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
.3740	3.233	.291	99.00	99.00	1.28	122	.202

Lehenengo lekuan DIFEREN eta KONSTANT aldagaien datu deskribatzaileak daude. KONSTANT delakoaren kasuan, logikoa denez, desbidazio standarda eta errore standarda 0 dira.

Beheko ezkerraldean "aldagaien" arteko diferentzien deskribapena agertzen zaigu. Lehenengo balioa 3 - 2.620, KONSTANT ken DIFERENen batezbestekoa. Desbidazio standarda DIFERENen desbidazio standardaren balio bera da, eta errore standardaren (Std. Error) balioa ere DIFERENi dagokiona da; KONSTANTen parametro hauek 0 bait dira.

Erdiko balioak ordea, 99.00 dira, hots, kalkula ezinezkoak izan dira. Izan ere aldagai bat konstantea izan da eta koerlazio-koefizientearen kalkulua ezinezkoa bait da.

Azkenik, eskuin aldean, Student-en T analisia, 3 balioa eta lagin-batezbestekoaren arteko diferentzi konparazioa agertzen da. Lagineko t (1.28) eta askatasun-graduak (122) direla eta, ezin dezakegu hipotesi nulua baztertu. Ezin ditzakegu DIFEREN aldagaia eta 3 konstantearen arteko diferentziak signifikatibotzat hartu, zeren eta diferentzia hauek zoriz lortzeko dagoen probabilitatea .202 bait da. Zoriaz eraginikoak direla onartu behar dugu beraz.

Hipotesi nulua ezin baztertu hori eta gehikuntza 3 dela baieztatu ahal izatea ez ditugu elkarrekin nahastu behar. 2.5 balioa KONSTANT bezala hartuko bagenu, eta agian 2 balioa emanez gero, berdin onartuko genuke hipotesi nulua. Edonola ere irakurlearen esku uzten dugu hori praktikan baieztatzea.

Azkenik, erlazionaturiko taldeen arteko instrukzioa honela ere idatz dezakegu:

T-TEST PAIRS=X_1 X_2 WITH X_3 X_4.

Hemen ondorengo batezbestekoen konparazioak egitea eskatzen diogu:

X_1, X_2-rekin alde batetik eta X_3, X_4rekin beste aldetik.

Erabilitako instrukzioa beste hau bada:

T-TEST PAIRS= X_1 X_2 X_3 X_4.

batezbestekoen konparazioak honako hauek izango lirateke: X_1 X_2-rekiko, X_1 X_3-rekiko, X_1 X_4-rekiko, X_2 X_3-rekiko, X_2 X_4-rekiko eta X_3 X_4-rekiko.

15. GAIA
BARIANTZ ANALISIA: ONEWAY ETA ANOVA AZPIPROGRAMAK

Aurreneko gaian independenteak edo erlazionatuak diren bi taldeko batezbestekoen konparaketa aztertu dugu.

Hala ere kasu askotan bi talde baino gehiagotan lortutako emaitzak aztertzea interesatuko zaigu. Adibidez, hiru hiritan lortutako TELEBIS1 aldagaiaren emaitzak aztertu eta konparatu nahi baditugu, binakako konparaketak egiten joan gintezke, baina erroreak pilatzen doazela kontutan hartzen badugu, prozedura hau, talde-kopurua handia izanik, ez da oso operatiboa izango; egin beharreko konparaketak eta talde-kopurua gehitzean erroreak pilatuko bait lirateke.

Bariantz analisia ere era baliagarria izango da gure lagina bi kriterioren bidez sailkatzen badugu; hau da, Faktore izeneko bi kriterioen arabera deskonposatzen dugu gure lagina, eta honen ondorioz sailkapen bikoitza lortzen dugu.

Gai honetan aztertuko ditugun bi ereduak ikusteko bi adibide aurkeztuko ditugu.

Faktore bati dagokio aurreneko eskema hau. Sailkatzeko kriterio bakarra dugu, hau da, ikerketaren populazio osoko elementuak aldagai kualitatibo baten emaitzen arabera sailkatzen ditugu. Gure adibidean sailkapen hau egiteko jatorrizko hiriak erabiliko ditugu.

Ataun, Eibar eta Gasteizen batezbeste telebistaren aurrean pasatzen duten ordu-kopurua berdina den ala ez aztertu nahi dugu.

Arazo honen konponketa FAKTORE BAKARREKO edo Bariantz sinplearen analisiaren bidez egingo dugu eta ONEWAY izango da eredu edo modelo honen emaitzak lortzeko SPSS/PC+-k eskaintzen digun azpiprograma.

Orokorki hartuta diseinuaren eskema hau izango litzateke:

ATAUN	EIBAR	GASTEIZ
Ta1	Te1	Tg1
Ta2	Te2	Tg2
Ta3	Te3	Tg3
Ta4	Te4	Tg4
Ta5	Te5	Tg5
Ta6	Te6	Tg6
Ta7	Te7	Tg7
.	.	.
.	.	.
.	.	.

Bigarren eskema, bi faktoreei dagokie. Hemen aztertu behar dugun populazioa, sailkapenaren bi kriterioen arabera deskonposatzen dugu, hau da, sailkapena egiteko bi aldagai kualitatibo hartzen ditugu oinarritzat. Gure adibidean, populazioa bi eratan deskonposatuko dugu: alde batetik jatorrizko herriari begiratuz, eta bestetik sujetuen sexuaren arabera.

Kasu honetan lortu nahi dugun helburua hirukoitza da:

1.- Ataun, Eibar edo Gasteizen telebistaren aurrean pasatzen duten ordu-kopurua berdina den ala ez aztertuko dugu. Hau aztertzea bakarrik izango balitz, aurreneko adibidean jarritako diseinuaren bidez egingo genuke, ONEWAY azpiprograma erabiliz.

2.- Neskek eta mutilek,batezbeste, telebistaren aurrean pasatzen duten ordu-kopuruak berdinak diren ala ez ere aztertu nahi dugu. Hau bakarrik aztertzea izango balitz, aurreneko kapituluan Student-en T testaren bidez lortzen genuen galdera honen emaitza.

3.- INTERAKZIOEN azterketa. Hau da bi faktorezko bariantz analisiaren zergatia ematen duena. Orain arte ikusi ditugun lehenengo bi helburuak beste analisi batzuen bidez aztertzeko posibilitatea bagenuen. Adibide baten bidez ikusiko dugu interakzioaren problematika. Hemen bi aldagai kualitatibo gurutzatzen ditugu eta emaitzak aztertuko ditugu:

HERRIA	ATAUN	EIBAR	GASTEIZ	
GIZONAK	$\overline{X}_{11} = 25$	$\overline{X}_{12} = 20$	$\overline{X}_{13} = 15$	$\overline{X}_{g} = 20$
EMAKUMEAK	$\overline{X}_{21} = 25$	$\overline{X}_{22} = 20$	$\overline{X}_{23} = 15$	$\overline{X}_{e} = 20$
	$\overline{X}_{a} = 20$	$\overline{X}_{e} = 20$	$\overline{X}_{e} = 20$	

Adibide honetan interakzioa dagoela nabari dezakegu. Hiru herrietan telebistaren aurrean egiten dituzten ordu-kopuruen batezbestekoak berdinak direla ikus ditzakegu. Bestetik, orokorki hartuta, gizonen orduak eta emakumeen orduak berdinak dira. Baina aldagai biak gurutzatzerakoan, Ataungo gizonek telebistaren aurrekaldean bertako emakumeek baino ordu gehiago pasatzen dutela ikus dezakegu. Gasteizen, berriz, alderantzizko prozesua gertatzen da. Horrelako egoerari INTERAKZIO deritzo.

Bi faktorezko bariantzaren analisia honelako erlazioak detektatzeko erabiliko dugu. SPSS/PC+ paketeak ANOVA azpiprograma eskaintzen digu arazo hau aztertzeko.

Bi faktoreen eredua N faktoreen orokortasunera eraman dezakegu, kontutan harturik, bi aldagai edo faktoreren interakzioaz gain, hiru faktoreren interakzioaz, lau... eta N faktoreren interakzioaz hitz egin beharra edukiko dugula, N faktore erabiltzen ditugunean.

EFEKTU FINKOAK ETA EFEKTU ALEATORIOAK

Hiru hiriko emaitzen konparaketa egiten dugunean, bi puntu oso desberdin era ditzakegu, edo beste era batera esanda, gure helburua bi eratakoa izan daiteke:

1.- Hiru hiri hauen emaitzak bakarrik konparatu nahi ditugu, ezaugarri berdineko hirietara hedatu gabe. Kasu honetan, hiri bakoitzetik aukeratutako laginak hiri horietan adierazgarriak izango lirateke. Horrelako azterketei EFEKTU FINKOKO deritze.

2.- Hiru hiri hauetako laginak ez dira hiri hauen adierazgarri bakarrik; baizik eta Euskal Herriko lurralde guztien adierazgarri dira. Hau da laginketa baten ondorioz Euskal Herriaren adierazgarri bezala hiru hiri hauetako elementuak izan dira aukeratuak. Beraz, kasu honetan ez gaude hiri hauetako bakarrik emaitzetan interesaturik; baizik eta hiri hauek adierazten duten populazio guztiko emaitzetan. Horrelako azterketei EFEKTU ALEATORIOZKO deritze.

Guk EFEKTU FINKOEN eredua aztertuko dugu, faktore bakarreko edo bariantz analisi sinpleen kasuan, efektu finkoen analisia efektu aleatorioen kasurako ere baliagarria izango dela kontutan harturik; batez ere talde-kopurua handia denean.

FAKTORE BATEN BARIANTZ ANALISIA. BARIANTZ ANALISI SINPLEA: ONEWAY azpiprograma

Bi talde baino gehiagotan lortutako emaitzen batezbestekoak konparatzeko baliagarria izango dela kontutan harturik, analisi honen bidez lortutako emaitzak baliozkotzat konsideratzeko beharrezkoak diren aplikazio-baldintzak kontutan hartu beharko ditugu, eta analisi hau aplikatu aurretik baldintza hauek betetzen diren baieztatu beharko dugu.

APLIKATZEKO BALDINTZAK

Talde desberdinetan konparatu nahi ditugun emaitzek, aldagai kuantitatibo batenak izan beharko dute. Hau da, konparatu nahi dugun aldagaiak arrazoi- edo tarte-eskala batean neurtua egon beharko du.

Taldeak egiteko erabaliko dugun aldagai aske edo independentea nominal edo kualitatibo bezala tratatua izango da. Taldeak egiteko aldagai kuantitatibo bat erabili nahi badugu, aurretik RECODE, IF edo antzerako agindu baten bidez aldagaia birkodetu egin beharko dugu.

Bete beharreko baldintzak hauek dira:

1.- Azpilagineko elementuak, banaketa normalari jarraitzen zaizkion azpipopulazio batzuetatik zoriz aukeratuak izan direla suposatzen da. Erroreak edo akatsak banaketa normalari jarraitu behar zaizkio.

2.- Azpilagin desberdineko bariantzek berdinak izan beharko dute. HOMOZEDASTIZITATE izeneko baldintza hau konprobatu ahal izateko, aztertzen dugun paketeak, COCHRAN-en C, BARTLETT-FOXENen F eta HARTLEYren F kontrasteak egiteko posibilitatea ematen digu.

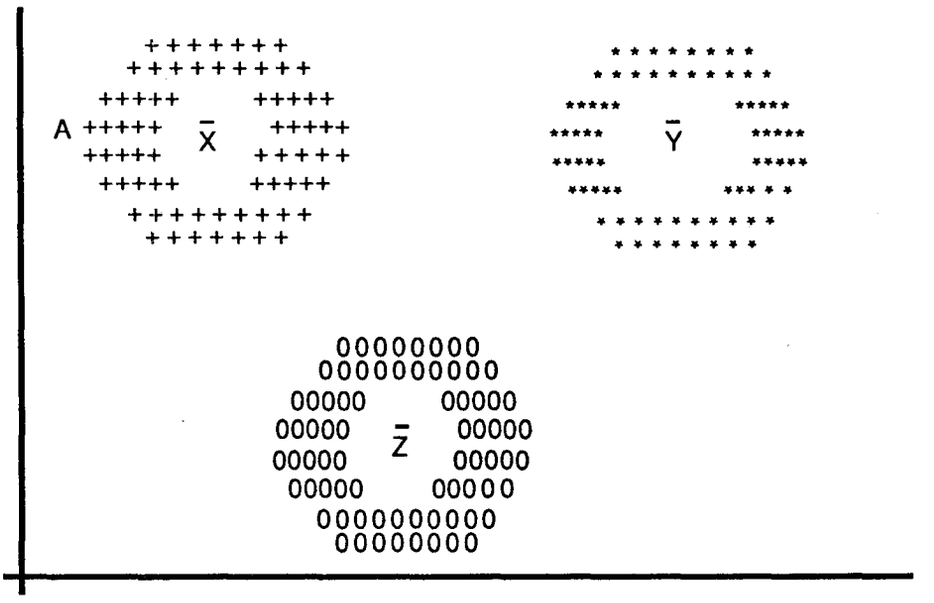
3.- Aztertzen ditugun azpilaginek, beraien artean aske edo independente izan beharko dute.

1. BARIANTZ ANALISIEN KALKULUAK

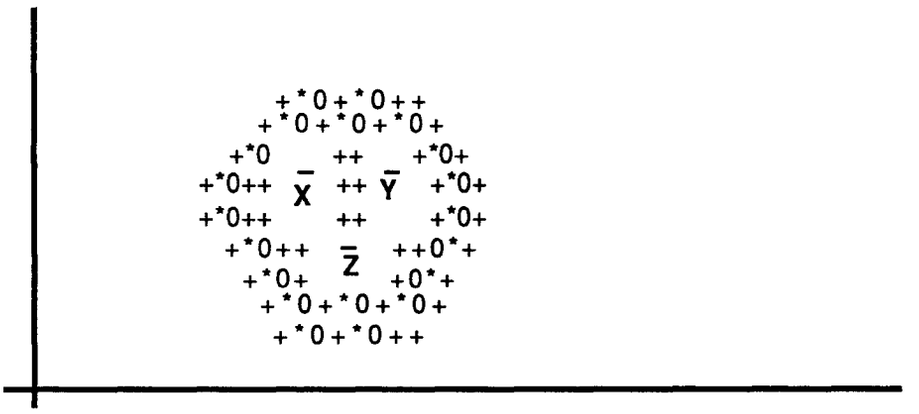
Bariantz analisiaren kalkulu-prozesua, Student-en T frogarekiko oso desberdina da.

Bariantz analisiaren planteamendu intuikorra ondorengo bi adibideen bitartez azalduko dugu.

TALDE DESBERDINAK



TALDE BERDINAK



Egiten den planteamendua bi distantzia desberdineko konparaketan datza: talde barruan dauden elementuen arteko distantzien eta taldeen arteko distantzien konparaketan.

Taldeen arteko distantziak, batezbestekoen arteko distantziak izango dira. BETWEEN GROUPS edo TALDEEN ARTEKO distantzia deritze.

Talde barruan dauden distantziak hartzeko, talde bakoitzaren batezbestekoa eta talde horren puntu edo elementu desberdinen arteko distantziak kontsideratuko dira. WITHIN GROUPS edo TALDE BARRUKO distantzia deritze.

Taldeko batezbestekoak desberdinak diren kasuan, taldeen artean dauden distantziak, hau da, taldeko batezbestekoen arteko distantziak (BETWEEN GROUPS) talde barruan dauden distantziekiko (WITHIN GROUPS) proportzionalki handiak izango dira.

Taldeko batezbestekoak berdinak diren kasuan, proportzioa inbertitu egiten da. Talde barruko distantziak (WITHIN GROUPS), taldeen arteko distantziekiko proportzionalki handiagoak izango dira.

Esan ditugun bi baieztapen hauek, aurreko irudian egiazta ditzakegu.

Ikus dezagun pausoz pauso bariantzaren analisiaren kalkulu-prozedura zein den. Honetarako, hiru taldeetako partaide diren elementuak dauzkagu. Talde horietan aplikazio–baldintzak betetzen direla suposatuko dugu.

A taldea	B taldea	C taldea
15	18	7
17	15	4
8	18	13
9	18	4
13	13	6
14	17	11

Kalkulu batzuk eginez hasiko gara, geroago erabiliko ditugun formuletarako beharrezkoak izango direlako. Lortu beharko ditugun emaitzak, talde bakoitzean hauek dira: aldagaien balioen karratuak, hauen baturak, talde bakoitzeko batezbestekoak, eta azkenik, elementu guztien batezbestekoa.

	X_a	X_b	X_c	X_a^2	X_b^2	X_c^2
	15	18	7	225	324	49
	17	15	4	289	225	16
	8	18	13	64	324	169
	9	18	4	81	324	16
	13	13	6	169	169	36
	14	17	11	196	289	121
$\Sigma =$	76	99	45	1024	1655	407

Hemendik

$$\overline{X}_a = 12.66 \ , \ \overline{X}_b = 16.50 \ , \ \overline{X}_c = 7.50$$

Batezbesteko orokorra:

$$\overline{X}_t = \frac{76 + 99 + 45}{18} = 12.22$$

Lehenengo aldagaien balioen berreduren batuketa orokorra kalkulatuko dugu. Horretarako formula hau erabiliko dugu:

$$\Sigma x^2 = \Sigma (X_i - \overline{X}_t)^2 = \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{N}$$

Kalkulu hauek egiterakoan bi era edo modu ezberdin aukera ditzakegu. Alde batetik aurreko formulak esaten dizkigun eragiketak egin ditzakegu, eta bestetik, kalkulu-prozesua errazten digun eta lehengoaren baliokidea den beste formula hau erabil dezakegu:

$$\Sigma x^2 = \Sigma X_a^2 + \Sigma X_b^2 + \Sigma X_c^2 - \frac{(\Sigma X_i)^2}{N}$$

eta balio hauek lehen eman ditugun tauletan kalkulaturik dauzkagunez:

$$\Sigma\ x^2 = 1024 + 165 + 407 - \frac{(76 + 99 + 45)^2}{18} = 3086 - 2688.88$$

eta hemendik KARRATUEN BATUKETA OSOA aterako dugu:

$$\Sigma\ x^2 = 397.12$$

Hau karratuen batuketa osoa da, eta lehen aipatu ditugun bi zatietan deskonposatuko dugu: TALDEEN ARTEKO KARRATUEN BATUKETAN eta TALDE BARRUKO KARRATUEN BATUKETAN.

TALDEEN ARTEKO karratuen batuketa lortzeko, ondoko formulak adierazten dituen kalkuluak egingo ditugu.

$$\Sigma\ X_e^2 = \Sigma\ (\overline{X}_i - \overline{X}_t)^2 * N_i =$$

$$= (12.66 - 12.22)^2 * 6 + (16.50 - 12.22)^2 * 6 + (7.50 - 12.22)^2 * 6 =$$

$$= 1.16 + 109.91 + 133.67 = 244.74 \text{ (TALDEEN ARTEKO).}$$

TALDE BARNEKO distantzien karratuen batuketa kalkulatzeko erabiliko dugun prozedura hau izango da. Lehenengo talde bakoitzeko barneko distantzien karratuak kalkulatuko ditugu, eta gero hauen batuketa egingo dugu.

A taldeko barneko distantzien karratuak:

$$\Sigma\ X_a^2 = (15-12.66)^2 + (17-12.66)^2 + (8-12.662 + (9-12.66)^2 +$$
$$+ (13-12.66)^2 + (14-12.66)^2 = 61.34$$

B taldeko barneko distantzien karratuak:

$$\Sigma\ X_b^2 = (18-16.50)^2 + (15-16.50)^2 + (18-16.50)^2 + (18-15.50)^2 +$$
$$+ (13-16.50)^2 + (17-16.50)^2 = 21.50$$

D taldeko barneko distantzien karratuak:

$$\Sigma\ X_d^2 = (7-7.50)^2 + (4-7.50)^2 + (13-7.50)^2 + (4-7.50)^2 +$$
$$+ (6-7.50)^2 + (11-7.50)^2 = 69.50$$

Hemendik lortuko dugu TALDE BARNEKO KARRATUEN BATUKETA, eta hau da: $61.34 + 21.50 + 69.50 = 152.34 = \Sigma\ X_d^2$

Ikus dezakegunez, KARRATUEN BATUKETA OSOA, TALDEEN ARTEKO KARRATUEN BATUKETA eta TALDE BARNEKO KARRATUEN BATUKETA batuz lor dezakegu:

K.Batuketa OSOA = K.B TALDE ARTEAN + K.B. TALDE BARNEAN.

Gure kasuan:

397.12 = 152.34 + 244.74

Ikus dezakegunez, 4 ehuneko desbidazio txiki bat dago, eta hau hamartar gutxirekin kalkuluak egin ditugulako gertatu da. Ez du inongo ondoriorik azken emaitzetan.

Bariantz analisi bat egiterakoan, F banaketa erabiliko dugu, eta 7. gaian ikusi genuen bezala, F banaketa erabiltzerakoan, honen askatasun-graduak kontutan eduki beharko ditugu. Beraz, hau da orain ikusiko duguna.

Gure adibidean 18 elementu dauzkagu, talde bakoitzean 6 elementu daudelarik. Talde bakoitzaren barruan dauden askatasun-graduak elementu-kopurua ken bat izango dira. Beraz, talde bakoitzeko askatasun-graduak 5 izango dira. Talde ezberdin guztien askatasun-graduak batzen baditugu, TALDE BARNEKO askatasun-graduen zenbakia lortuko dugu. TALDEEN ARTEKO askatasun-graduak, talde-kopurua ken bat izango da. Beraz gure kasuan taldeen arteko askatasun-graduak 2 izango dira.

Orokorki hartuta, kasu bakoitzeko askatasun-graduak hauek izango dira:

A.G. OSOA = elementu-kopurua - 1 = 18 - 1 = 17
A.G. TALDE ARTEAN = talde-kopurua - 1 = 3 - 1 = 2
A.G. TALDE BARNEAN = elementu-kop. - talde-kop. = 18 - 3 = 15

Jarraian taula laburtua aurkezten dugu. Bertan, KARRATUEN BATEZBESTEKOAK, karratuen batuketa eta askatasun- graduen arteko zatiketa eginda lortu ditugu.

ALDAKUNTZAREN JATORRIA	KARRATUEN BATUKETA	ASKATASUN- -GRADUAK	KARRATUEN BATEZBEST.	SNEDECKOR F
TALDEEN ARTEKOA	244.74	2	122.37	12.05
TALDE BARNEKOA	152.34	15	10.15	
OSOA	397.12	17		

Bariantz analisiko kalkuluak eskuz egiten baditugu, tauletan lortutako balioekin konparatu behar dugun balioa, F-rena da, hau da, laginetik lortu dugun F-ren balioa 12.05 da

eta hau 2 eta 15 askatasun-graduko F-ren tauletan, errorea finkatuz, lortuko dugun balioarekin konparatuko dugu. Gure kasuan, F-ren taulan %5eko errore-mailan lanean ari bagara, lortzen dugun balioa 3.68 da. Ostera %1eko errorearekin ari bagara, tauletan aurkituko dugun balioa 6.38 izango da, beti noski 2 eta 15 askatasun-graduko F-ren tauletan. Bi kasu hauetan laginetik lortu dugun balioa, 12.05, tauletan lortutakoak baino handiagoa da. Be-raz batezbestekoak berdinak direla dioen hipotesi nulua baztertu egingo dugu. Beraz, diferentziak signifikatiboak direla kontsidera dezakegu, %5 edo %1-eko errorearekin lanean ari bagara.

ONDORENGO ANALISIAK

Logikoki, bariantzaren analisia ez litzateke hemen amaituko. Beraien artean batezbesteko desberdinak dituzten taldeak zeintzuk diren analizatu beharko genuke. Lehen erabaki duguna zera da: hiru taldeko batezbestekoak ez direla berdinak. Orain begiratu beharko duguna zera da: beren artean desberdinak zeintzuk diren. Hau ondorenean egindako kontraste bidez burutzen da. Metodo hauen artean ezagunenak ondokoak dira: SCHEFFE, LSD, DUNCAN, TUKEY..... Ez ditugu zehatz-mehatz metodo hauek aztertuko, baina SPSS/PC+ paketeari analisi horiek nola eska diezazkiokegun ikusiko dugu, eta irteerakoren bat ere aztertuko dugu.

2. ONEWAY AZPIPROGRAMA

Lehengo adibidea egiteko erabili ditugun datuak analizatzen hasiko gara. Datu hauek LIBURU5.DAT izendatu dugun fitxategi batean sartuak izan dira, eta erabili dugun programa hau izan da:

```
SET MORE OFF.
SET LIST 'LIBURU5.LIS'.
SET LOG 'LIBURU5.LOG'.
DATA LIST FILE='LIBURU5.DAT'/X 1-2 TALDEA 3.
ONEWAY VAR=X BY TALDEA(1,3)/STATISTICS=ALL.
FIN.
```

Beraz, jarraian aurkezten ditugun emaitzak lortzeko erabili dugun azpiprograma hau da:

```
ONEWAY VAR=X BY TALDEA(1,3)/STATISTICS=ALL.
```

Ikusten dugunez, azpiprogramen eskema aurreko kapituluetan ikusi ditugunen tankerakoa da. Lehenengo, beti bezala, azpiprogramaren izena ezarri behar dugu, gero lanerako erabiliko ditugun aldagaiak indikatuz. Kasu honetan aldagai dependenteari (kuantitatiboari) X deitu diogu, eta taldeak eratzeko erabili dugun aldagaiari (kualitatiboari) TALDEA

deitu diogu. TALDEA aldagaiak hiru balio hartzen ditu: 1, 2, eta 3. Horregatik aldagaien atzekaldean, parentesi artean, balio txikiena eta handiena, 1 eta 3, ezarri ditugu.

Jarraian, estatistikoak, aukerak eta ondorengo kontrasteak indika diezazkiokegu.

Formatu laburragoa ere erabil dezakegu. Nahikoa izango da hau idaztea:
ONEWAY VAR=X BY TALDEA(1,3).

Hala ere, aurrerago espezifikatuko ditugun estatistiko denak eskatu dizkiogu.

Lortutako emaitzak hauek dira:

The raw data or transformation pass is proceeding
18 cases are written to the uncompressed active file.

----------ONEWAY----------

Variable X
By Variable TALDEA

Analysis of Variance

Source	D.F.	Sum of Squaren	Mean Squares	F Ratio	F Prob.
Between Groups	2	244.7778	122.3389	12.0514	.0008
Within Groups	15	152.3333	10.1556		
Total	17	397.1111			

----------ONEWAY----------

Group	Count	Mean	Standard Dev.	Std Error	95 Pct Conf Int for Mean
Grp 1	6	12.6667	3.5024	1.4298	8.99 To 16.34
Grp 2	6	16.5000	2.0736	.8466	14.33 To 18.67
Grp 3	6	7.5000	3.7283	1.5221	3.85 To 11.41
Total	18	12.2222	4.8332	1.1392	9.81 To 14.62

Fixed Effects Model	3.1868	.7511	10.62	To 13.82
Radom Effects Model		2.6076	1.00	To 23.44
Ranmod Efeccts Model -				
Estimate of Between Component Variance				18.7056

Group	Minimum	Maximum
Grp 1	8.0000	17.0000
Grp 2	13.0000	18.0000
Grp 3	4.0000	13.0000
Total	4.0000	18.0000

Test for Homogeneity of Variances
Cochrans C = Max. Variance/Sum (Variances) = .4562, P = .681
Bartlett-Box F = .819 , P = .442
Maximum Variance/Minimum Variance 3.233

Lortutako emaitzak aztertzen baditugu, bereiztu samarreko hiru zati agertzen direla ikusiko dugu. Lehenengo zatian bariantz analisia aurkezten du, eta hemen lehenengo zutabean askatasun-graduak agertzen dira, D.F. izenpean. Gero karratuen batuketak aurkezten ditu; bai taldeen artekoak (BETWEEN GROUPS), bai talde barnekoak (WITHIN GROUPS). Azkenik karratuen batuketa osoa agertzen da. Hurrengo zutabean karratuen batezbestekoak agertzen dira, MEAN SQUARES izenpean, eta ondoren, aurreko zutabeko bi balioen zatiketa den SNEDECKOREN F (F-RATIO). Azkenik, F-RATIO balio hau zoriz lortzeko probabilitatea eskaintzen digu paketeak, eta kasu honetan probabilitate hau .0008 dela adierazten du. Emaitza hau ikusi ondoren, gure erabakia zera izango da: Zoriz emaitza hau lortzea oso arraroa denez gero, hipotesi nuluak esaten duena baztertu egingo dugu. Hau da, batezbesteko berdinak direla esaten duen oinarrizko hipotesia ezeztatu egingo dugu, guk aurrez ezarritako goreneko errorea probabilitate hau baino handiagoa baldin bada, noski.

Lehenengo zati hau da, SPSS/PC+ paketeak aurkezten duena, inongo estatistikorik eskatu ez bagenu. Baina gure adibidean estatistiko guztiak eskatu dizkiogunez, beste bi zati ere aurkezten ditu.

Bigarren zatian talde bakoitzeko estatistiko deskribatzaileak aurkezten ditu; hau da, talde bakoitzeko elementu-kopurua, batezbestekoa, desbiderapen estandarda, errore estandarda, eta azkenik %95 mailako batezbestekoen konfidantza-tarteak. Azken lerroan TOTAL izenarekin, elementu guztien estatistikoak aurkezten ditu.

Emaitza hauek eta 5 lerro beherago agertzen diren txikiena (MINIMUM) eta handiena (MAXIMUM) balioak, 1 estatistikoa eskatu diogulako eskaintzen digu paketeak.

2 estatistikoa eskatu diogulako, efektu finko eta aleatorioen desbidazio estandarda eta errore estandarda, aurkezten ditu.

3 estatistikoa eskatzeagatik, talde ezberdinen bariantzen homogenotasuna aztertzen du. Cochran-en C eta Bartlett-en F analisia eta talde ezberdinen bariantzetatik handien eta txikien arteko proportzioa ere aurkezten digu. Kasu hontan analisi hauen emaitzak zoriz lortzeko probabilitateak .01 eta .05 baino handiagoak direnez, homogenotasunen baldintza betetzen dela onartuko dugu.

ONEWAY-en estatistikoak

Estatistikoak eskatzeko era, ondoko hau izango da:

/STATISTICS=estatistikoen zenbakiak.

Estatistikoak eta hauei dagozkien zenbakiak hauek dira:

Zenbakia

1 Talde bakoitzerako ondoko estatistikoak agertzen ditu: maiztasun absolutua, batezbestekoa, desbiderapen estandarda, estimazioaren errore estandarda, balio handiena eta txikiena, eta batezbestekoentzat %95eko probabilitateko konfidantza-tartea.

2 Efektu finko eta aleatorioentzat, desbiderapen estandarda, estimazioaren errore standarda eta batezbestekoentzaz %95eko probabilitateko konfidantza-tartea.

3 Homozedastizitatea egiaztatzeko hiru test egiten ditu: Cochran-en C, Bartlett-Boxen F eta Hartley-en F.

Estatistiko guztiak nahi baditugu, honela eska dezakegu:

/STATISTICS=ALL.

AUKERAK edo HAUTABIDEAK

Aurreko azpiprogrametan bezala, aukerak eskatzeko era hau izango da:

/OPTIONS=aukeren zenbakiak.

Azpiprograma honen aukera garrantzitsuenak eta hauen zenbakiak ondokoak dira:

Zenbakiak

1 Erabiltzaileak espezifikatutako balio desagertuak ere analisian sartu egiten ditu.

2 Analisian parte hartzen duten aldagaietan, balio desagertuak dauzkaten elementuak baztertu egiten ditu.

3 Aldagaien etiketak ez ditu azaltzen.

4 Talde bakoitzaren maiztasun absolutuak batezbestekoak eta desbiderapen standardak matrize erara aurkezten ditu.

5 Aldagai kualitatiboen balioen etiketak agertzen dira.

6 Aurretik 4 aukera erabiliz lortutako maiztasun absolutu, batezbesteko eta desbiderapen standard den matrizea irakurtzen du eta bariantz analisia egiten du.

ONDORENGO KONTRASTEAK

Lehen aipatu dugunez, bariantz analisiaren bidez taldeen arteko batezbestekoak diferenteak direla erabakitzen badugu, hurrengo pausoa zein taldetako emaitzak diren desberdinak ikertu beharko dugu. Horretarako ONDORENGO KONTRASTEAK erabiliko ditugu. Kontraste hauek ugari dira, eta hauetatik hemen ikusiko dugun bakarra, SCHEFFEren kontrastea izango da.

SPSS/PC+ paketeak, beste posibilitate bat ere eskaintzen digu, eta hau da: talde ezberdinetako elementuak batu beste talde berri batzuetan eta gero talde berri hauen batezbestekoen arteko diferentziak aztertzea. Eman dezagun ATAUN, GASTEIZ eta EIBAR hirietako pertsonak ULERMENO aldagaian lortu dituzten balioen batezbestekoak aztertu nahi ditugula. Hurrengo azpiprogramaren bidez aipatu ditugun emaitzak lortuko ditugu.

```
ONEWAY VAR= ULERMENO BY HERRIA(1,3)/
CONTRAST= -1 .5 .5/
RANGES=SCHEFFE/
STATISTICS=ALL.
```

Azpiprogama honetan, lehenengo pausoan hiru herrietako pertsonen emaitzak konparatuko ditugu. Bigarren pausoan, CONTRAST= -1 .5 .5, idatzi dugulako, pertsona guztiekin bi talde egin ditugu. Talde hauek egiteko erabili dugun era hau izan da: alde batetik, lehenengo herriko pertsonak, eta bestetik bigarren eta hirugarren herrikoak, zeren eta ezarri ditugun balioak: -1 .5 eta .5 izan bait dira. Zenbaki hauek baldintza pare bat bete behar dute. Alde batetik beren baturak 0 izan behar du. Bestetik, bi talde egiterakoan, lehenengoan zeinu positiboak daukaten taldeen pertsonak egongo dira eta bigarrenean zeinu negatiboa erantsi diegun taldeko pertsonak. Hau da, gauza bera adierazteko beste zenbaki hauek ere erabil ditzakegu: -2 1 1.

Bestalde, RANGES=SCHEFFE, idatzi dugulako, ondorengo kontraste hau ere egiteko eskatzen diogu.

Azkenik estatistiko guztiak eskatu ditugu.

Lortu ditugun emaitzak hauek izan dira:

----------ONEWAY----------

Variable ULERMENO AHOZKO ULERKUNTZA
By Variable HERRIA

Analysis of Variance

Source	D.F.	Sum of Squaren	Mean Squares	F Ratio	F Prob.
Between Groups	2	414.2851	207.1425	25.7002	.0000
Within Groups		120	967.1946		8.0600
Total		122	1381.4797		

Group	Count	Mean	Standard Deviation	Std. Error	95 Pct Conf Int for Mean
Grp 1	46	15.9565	2.7322	.4028	15.14 To 16.76
Grp 2	30	12.5667	2.8730	.5245	11.49 To 13.63
Grp 3	47	11.9574	2.9189	.4258	11.10 To 12.81
Total	123	13.6016	3.3651	.3034	13.00 To 14.20

Test for Homogeneity of Variances
Cochrans C = Max. Variance/Sum (Variances) = .3515, P = 1.000
Bartlett-Box F = .104 , P = .902
Maximum Variance/Minimum Variance 1.141

----------ONEWAY----------

Variable ULERMENO AHOZKO ULERKUNTZA
By Variable HERRIA

Contrast Coefficient Matrix

	Grp 1	Grp 2	Grp 3
Contrast 1	-1.0	.5	.5

Pooled Variance Estimate

	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	-3.6945	.5341	-6.917	120.0	.000

Separate Variance Estimate

	Value	S. Error	T Value	D.F.	T Prob.
Contrast 1	-3.6945	.5257	-7.027	96.3	.000

```
----------------ONEWAY----------------

     Variable  ULERMENO   AHOZKO ULERKUNTZA
By   Variable  HERRIA

Multiple Range Test

Scheffe Procedure
Ranges for the    0.50   level  -
                3.51    3.51
The ranges above are table ranges.
The value actually compared with Mean (J)-Mean (I) is..
                2.0075 * Range * Sqrt  (1/N(I)  +  1/N(J) )

(*) Denotes pairs of groups significantly diferent at the 0.50 level

                                G  G  G
                                r  r  r
                                p  p  p

     Mean       Group           3  2  1

   11.9574      Grp 3
   12.5667      Grp 2
   15.9565      Grp 1           *  *
```

Emaitzak aztertzeko orduan, lehenengo zatiaren egitura aurreko adibidean aztertutakoarekin bat datorrela esan dezakegu. F-ren balioa F PROB =.0000 izanik, taldeen artean dauden ezberdintasunak adierazgarriak edo signifikatiboak direla izango da gure erabakia.

Ondoren, eskatu dizkiogun estatistikoaren emaitzak ematen dizkigula ikusten dugu, hau da, ikerketan parte hartu duten herri ezberdinen deskribapena egiten du.

Eskatu ditugun kontraste-emaitzak komentatuko ditugu. Lehenengo pausoan CONTRAST aginduan eskatu diogun analisiak, T-test analisiaren bidez egiten ditu, Ataungo emaitzak Eibar eta Gasteizekoekin konparatuz. T-test azpiprograman ikusi genuen bezala, bariantza berdin eta bariantza ezberdinen kasuak bereizten ditu. Bariantzak berdinak edo ezberdinak diren erabakitzea guri dagokigu, bariantzaren homogenotasunaren testaren emaitzetan oinarrituz. Kasu honetan, adibidez, berdintzat joko ditugu, zeren eta COCHRANS C eta BARTLETT-BOX F-ren balioak zoriz lortzeko dagoen probabilitateak (1.000 eta .902) oso handiak bait dira.

Beraz, emaitzak POOLED VARIANCE ESTIMATEri dagozkion bi lerroetan aztertu

beharko ditugu.

Bi lerro hauetan, aztertzen ari garen bi taldeen batezbestekoen ezberdintasuna azaltzen da, hurrengo pausoan errore estandarda, gero T estatistikoen balioa, askatasun-graduak eta azkenik t estatistikoa zoriz lortzeko dagoen probabilitatea. Kasu honetan probabilitate hau .000 denez gero, gure erabakia konparatzen ari garen bi taldeko batezbestekoak diferenteak direla izango da.

Azkenik, SCHEFFE frogaren emaitzak azaltzen dizkigu, honek beraien artean ezberdintasun adierazgarriak dituzten taldeak zeintzuk diren aztertzen duelarik. Ezberdintasunen adierazgarritasuna, kontrako agindurik ematen ez badiogu, .05-eko errorearekin egiten du. Hau da, signifikatibo bezala ezartzen ditu taldeen arteko batezbestekoen ezberdintasunak zoriz lortzeko dagoen probabilitatea .05 baino txikiagoa denean. Errore-maila aldatzeko honela eskatuko genioke: SCHEFFE (.01), kasu honetan errore-maila %1ekoa izanik.

Scheffe-ren frogaren emaitzak azaltzeko era oso sinplea da, eta batezbestekoen diferentziak signifikatiboak dituzten talde-pareak adierazteko izartxoak erabiltzen ditu. Kasu honetan, 1 eta 2, eta 1 eta 3 taldeen arteko batezbestekoak ezberdinak direla azaltzen digu.

3. *BARIANTZAREN ANALISI MULTIFAKTORIALA: ANOVA AZPIPROGRAMA*

Faktore bakarreko edo sailkapen sinpleko bariantzaren analisia ikusi dugu ONEWAY azpiprograma azalduz.

Hemendik aurrera, sujetu edo elementuak sailkapen-erizpide bikoitz, hirukoitz... batean oinarrituz sailkatzen diren kasuak aztertuko ditugu.

Lehen adierazi dugu analisi honen funtsezko zatia faktoreen gurutzaketak sor dezakeen elkarrekiko eraginaren azterketa dela.

Abiapuntua, beraz, faktore izeneko bi sailkapen-erizpide edo gehiago izatea da. Hauek aldagai aske bezala kontsideratzen dira, eta kuantitatiboa (tarte edo arrazoi eskala baten neurtua) menpeko bezala kontsideratuko dugu.

Ondorengo eskema, bi faktore dituen diseinu bati dagokio; 2 * 3 diseinu faktorial izenekoari.

	B_1	B_2	B_3
A_1	X_{11}	X_{12}	X_{13}
A_2	X_{21}	X_{22}	X_{23}

Analisi hau erabiltzeko bete behar diren baldintzak, bariantza sinplearen analisiari zegokion zatian adierazitako berberak dira.

Oinarrizko hipotesia edo HIPOTESI NULUA hau da: Erabilitako sailkapen-erizpideen arabera lortu ditugun talde guztien batezbestekoak berdinak dira.

Bi faktore edo gehiagoko bariantzaren analisien kalkulu-prozesuan gelditu gabe, SPSS/PC+ paketeko ANOVA azpiprogramarekin lor dezakegun emaitzak aztertzeari hasiera emango diogu.

Oinarrizko azpiprograma honako hau izango da:

```
ANOVA VAR=ULERMENO BY SEXUA(1,2) HERRIA(1,3)/STATISTICS=3.
```

Azpiprograma honen eskema Oneway-ren antzekoa da. Azpiprogramaren izenaren ondoren aldagaiak idazten dira, menpeko edo aldagai kuantitatibotik hasiz; jarraian, aldagai nominalak, aztertu nahi diren taldeko balio handiena eta txikiena adieraziz. Bukatzeko, estatistikoak eta gero aipatuko ditugun aukerak edo opzioak eska ditzakegu.

Lortu ditugun emaitzak ondorengoak dira:

```
                    *** CELL MEANS ***

     Variable  ULERMENO   AHOZKO ULERKUNTZA
By   Variable  HERRIA

TOTAL POPULATION
     13.62
  (  121)

SEXUA
        1           2
     12.67       14.76
  (    66)    (    55)

HERRIA
        1           2           3
     15.96       12.41       12.04
  (    46)    (    29)    (    46)

          HERRIA
                  1          2          3
SEXUA   1     15.29      13.44      11.18
                (14)   (    18)   (    34)
        2     16.25      10.73      14.50
            (    32)   (    11)   (    12)
```

*** ANALYSIS OF VARIANCE ***

By ULERMENO AHOZKO ULERKUNTZA
SEXUA
HERRIA

Source of Variation	Sum of Squares	DF	Mean Spuare	F	Signif of F
Main Effects	423.881	3	141.294	21.009	.000
SEXUA	16.229	1	16.229	2.413	.123
HERRIA	291.963	2	145.981	21.706	.000
2-way Interactions	141.207	2	70.603	10.498	.000
SEXUA HERRIA	141.207	2	70.603	10.498	.000
Explained	565.088	5	113.018	16.805	.000
Residual	773.425	115	6.725		
Total	1338.512	120	11.154		

125 Cases were processed.
4 CASES (3.2 PCT) were missing.

Aurkezturiko emaitzak oinarrizko bi zatitan agertzen dira. Lehenengo zatian erabili diren erizpideen arabera lorturiko talde bakoitzeko emaitzen deskribapen bat egiten du.

Deskribapen honetan talde bakoitzaren batezbestekoak azaltzen dizkigu, lau bloke ezberdinetan. Lehenengoan, lagineko batezbesteko orokorra agertzen zaigu. Bigarrenean lehenengo faktoreen arabera lorturiko taldeen batezbestekoak aurkezten dizkigu. Hirugarren blokean, bigarren faktoreen arabera lortutakoenak. Azkenik, sailkapen-erizpide biak gurutzatuz lorturiko sei taldeen batezbestekoak aurkezten dizkigu. Emaitza hauek 3. estatistikoa eskatu dugulako azaltzen dizkigu.

Bigarren zatian, bariantzaren analisiaren emaitzak ikus ditzakegu. Aldaketaren efektuak lau atal ezberdinetan azaltzen dizkigu: MAIN EFFECTS, 2-WAY INTERACTIONS, EXPLAINED eta RESIDUAL.

Aipatzekoa da SUM OF SQUARES, hau da, karratuen baturaren zutabean, MAIN EFFECTS eta 2-WAY INTERACTIONSen batura EXPLAINEDen balioa dela, eta EXPLAINED eta RESIDUALen batura, TOTAL balioa dela. Bariantza analisiaren teorian bila dezakegu propietate honen arrazoia, baina ez gara hemen hau esplikatzen gelditukо.

MAIN EFFECTS balioak, erabili ditugun bi faktoreen, efektu zuzenak adierazten du. Gure kasuan efektu hauek SEXUA eta HERRIA aldagaiari dagozkionak dira.

Faktore bakoitzaren efektuak ONEWAY AZPIPROGRAMAN aurkeztutakoekin bat datoz. Hau da, zutabe ezberdinetan, karratuen batura, askatasun-graduak, karratuen baturaren batezbestekoa, Snedeckor-en F-ren lagineko balioa eta balio hau zoriz lortzeko dagoen probabilitatea agertzen dira. Kontutan hartu behar dugu balio hauek kalkulatzeko prozesua eta Oneway azpiprograman erabiltzen den prozesua berdinak direla.

Gure kasuko emaitzetan, sexuarekiko egindako taldeetan lortutako batezbestekoen arteko ezberdintasunak ez direla signifikatiboak ikus dezakegu. Ostera herriari begiratuz lortu ditugun taldeen batezbestekoak garrantzitsuak dira. Noski, ohizko errore-mailak (.05, .01 eta .001) erabiliz.

Faktore bien arteko interakzioa ere signifikatiboa dela ikus daiteke. Hau aztertzea komenigarria izango da, zeren eta ANOVA azpiprograma erabiltzea interakzio honek justifikatzen bait du. Kontutan eduki behar dugu orain arte ikusi ditugun emaitza guztiak, T-TEST eta ONEWAY azpiprogramak erabiliz lor ditzakegula.

Bi faktoreen arabera lorturiko sei taldeko batezbestekoak hauek izan dira:

		HERRIA 1 ATAUN	2 EIBAR	3 GASTEIZ	
SEXUA					
EMAKUMEA	1	15.29	13.44	11.18	$Y_1 = 12.67$
		(14)	(18)	(34)	
GIZONA	2	16.25	10.73	14.50	$Y_2 = 14.76$
		(32)	(11)	(12)	
		$X_1 = 15.96$	$X_2 = 12.41$	$X_3 = 12.04$	

Lehen ikusi dugunez emakume eta gizonen batezbestekoen arteko ezberdintasunak ez dira signifikatibo edo adierazgarriak. Ostera herri ezberdinekoak adierazgarriak dira. Sei taldeko batezbestekoak aztertzen baditugu, alde batetik, orokorki hartuta, emakumeen batezbestekoa gizonezkoena baino txikiagoa dela (12.67 eta 14.76) ikus dezakegu. Hau Ataun eta Gasteiz herrietan ere gertatzen da. Eibarreko batezbestekoak kontutan hartzen baditugu, ezberdintasunen zentzua aldatu egiten da. Hau da interakzioen adibide tipiko bat. Herri edo talde bateko ezberdintasunen zentzua, ezberdintasun orokorren zentzuen kontrakoa da. Zentzu honen aldaketa adierazkorra den ala ez ikusteko, Bariantz analisiko emaitzetan agertzen den 2-WAY INTERACTIONS lerroan ikusiko dugu. Kasu honetan, lerro honetako azken zutabeko balioa .000 denez geroz, interakzioari buruz hartuko dugun era-

bakia adierazkorra dela izango da.

Pasa gaitezen azpiprograma honen estatistikoak aztertzera:

ESTATISTIKOAK

Ondoren aipatzen ditugun estatistikoak eskatzeko era hau da:

/STATISTICS= estatistikoen zenbakiak.

Estatistiko 1 : sailkapen anizkoitzaren analisien emaitzak aurkezten ditu.

Estatistiko 2 : aldagai konkomitanteen erregresioko koefizienteak aurkezten ditu.

Estatistiko 3 : faktore guztiei begiratuz lorturiko talde bakoitzeko batezbestekoak eta maiztasun absolutuak aurkezten ditu.

OPZIOAK

Aukera edo opzioak eta eskatzeko era ondorengo hau da:

/OPTIONS= aukeren zenbakiak.

OPZIOA 1 : datu galduen etiketak agertzen dira.
OPZIOA 2 : etiketak ez dira agertzen.
OPZIOA 3 : interakzioen diseinua ez da agertzen.
OPZIOA 4 : interakzio hirukoitzak ez dira agertzen.
OPZIOA 9 : erregresioa egiten du.
OPZIOA 10 : egiten duen bariantz analisia hierarkikoa da.
OPZIOA 11 : estatistiko 3 eskatzen badiogu, aukera hau erabiliz, emaitzak formato laburtuan azaltzen ditu.

16. GAIA
ERREGRESIO ANIZKOITZA: REGRESSION AZPIPROGRAMA

Ikerketa enpirikoetan gehien erabiltzen diren metodoetan, behar bada ezagunena erregresio-metodoa da. Metodo honen helburua zera da: bi datu-multzo edo gehiago edukiz, lehenengo pausoan, beren artean erlazionatuta dauden ala ez jakitea, eta baiezko kasuan erlazio hau kuantifikatu eta zein zentzutakoa den aztertzea.

Metodo hau erabiltzeko, aztertu nahi ditugun aldagaiek kuantitatiboak izan behar dute. Hala ere batzuetan aldagai dikotomizatuak ere erabiltzen dira.

Analisi honetan aldagaiak bi sailetan banatuko ditugu. Alde batetik, menpeko aldagaia edo dependentea; guk aztertu nahi duguna. Beste aldean gelldituko lirateke beste aldagai guztiak, aske edo independente deituak. Esan beharra dago erregresio-analisian menpeko aldagaien aldaketak aztertzen direla aldagai askeen aldaketetan oinarrituz. Aldagai askea bakarra denean, erregresio Sinplea izango da. Ostera aldagai askeak bi edo gehiago direnean, erregresio anizkoitza izango da. Guk aztertuko ditugun kasuetan, beti menpeko aldagaia bakarra izango da.

Koerlazio eta erregresioen arteko ezberdintasun nagusiena zera da: koerlazioen kasuan ez da posible aldagai baten aldaketak besteen aldaketen arabera sortzen direla esatea, hau da, ez da posible izango kausalitateaz hitz egitea. Ostera, erregresio kasuan bai. Aldagai bat (menpekoa) ondoriozkoa izango da, eta besteak (askeak) zergatiak edo kausalak.

Erregresio-analisiak, batez ere lau helburu hauek
beteko ditu:

1.- Aldagaien arteko harremanak aztertzea.

2.- Aldagaien arteko erlazioa hobeto aztertu edo esplikatzen duen funtzio matematikoaren mota erabakitzea.

3.- Funtzio matematikoa zein eratakoa den jakin ondoren, funtzio honen parametroak kalkulatzea.

4.- Aldagai dependentearen aurresanak , aldagai askeekiko egitea.

Kapitulu honetan, aldagaien arteko erlazioak, linealak direla suposatuko dugu. Hau da, gure lagineko hodei-puntura gehien hurbiltzen den funtzio matematikoa lineala dela onartuko dugu oinarrizko hipotesi bezala. Beraz aipatutako 2. puntua ez dugu aztertuko.

Hala ere, esan beharra dago, aldagaien arteko erlazioa lerromakurra denean, eman behar diren pausoak, berdinak direla.

1. ERREGRESIO ANIZKOITZEN ANALISIEN SARRERA

Metodo honen oinarriak aztertzeko, 4. gaian egin genuen bezala adibide baten bidez eratuko dugu. Komenigarria izango da irakurleentzat gai honetan hasi baino lehen, 4.gaia ondo menperatzea.

Eman dezagun 10 ikasleei, ondorengo 4 aldagaien ezaugarriak neurtu dizkiegula:

Y : Zapuzte- edo frustrazio-maila: FRUSTRAZ

X_1 : Telebistaren aurrean astean pasatzen dituen orduak: TELEBIS1.

X_2 : Antsietate-maila. ANTSIETA

X_3 : Maila sozioekonomikoa. MAILASOZ

Gure helburua zera izango da: FRUSTRAZ aldagaian (menpeko aldagaia) dauden aldaketak TELEBIS1, ANTSIETA eta MAILASOZ aldagaietan (aldagai independente edo askeak) dauden aldaketetan oinarrituz ulertzen edo esplikatzen saiatzea.

Eman dezagun lortu dugun datu-matrizea hau dela:

	Y	X_1	X_2	X_3
	80	24	75	3
	73	20	64	3
	84	31	81	2
	40	12	39	4
	47	17	50	2
	56	16	60	2
	54	9	47	4
	90	32	83	1
	70	13	60	4
	43	19	40	2
Σ =	637	193	599	27

Logikoa dirudi Y aldagaia (FRUSTRAZ), beste aldagaiekin erlazionatuta dagoela esatea. Aldagai hauen arteko harremanak aztertzeko CORRELATION azpiprograma erabil dezakegu, aurreko gaietan ikusi dugun bezala. Hala ere, honelako galdera hau plantea dezakegu: X_1, X_2 eta X_3 aldagaien balioak aldatzen direnean, zein eratako aldaketa gertatzen da Y aldagaian? Y aldagaien balio bat noraino ezagut dezakegu, X_1, X_2 eta X_3 aldagaien balioak ezagutuz? Bi galdera hauen erantzunak erregresio-analisien bidez lortuko ditugu.

Erregresio-analisiko emaitzak lortzeko erabiliko dugun azpiprograma REGRESSION izango da.

Aurreko adibidea aztertuko dugu kalkulu guztiak eskuz eginik, gero emaitza hauek lortzeko idatzi beharko dugun instrukzioa zein izango den ikusiko dugu. Eta azkenik, SPSS/PC(+) paketeak ematen dizkigun emaitzak aztertuko ditugu.

Kalkuluei hasiera erregresio lineal sinpleko kalkuluak eskuz eginez emango diegu.

Lehenengo pausoan, 4. gaian ikusi dugun bezala, ondorengo kalkulu guztiak egin behar ditugu:

	Y	X_1	X_2	X_3	Y^2	X_1^2	X_2^2	X_3^2	YX_1	YX_2	YX_3	X_1X_2	X_1X_3	X_2X_3
	80	24	75	3	6400	576	5625	9	1920	6000	240	1800	72	225
	73	20	64	3	5329	400	4096	9	1460	4672	219	1280	60	192
	84	31	81	2	7056	961	6561	4	2604	6804	168	2511	62	162
	40	12	39	4	1600	144	1521	16	480	1560	160	468	48	156
	47	17	50	2	2209	289	2500	4	799	2350	94	850	34	100
	56	16	60	2	3136	256	3600	4	896	3360	112	960	32	120
	54	9	47	4	2916	81	2209	16	486	2538	216	423	36	188
	90	32	83	1	8100	1024	6889	1	2880	7470	90	2656	32	83
	70	13	60	4	4900	169	3600	16	910	4200	280	780	52	240
	43	19	40	2	1849	361	1600	4	817	1720	86	760	38	80
$\Sigma=$	637	193	599	27	43495	4261	38201	83	13252	40674	1665	12488	466	1546

$$\overline{Y} = 63.7 \qquad \overline{X_1} = 19.3 \qquad \overline{X_2} = 59.9 \qquad \overline{X_3} = 2.7$$

$$S_y^2 = 324.33 \qquad S_{x1}^2 = 59.567 \qquad S_{x2}^2 = 257.878 \qquad S_{x3}^2 = 1.122$$

Hemen lortu ditugun emaitzak, geroago erabiliko ditugu. Kontutan eduki behar dugu aurkeztutako lagineko bariantzaren balioak lagineko kuasibariantzarenak direla, zeren eta SPSS/PC+ paketeak balio hauek erabiltzen bait ditu estimatzaile bezala. Arrazoia zera da: estimatzaile hau alboragabea dela.

Suposa dezagun zapuzte-maila neurtzen duen aldagaia (FRUSTRAZ), telebistaren aurrean (TELEBIS1) pasatzen duten orduen arabera nola aldatzen den aztertu nahi dugula. Lehenengo pausoan koerlazio-koefizientea kalkula dezakegu. Kasu honetan Pearson-en formula aplikaturik, zera lortzen dugu:

$$R_{yx1} = \frac{N\,\Sigma(YX_1) - (\Sigma Y)(\Sigma X_1)}{\sqrt{N\Sigma(Y)^2 - (\Sigma Y)^2}\,\sqrt{N\Sigma(X_1)^2 - (\Sigma X_1)^2}}$$

$$R_{yx1} = \frac{10 * 13252 - 637 * 193}{\sqrt{10*43495 - 637^2}\,\sqrt{10*4261 - 193^2}} = \frac{9579}{12507} = .7659$$

Aldagai biak positiboki eta kantitate handian erlazionatuta daudela ohartzen gara. Hala ere, aldagaien arteko erlazioa handia edo txikia den erabakitzeko, edo beste era batera esanda erlazio hau kuantifikatzeko, erregresio-analisian MUGATZE-KOEFIZIENTEA erabiltzen da. Honen balioa koerlazio-koefizientea berbituz kalkulatzen da. Gure kasuan:

MUGATZE-KOEFIZIENTEA: $R^2 = .7659^2 = .5866$

Mugatze-koefizientea honela defini dezakegu: "Menpeko edo aldagai dependentearen bariantzaren proporzioa da, aldagai aske edo independentetik azaldu dena".

Gure kasuan, FRUSTRAZ aldagaien aldaketetatik, %58.66 TELEBIS1 aldagaian dauden aldaketagatik azaldua dela baiezta dezakegu. Emandako aldakortasunaren azalpena, honela esan dezakegu: Aldagai dependentearen bariantzaren %58.66, aldagai independenteen bariantzan oinarrituak daude.

Koerlazio-koefiziente guztietan, zati bat zoriari dagokiona izaten da, hau da, inolako erlaziorik datu-multzo batean aztertzen badugu, zoriaren ondorioz, aldagaien arteko koerlazio-koefizientea ez da 0 izango, balio honetatik oso hurbil egon arren. Efektu hau zuzentzeko, SPSS/PC+ paketeak, mugatze-koefizienteekin batera, ZUZENDUTAKO MUGATZE-KOEFIZIENTEA aurkezten du. Balio berri hau kalkulatzeko ondoko formula erabil dezakegu:

$$R_z^2 = R^2 - \frac{k\,(1 - R^2)}{(N - k - 1)}$$

non : k erabilitako aldagai askeen kopurua

N pertsona edo elementu-kopurua

R^2_z Zuzendutako mugatze-koefizientea

Gure kasuan:

$$R_z^2 = .5866 - \frac{1\,(1 - .5866^2)}{(10 - 1 - 1)} = .5349$$

Zuzendutako koefizientean, aurreko interpretazioa zuzendu beharko genuke: %58.66tik %53.49ra mugatze-koefizientea jaitsiz azaldutako errealitateari gehiago egokitzen zaiola esan genezake.

Hurrengo pausoa, FRUSTRAZ aldagaian espero diren balioak aurresatea izango litzateke. Eman dezagun haur batek astean zehar telebistaren aurrean 20 ordu pasatzen dituela. Haur honek FRUSTRAZ aldagaian hartzen dituen balioak aurresateko, erregresio zuzena honela erabil dezakegu.

Aldagai independenteari buruz hitz egiten dugunean, erregresio zuzenaren oinarria hau da: Matrizearen datu-bikote bakoitza, planoan puntu bat da (X,Y). Erregresio zuzena kalkulatzea, planoan jarritako puntu-multzora gehien hurbiltzen den zuzenaren ekuazioa kalkulatzea izango da.

Erregresio zuzena, hau da, puntu-multzoari gehien hurbiltzen zaion zuzena, ondorengo formularen bidez kalkulatzen da. Edozein zuzenen ekuazioa era honetakoa denez: $Y = a + b X$

$$b = \frac{N\ \Sigma(YX_1) - (\Sigma Y)(\Sigma X_1)}{\Sigma(X_1)^2 - (\Sigma X_1)^2}$$

$$a = \overline{Y} - b\ \overline{X_1}$$

Gure kasuan ekuazioko parametroen estimazioak hauek dira:

$$b = \frac{10 * 13252 - 637 * 193}{10*4261 - 193^2} = \frac{9579}{5361} = 1.7867$$

$$a = \overline{Y} - b\overline{X_1} = 63.7 - 1.7867 * 19.3 = 29.21$$

Hemendik, datu-multzoari gehien hurbiltzen zaion zuzenaren ekuazioa, hau da:

$$Y = 29.21 + 1.7867 * X_1$$

FRUSTRAZ = 29.21 + 1.7867 * TELEBIS1

Orain arte zuzen hau gehien hurbiltzen dena dela esan dugu. Hala ere komenigarria da noraino hurbiltzen den jakitea.

Noraino hurbiltzen den jakiteko, mugatze-koefizientea erabil dezakegu, edota beste

batzuetan ESTIMAZIOKO ERRORE TIPIKOA: S_{ee}. Mugatze-koefizientea eta errore tipikoa honela erlazionatzen dira:

$$S^2_{ee} = (1 - R^2_{xy})\, S^2_y \text{ (1. formula)}$$

Gure kasuan:

$$S^2_{ee} = (1 - 0.5349) * 324.33 = 150.85$$

Kalkulu hauek egiteko, zuzendutako mugatze-koefizientea eta lagineko kuasibariantza erabili ditugu.

Dena dela, egokitze edo hurbiltzearen ontasunaren interpretazioa, mugatze-koefizientearen arabera errazagoa izango da. Azken batean 1. formulak adierazten duen bezala estatistiko bien arteko erlazioa garbia da. Erlazio hau 4. gaian ikusten genuen.

Erregresio zuzena kalkulatu eta hurbilketa-prozesuan egindako errorea aztertu ondoren, zuzen honen bidez aurresana egiteko era ikusiko dugu.

Telebistaren aurrean 20 ordu ematen dituen mutikoarentzat bere zapuzte- edo frustrazio-mailarako egingo diogun aurresana hau izango da:

FRUSTRAZ=29.21 + 1.7867 * TELEBIS1=29.21 + 1.7867 * 20=64.94

Hau estimazio puntual bat da, eta tarte eran planteatu nahi badugu, ondoko formulan oinarrituz egingo dugu:

$$Y'i \pm T\alpha \sqrt{S^2_{ee}\left[1 + 1/n + \frac{(X_{1i} - \overline{X}_1)^2}{(n-1)\, S^2_x}\right]}$$

Gure kasuan eta, $T\alpha$ Student-en T tauletan, n - 2 askatasun-graduko lerroan lortzen dugun balioa dela kontutan harturik, hau da, 8 askatasun-graduko lerroan, %5 mailako erroreaz lana egiten baldin badugu, lortuko dugun balioa 2.306 izango da, eta S^2_{ee} = 120.63 dela kontutan hartuz, hau lortuko dugu:

$$64.94 \pm 2.306 \sqrt{120.63\left[1 + 1/10 + \frac{(20 - 19.3)^2}{(10-1)\, 53.61}\right]} =$$

$$64.94 \pm 2.306 * 11.52 = 64.94 \pm 26.56 = [38.38, 91.50]$$

Egia esan, lortu dugun konfidantza-tartea ez da oso zehatza; zabala baizik. Zabaltasun honen arrazoiak bi dira. Bata erregresio zuzen honen bidez lortu dugun hurbiltasuna

oso txikia izatea. TELEBIS1 aldagaien aldaketak, FRUSTRAZ aldagaien aldaketatik %58.66 esplikatzen du. Honek zera esan nahi du: beste %42 esplikatu gabe gelditzen dela. Bigarren arrazoia zera da: menpeko aldagaien (FRUSTRAZ) bariantza oso handia izatea, hau da, pertsona ezberdinek lortzen dituzten puntuazioak oso aldakorrak izatea.

Erregresio-ekuazioen parametroen konfidantza-tarteak ere lor ditzakegu, hots, "a" eta "b" balioenak. Horretarako "a" eta "b" balioen desbiderapen estandardetik abiatuko gara:

$$S_a = S_{ee}\left[\sqrt{\frac{1}{N} + \frac{(\bar{X})^2}{(N-1)\,S_x^2}}\right]$$

$$S_b = \frac{S_{ee}}{\sqrt{(N-1)\,S_x^2}}$$

Konfidantza-tarteak eraikitzea ahalbidetuko diguten estatistikoak ondoko hauek dira:

$$ta = \frac{a}{S_a} \qquad tb = \frac{b}{S_b}$$

ta eta tb estatistikoak Student-en T banaketari jarraitzen zaizkio, N-k-1 askatasun-graduarekin. Gure kasuan N-2 askatasun-graduarekin, zeren eta k balioak erregresio-ekuazioan sartu ditugun aldagai aske edo independenteen kopurua adierazten bait du.

2. *BARIANTZ ANALISIAREKIN DUEN ERLAZIOA*

Aipatu dugun bezala, erregresio sinplearen metodo linealaren funtsezko oinarria puntu-multzoari zuzen bat ahalik eta hoberen hurbiltzea litzateke.

Erregresio-ekuazioaz baliatuz, menpeko aldagaiaren aldakortasunaren zati bat besterik ez dugu azalduko. Beste zati bat azaldu gabe geratuko da.

Erregresio-analisiaren funtsezko erlazioa ondorengo bariantzaren deskonposaketa izango da.

Y-ren BARIANTZA = AZALDUTAKO BARIANTZA + AZALDU GABEKO BARIANTZA

Hemen ikusi dugun deskonposaketa eta bariantz analisian egiten genuena berdinak dira. Berdintasun hau aprobetxatuz, Snedeckor-en F hitzetan lortutako hurbiltasuna azter dezakegu, aurreko berdinketan oinarrituz. Azaldutako bariantza erregresio-ekuazioaren baitan dago. Bestea, azaldu gabekoarena edo erroreena (hondarrena) da.

Aurreko erlazioetatik eta bariantz analisiarekin dagoen antzekotasunagatik, ondoko erlazio hauek lortuko ditugu:

$$F = \frac{\text{BATEZBESTEKO KOADRATIKOA ERREGRESIOAREN ONDORIOZ}}{\text{BATEZBESTEKO HONDAR-KOADRATIKOA}}$$

$$R^2 = 1 - \frac{\text{HONDAR-KARRATUEN BATURA}}{\text{KARRATUEN BATURA OROKORRA}}$$

Garbi dago aldagaien artean zenbat eta erlazio handiagoa egon hainbat eta zati handiagoa esplikatzen duela aldagai askeak, eta horregatik batezbesteko hondar-koadratikoa txikiagotu egingo da. Honen ondorioz F balioa handiagotu egingo da. Badira erregresioren bidez lorturiko hurbiltze-egokitasuna F parametroen arabera aztertzen duten idazleak. Baina normalean, errazagoa izaten da mugatze-koefizientea erabiltzea.

3. ERREGRESIO SINPLEA SPSS/PC+ erabiliz

Datu-multzo bat erregresio sinplearen bidez aztertzerakoan, erabiliko dugun oinarrizko instrukzioa honako hau izango da:

```
REGRESSION VARIABLES=FRUSTRAZ TELEBIS1/DEPENDENT=FRUSTRAZ/
METHOD=ENTER.
```

Aurreko instrukzioan azpiprogramen izena idatzi ondoren hiru zati ezberdin bereiz ditzakegu. Lehenengoan, ekuazioan parte hartuko duten aldagaien izenak idatzi behar dira (gure kasuan: FRUSTRAZ eta TELEBIS1). Ondoren aldagai dependentea zein den adierazi behar da (gure kasuan FRUSTRAZ). Azkenik, erregresioko ekuazioa kalkulatzeko erabili dugun metodoa zehaztu behar da. Erregresioa sinplea denean, beti ENTER izango da. Erregresio anizkoitzen kasuan, aurrerago ikusiko dugun bezala metodo ezberdinak erabiltzeko posibilitatea edukiko dugu.

Aurreko instrukzioa erabiliz lortu ditugun emaitzak hauek dira:

Listwise Deletion of Missing Data
Equation Number 1 Dependet Variable.. FRUSTRAZ

Variable(s) Entered on Step Number
1.. TELEBIS1

Multiple R	.76586
R Square	.58654
Adjusted R Square	.53485
Standard Error	12.28073

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	1711.56950	1711.56950
Residual	8	1206.53050	150.81631

F = 11.34870 Sgnif F = .0098

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
TELEBIS1	1.78679	.53040	.76586	3.369	.0098
(Constant)	29.21489	10.94856		2.668	.0284

End Block Number 1 All requested variables entered.

Emaitzak aztertuz gero, zera ikusten dugu: Lehenengo, koerlazio-koefizientea .76586 dela (MULTIPLE R izenpean agertzen da). Bere berbidura, mugatze-koefizientea: .58654 eta lehen esan dugun bezala balio honen bidez egin dugun hurbilketaren egokitasuna aztertuko dugu. Kasu honetan errore estandarda S_{ee} = 12.28073, zeinaren bidez hurbilketaren Y-ren bariantza, mugatze-koefizientea eta errore estandarden arteko erlazioa azter bait dezakegu:

$$S^2_{ee} = (1 - R^2)\, S^2_y = (1 - .53485^2)\, 324.233 = 150.81$$

Kontutan eduki behar dugu Y-ren bariantzaren estimazio balioa lagineko quasibariantza dela, hau da:

$$S^2_y = \frac{\Sigma\,(Y_i - \bar{Y})^2}{N - 1}$$

Emaitzen hurrengo atalean, erregresio-ekuazio honi dagozkion bariantz analisiaren emaitzak agertzen dira.

Egindako deskonposaketa honako hau da:

BERBIDUREN BATURA OROKORRA = ERREGRESIO-BATURA + HONDAR-BATURA

2918.1 = 1711.5695 + 1206.5305

Azkenik, VARIABLES IN THE EQUATION izenpean bereizten den atalean erregresio-ekuazioa azaltzen zaigu. Lehenengo zutabean, aldagai aske edo independentearen izena, hurrengoan "a" eta "b" parametroen balioak erregresio-ekuazioan, zeina gure kasuan hau den:

FRUSTRAZ = 29.21489 + 1.78679 * TELEBIS1

Hurrengo zutabeetan parametroen desbiderapen estandardak eta parametroen estimazioak azaltzen zaizkigu, ikerketaren aldagaiak puntuazio tipikoetan konsideratzen baditugu. Kasu honetan, hau da, erregresio sinplean, Beta balioa eta koerlazio-koefizientea berdinak dira. Azken bi zutabeetan T estatistikoa azaltzen zaigu. Balio hauen konfidantza-tarteak eraikitzeko erabil dezakegu. Azken zutabean berriz, T balioen signifikazioa, hau da, balio hauek zoriz lortzeko dagoen probabilitatea.

Erregresio sinplearen adibide bat aztertuz gero, berehala sortzen da hurrengo galdera: nola hobe dezakegu aldagai askearen bidez menpeko aldagaiarekiko lortu dugun hurbilketa?. Hau da, nola lor dezakegu mugatze-koefizientea handitzea? Gure helburua zera da: menpeko aldagaien, FRUSTRAZ aldagaien, bariantzaren proportzio handiagoa esplikatzea. Horretarako, neurtuak ditugun beste aldagaiak erabiliko ditugu (aldagai aske batzuk) menpekoa nola eragiten duten ikertuz. Prozesu honi ERREGRESIO ANIZKOITZ deritzogu.

4. ERREGRESIO ANIZKOITZA

Teorikoki erregresio sinple eta anizkoitzaren planteamenduak berdinak dira. Erregresio sinplean bi aldagaiekin ari gara lanean (bat askea eta bestea menpekoa). Erregresio anizkoitzean berriz, menpeko aldagaia ere bakarra izango da, baina aldagai askeak nahi ditugun beste izan daitezke. Hau izango da izenaren arrazoia.

Ekuazio orokorra hau izango da:

$$Y = A + B_1*X_1 + B_2*X_2 + ... + BK*XK$$

SPSS/PC+ paketearen emaitzak eta interesatzen zaizkigun estatistikoak aztertzeko lehengo adibidearekin segituko dugu.

Erregresio-analisien helburuak ondorengo bi atal konkretuetan finka ditzakegu:

1.- Ekuazioen parametroen balioak kalkulatzea: Gure adibidean:

$$Y = A + B_1*X_1 + B_2*X_2 + B_3*X_3$$

2.- Aurreko puntuan kalkulatu dugun ekuazioaren bidez, menpeko aldagaiarekiko lortu dugun hurbiltasuna kuantifikatzea.

Orain arte koerlazio eta mugatze-koefizienteez hitz egin dugu. Dena den, kasu honetan hiru aldagai aske dauzkagu: TELEBIS1, MAILASOZ eta ANTSIETA, eta lehengo menpeko aldagaia FRUSTRAZ.

Orain arte ikusi ditugun formulekin ondorengo koerlazioen matrizea lor dezakegu. Emaitza hauek Correlation azpiprograma erabiliz lortu ditugu:

	FRUSTRAZ	TELEBIS1	ANTSIETA	MAILASOZ
FRUSTRAZ	1.0000	.7659	.9674	-.3198
TELEBIS1	.7659	1.0000	.8313	-.7488
ANTSIETA	.9674	.8313	1.0000	-.4657
MAILASOZ	-.3198	-.7488	-.4657	1.0000

Menpeko aldagaiarekin erlazio handiena duen aldagai askea ANTSIETA da. Aurreko matrizean ikus dezakegun bezala, aldagai guztiak beren artean erlazionatuta daude.

Koerlazio-matrize hau aztertuz gero, FRUSTRAZ eta TELEBIS1 aldagaien arteko harremanak plantea ditzakegu, ANTSIETA aldagaiaren eragina jasatzen duela, hau da, telebistak haurrengan sortzen duen frustrazioa haurraren antsietateagatik izan daitekeela. Estatistikak ahalbidetzen du erlazio berri hau kuantifikatzea, KOERLAZIO PARTZIALAREN bidez. Koerlazio partzialaren koefizienteez hitz egin ahal izateko gutxienez hiru aldagai beharko ditugu, zeren adibidez bi aldagairen arteko koerlazio partziala kalkulatzen ari garenean, partziala izateko erlazio honetatik hirugarren aldagaiaren eragina baztertu behar bait dugu. Hau da, eman dezagun FRUSTRAZ eta TELEBIS1 aldagaien arteko erlazioa kalkulatu nahi dugula, baina erlazio honetan ANTSIETA aldagaiak daukan eragina baztertuz. Erlazio hau kuantifikatzen duen koefizientea koerlazio partziala izango da. Erlazio hau adierazteko honela idatziko dugu: $R_{y1.2}$

$R_{y1.2}$ idazten badugu, Y eta X_1-en arteko erlaziotik X_2 aldagaiak duen eragina baztertuz, gelditzen den erlazioa adierazten du. Kalkulaketa formula honen bidez egingo dugu, non R_{y1}, Y eta X_y-en arteko koerlazio-koefizientea den; R_{12}, X_1 eta X_2-en artekoa, etab.

$$R_{y1.2} = \frac{R_{y1} - R_{y2} * R_{12}}{\sqrt{(1 - R^2_{y2})(1 - R^2_{12})}}$$

Gure kasuan koerlazio partzialen koefizientea hau izango da:

$$R_{y1.2} = \frac{.7659 - .9674 * .8313}{\sqrt{(1 - .9674^2)(1 - .8313^2)}} = \frac{-.0383}{.1407} = -.2721$$

Orain arte lortu ditugun emaitzen interpretazioa oso argigarria izan daiteke.

FRUSTRAZ eta TELEBIS1 aldagaien arteko koerlazio-koefizientea .7659 da. Beraz, esan genezake haur batek zenbat eta ordu gehiago telebista aurrean pasa, orduan eta frustrazio handiagoa edukiko duela. Hala ere, harreman honetatik haurraren antsietatea baztertzen badugu, hau da, haur honek inongo antsietaterik edukiko ez balu, gure aldagai bien arteko erlazioa alderantzizkoa izango litzateke, hau da, koerlazio partziala -.2721 da. Honek zera esan nahi du: FRUSTRAZ eta TELEBIS1 aldagaien arteko erlazioan, telebistak sortzen duen antsietatea dagoela erdian. Gogora dezagun datu simulatuekin ari garela lanean.

Bi aldagaien arteko erlazioan, hirugarrenen eragina baztertu dugu aipatutako kasuan, era berean laugarren aldagaien eragina bazter dezakegu, bostgarrenena...

Bi aldagaien eragina baztertzen badugu (adibidez eman dezagun Y eta X_1-en arteko erlazioa kalkulatu nahi dugula, baina X_2 eta X_3-en eragina baztertuz) koefiziente hau era honetan izendatuko dugu: $R_{y1.23}$, eta beraren balioa lortzeko erabiliko dugun formula hau izango da:

$$R_{y1.23} = \frac{R_{y1.2} - R_{y3.2} * R_{13.2}}{\sqrt{(1 - R^2_{y3.2})(1 - R^2_{13.2})}}$$

KOERLAZIO-KOEFIZIENTE ANIZKOITZA (multiplea)

Orain arte kalkulatu ditugun koerlazio-koefizienteak, bi aldagaien arteko erlazioa neurtzen dituztenak izan dira. Kasu honetan ikusiko dugun koefizienteak, berriak, adieraziko duen harremana hau izango da: alde batetik aldagai bakar bat (Y deituko dioguna) eta bestetik aldagai-talde bat. Eman dezagun FRUSTRAZ aldagaia eta TELEBIS1, ANTSIE-

TA eta MAILASOZ aldagaien arteko erlazioa kalkulatu nahi dugula, baina azken hiru aldagaiak elkarturik hartuta. Erlazio hau adierazteko erabiliko dugun koefizientea KOERLAZIO ANIZKOITZA izango da.

Lehen bezala koerlazio-koefizienteen berbidura MUGATZE-KOEFIZIENTEA izango da; kasu honetan ANIZKOITZA.

Garrantzi handiko koefizientea izango da mugatze-koefiziente hau, zeren eta erregresio anizkoitzaren bidez lorturiko hurbiltasuna kalkulatzeko erabiliko bait dugu.

Mugatze-Koefiziente Anizkoitzak adierazten diguna zera da: Aldagai askeko multzo baten bidez esplikatzen den menpeko aldagaien bariantzaren portzentaia.

Beraren kalkulua formula honen bidez egingo dugu. Eman dezagun Y aldagaia eta X_1 eta X_2 aldagaien arteko mugatze-koefizientea kalkulatu nahi dugula. X_1 eta X_2-k bien ar-tean, esplikatzen duen Y aldagaien bariantzaren portzentaia kalkulatu nahi dugu. Beste era batera esanda, X_1 eta X_2 aldagaiek, Y aldagaien aldaketak noraino esplikatzen dituzten.

$$R^2_{y.12} = 1 - [(1 - R^2_{y1}) * (1 - R^2_{y2.1})]$$

Y aldagaia eta X_1, X_2 eta X_3 aldagaien arteko mugatze-koefizientea kalkulatu nahi badugu, formula hau izango litzateke:

$$R^2_{y.123} = 1 - [(1 - R^2_{y1}) * (1 - R^2_{y2.1}) * (1 - R^2_{y3.12})]$$

Gure adibideko kasuan, gero ikusiko dugun bezala Y eta X_1, X_2 eta X_3-ren arteko: $R^2_{y.123} = .96118$ da. Beraz, honek adibidean erabili ditugun hiru aldagai askeek, FRUSTRAZ menpeko aldagaien bariantzaren %96.118 esplikatzen dutela esan nahi du.

Erregresio sinplearen kasuan bezala, gehienetan erabiliko ditugun mugatze-koefizienteak, ZUZENDUAK izango dira, eta zoriari dagokion zatia orekatzeko erabiltzen den faktorea ezarriz, MUGATZE-KOEFIZIENTE ANIZKOITZ ZUZENDUA lortzeko erabiliko dugun formula, hau izango da:

$$R^2_a = R^2 - \frac{k\,(1 - R^2)}{(N - k - 1)}$$

non : k erabili ditugun aldagai askeen kopurua
N lagineko elementuen kopurua diren.

Gure kasuan:

$$R_a^2 = .96118 - \frac{3\,(1 - .96118^2)}{(10 - 3 - 1)} = .94177$$

Orain arte aurkeztu ditugun emaitzak, SPSS/PC+ paketeak aurkeztu dizkigu. REGRESSION azpiprograma erabiltzeko, era honetako agindu bat idatzi beharko dugu:

REGRESSION VARIABLES= FRUSTRAZ TELEBIS1 ANTSIETA MAILASOZ/
STATISTICS=R COEFF OUTS F/DEPENDENT=FRUSTRAZ/METHOD=ENTER.

Aurreko agindua erabiliz lortu ditugun emaitzak hauek dira:

```
Equation Number 1          Dependet Variable..        FRUSTRAZ
Beginning Block Number     1. Method:   Enter
Variable(s) Entered on Step Number
    1..    TELEBIS1
    2..    MAILASOZ
    3..    ANTSIETA

Multiple R          .98040
R Square            .96118
Adjusted R Square   .94177
Standard Error     4.34511
F  =  49.52026                  Sgnif F  =     .0001

-----------------------Variables in the Equation-----------------------

Variable          B          SE  B          Beta          T  Sig  T
TELEBIS1     .36142         .49821         .15491        .526   .4955
MAILASOZ    4.05384        2.27966         .23849       3.162   .1257
ANTSIETA    1.06493         .17934         .94973      .35262   .0010
(Constant) -18.01001       10.41132                     2.992   .1344

End Block Number    1    All requested variables entered.
```

Aurreko emaitzek, lehenengo atalean erregresio anizkoitzaren bidez lortzen dugun hurbiltasuna adierazten digute. Koerlazio-koefiziente anizkoitza .98040 da, bere karratua, mugatze-koefizientea .96118 da, mugatze-koefiziente zuzendua .94117, eta errore estandarda 4.34511. Azaltzen den hurrengo balioa F-rena da eta ondoren beraren signifikazioa, hau da, F-ren balioa zoriz lortzeko dagoen probabilitatea.

Hurrengo blokean erregresioaren ekuazioa aztertzen digu. Lehenengo zutabean erregresio-ekuazioko parametroen estimazioak. Hemendik gure erregresio-ekuazioa atera dezakegu.

FRUSTRAZ= -18.01+ 1.06*ANTSIETA+ 4.05*MAILASOZ+ .36*TELEBIS1

Ekuazio honetatik abiatuz, FRUSTRAZen aurresanak egin ditzakegu. Eman dezagun pertsona baten aldagaien balioak hauek direla:

ANTSIETA = 70 TELEBIS1= 15 MAILASOZ = 3

Pertsona honen FRUSTRAZ aldagaien balioentzako egingo dugun aurresana hau izango da:

FRUSTRAZ = -18.01 + 1.06 * 70 + 4.05 * 3 + .36 * 15 = 73.74

ALDAGAI ASKEEN AUKERAKETA. EKUAZIOA OSATZEKO METODOAK

Orain arte erregresioan lortzen ziren emaitzak aztertzen saiatu gara, eta lortzen dugun hurbiltasuna handia ala txikia den mugatze-koefizientea erabiliz ikusi dugu.

Erregresio ekuazioen helburu garrantzitsu bat R^2 goreneko balioa lortzea izango da. Koefiziente hau handitzeko erabiltzen den metodo bat aldagai gehiago ekuazio barruan sartzea izango da. Baina batzuetan ohartuko gara hau ez dela baliagarria izango. Honen zergatia zera da: aldagai aske berria, menpeko aldagaia esplikatzen duena, lehendik ekuazioan sartuta dauzkagun aldagaiek ia esplikatuta edukitzea. Bostgarren aldagaitik aurrera oso kasu gutxitan handituko da mugatze-koefiziente zuzendua. Bestetik ekuazio barruan zenbat eta aldagai gehiago sartu, orduan eta zailagoa izango da emaitzak interpretatzea.

SPSS/PC+ paketeak, orain arte aipatu ditugun arrazoiak kontutan harturik, aldagaiak ekuazioan sartzeko orduan metodo ezberdinak erabiltzeko posibilitatea ematen digu. Gehien erabiltzen diren biak aipatuko ditut. STEPWISE eta ENTER.

STEPWISE METODOA. Agian gehien erabiltzen dena izango da, nahiz eta prozesuaren bukaeran ENTER metodoaren emaitza berdinak lortu; bidean lortuko dugun informazioa handiagoa izango bait da. Esan beharra dago etapazko metodoa dela. Hau da, metodo hau erabiliz, aldagai guztiak ez dira denak batera ekuazioan sartzen banan-banan baizik.

Aldagai askeetatik ekuazio barruan sartzen den lehenengo aldagaia, menpeko alda-

gaiarekiko koerlazio handiena daukana izango da. Aldagai honekin kalkulu guztiak egin eta azaldu ondoren, bigarren aldagaia sartuko du. Bigarren aldagai hau sartzerakoan, kontutan hartzen ditu koerlazio-koefiziente partzialak. Hau da, lehenengo pausoan kanpoan gelditu diren aldagai aske eta menpeko aldagaien arteko koerlazio-koefiziente partzialak kalkulatzen ditu, ekuazio barruan dagoen aldagaien eragina baztertuz. Hau da, bigarren aldagaia ekuazioan sartzerakoan komeni dena zera da: sartuko denak menpeko aldagaien bariantzaren portzentaia handiena esplikatzea, baina ekuazio barruan sartuta dagoenak esplikaten ez duena esplikatzea. Beraz hau da koerlazio-koefiziente partzialak adierazten duena. Bigarren aldagaia zein izango den erabaki ondoren, berriz kalkulu guztiak egin eta azaltzen ditu. Ondoren hirugarren aldagaia sartuko du ekuazioan. Hirugarren aldagaia sartzerakoan, orain arte kanpoan gelditu diren aldagai aske eta menpeko aldagaien arteko koerlazio-koefiziente partzialak kalkulatzen ditu, ekuazio barruan sartuta dauden aldagaien eragina baztertuz, eta horrela jarraitzen zaio prozesua bukatu arte.

Prozesua bi era ezberdinetara buka dezake. Lehenengoa, aldagai guztiak sartzen direnean. Baina batzuetan SPSS programa-paketeak ez ditu aldagai guztiak ekuazioan sartzen, hau da azkenak sartu baino lehen prozesua gelditu egiten da. Honen arrazoia zera da: aldagai berri bat sartzerakoan, lehen esan dugunez, koerlazio partzialak kalkulatzen ditu, baina batzuetan gerta daiteke koerlazio-koefiziente partzial hauek oso txikiak izatea, hau da, ekuazioan sartu gabe gelditzen diren aldagaiak, menpeko aldagaiari buruz esplikatzen duena ia barruan dauden aldagaiak esplikatuta egotea. Orduan aldagai berririk ez du sartzen, eta prozesua gelditu egiten da. Eman dezagun bi aldagai aske berdinak ditugula. Beraz ekuazioan bat sartzen badugu, bigarrena sartzea kalterako izango da, zeren eta ez du ezer berririk esplikatuko. STEPWISE metodoa erabiltzen dugunean, zentzu honetan egiten du lana SPSS/PC+ paketeak. Koerlazio-koefiziente partzialak txikiak edo handiak noiz diren erabakitzeko mugatze-koefizienteen gorakuntza azaltzen duen F balioa eta balio honen signifikazioa erabiltzen du. Aipatzekoa da prozesu honetan TOLERANCE koefizientea ere erabiltzen duela. Koerlazio-koefiziente berria da, baina hemen ez dugu aztertuko.

ENTER METODOA: Metodo hau erabiliz, aldagai aske guztiak batera sartzen dira ekuazioan.

5. ERREGRESIOA APLIKATZEKO BALDINTZAK

Metodo hau oso erabilgarri eta erabilia da, baina askotan metodo hau erabiltzeko bete behar diren baldintzak kontutan hartu gabe. Egia da, batzuetan nahiz eta hemen aipatuko ditugun baldintzak ez bete lortuko ditugun emaitzak baliagarriak izango direla, baina komenigarria izango da metodoa aplikatu baino lehen ea baldintza hauek betetzen diren ala ez kontutan hartzea.

Lehenengo eta oinarrizko baldintza, aldagai kuantitatiboak behar izatea da.

Eredu teorikoak bete behar dituen baldintzak hauek dira:

i) LINEALTASUNA: menpeko aldagai eta aldagai askeen arteko erlazioek linealak izan behar dute.

ii) EREDUAREN ZEHAZPENA: Eredu terorikoak ondo zehaztua egon behar du. Honek, menpeko aldagaiarekin erlazio handia duten aldagaiek ereduan egon behar dutela esan nahi du.

iii) ERROREEN INDEPENDENTZIA: Aldagai askeak errorerik gabe neurtzen dira.

iv) EREDU ALBORAGABEA da.

Erregresioa (inferentzi ikuspuntutik, hau da, aurresan zehatzak egiteko) erabiltzeko bete behar diren baldintzak aurrekoak eta ondorengoak izango dira.

v) HOMOZEDASTIZITATEA. Aldagai askeen balio ezberdinetarako, menpeko aldagaien bariantzek berdinak izan behar dute.

vi) ASKATASUNA. Elementu baten balioek, beste elementuen balioekin ez dute erlaziorik.

vii) MULTIKOLINEALTASUN EZA. Aldagai askeek beren artean aske izan behar dute. Hau da, aldagai askeen arteko koerlazio-koefizienteek 0 edo oso txikiak izan behar dute.

viii) NORMALTASUNA. Egiten diren hondar edo erroreen banaketak Normala izan behar du.

Hipotesi- edo aplikazio-baldintza hauetan oinarrituz, erregresioaren eredu teorikoa erabil dezakegu.

6. ERREGRESIO ANIZKOITZA SPSS/PC+ PAKETEA ERABILIZ

Erregresioa erabiltzeko azpiprogramaren instrukzio orokorra era honetakoa izango da:

```
REGRESSION VARIABLES=
          /STATISTICS=
          /DEPENDENT=
          /METHOD=
          /RESIDUALS=
```

Gure kasuan, honelako instrukzioa idatz dezakegu:

REGRESSION VARIABLES= FRUSTRAZ TELEBIS1 ANTSIETA MAILASOZ/
STATISTICS=R COEFF OUTS F/DEPENDENT=FRUSTRAZ/
METHOD=ENTER/RESIDUALS=HISTOGRAM.

Aipatu ditugun atal guztiak aztertuko ditugu:

VARIABLES : Erabiliko ditugun aldagai guztiak idatzi beharko ditugu. Menpekoa eta askeak.

DEPENDENT : Menpeko aldagaien izena idatzi behar dugu berdin zeinuaren ondoren.

METHOD : Zein metodo erabiliko dugun zehaztu behar dugu. Lehen ikusi dugun bezala gehien erabiltzen direnak STEPWISE eta ENTER izango dira.

RESIDUALS : Hondarrak aztertzeko erabiliko ditugun metodoak zehaztu behar ditugu atal honetan. Honen garrantzia, aplikazio-baldintzak betetzen diren ala ez ikustea izango da.

STATISTICS : Erregresio-metodoen estatistikoen izenak idatzi beharko ditugu atal honetan.

REGRESSION azpiprogramaren ESTATISTIKOAK

Erregresioaren kasuan, ez dira aurreko kasuetan bezala zenbakiak ezartzen; baizik eta hemen ikusiko ditugun izenak:

R : Koerlazio-koefiziente anizkoitza.

ANOVA : Bariantz analisiko taula aurkezten du.

F : B erregresio-koefizienteari dagokion F balioa eta bere signifikazioa aurkezten du.

CHA : Etapazko metodoetan, etapa batetik bestera mugatze-koefizienteen gorapena azaltzen du.

COEFF : Erregresio-ekuazioan sartu diren aldagaien estatistikoak azaltzen ditu.

OUTS : Oraindik ekuazioan sartu ez diren aldagaien estatistikoak azaltzen ditu.

TOL : Perdoiaren emaitzak aurkezten ditu.
ALL : Aurreko estatistiko guztiak azaltzen ditu.

Estatistikoak, bereziki ez badiogu eskatzen, beti R, ANOVA, COEFF eta OUTS, estatistikoen emaitzak ematen ditu.

ADIBIDE OROKORRA

Liburu osoan aztertzen ari garen adibide orokorrari jarraituz, erregresio-metodoarekin lortu ditugun emaitzak aztertuko ditugu.

FRUSTRAZ aldagaian, ULERMENO, TELEBIS1, INTELIGE eta ANTSIETA aldagaiek duten eragina aztertuko dugu.

Erabili dugun instrukzioa hau izan da:

```
REGRESSION VARIABLES=ULERMENO INTELIGE TELEBIS1 ANTSIETA
FRUSTRAZ/DEPENDENT=FRUSTRAZ/METHOD=STEPWISE/
RESIDUALS=HISTOGRAM(SRESID) NORMPROB OUTLIERS(MAHAL)
ID(ZENBAKIA).
```

Aurreko azpiprogrametan bezala aldagaiak DATA LIST FILE, aginduan kontsekutiboki idatzita egon baziren, TO hitza erabil dezakegu lehenengotik azkeneraino aldagaiak adierazteko.

Gure kasuan DEPENDENT=FRUSTRAZ idatzi dugunez, menpeko aldagaia hau izango da. METHOD=STEPWISE, ezarri dugulako, emaitzak azaltzerakoan etapaka joango da.

RESIDUALS, aginduan idatzi duguna, emaitzak aztertzerakoan ikusiko dugu.

Emaitzak aztertzen hasi aurretik, beti komenigarria izango da koerlazio-koefizienteen matrizea ezagutzea. Horregatik, hemen idatzi dugu:

Correlations:

	ULERMENO	INTELIGE	TELEBIS1	ANTSIETA	FRUSTRAZ
ULERMENO	1.0000	.0564	-.4157**	-.1730	-.3405**
INTELIGE	.0564	1.0000	.0576	-.546	-.1120
TELEBIS1	-.4157**	.0756	1.0000	.4300**	.7575**
ANTSIETA	-.1730	-.0546	.4300**	1.0000	.5287**
FRUSTRAZ	-.3405**	-.1120	.7575**	.5287**	1.0000

N of cases: 121 1-tailed Signif: * - .01 ** - .001

Idatzi dugun azpiprogramaren emaitzak, guk ere etapaka aztertuko ditugu. Lehenengo etapako emaitzak hauek izan dira:

1. ETAPA

```
**** MULTIPLE REGRESSION ****
Listwise Deletion of Missing Data
Equation Number 1          Dependent Variable..        FRUSTAZ
Beginning Block Number     1. Method:   Stepwise

Equation Number 1          Dependet Variable..         FRUSTAZ

Variable (s)  Entered on Step Number

1..   TELEBIS1 TELEBISTA AURREAN ORDUAK LEHENEGO NEURK

Multiple R              .75750
R Square                .57380
Adjusted R Square       .57022
Standard Error        11.06454

Analysis of Variance
                        DF      Sum of Squares      Mean Square
Regression               1        19613.91226      19613.91226
Residual               119        14568.46791        122.42410

F  =        160.21284          Signif F =     .0000

**** MULTIPLE REGRESSION ****

Equation Number 1       Dependet Variable..       FRUSTAZ

------------ Variables in the Equation ------------

Variable          B          SE B          Beta          T          Sig T

TELEBIST1     1.76560       .13949        .75750       12.658       .0000
(Constant)    28.97302      2.86174                    10.124       .0000

------------ Variables not in the Equation ------------

Variable        Beta In       Partial      Min Toler        T          Sig T

ULERMENO      -.03094      -.04311        .82716        -.469        .6401
INTELIGE      -.17016      -.25990        .99429      -.2.924        .0041
ANTSIETA       .24904       .34439        .81508        3.985        .0001
```

Ikusten dugunez, lehenengo etapa honetan ekuazioan sartu den aldagai askea TELEBIS1 izan da; aurreko koerlazio-matrizean ikus dezakegunez, FRUSTRAZ aldagaiekin koerlazio handiena daukana bait da. Lehenengo azaltzen den estatistikoa koerlazio anizkoitzena da (Multiple R), nahiz eta kasu honetan koerlazio sinplea izan. Ondoren, R Square balioa, mugatze-koefizienteena da, hau da, koerlazio-koefizienteen berbidura. Hurrengo azaltzen dena, mugatze-koefiziente zuzendua da (Adjusted R Square). Lehen egin dugun adibideetan bezala formula honekin kalkulatzen du:

$$R_a^2 = R^2 - \frac{k(1 - R^2)}{(N - k - 1)}$$

Logikoki, kasu honetan mugatze-koefiziente eta zuzendutako mugatze-koefizienteen artean dagoen ezberdintasuna lehengo adibideko kasukoa baino txikiagoa da, zeren eta lagineko tamaina handiagoa bait da.

Lehenengo zutabe honetako azkeneko balioa errore estandarden balioa da.

Emaitzen bigarren zatian, bariantz analisiko taula aurkezten du: alde batetik, esplikatuta gelditzen den menpeko aldagaien bariantzaren zatia eta bestetik esplikatu gabekoa. Irakurleak baiezta dezake lehen aipatu ditugun determinazio koefiziente, errore standard eta bariantzaren arteko egokitzapenak.

Esan beharra dago, gure laginak zituen 125 elementuetatik, 121 bakarrik erabili dituela, zeren eta 4 pertsonen datuak galduak edo ausenteak bait ziren. Hau bariantza analisiko askatasun-graduetan ikus dezakegu; ONEWAY azpiprogramaren emaitzen berdinak bait dira hemen aurkezten diren emaitzak.

Ondorengo blokean, erregresioaren emaitzak agertzen dira. Lehen aipatu dugun bezala, lehenengo etapa honetan ekuazioan sartu den aldagai askea TELEBIS1 izan da. Beraz aldagai hau erabiliz lorturiko emaitzak agertzen dira:

FRUSTRAZ = 28.97302 + 1.76560 TELEBIS1

Beste zutabeetan agertzen diren balioak, erregresio-ekuazioaren parametroen estimazioen estatistikoak dira. SE B, B estatistikoen desbiderapen standarda da. BETA, B balioen tipifikazioen emaitza. Kasu honetan BETA balioa eta koerlazio-koefizientea berdinak dira. Azkenik, T balioak eta balio hauen signifikazioak agertzen dira.

Azken blokean berriz, ekuazioan sartu gabe gelditu diren aldagai askeen estatistikoak agertzen dira. Bloke honetatik garrantzi handiena daukan zutabea Partial izenpean dagoena da. Balio hauek ikusi ondoren badakigu ekuazioan sartuko den hurrengo aldagaia ANTSIETA izango dela, zeren eta denetan koerlazio-koefiziente partzial handiena bait dauka .34439. Balio honek zera esan nahi du: FRUSTRAZ eta ANTSIETA aldagaien arteko

harremanean, TELEBIS1 aldagaiak duen eragina alde batera uzten badugu, lortzen dugun koerlazio-koefizienteak hori balio duela. Balio honen signifikazioa oso txikia da, .0001. Beraz balio hau zoriz lortzeko dagoen probabilitatea oso txikia denez, ekuazioan sartuko du hurrengo pausoan, ondoren aurkezten dugun emaitzetan ikus dezakegun bezala:

2. ETAPA

**** MULTIPLE REGRESSION ****

Variable (s) Entered on Step Number

2.. ANTSIETA

Multiple R	.79016
R Square	.62435
Adjusted R Square	.61798
Standard Error	10.43160

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	2	21341.82453	10670.91227
Residual	118	12840.55563	108.81827

F = 98.06177 Signif F = 0.0

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependet Variable.. FRUSTAZ

---------- Variables in the Equation ----------

Variable	B	SE B	Beta	T	Sig T
TELEBIST1	1.51599	.14567	.65040	10.407	.0000
ANTSIETA	.28871	.07245	.24904	3.985	.0001
(Constant)	15.33548	4.35798		3.519	.0006

---------- Variables not in the Equation ----------

Variable	Beta In	Partial	Min Toler	T	Sig T
ULERMENO	-.03269	-.04851	.69495	-.525	.6003
INTELIGE	-.14975	-.24249	.80524	-.2.704	.0079

Emaitzak lehen etapako formatu berdinean agertzen dira; koerlazio-koefiziente anizkoitza, eta logikoki, honen ondorioz, mugatze-koefizientea ere bai. Hemen zera ikus dezakegu: mugatze-koefizientea handitzen denean errore standarda txikiagotu egiten dela. Zenbat eta zati handiagoa esplikatu, hainbat eta errore txikiagoa edukiko dugu. Hau bariantz analisien emaitzetan ere ikus dezakegu.

Hurrengo zatian erregresio-ekuazioa azaltzen da. Kasu honetan bi aldagai askerekin osatuta: TELEBIS1 eta ANTSIETA.

FRUSTRAZ = 15.33548 + 1.51599 TELEBIS1 + .28871 ANTSIETA

Ikusten dugunez, erregresio-ekuazioko parametroak, bigarren aldagaia sartu denean, aldatu egin dira. Hau etapa guztietan gertatuko da.

Parametroen estimazioen desbiderapen estandardak, parametro standarizatuak, dagozkion T balioak eta hauen signifikazioak ere agertzen dira.

Azkenik, bigarren etapa honetan ekuaziotik kanpo gelditu diren aldagaien estatistikoak ere agertzen dira VARIABLES NOT IN THE EQUATION atalean.

Berriz aurreko etapako zutabe berdinak azaltzen dira, eta hemen, hirugarren etapan sartuko den aldagaia INTELIGE izango dela ikus dezakegu; koerlazio-koefiziente partzial handiena berarena bait da.

Goazen bada, hirugarren etapako emaitzak aztertzera:

3. ETAPA

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependent Variable.. FRUSTAZ

Variable (s) Entered on Step Number

3.. INTELIGE ADIMEN-KOEFIZIENTEA

Multiple R	.80401
R Square	.64644
Adjusted R Square	.63737
Standard Error	10.16342

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	22096.85438	7365.61813
Residual	117	12085.52578	103.29509

F = 71.30656 Signif F = .0000

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependet Variable.. FRUSTAZ

---------- Variables in the Equation ----------

Variable	B	SE B	Beta	T	Sig T
TELEBIST1	1.55840	.14279	.66860	10.914	.0000
ANTSIETA	.27016	.07092	.23304	3.809	.0002
INTELIGE	-.46092	.17048	-.14975	-2.704	.0079
(Constant)	62.01749	17.78101		3.519	.0006

---------- Variables not in the Equation ----------

Variable	Beta In	Partial	Min Toler	T	Sig T
ULERMENO	-.01684	-.02563	.68153	-.276	.7829

End Block Number 1 PIN = .050 Limits reached.

Azken etapa honetan, koerlazio-koefiziente anizkoitza ere handitu egin da. Kasu honetan .80401 da, mugatze-koefizientea .64644, eta zuzendutako determinazio-koefizientea .63737. Honek zera esan nahi du: ekuazioan sartu diren hiru aldagai askeen artean, menpeko aldagaien bariantzaren % 63.737 esplikatzen dutela. Errore standarda txikiagotu egin da, mugatze-koefizientea handiagotu den neurrian.

Erregresioko ekuazioa hau da:

FRUSTRAZ=62.01 + 1.55*TELEBIS1 + .27*ANTSIETA - .46*INTELIGE

Ikusten dugunez, ULERMENO aldagaia ez da ekuazioan sartu, eta gero ikusiko dugunez ez da sartuko. Beraz hau da azken erregresio-ekuazioa. Ekuazio honen bidez egingo ditugu gure aurresanak, beti ere ea metodo hau erabiltzeko baldintzak betetzen diren ala ez kontutan harturik.

Aipatu dugunez, prozesua hemen gelditukо da. Arrazoia zera da: ULERMENO eta FRUSTRAZ aldagaien arteko koerlazio-koefiziente partziala kalkulatu du, eta beraren balioa -.02563 da. Balio honi dagokion T balioa -.276 da, eta balio hau zoriz lortzeko dagoen probabilitatea, SIG T atalean azaltzen den bezala, .7829 izango da. Honek zera esan nahi du: balio hau ez dela esanguratsua. Beraz programak erabaki du balio hau ez sartzea. Hori adierazteko, esaldi hau ezartzen du: End Block Number 1 PIN = .050 Limits reached.

Beraz, hemen bukatzen da erregresio-ekuazioari dagokion zatia.

Ekuazioaren prozesu hau burututakoan, beste zati bat azaltzen da, eta aplikatzeko baldintzak aztertzeko eskatu diogun RESIDUALS aginduan aipatutako estatistikoen emaitzena da.

Gure agindua hau izan da:

RESIDUALS=HISTOGRAM(SRESIS) NORMPROB OUTLIERS(MAHAL)
ID(ZENBAKIA).

Aplikatzeko baldintzetan erregresioa aurresanak egiteko erabili nahi badugu, garrantzi handiena daukana zera da: hondarrak edo erroreak, banaketa normalari jarraitu behar zaizkiola.

Baldintza hau betetzen den ikusteko HISTOGRAM(SRESID) eta NORMPROB idatzi dugu.

Agindu honen bitartez hondarren adierazpen grafiko bat lortuko dugu. Hondarren analisi zehatzagoa egin nahi badugu, test ez-parametrikotan ikusiko dugun K-S testa egin beharko genuke.

HISTOGRAM(SRESID) eskatu dugulako ondoko adierazpena lortuko dugu. Hemen ikusi behar duguna zera da: ea izartxoz eginda dagoen irudia eta puntuz marratuta dagoen kurba elkarrekiko hurbil dauden ala ez. Beraz ez da metodo estatistikoa; baizik eta baldintza hau betetzen den jakiteko lehen hurbilketa.

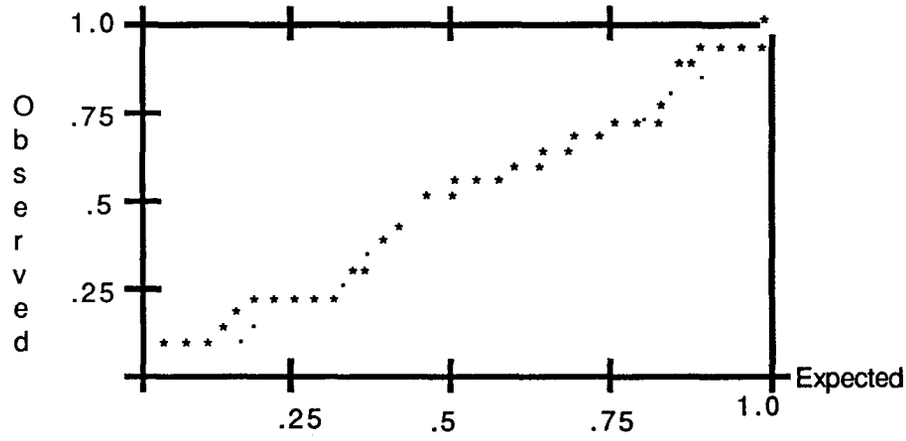
Gure irudian ikusten dugunez, lortzen den hurbilketa nahikoa ona da:

```
Histogram - Studentized Residual
NExp  N     (* = 1 Cases,   . : = Normal Curve)
 0   .09      Out
 0   .19     3.00
 0   .47     2.67
 0   1.08    2.33  .
 5   2.21    2.00  *:***
 2   4.05    1.67  ** .
 6   6.64    1.33  ******.
 *   9.76    1.00  *********:*****
 *   12.9     .67  ************.
 *   15.2     .33  ************* .
 *   16.0     0.0  ***************:
 *   15.2    -3.3  **************:
 *   12.9    -.67  ************:*
 5   9.76   -1.00  *******  .
 9   6.64   -1.33  ******:
 4   4.05   -1.67  ***:
 4   2.21   -2.00  *:**
 1   1.08   -2.33  :
 0   .47    -2.67
 0   .19    -3.00
 0   .09      Out
```

NORMPROB eskatu dugulako, ondoko beste irudi hau lortuko dugu. Metodo honekin erabili ditugun aldagaiak, normaltasunean baldintza betetzen duten ala ez jakiteko ere erabiliko dugu. Hau ere ez da metodo estatistikoa, baina askotan lagungarria izango da. Lehen esan dugun bezala, analisi estatistiko baten bidez baldintza hau betetzen den konprobatzeko, K-S test ez-parametrikoa erabili beharko genuke.

Lortu dugun irudia ondokoa da, eta hemen laukiaren diagonalak banaketa normal teorikoa adieraziko du, eta izartxoek, guk lortu ditugun hondarren adierazpena. Beraz gure hondarrak zenbat eta diagonaletik hurbilago egon, hainbat eta normalagoak izango dira. Ikusten dugunez, lortu dugun hurbiltasuna nahikoa ona da.

Normal Probability (P-P) Plot
Standardized Residual



Datu aberrante edo "outliers"ak egoteak edo izateak, erregresio-ekuazioaren distorsio handia sor dezake. OUTLIER bat definitzerakoan, idazle ezberdinak ez dira ados jartzen, baina esan dezakeguna zera da: hodei-puntutik at gelditzen diren puntuak direla. Hau da puntu gehienak dauden tokitik nahikoa desbideraturik gelditzen direla. Kaleko hizkuntzan esaten diren pertsona edo elementu arraroak, edo normatik kanpo daudenak. Garrantzi handia edukiko du pertsona edo elementu hauek zeintzuk diren jakiteak. Horretarako ondoko agindua erabili dugu:

RESIDUALS=OUTLIERS(MAHAL) ID(ZENBAKIA).

Eta lortu ditugun emaitzak hauek izan dira:

Case #	ZENBAKIA	*MAHAL
83	83	8.52928
67	67	7.92002
64	64	7.60497
58	58	7.01695
123	123	6.92560
103	103	6.92560
63	63	6.92560
43	43	6.79993
28	28	6.46445
118	118	6.30587

Aurreko laukian ikusten dugunez, aldagaien artean dauden harremanak aztertu ondoren, harreman arruntetik kanpora edo urrunago gelditzen diren elementuak hor agertzen diren zenbakidunak dira. Honek zera esan nahi du: elementu hauek, beren aldagaien balioetan ez dituztela betetzen beste elementu guztiek betetzen dituzten erlazioak. Estatistikaren lana hemen bukatzen da. Beste kasu askotan bezala ikertzailea izango da elementu hauekin zer egin behar den erabaki beharko duena.

Beraz ikusi dugunez, erregresio-metodoa asko erabiltzen den metodoa da, eta askotan oso emaitza interesgarriak lortzeko bide egokia izango da, baina kontutan eduki beharko dugu metodo hau erabiltzeko baldintza batzuk bete behar direla, eta baldintza hauek betetzen ez badira, lortuko ditugun emaitzak kontu handiz erabili beharko ditugula.

17. GAIA
TEST EZ-PARAMETRIKOAK. NPAR-TESTS

Test ez-parametriko izenaz, parametrikoak erabiltzeko bete behar diren aplikazio-baldintzak bete behar ez dituen test pilo bat ezagutzen da. Bi eratako frogen arteko ezberdintasun nagusiena, zera da: test parametrikoak aplikatzerakoan aldagaien normaltasuna eskatzen da eta beraz baldintza hau test ez-parametrikoak erabiltzeko ez da bete behar.

Aldagaien normaltasunetan ez da izango bete behar den aplikazio-baldintza bakarra. Horretatik aparte, test parametriko bakoitza aplikatzeko berari dagokion beste baldintza batzuk bete beharko dituzte erabiliko ditugun aldagaiek. Batzuetan, baldintza hauek oso gogorrak izango dira, eta nahikoa zaila izango da denak betetzea. Hala ere, esan beharra dago baldintza guztiak betetzen ez badira ere, batzuetan behintzat, lortuko ditugun emaitzak baliagarriak izango direla. Aplikazio-baldintzak betetzen diren ala ez jakiteko askotan test ez-parametrikoez baliatuko gara.

Beraz, badakigu, gutxienez bi kasu ezberdinetarako test ez-parametrikoak baliagarriak izango direla:

- Test parametrikoak erabiltzeko bete behar diren aplikazio-baldintzak betetzen ez badira, test hauek, alternatiba egokiak izango dira.

- Test parametrikoak erabiltzeko bete behar diren baldintzak egiaztatzeko.

Test hauek, denbora askoan, test parametrikoen ahizpa zakarrak izan dira, baina azken bolada honetan eta batez ere ordenadoreak eta pakete estatistikoak oso erabilgarri bihurtu direnetik, test hauek oso garrantzi handia hartu dute, eta ondorioz asko erabiltzen dira. Aldaketa honen arrazoiak, 1968an Bradley jaunak ematen zituen, eta gaur egun ere ontzat har ditzakegu, nahiz eta batzuk apur bat desfasatuta gelditu (batez ere programa-paketeak agertu direnetik).

1.- DEDUKZIOAREN SINPLETASUNA. Test parametrikoak ulertu eta erabiltzeko behar diren ezagupen matematikoak, batzuetan oso astunak izango dira. Test ez-parametrikoen atzetik dagoen teoria matematikoa askoz ere errazagoa da.

2.- ERABILTZEKO SAMURTASUN ETA ERRAZTASUNA. Askotan test ez-parametrikoak erabiltzeko egin behar diren eragiketak hauek izango dira: datuak ordenatu, eta gero batuketak eta kenketak. (Arrazoi hau adibidez, gaur egun, eta liburu honen ikuspegitik batez ere, nahikoa "pasatuta" dago, baina hala ere eta, oraindik pertsona asko ordenadoreen beldur dela kontutan harturik, hemen mantendu egin dut).

3.- APLIKAZIO-ARLOAK. Test ez-parametrikoak erabiltzeko bete behar diren aplikazio--baldintzak gutxiago eta betetzen errazagoak izango dira. (Hau izan daiteke gaur egun test hauek erabiltzeko edukiko dugun arrazoi garrantzitsuenetako bat; batzuetan aldagai-motagatik eta beste batzuetan test parametrikoak erabiltzeko aldagai hauek bete behar dituzten baldintzengatik. Bestalde, batez ere ikerketa-motagatik, askotan ezinezkoa izango da lagin handiak lortzea, eta esan beharra dago lagin-kopurua N=10 baino txikiagoa denean ez dagoela test parametrikorik erabiltzerik (populazioko banaketa ezaguna ez denean behintzat).

4.- APLIKAZIO-BALDINTZAK BETETZEN EZ DIRENEAN. Kasu honetan esan beharra dago, alde batetik aplikazio-baldintzak oso arruntak direnez gero, betetzen diren ala ez konprobatzea askoz ere errazagoa izango dela. Bestalde, baldintza hauek betetzen ez badira, gehienetan beste test ez-parametriko bat erabiltzeko posibilitatea edukiko dugu (parametrikoetan gertatzen ez dena).

5.- BEHAR DIREN NEURKETA-MOTAK. Test ez-parametrikoak erabiltzeko, nahiz eta tartezko datuentzat ere baliagarriak izan, nahikoa izango da gure datuak neurri ordinal edo kualitatibo baten bidez neurtuak izatea. Test parametrikoak erabiltzeko, arrazoi edo tartezko eskala batean neurtuak egon beharko dute kasu gehienetan.

6.- LAGIN-TAMAINA. Lagin-tamaina 10 baino txikiagoa denean, alde batetik test parametrikoak aplikatzea nahikoa arriskutsua izango da, eta bestalde, edozein aplikazio--baldintza betetzen ez bada, test parametrikoak erabiliz lortuko ditugun emaitzak ez dira baliagarriak izango. Lagina handitzen joaten denean, test parametrikoak gero eta baliagarriagoak izango dira, eta beraz, ez-parametrikoek daukaten garrantzia txikiagotzen joango da.

Aurreko arrazoiak, beste batekin burutuko ditugu, eta hau kontzeptu berri batean oinarritzen da: TEST BATEN AHALMEN edo POTENTZIAn. 12. gaian ikusten genuenez, edozein test estatistiko aplikatzerakoan bi eratako erroreak edukitzeko arriskua daukagu: I ERROREA, hau da, oinarrizko hipotesia egia denean ezeztatzea eta II ERROREA, oinarrizko hipotesia gezurra denean onartzea. I ERROREA edukitzeko probabilitatea α letraz izendatzen genuen eta II ERROREA edukitzeko probabilitatea β letraz.

Test baten AHALMENA, π letraz izendatuko duguna, test honen oinarrizko hipotesia gezurra denean ezeztatzeko daukagun probabilitatea izango da.

Beraz aurreko definiziotik, ondoko berdintasuna ezar dezakegu:

$$\pi = 1 - \alpha$$

Interesatuko zaizkigun froga estatistikoak, baldintza hau betetzen dutenak izango dira: ezarri dugun α probabilitatea baino handiagoa ez dutenak eta honekin batera β probabilitatea txikiena daukatenak, edo beste era batera esanda, I erroreari ezarri diogun muga gainditzen ez duten testetan, testaren AHALMEN handiena daukana. Logikoki test baten AHALMENA test horren prozedurarekin erlazionatuta egongo da.

Aurreko kontzeptu berri honetatik test ez-parametrikoak erabiltzeko 7. arrazoia aterako dugu:

7.- AHALMEN ESTATISTIKOA. Test ez-parametrikoek gutxienez parametrikoek beste ahalmen edukiko dute, hauen baldintzak betetzen ez direnean.

Beraz aurreko arrazoi guztiak ikusi ondoren, laburpen gisa hau esan dezakegu:

1.- Test parametrikoen aplikazio-baldintzak betetzen direnean, hauek ez-parametrikoak baino hobeak izango dira.

2.- Test parametrikoen baldintzak betetzen ez badira, BETI HOBE izango da test ez-parametrikoak erabiltzea.

3.- Test parametrikoak erabiltzeko bete behar diren aplikazio-baldintzak aztertzeko, test ez-parametrikoak erabiliko dira.

Beraz aurreko arrazoiak ikusi ondoren, test ez-parametriko ezberdinak aztertzera pasatuko gara.

Test hauek aurkezterakoan eramango dugun ordena, SPSS/PC+ paketeak eramaten duenari jarraituko zaio. Egingo dugun sailkapena, ondoko laukian agertzen dena, aldagai-mota eta ikerketaren diseinuari begiratuz izango da.

ESPERIMENTUEN DISEINUA	ESKALA NOMINALA	ESKALA ORDINALA
LAGIN BAT	CHI-SQUARE RUNS (SEGIDAK) BINOMINAL	KOLMOGOROV-SMIRNOV
BI LAGIN ERLAZIONATUTA	Mc NEMAR	SIGN (ZEINUAK) WILCOXON
K LAGIN ERLAZIONATUTA	COCHRAN-en Q	FRIEDMAN KENDALL
BI LAGIN ASKE EDO INDEPENDENTE		MEDIAN MANN-WITNEY KOLMOGOROV-SMIRNOV WALD-WOLFOWITZ MOSES
K LAGIN ASKE EDO INDEPENDENTE		MEDIAN KRUSKAL-WALLIS

Aurreko laukian egin dugun sailkapena ikusirik, test bat aukeratzerakoan bi gauza kontutan eduki beharko ditugu:

- Aldagai-mota

- Esperimentuaren diseinua. Hau esaterakoan, zera esan nahi dugu: zein eratako hipotesia baieztatu nahi dugun kontutan eduki behar dugula. Gure hipotesian, bi aldagai elkarrekin erlazionatuta baldin baditugu, SIGN edo WILCOXONen frogak erabili beharko ditugu, noski.

Lagin bati dagozkion hipotesiak baieztatzeko erabil ditzakegun test ez-parametrikoak ikusten hasiko gara.

1. LAGIN BATI DAGOKION INFERENTZIA

Aztertu nahi dugun aldagai-mota kontutan harturik, bi zatitan banatuko ditugu test hauek. Gure aldagaia eskala nominal batean neurtuta baldin badaukagu, hiru froga erabil ditzakegu: CHI-SQUARE, RUNS eta BINOMIAL. Eskala ordinal batean neurtuta baldin baditugu, KOLMOGOROV-SMIRNOVen froga erabili beharko dugu. Kasu bakoitzean, froga ezberdinak hipotesi ezberdinak baieztatzeko erabiliko ditugu.

1.1. KOLMOGOROV-SMIRNOVen froga. Aldagai ordinalak

Gure aldagaian dauzkagun datuak banaketa teoriko bati noraino hurbiltzen zaizkion

ikusteko erabiliko dugu froga hau. Beraz, froga hau aplikatu aurretik, banaketa teorikoa zein izango den erabakita eduki beharko dugu.

SPSS/PC+ paketeak, hiru banaketa teorikotarako hurbiltasunak aztertzeko posibilitatea ematen digu: POISSON, NORMALA eta UNIFORMEA. Banaketa bakoitzean, nahi izanez gero, ondoko parametroak ezar ditzakegu.

NORMALA : batezbestekoa eta desbidazio estandarda.
POISSON : batezbestekoa.
UNIFORMEA : minimoa eta maximoa.

K-S testa erabiltzeko idatzi behar dugun instrukzioa honelakoa izango da:

NPAR TESTS K-S (banaketa) = aldagaia.

Gure datuak 100 eta 15 parametroak dituen banaketa normalari egokitzen zaizkion ala ez erabakitzeko honela idatzi beharko dugu:

NPAR TESTS K-S (NORMAL,100,15)=ALDAGAIA.

Adibidez, liburuan zehar aztertzen ari garen INTELIGE aldagaien datuak banaketa normalari jarraitzen zaizkion ala ez erabaki nahi badugu, honela idatz dezakegu:

NPAR TESTS K-S (NORMAL) = INTELIGE.

Kasu honetan, gure datuak INTELIGE aldagaien lagineko batezbestekoa eta desbiderapen standarda duen banaketa normalari hurbiltzen zaion ala ez erabakitzeko erabili dugu K-S testa.

Oinarrizko hipotesia ondo egokitzen dela izango da, hau da, banaketa teoriko eta lagineko datuen artean dauden ezberdintasunak zoriz lortutakoak direla izango da.

Lortu ditugun emaitzak hauek dira:

```
- - - - - Kolmogorov - Smirnov Goodness of Fit Test
    INTELIGE ADIMEN-KOEFIZIENTEA

Test Distribution  -  Normal                    Mean:  100.48
                                   Standard Deviation:    5.48
          Cases :   121
          Most Extreme Differences
 Absolute     Positive     Negative     K-S Z      2-tailed P
  .09327      .08855      -.09327      1.026        .243
```

Ikusten dugunez, gure datuak, 100.48 batezbestekoa eta 5.48 desbidazio standarda duen banaketa normalari egokitzen zaizkion aztertu dugu.

Balio hauek aurkeztu ondoren, zenbat elementu dauzkagun adierazten du, eta azken zatian hipotesi-testaren kalkuluak agertzen ditu.

Froga honetan, gure datuak ia banaketa normal bati egokitzen zaizkion eztabaidatzen dugu. Horretarako egiten duena zera da: Banaketa teoriko eta ikusitako maiztasunen arteko ezberdintasunak kalkulatu eta hemendik K-S-ren Z kalkulatu ondoren balio hau zoriz lortzeko probabilitatea adierazi. Ez gara hemen pauso guztiak aztertzen hasiko. Azken emaitza bakarrik ikusiko dugu. Gure kasuan, lagin eta banaketa teorikoen artean dauden ezberdintasunak zoriz lortzeko dagoen probabilitatea .243 da, beraz, gure gehieneko errore bezala ohizkoak diren: .05, .01 edo .001 hartuta, gure erabakia hau izango da: Dauden ezberdintasunak zoriz lortutakoak dira, eta gure datuak banaketa normalari N(100.48, 5.48) jarraitzen zaizkiola ezin dugu ezeztatu.

1.2. CHISQUARE froga. Aldagai nominalak (kualitatiboak)

Aldagai kualitatibo edo nominalak direnean, gure lagineko datuak aurretik ezarritako maiztasun teoriko batzuetara egokitzen diren ala ez baieztatzeko, Ji karratuaren testa erabiliko dugu. Froga hau egiteko SPSS/PC+ paketeak CHISQUARE azpiprograma eskaintzen digu.

Instrukzio orokorra hau izango da:

NPAR TESTS CHISQUARE=ALDAGAIA/EXPECTED=f_1,f_2,.....,f_k.

Non f_1, f_2, ..., f_k aldagaien atributu bakoitzentzat espero diren maiztasun erlatiboak adierazteko zenbakiak diren. Zenbaki hauek frakzio batzuen zenbakitzaileak dira. Beraz honelako agindu bat idazten badugu:

NPAR TESTS CHISQUARE=ALDAGAIA/EXPECTED=1,2,3,4.

Kasu honetan ALDAGAIA kualitatiboa da, eta nik espero ditudan maiztasun erlatiboak: 1/10, 2/10, 3/10, eta 4/10 izango dira. Zatitzailea, zenbakitzaileen batuketa eginda lortu dut.

Adibide gisa, har dezagun ea datuak bildu ditugun herrietan SEXUA aldagaiari dagozkion maiztasunak, gizonezkoenak eta emakumezkoenak, berdinak diren ala ez. Hau ikusteko idatzi behar dugun agindua ondokoa izango da:

NPAR TESTS CHISQUARE=SEXUA/EXPECTED=1,1.

Lortu ditugun emaitzak hauek izan dira:

– – – – – Chi-square Test

SEXUA

	Cases			
	Category	Observed	Expected	Residual
EMAKUMEA	1	68	62.50	5.50
GIZONA	2	57	62.50	-5.50
	Total		125	

Chi-Square	D.F.	Significance
.968	1	.325

Aurreko emaitzak Ji karratuaren frogan lortzen genituenen antzekoak dira. Oinarrizko hipotesia, dauden ezberdintasunak zoriz lortutakoak direla izango da. Puntu honetatik abiatuz, kalkulu guztiak egiten ditu. Ikusitako maiztasunak, 68 eta 57 dira. Oinarrizko hipotesia betetzen bada espero diren maiztasunak 62.50 eta 62.50 izango dira. Beraz dauden ezberdintasunak, 5.50 eta -5.50. Balore hauetatik lagineko Ji karratuaren balioa lortzen du: .968, eta kasu honetako askatasun-graduak 1 (atributuak - 1) direla kontutan harturik, lagineko Ji karratuaren balioak zoriz lortzeko dagoen probabilitatea .325 dela adierazten du. Beraz emaitza hau ikusi ondoren gure erabakia hau izango da: ikusitako maiztasun eta espero diren maiztasunen artean dauden ezberdintasunak zoriz lortutakoak direla ezin dugu ukatu.

1.3. RUNS froga. Eskala nominalak (kualitatiboak)

RUNS edo segidaren froga, gure datuak zoriz aukeratutako lagin batetik lortutakoak diren ala ez erabakitzeko oso froga interesgarria izango da. Garbi dago gure datuak ordenadorera sartzen ditugunean ez ditugula zoriz sartuko, zeren eta gauza normala izango bait da adibidez ikastola bakoitzeko datuak elkarrekin edukitzea eta beraz denak batera sartzea. Hala ere askotan interesgarria izan daiteke aldagai bateko datuak zoriz sartuta dauden ala ez jakitea. Honetarako froga hau erabiliko dugu.

Froga hau erabiltzeko idatzi behar dugun instrukzioa honelakoa izango da:

NPAR TESTS RUNS(ebaki-puntua) = ALDAGAIA.

Frogaren planteamendu orokorra, hau izango da. Lehenengo pausoan gure aldagaia dikotomizatu egingo dugu. Horretarako EBAKI-PUNTU bat aukeratu beharko dugu. Puntu honetaz baliatuz orain aldagaien balioak 0 edo 1 izango dira, segun eta aldagaien ba-

lioa ebaki-puntua baino txikiagoa berdina edo handiagoa den.

Gure adibiderako, MAILASOZ aldagaia hartu dugu, eta ebaki-puntu bezala 3 balioa hartu dugu. Froga hau egiteko idatzi dugun agindua hau izan da:

NPAR TESTS RUNS(3) = MAILASOZ.

Lortu ditugun emaitzak hauek izan dira:

```
- - - - - Runs Test
  MAILASOZ MAILA SOZIOEKONOMIKOA

     Runs:      98              Test Value = 3
     Cases:     64    Lt 3
                59    Ge 3                Z = 6.4574
               123    Total  2-tailed     P = .0000
```

Emaitzak aztertzerakoan zera ikusten dugu: lortu dugun segida-kopurua 98 dela eta ebaki-puntua (TEST VALUE) 3 hartu dugula. Aldagai honetan 3 balioa baino balio txikiagoak hartzen dituztenen kopurua 64 dela esaten du, eta 3 edo balio handiagoak hartzen dituztenak 59. Balio hauek banaketa normal batera bihurtzen ditu, eta lagineko emaitza hauei banaketa normal tipikoan dagokion balioa 6.4574 denez geroz, balio hau zoriz lortzeko dagoen probabilitatea .0000 dela adierazten du.

Emaitza hauek ikusi ondoren, gure erabakia hau izango da: Emaitza hauek zoriz lortzeko dagoen probabilitatea .0000 denez gero, gure erabakia oinarrizko hipotesia ez dela betetzen izango da. Beraz gure datuak ez daude zoriz sartuta.

1.4. BINOMIAL froga. Aldagai dikotomikoak

Froga hau aldagai dikotomikoak direnean erabiliko dugu, eta CHISQUARE frogaren errepikapen bat bezala kontsidera dezakegu.

Froga honen oinarria, bere izenak adierazten duen bezala, banaketa binomialean dago. Beraz arrakasta lortzeko dagoen probabilitatea adierazi beharko dugu. Planteamendu orokorra hau da. Eman dezagun txanpon bat 100 aldiz airera botatzen dugula. Oinarrizko hipotesi bezala txanpona ondo eginda dagoela onartzen badugu, espero diren maiztasunak 50 aurpegi eta 50 gurutze izango dira. Gure esperimentuan lortu ditugun maiztasunak 55 eta 45 izan badira, froga honen bidez lortuko dugun probabilitatea, emaitza hauek zoriz lortzeko dagoena izango da.

Froga hau erabiltzeko ondoko aginduen bezalakoa idatzi beharko dugu:

NPAR TESTS BINOMIAL(.4) = SEXUA(1,2).

Aurreko aginduan oinarrizko hipotesia betetzen bada, arrakasta (aldagaien lehenengo balioa) lortzeko dagoen probabilitatea .4 dela esaten diogu, eta SEXUA aldagaiak, 1 eta 2 balioak hartzen dituela.

Probabilitaterik ezartzen ez badiogu, programak, probabilitate hau .5 dela suposatuko du. Beraz guk erabili dugun agindua hau izan da:

NPAR TESTS BINOMIAL = SEXUA(1,2).

Lortu ditugun emaitzak hauek izan dira:

```
- - - - - Binomial Test
   SEXUA

      Cases
                                  Test Prop. =     .5000
                 64    = 1        Obs. Prop. =     .5440
                 57    = 2        Z Approximation
                123    Total      2-tailed  P =    .3711
```

Ikusten dugunez, emaitzen aurkezpena oso laburra da. Probabilitate teorikoa .5000 dela eta ikusitako maiztasun erlatiboa .5440 dela adierazi ondoren, ezberdintasun hauek zoriz lortzeko dagoen probabilitatea eskaintzen digu. Probabilitate hau .3711 denez gero, gure erabakia hau izango da: Oinarrizko hipotesia, ezberdintasun hauek zoriz lortutakoak direla, ezin dugu ukatu.

2. BI LAGINETARAKO INFERENTZIAK

Bi lagineko batezbestekoen arteko diferentziak aztertzeko erabiltzen den ohizko metodoa T-testarena da. Metodo hau, bai lagin askeak direnean bai erlazionatuak direnean erabiltzen da. Baina test parametrikoa den T-testa erabiltzeko, ondoko aplikazio baldintzak bete behar dira:

- Elementu ezberdinen neurketek askeak izan behar dute. Lagin-aukeraketak zorizkoa izan behar du, eta elementu bakoitzaren puntuazioak bestearekiko askea izan behar du.

- Lagin biei dagozkien populazio-banaketek, normalak izan behar dute.

SPSS/PC+ paketeak zuzenketa baten bidez konpontzen du.

- Aztertu nahi dugun aldagaiak jarraia izan behar du. Gutxienez tartezko eskala batean neurtuta egon behar du.

Gauza nabarmena da baldintza hauek oso kasu gutxitan konprobatzen direla. Honen ondorioz, askotan, lortzen diren emaitzak nahikoa eztabaidagarriak izango dira.

Aplikazio-baldintza hauek betetzen ez direnean, oso interesgarriak diren froga ez-parametrikoak atal honetan aurkeztuko ditugu.

2.1. Bi lagin askerentzako froga ez-parametrikoak

SPSS/PC+ paketeak, bi lagin askeren emaitzak konparatzeko bost froga ezberdin erabiltzeko posibilitatea eskaintzen digu. Hauetatik hiru bakarrik ikusiko ditugu (hain zuzen AHALMEN handiena daukatenak) eta ondoko hauek izango dira: MEDIANA, MANN-WHITNEY eta KOLMOGOROV-SMIRNOVen frogak.

Froga hauek, lagin handi zein txikiekin erabiltzeko baliagarriak izango dira. Froga bakoitzak erabiltzen duen metodoa hobeto ikusteko, PUNTUAK izeneko aldagaien asmatutako ondoko datuak erabili ditugu:

1 TALDEA : 16, 12, 13

2 TALDEA : 14, 11, 8, 10

Bi talde hauek askeak direla suposatuko dugu.

2.1.1. MEDIANAREN froga. Eskala ordinalak

Froga honen helburua zera da: dauzkagun bi laginak mediana berdinak dituzten bi populaziokoak diren ala ez erabakitzea. Gure datuek, gutxienez, eskala ordinal batean neurtuak egon beharko dute.

Froga hau erabiltzen dugunean, oinarrizko hipotesia populazio bien medianak berdinak direla izango da, eta erabiliko duen metodoa hau izango da: lehenengo pausoan lagin biak batu ondoren, puntuazio denen mediana kalkulatzen da, hau da, gure kasuko bi lagin txikietan lortuko dugun mediana 12 izango da. Balio hau kalkulatu ondoren bi sarrerako taula bat osatzen da; alde batetik talde biak ezarriz, eta bestetik kalkulatu dugun medianarena baino puntuazio handiagoak direnak eta txikiagoak direnak ezarriz. Bi sarrerako taula hau osatu ondoren, datu-kopurua txikia denean (Fisher-en froga erabiliz) taula hau zoriz

lortzeko dagoen probabilitatea kalkulatzen da. Datu-kopurua handia denean, Ji karratuaren testa erabiliz probabilitate berdina kalkulatzen da.

Froga hau erabiltzeko ondoko agindua idatzi dugu:

NPAR TESTS MEDIAN = PUNTUAK BY TALDE(1,2).

Lortu ditugun emaitzak hauek izan dira:

– – – – – Median Test

TEST
by TALDE

		TALDE 1	TALDE 2
TEST	Gt Median	1	1
	Le Median	1	2

Cases	Median	Exact Probability
7	12	.4857

Orokorrean 20 elementu baino gutxiago dauzkagunez, probabilitatea kalkulatzeko Fisher-en froga erabili du.

Emaitzen interpretazioa oso erraza da. Aipatutako bi sarrerako taula eraiki ondoren, eta 7 kasu direla, eta mediana 12 dela adierazi ondoren, gure taulan dauzkagun maiztasunak zoriz lortzeko dagoen probabilitatea adierazten du. Gure kasuan probabilitate hau .4857 da. Beraz hemendik ezingo dugu oinarrizko hipotesia ezeztatu. Hau da, lagin biak mediana berdina duten populaziotik datozela ezingo dugu ukatu.

Froga berdina, liburuko adibide orokorren TELEBIS1 aldagaiari aplikatzen badiogu, SEXUA aldagaiak egiten dituen bi taldeetan, ondoko instrukzioa idatzi beharko dugu:

NPAR TESTS MEDIAN =TELEBIS1 BY SEXUA(1,2).

Lortu ditugun emaitzak hauek izan dira:

– – – – – Median Test

TELEBIS1 TELEBISTA AURREAN ORDUAK LEHENENGO NEURK
by SEXUA

		SEXUA 1	SEXUA 2
TEST	Gt Median	37	24
	Le Median	31	31

Cases	Median	Chi-Square	Significance
123	17	1.0141	.3139

Ikusten dugunez, medianaren balioa (17) kalkulatu ondoren bi sarrerako taula eraiki du, eta hemengo emaitzak Ji karratuaren testaren bidez aztertu ditu. Lagineko X^2 balioa 1.0141 da eta balio hau zoriz lortzeko dagoen probabilitatea .3139. Hemendik gizon eta emakumeek aldagai honetan dauzkaten medianak berdinak direla erabaki dezakegu.

2.1.2. MANN-WHITNEYen U froga. Eskala ordinalak

Ez-parametrikoak diren frogetatik, ahalmen handienetako bat izango da eta, KOLMOGOROV-SMIRNOVen frogarekin batera, T-testaren ordez erabiltzeko oso baliagarria izango da.

Gure datuek, gutxienez eskala ordinal batean neurtuak egon beharko dute.

Erabiliko den prozedura laginen tamainaren arabera aldatu egingo da. Talde bakoitzeko elementu-kopurua 8 baino txikiagoa baldin bada, formula batzuk erabiliko dira. Besteetan, elementu-kopurua 8 baino handiagoa denean beste batzuk erabiliko dira.

Talde bakoitzeko elementu kopurua 8 baino txikiagoa den kasuarekin hasiko gara.

Froga hau erabiltzeko idatzi behar dugun aginduak era honetakoa izan beharko du:

NPAR TESTS M-W= Menpeko aldag. BY aldag.askea(txik.,hand.).

Gure PUNTUAK aldagaiaren 7 neurketak aztertzeko, ondoko agindua erabili dugu:

NPAR TESTS M-W = PUNTUAK BY TALDE(1,2).

Lortu ditugun emaitzak hauek izan dira:

```
- - - - -  Mann-Whitney  U - Wilcoxon Rank Sum W Test

      PUNTUAK
   by TALDE

       Mean Rank       Cases
            5.33          3 TALDE  =  1
            3.00          4 TALDE  =  2
                          -
                          7 Total

                         EXACT              Corrected for Ties
   U          W         2-tailed P          Z          2-tailed P
   2.0       16.0         .2286            -1.4142        .1573
```

Emaitzak aztertuz, erabili duen prozedura aztertuko dugu:

1.- Gure 7 datuak ordenatuko ditugu, zein taldetakoak diren adieraziz:

ORDENA :	1	2	3	4	5	6	7
PUNTUAK :	8	10	11	12	13	14	16
TALDE :	2	2	2	1	1	2	1

1. Taldeko ordenen batura: 4+5+7 = 16 batezbestekoa = 5.33

2. Taldeko ordenen batura: 1+2+3+6 = 12 batezbestekoa = 3

U estatistikoen kalkulua ondoko formula erabiliz kalkulatuko dugu:

$$U = N_1 * N_2 + \frac{N_1 \ (N_1 + 1)}{2} - R_1$$

Non R_1, lehenengo taldeko ordenen batura den, eta N_1 eta N_2 talde bien lagin-kopuruak.

Gure kasuan:

$$U = 3 * 4 + \frac{3\,(3+1)}{2} - 16 = 2$$

W, 1.taldeko elementuen ordenen batura. Gure kasuan 16.

Beraz balio guzti hauek lortu ondoren, balio hauek zoriz lortzeko dagoen probabilitatea aurkezten du. Probabilitate hauek taula batzuen bidez lortzen dira. Inork taula hauek ikusi nahiko balu, SIDNEY-SIEGEL-en "ESTADISTICA NO PARAMETRICA" liburuan aurki dezake. Liburu hau TRILLAS argitaletxeak argitaratu du. Beraz gure probabilitatea .2286 da, eta aldamenean, banaketa normalaren bidez probabilitate honi buruz egin duen hurbiltasuna dago. Kasu honetan balio hau ez dugu kontutan hartu behar.

Lortu dugun probabilitatea .2286 denez geroz, ezingo dugu oinarrizko hipotesia ezeztatu. Beraz gure erabakia, talde bien emaitzak berdinak direla ez ezeztatzea izango da.

Talde bakoitzeko elementu-kopurua handiagoa denean, probabilitateak beste era batera kalkulatzen dira. Gure adibide orokorrean daukagun TELEBIS1en emaitzak gizon eta emakumeen taldeetan berdinak diren ala ez ikusteko, ondoko agindua idatzi dugu:

NPAR TESTS M-W = TELEBIS1 BY SEXUA(1,2).

Lortu ditugun emaitzak hauek izan dira:

```
- - - - -  Mann-Whitney  U - Wilcoxon Rank Sum W Test

     TELEBIS1    TELEBISTA AURREAN ORDUAK LEHENENGO NEURK
  by  SEXUA

      Mean  Rank       Cases
            64.28         68  SEXUA  =  1   EMAKUMEA
            59.18         55  SEXUA  =  2   GIZONA
                         ---
                         123  Total

                                                Corrected for Ties
          U                W                  Z       2-tailed  P
       1715.0           3255.0             -.7921        .4283
```

Kasu honetan, aurreko emaitzetan agertzen zen EXACT 2-TAILED P, ez da agertzen. Beraz probabilitatearen kalkulua, banaketa normalaren bidez egiten du, aurreko pauso guztiak berdinak direlarik. Gure kasuan, U eta W estatistikoen balioak, zoriz lortzeko dagoen probabilitatea .4283 denez gero, talde bien batezbestekoak berdinak direla erabakiko dugu, edo gutxienez, ezingo dugu ezberdinak direla esan.

2.1.3. KOLMOGOROV-SMIRNOVen froga. Eskala ordinalak.

Bi lagin asketan lortu ditugun emaitzak konparatzeko baliagarria den froga da. Lehen, izen bereko beste froga bat ikusi dugu. Kasu hartan lortutako emaitzak ea banaketa teoriko batera egokitzen ziren ala ez ikusi dugu. Kasu honetan ematen dituen pausoak eta hangoak nahikoa berdinak dira.

Froga hau erabiltzeko, konparatu nahi ditugun aldagaien emaitzek gutxienez eskala ordinal batean neurtuak egon beharko dute.

Oinarrizko hipotesia, talde bien emaitzak , batezbestekoak eta bariantzak, berdinak izatea izango da.

Kasu honetan aztertuko diren ezberdintasunak, talde bien arteko maiztasun metatuenak izango dira, eta ezberdintasun hauetan oinarrituz, puntu ezberdinetako maiztasun metatuen arteko gehieneko distantzia absolutua, D_m izendatutakoa kalkulatzen da. D_m honela definituko dugu:

D_m = distantzia maximoa $[F_{n1}(x) - F_{n2}(x)]$

D_m kalkulatu ondoren, hartu duen balioa zoriz lortzeko dagoen probabilitatea, banaketa normala erabiliz, kalkulatzen da.

Hau da SPSS/PC+ paketeak erabiltzen duen prozedura. Eta hau egiteko ondoko baliokidetasuna erabiltzen du:

$$4 D_m^2 \frac{N_1 * N_2}{N_1 + N_2}$$ 2 askatasun-graduko X^2-ari jarraitzen zaio.

Atal honetan aztertzen ari garen 7 elementuko bi taldeen emaitzak berdinak diren ikusteko idatzi dugun agindua hau da:

NPAR TESTS K-S = PUNTUAK BY TALDE(1,2).

Lortu ditugun emaitzak hauek izan dira:

```
- - - - - Kolmogorov - Smirnov 2-Sample Test
      TEST
   by TALDE

            Cases
               4    TALDE = 1
               3    TALDE = 2
              ---
               7    Total

                        Most Extreme Differences
      Absolute        Positive        Negative        K-S Z      2-tailed P
      .75000          .75000          .00000          .982          .290
```

ABSOLUTE izenarekin, D_m-ren balioa agertzen zaigu, hau da, maiztasun metatuen arteko ezberdintasun handiena. POSITIVE izango da ezberdintasun positiboen arteko handiena eta NEGATIVE alderantzizkoena. Azkenik, beti bezala, balio hau zoriz lortzeko dagoen probabilitatea ematen du, horretarako erabili duen Z-ren balioa erdian ezarririk. Beraz gure probabilitatea .290 denez geroz, talde bien arteko emaitzak ezin ditugu ezberdin bezala kontsideratu.

TELEBIS1 aldagaiarekin froga berdina egin ondoren, hau da, sexuek eratzen dituzten bi taldeko emaitzen arteko ezberdintasunak aztertzen baditugu, ondoko instrukzioa idatziko dugu:

NPAR TESTS K-S = TELEBIS1 BY SEXUA(1,2).

Emaitzak hauek izan dira:

```
- - - - - Kolmogorov - Smirnov 2-Sample Test

      TELEBIS1  TELEBISTA AURREAN ORDUAK LEHENENGO NEURK
   by SEXUA

            Cases
              68    TALDE = 1
              55    TALDE = 2
             ---
             123    Total

                        Most Extreme Differences
      Absolute        Positive        Negative        K-S Z      2-tailed P
      .27807          .08850          -.27807         1.533         .018
```

Aurreko kasuko pauso berdinak eman ondoren, Z balioa lortzen du: 1.533, eta balio hau zoriz lortzeko dagoen probabilitatea .018 dela adierazten du.

Kasu honetan, gure gehieneko errorea .05 baldin bada, aldagai honetan bi taldetan lortu ditugun emaitzak ezberdinak direla esan dezakegu, emaitza hauek zoriz lortzeko dagoen probabilitatea .05 baino txikiagoa delako. Ostera gure errore-marjinak .01 edo .001 baldin badira, ezingo dugu oinarrizko hipotesia ezeztatu.

Ikusi dugunez, froga estatistiko honek duen sentikortasuna aurrekoen aldian handia da.

2.2. ERLAZIONATUTAKO BI LAGINENTZAKO FROGA EZ-PARAMETRIKOAK

Hemendik aurrera, adibide berri bat aztertuko dugu. Hauek ere asmatutako datuak izango dira, baina aztertu behar ditugun frogak kontutan harturik ondoko ikerketa egiten ari garela suposatuko dugu. Eman dezagun, publizitate edo marketineko enpresa batek ondoko ikasketa bat egitea erabaki duela. Salmentan jarri den produktu berri bati buruz bi momentu ezberdinetan galdera ezberdinak dauzkaten bi inkesta pasatu ditu. Inkesta hauek, populaziotik zoriz aukeratu dituen lagin bateko 40 pertsonei pasatu dizkiete.

Neurtutako aldagaiak eta lortutako datuak, hemen azaltzen direnak izan dira:

11191011121111 22181212131111 33171413100101	1-2	zutabea	ZENBAKIA, pertsona bakoitzari ezarri diogun zenbakia.
41180910 80111 531507 9 60111	3	zutabea	ADINA. Eskala nominalean neurtuta
621611 7 70011 731406 8120010	4	zutabea	SEXUA
82131314141010 91151412171010 1011610 7 91010 11217 9 7131001	5	zutabea	LEHENA, lehenengo momentuan produktuari buruz inkestatuak zeukan eritzia.
12318 711120001 13116 912150101 142161113120101 153171310120111	6-7	zutabea	IRAGARTB, telebistako iragarkia ikusi ondoren produktuari buruz zeukan eritzia
1611510 9130111 1731610 7100111 1821412 8 80111 193251414 61110	8-9	zutabea	IRAGIRRA, irratiko iragarkia ikusi ondoren produktuari buruz zeukan eritzia.

202260912 71110 2112707 7120010 2212511 7140010 232260611170001 243241312 90001	10-11	zutabea	IRAGEGUN, egunkaritako iragarkia ikusi ondoren produktuari buruz zeukan eritzia.
251251413130001 262261010120011 27327 9 9150011 28123 7 7120010 29324 9 8121010	12	zutabea	IRITSITB, telebistan iragarkiak egotearen alde edo aurka dauden. Aldagai dikotomikoa 1: alde, 2 aurka.
302221114131010 313241312101010 3212510 7 80011 33126 8 7 60111 34124 4 8 70111	13	zutabea	IRITSIRR, irratian iragarkiak egotearen alde edo aurka dauden.Aldagai dikotomikoa. 1: alde, 2 aurka.
35123 7 9120111 363271211140111 3712815 8170011 38129 913 91011 393221115131110	14	zutabea	IRIEGUNK, egunkaritan iragarkiak egotearen alde edo aurka dauden. Aldagai dikotomikoa. 1:alde, 2 aurka.
401231312121110	15	zutabea	AZKENIRI, prozesu guztia bukatzerakoan produktua erosteko asmoa duen ala ez.

Atal honen izenburuan esan dugun bezala, bi lagin erlazionatutan lortutako emaitzak konparatuko ditugu. Lagin erlazionatuak kontsideratuko ditugu, alde batetik, pertsona berez osatutakoak, hau da, pertsona berei bi momentu ezberdinetan neurketak egin dizkiegunean. Hau da kasu arruntena, eta ziur asko era honetako frogen abiapuntua. Hala ere lagin erlazionatuak kontsideratuko ditugu ondoko era honetan lortuko ditugun emaitzak ere: Eman dezagun populazio bateko pertsonak binaka ezarri ditugula. Pare bakoitza ezaugarri berdinak dituzten pertsonez osatu dugu. Parekatze hau egin ondoren, laginketa egiterakoan ez ditugu pertsonak aukeratuko, baizik eta bikoteak, eta ondorengo pausoa bikote bakoitzeko pertsona bat talde batean ezarriko dugu eta bestea bigarren taldean. Honela aukeratutako taldeak ere erlazionatutzat kontsideratuko ditugu. Parekatze honen zergatia hau da: talde parekatuak aukeratu ondoren, adibidez bi tratamendu ezberdinen eragina konparatzeko, ezaugarri berdinak dituzten taldeak aukeratu ondoren talde bakoitzeko elementuei tratamendu bat ezarriko diegu, eta talde bakoitzean lortu ditugun emaitzak konpara ditzakegu.

2.2.1. MCNEMARen froga. Eskala nominala edo ordinala

McNemaren froga, aldaketen signifikazioen froga izenez ere ezagutua, tratamendu baten aurretik eta ondoren lortutako emaitzak konparatzeko erabiltzen da.

Froga hau aldagai nominal eta ordinalekin erabil dezakegu.

Planteamendu orokorra, X^2-aren testarena izango da. Beraz froga hau erabiltzeko ondoren aurkezten dugun bi sarrerako taula eraiki beharko dugu. Taula honetan 1.momentuko eta 2. momentuko emaitzak ikusi ondoren (alde edo aurka bi momentu ezberdinetan), per-tsona bakoitzak taulako lauki batean egon beharko du. Beraz ondoko taularen antzeko maiztasun-taula bat eraiki dezakegu. Taula honetan oinarrituz, X^2 aplikatzen da.

		2. MOMENTUAN	
		-	+
1. MOMENTUAN	+	A	B
	-	C	D

Erabiltzen diren formulak, hauek izango dira:

$$X^2 = \frac{(A - D)^2}{(A + D)} \qquad 1 \text{ ASKATASUN-GRADUEKIN}$$

Yates-en zuzenketa egiten denean, beste formula hau erabiliko dugu:

$$X^2 = \frac{(|A - D| - 1)^2}{(A + D)} \qquad 1 \text{ ASKATASUN-GRADUEKIN}$$

Adibide bezala, telebistako iragarki aurretik eta ondoren, gure produktuari buruzko eritzia konparatu nahi dugu.

Horretarako idatzi dugun agindua hau izan da:

NPAR TESTS MCNEMAR =IRITSITB WITH AZKENIRI.

Lortu ditugun emaitzak hauek izan dira:

– – – – – McNemar Test
IRITSITB
With AZKENIRI

		AZKENIRI 1	AZKENIRI 0		
IRITSITB	0	22	4	Cases	40
	1	4	10	Chi-Square	3.7813
				Significance	.0518

Ikusten dugunez, lehenengo pausoan taula eraiki du, eta gero lehen aipatu ditugun formulak erabiliz, kasu honetan, Yates-en zuzenketa egin behar izan du, Chi-Square balioa kalkulatu du, eta ondoren, balio hau zoriz lortzeko dagoen probabilitatea. Beraz gure kasuan, nahiz eta aldaketa asko egon, hauek zentzu bietan dira. X^2-aren balioa zoriz lortzeko dagoen probabilitatea .0518 denez gero, ezin dezakegu aldaketa hauek zentzu batean signifikatiboak direnik erabaki.

2.2.2. ZEINUEN froga. Eskala ordinalak.

Froga honen prozedura, hau da: egon diren aldaketak, eta aldaketa hauen zeinuak ikusten dira, hau da, zenbat positibo eta zenbat negatibo diren. Eta bi zenbaki hauen diferentzia zoriz lortutakoa izan daitekeen aztertzen da. Oinarrizko hipotesia, diferentzia hau 0 izaten izango da.

Diferentzien signifikazioa aztertzeko, SPSS/PC+ paketeak, lagin-kopurua 25 baino txikiagoa denean, banaketa binomiala erabiltzen du. Besteetan, lagin-kopurua 25 baino handiagoa denean, banaketa normala erabiliko du.

Kasu honetan adibide bezala, bi erabili ditugu. Lehenengo adibidean telebistako iragarkia ikusi ondoren produktuari buruz zeukaten eritzia, egunkariko iragarkia ikusi ondoren zeukaten eritziarekin konparatu dugu. Bigarren adibidean aldagai hauek erabili ditugu: produktuari buruzko eritzia, telebistako iragarkia ikusi aurretik eta ondoren, hau da, IRAGARTB eta LEHENA.

Erabili ditugun aginduak hauek izan dira:

NPAR TESTS SIGN=IRAGARTB WITH IRAGEGUN.

NPAR TESTS SIGN=IRAGARTB WITH LEHENA.

Hala ere, agindu biak honela idazteko posibilitatea daukagu:

NPAR TESTS SIGN=IRAGARTB WITH IRAGEGUN LEHENA.

Lortu ditugun emaitzak hauek izan dira:

```
- - - - - Sign   Test
     IRAGARTB
with IRAGEGUN

Cases
   15    - Diffs  (IRAGEGUN Lt IRAGARTB)          Z = 1.1355
   23    + Diffs  (IRAGEGUN Gt IRAGARTB)
    2      Ties                              2-tailed P =  .2561
   40      Total
```

```
- - - - - Sign   Test
     IRAGARTB
with LEHENA

Cases
   35    - Diffs  (LEHENA Lt IRAGARTB)            Z = 5.5000
    1    + Diffs  (LEHENA Gt IRAGARTB)
    4      Ties                              2-tailed P =  .0000
   40      Total
```

Aurreko bi adibideetan, froga hau burutzeko erabiltzen den prozedura ikus dezakegu. Lehenengo, aldagai bien diferentzien zeinuak ikusten dira, (esan beharra dago Ties hitzak errepikatzen diren balioak adierazten dituela), eta gero diferentzia hauen zeinuen arteko diferentziak ea zoriz lortutakoak diren ala ez aztertzen da. Azterketa hau banaketa normalaren bidez egiten da, 25 elementu baino gehiago dauzkagulako. Lehenengo kasuan, diferentzien zeinuen arteko diferentziak zoriz lortutakoak direla erabakiko dugu. Bigarrenean ostera, gure erabakia diferentzi hauek signifikatiboak izatea izango da.

2.2.3. WILCOXONen froga. Aldagai ordinalak

Zeinuen froga baino AHALMEN HANDIAGOA duen froga bat izango da hau. Froga honetan kontutan hartuko ditugu ez bakarrik diferentzien zeinuak; baizik eta diferentzien balioak ere bai.

Froga hau erabiltzerakoan ematen diren pausoak hauek dira:

1.- Aldagai bien arteko diferentziak kalkulatzen dira.
2.- Lortutako diferentziak, zeinua kontutan hartu gabe, ordenatu egiten dira.
3.- Aurreko pausoan elementu bakoitzak hartu duen ordenen balioari, lehenengo pausoan zeukan zeinua erantsiko diogu.
4.- 3.pausoan lortutako balioen batuketa lortuko dugu.

Oinarrizko hipotesia betetzen denean, hau da, aldagai biak estatistiko berdinak dauzkatenean, zeinu positiboak dauzkatenen ordenen batuketek eta zeinu negatiboak dituztenenek berdinak izan beharko dute. Beraz 4.pausoan egindako batuketak 0 edo 0-tik oso hurbil egon beharko du.

Aurreko hipotesia baieztatzeko, lagin-kopurua 25 edo txikiagoa denean, WILCOXONen taulei begiratu beharko dugu. Lagin-kopurua 25 baino handiagoa denean, banaketa normala erabiltzen da. Banaketa normal honen parametroak hauek izango dira:

$$\mu = \frac{N(N+1)}{4}$$

$$\sigma^2 = \frac{N(N+1)(2N+1)}{24}$$

Wilcoxon-en froga erabiliz lortu ditugun emaitzak aztertuko ditugu. Erabili ditugun aginduak hauek izan dira:

NPAR TESTS WILCOXON=IRAGARTB WITH IRAGEGUN.

NPAR TESTS WILCOXON=IRAGARTB WITH LEHENA.

Hala ere, zeinuen frogan egin dugun bezala, aurreko bi aginduak ondoren idazten dugun eran batzeko posibilitatea edukiko dugu:

NPAR TESTS WILCOXON=IRAGARTB WITH IRAGEGUN LEHENA.

Lortu ditugun emaitzak hauek izan dira:

```
- - - - - Wilcoxon Matched-pairs Signed-ranks Test
      IRAGARTB
with  IRAGEGUN

 Mean Rank     Cases
     15.90      15    - Ranks  (IRAGEGUN Lt IRAGARTB)
     21.85      23    + Ranks  (IRAGEGUN Gt IRAGARTB)
                 4      Ties   (IRAGEGUN Eq IRAGARTB)
                40     Total
Z = -1.9143                     2-tailed P = .0556
```

```
- - - - - Wilcoxon Matched-pairs Signed-ranks Test
      IRAGARTB
with  LEHENA

 Mean Rank     Cases
     18.97      35    - Ranks  (LEHENA Lt IRAGARTB)
      2.00       1    + Ranks  (LEHENA Gt IRAGARTB)
                 4      Ties   (LEHENA Eq IRAGARTB)
                40     Total
      Z = -5.2002               2-tailed P = .0000
```

Froga hau erabiliz lortu ditugun emaitzak zeinuen frogakoekin konparatzen baditugu, lehenengo kasuko ezberdintasunak zoriz lortzeko dagoen probabilitatea txikiagotu egin dela ikus dezakegu. Honen arrazoia zera da: froga honek diferentzien zeinuaz gain diferentzien balioak ere kontutan hartzen dituela.

3. K LAGINETARAKO INFERENTZIAK

Bi lagineko kasuan ikusten genuen bezala, hemen ere K laginetan lortutako emaitzak konparatzeko badaude froga parametriko batzuk. Ezagunena eta gehien erabiltzen dena BARIANTZ analisia da. Ikusi dugunez, SPSS/PC+ paketeak konparaketa hauek egiteko ONEWAY eta ANOVA azpiprogramak eskaintzen dizkigu.

Baina Bariantz analisia erabiltzeko bete behar diren aplikazio-baldintzak hauek dira:

- Elementu ezberdinen neurketek askeak izan behar dute. Lagin-aukeraketak zorizkoa izan behar du, eta elementu bakoitzaren puntuazioak bestearekiko askea izan behar du.

- Lagin guztiei dagozkien Populazioko banaketek, normalak izan behar dute.

- Populazio guztietako bariantzek berdinak izan behar dute.

- Aztertu nahi dugun aldagaiak jarraia izan behar du. Gutxienez tartezko eskala baten neurtuta egon behar du.

- Populazio ezberdinetako batezbestekoek zutabe-lerro eta bien arteko interakzioen efektuen konbinaketa lineal izan beharko dute.

K laginetan lortutako emaitzak konparatzerakoan, bi laginetakoak konparatzerakoan ikusten genituen oztopo berdinarekin agertuko gara. Hau da, baldintza hauek oso kasu gutxitan konprobatzen direla. Honen ondorioz, askotan, lortzen diren emaitzak nahikoa eztabai-dagarriak izango dira.

Beraz, aplikazio-baldintza hauek betetzen ez direnean, oso interesgarriak diren froga ez-parametrikoak atal honetan aurkeztuko ditugu.

Bi laginetako kasuan egin dugun bezala, K laginei dagozkien froga ez-parametrikoak aurkezterakoan, bi zatitan plazaratuko ditugu: talde askeei dagozkienak eta talde erlazionatuei dagozkienak.

3.1. K lagin askerentzako froga ez-parametrikoak

3.1.1. MEDIANAREN froga. Eskala ordinalak.

K lagineko emaitzak konparatzeko erabiliko dugun Medianaren froga, bi lagineko emaitzak konparatzeko erabili dugunen ondorio edo luzapen bat izango da.

K laginetan neurtu dugun aldagaia ea mediana berdinak dauzkaten populazioetatik aukeratuak izan den baieztatzeko baliagarria izango da froga hau.

Froga hau aplikatzeko ematen diren pausoak hauek izango dira. Lehenengo pausoan k laginetan lortu ditugun emaitza guztiak, ordena bakar batean ordenatuko ditugu. Hau egin ondoren, talde guztiko mediana kalkulatuko dugu, eta oinarrian geneukan talde bakoitzean kontatuko ditugu zenbat elementuk talde osoko mediana baino ordena handiago eta txikiagoa daukaten. Maiztasun hauekin ondoko erako taula bat eraikiko dugu, eta taula honetan X^2-aren testa egingo dugu.

	TALDE 1	TALDE 2	.....	TALDE K
MEDIANA BAINO HANDIAGOAK	A	B	...	D
MEDIANA BAINO TXIKIAGOAK	E	F	...	H

Froga hau erabiltzeko idatzi behar dugun agindua era honetakoa izango da:

NPAR TESTS MEDIAN = menpeko aldagaia BY askea (min,max).

Adibide bezala, telebistako iragarkia ikusi ondoren gure produktuari buruzko eritzia, adinari begiratuz egin ditugun hiru taldetan, ea berdina den ala ez aztertu nahi izan dugu. Esan beharra dago adina eskala nominalean neurtuta daukagula.

Horretarako idatzi dugun instrukzioa hau izan da:

NPAR TESTS MEDIAN = IRAGARTB BY ADINA(1,3).

Lortu ditugun emaitzak hauek izan dira:

– – – – – Median Test

IRAGARTB
by ADINA

		ADINA		
		1	2	3
IRAGARTB	Gt Median	5	6	7
	Le Median	12	4	6

Cases	Median	Chi-Square	D.F.	Significance
40	10.0	2.9892	2	.2243

Beraz, emaitzak ikusi ondoren, gure erabakia dauden ezberdintasunak zoriz lortutakoak direla izango da.

Erabaki honetara ailegatzeko, emaitzetan ikus dezakegunez, talde guztietako mediana 10 da, eta oinarrizko talde bakoitzean talde guztiko mediana baino handiagoak eta txikiagoak direnen maiztasunak agertzen dira. Ondoren X^2-aren testaren emaitzak agertzen dira.

3.1.2. KRUSKAL-WALLISen froga. Eskala ordinalak

K laginetako emaitzak konparatzeko gehien erabiltzen den froga ziur asko hau izango da. Aurreko kasuan bezala oinarrizko hipotesia, lagin ezberdinei dagozkien populazioak zentralizazio-neurri berdinak edukitzea izango da.

Medianaren froga baino AHALMEN handiagoa dauka. Bestetik, froga hau aplikatu nahi baldin badugu, gutxienez eskala ordinal batean neurtutako aldagaiak izan beharko dute.

Aurreko frogan bezala, hau aplikatzerakoan, lehenengo pausoan, lagin ezberdinetako elementu guztiak talde batean sartu ondoren, talde hau ordenatu egiten da. Frogaren oinarria, lagin bakoitzeko elementuen ordenen baturen konparaketa egitea da.

Oinarrizko hipotesia betetzen bada, lagin ezberdinetan lortu ditugun baturek berdinak edo berdintsuak izan beharko dute. Hau da froga honek baieztatzen duena.

Horretarako, ondoren definituko dugun H estatistikoa erabiltzen du:

$$H = \frac{12}{N\ (N+1)} \Sigma \frac{R_j^2}{N_j} - 3\ (N+1)$$

non K lagin edo talde-kopurua
N_j j taldeko elementu-kopurua
R_j j taldeko elementuen ordenen batura
N talde guztietako elementu-kopurua

Nahikoa erraza da, H estatistikoak k-1 askatasun-gradu dituen X^2-aren banaketari jarraitzen zaiola frogatzea. Honetan oinarrituz, test hau burutzen da.

Froga hau aplikatzeko idatzi behar dugun instrukzioak honelakoa izan behar du:

NPAR TESTS K-W = menpeko aldagaia BY askea (min,max).

Guk, adibide bezala aurreko frogan aztertu duguna errepikatu dugu, hau da, telebistako iragarkia ikusi ondoren gure produktuari buruzko eritzia, adinari begiratuz egin ditugun hiru taldeetan, ea berdina den ala ez.

Idatzi dugun instrukzioa hau izan da:

NPAR TESTS K-W =IRAGARTB BY ADINA(1,3).

Eta lortu ditugun emaitzak hauek izan dira:

```
- - - - - Kruskal-Wallis 1-way ANOVA

     IRAGARTB
  by ADINA

      Mean Rank      Cases
          18.71         17  ADINA = 1
          21.15         10  ADINA = 2
          22.15         13  ADINA = 3
                        40  Total

                                          Corrected for Ties
  CASES    Chi-Square    Significance    Chi-Square    Significance
   40        .7198         .6977          .7312          .6938
```

Ikusten dugunez, aurreko frogan lortu dugun probabilitatea eta oraingoa ez dira berdinak. Froga honen ahalmena handiagoa denez, hau erabiltzea gomendatzen da.

Emaitzak aztertzeko ez dago inongo zailtasunik; X^2-aren froga egiten bait du.

3.2.- K lagin erlazionaturentzako froga ez-parametrikoak

Bi lagin erlazionatuentzako egin ditugun kontsiderazio guztiak, kasu honetarako ere baliagarriak izango dira, hau da, talde erlazionatuak ez dira bakarrik izango k momentu ezberdinetan neurtu ditugun elementu berdinak osatzen dituztenak bakarrik izango, baizik eta baita ezaugarri berdinak dituzten k-koteak egin ondoren hauek k taldetan banatzen ditugunean ere.

SPSS/PC+ paketeak, kasu hauetarako 3 froga ezberdin eskaintzen dizkigu; garrantzirik handiena daukatenak, edo gehien erabiltzen direnak.

3.2.1.- COCHRANen Q froga. Eskala nominala

K talde erlazionatu dauzkagunean eta neurtu dugun aldagaia eskala nominalean daukagunean, hau da, aldagai kualitatiboa denean, froga hau erabiliko dugu.

Item ezberdinen zailtasun-indizeak kalkulatzeko ere erabil daiteke.

Lehenengo pausoan gure eskala nominala dikotomizatu egin beharko dugu. Hau da, aldagaien atributu edo kualitate bati arrakastaren tratamendua emango diogu, eta besteei porrotarena.

Oinarrizko hipotesiak, talde ezberdinetan dauden diferentziak zoriari dagozkionak direla esango du.

Ez gara sartuko froga honen oinarri teoriko-matematikotan, baizik eta nola aplikatzen den ikusiko dugu.

Froga hau erabiltzerakoan Q estatistikoa kalkulatu behar da lehenengo pausoan. Q estatistiko hau ondoko formularen bidez definitzen da, eta estatistika teorikoak dioenez, estatistiko hau, k-1 askatasun gradu dituen X^2-aren banaketari jarraitzen zaio.

$$Q = \frac{k\ (k-1)\ \Sigma\ (G_j - \overline{G})^2}{k\,\Sigma L_i - \Sigma L_i^2}$$

non: k : talde-kopurua.
G_j : j taldeko arrakasta-kopurua.
G : G_j balioen batezbestekoa.
L_i : i elementuen arrakasta-kopurua.

Eman dezagun, gure lagina osatzen duten 40 pertsonen ondoko hiru eritzi hauek konparatu nahi dugula: telebistan, irratian eta egunkarietan egoten diren iragarkien alde edo kontra dauden.

Beraz, hemen hiru talde erlazionatu dauzkagu, eta bakoitzean aldagai dikotomiko bat neurtu dugu.

Aipatu dugun konparaketa egiteko, ondoko agindua idatzi dugu:

NPAR TESTS COCHRAN=IRITSITB IRITSIRR IRIEGUNK.

Emaitzak hauek izan dira:

```
- - - - - Cochran   Q   Test

   Cases
   =   1  =   0    Variable
      14     26    IRITSITB
      19     21    IRITSIRR
      32      8    IRIEGUNK

   Cases        Cochran Q          D.F.        Significance
    40           17.2667             2             .0002
```

Emaitzen interpretazioa ere oso erraza da. Significance balioak dioenez, lortu ditugun emaitzak zoriz lortzeko dagoen probabilitatea .0002 denez geroz, hiru eritzi ezberdinen artean dauden diferentziak garrantzitsuak direla erabakiko dugu. Beraz, oinarrizko hipotesia ezeztatu egingo dugu.

3.2.2. FRIEDMANen froga. Eskala ordinala

X^2-aren testa alde batera utzita, ziur asko test ez-parametrikoetatik gehien erabiltzen dena hau izango da.

Bariantz analisiarekin konparatuz gero, froga honen abantaila handiena, askoz ere aplikazio-baldintza apalagoak bete beharra izango da.

Froga hau K talde erlazionatutan eskala ordinal batean neurtuta daukagun aldagai baten emaitzak konparatzeko baliagarria izango da. Beraz laginek tamaina berdinekoak izan beharko dute.

Froga hau erabiltzerakoan eman behar dugun lehenengo pausoa, elementu bakoitzaren emaitzak ordenatzea izango da. Kontutan eduki behar dugu orain arte taldeko elementuen balioak ordenatu ditugula. Ez da hau froga honen kasua. Hemen lehen esan dugunez, elementu bakoitzaren puntuazioak ordenatuko ditugu.

Beraz, honelako datu-matrize bat lortuko dugu:

ELEMENTUA	IRAGEGUN	IRAGIRRA	IRAGARTB
1	3	2	1
2	3	1.5	1.5
3	1	2	3
4	1	3	2
5	1	3	2
6	1.5	1.5	3
.	.	.	.

Aurreko datu-matrizean, zutabe bakoitzeko lehentasun-ordenak berdinak badira, zutabe bakoitzeko baturak ere berdinak izango dira, eta batura hauen artean dauden ezberdintasunak zoriari dagozkionak izango dira. Hipotesi honetan oinarrituz, k-1 askatasun-gradu dituen X^2_r estatistikoa ondoko formularen bidez kalkulatuko dugu.

$$X_r^2 = \frac{12}{K\ (K+1)\ N} \Sigma\ R_j^2 - 3N\ (k+1)$$

non N = lerro-kopurua (elementu-kopurua).
K = zutabe-kopurua (aldagaiak).
R_j = j zutabeko ordenen batura.

Estatistikoa kalkulatu ondoren, hurrengo kalkuluak X^2-aren testarenak izango dira:

Kalkulu hauek SPSS/PC+ erabiliz egiteko, idatzi dugun azpiprograma ondokoa izango da:

NPAR TESTS FRIEDMAN=IRAGEGUN IRAGIRRA IRAGARTB.

Lortu ditugun emaitzak hauek izan dira:

- - - - - Friedman Two-way ANOVA

Mean Rank	Variable
2.21	IRAGEGUN
1.84	IRAGIRRA
1.95	IRAGARTB

Cases	Chi-Square	D.F.	Significance
40	2.9625	2	.2274

Emaitzak aurreko frogetakoen berdintsuak dira. MEAN RANK izenpean agertzen diren balioak, zutabe bakoitzeko ordenen batezbestekoak direla bakarrik hartu behar dugu kontutan.

3.2.3. KENDALLen froga. Eskala ordinala

Nahiz eta SPSS/PC+ paketeak Kendallen froga hipotesi-test bezala aurkeztu, guk froga hau bi era ezberdinetara erabil dezakegu. Lehenengoa, Kendallen konkordantzi koefiziente den W kalkulatzeko, eta bigarrena hipotesi-test bezala erabiltzeko.

Era bietan ikusiko dugu froga hau.

Aurreko frogako adibide berdinari jarraituko gatzaizkio, hau da, IRAGEGUN, IRAGIRRA eta IRAGARTB aldagaien balioak konparatuko ditugu, eta hiru aldagai hauen artean dagoen konkordantzi (komunztadura) indizea ere kalkulatuko dugu.

Horretarako lehenengo pausoa, Friedmanen frogan egin dugun bezala, elementu bakoitzeko aldagaien puntuazioak ordenatu egingo ditugu. Beraz lortuko dugun datu-matrizea, era honetakoa izango da:

ELEMENTUA	IRAGEGUN	IRAGIRRA	IRAGARTB
1	3	2	1
2	3	1.5	1.5
3	1	2	3
4	1	3	2
5	1	3	2
6	1.5	1.5	3
.	.	.	.

Elementu guztien eritzia berdina baldin bada, zutabe ezberdinen baturek, 40 elementu dauzkagunez, era honetakoak izan beharko lukete: 40, 80 eta 120, zeren eta elementu guztiak ados egonez gero, elementu guztiek zutabe ezberdinetan balio berdinak eduki beharko bait lituzkete.

Balio hauei lagineko emaitzak noraino hurbiltzen zaizkien aztertuko du froga ez-parametriko honek.

Horretarako erabiliko duen estatistikoa, Kendallen W izango da, eta estatistiko hau ere, k-1 askatasun- gradu dituen X^2-aren banaketari jarraituko zaio. W estatistikoa era honetan definitzen da:

$$W = \frac{12 \sum (R_j - \overline{R})^2}{N^2 \; K \; (K^2 - 1)}$$

non: R_j : j zutabeko ordenen batura.
R : R_j balioen batezbestekoa
K : aldagai-kopurua
N : elementu-kopurua

Gure adibideko emaitzak lortzeko idatzi dugun instrukzioa hau izan da:

NPAR TESTS KENDALL=IRAGEGUN IRAGIRRA IRAGARTB.

Lortu ditugun emaitzak:

- - - - - Kendall Coefficient of Concordance

Mean Rank	Variable
2.21	IRAGEGUN
1.84	IRAGIRRA
1.95	IRAGARTB

Cases	W	Chi-Square	D.F.	Significance
40	.0403	2.9625	2	.2274

Emaitzetan ikusten dugunez, konkordantzi koefizienteen balioa, W = .0403 da. Honek zera esan nahi du, elementu ezberdinek egiten dituzten ordenak ez direla berdinak, hau da, ez direla konkordanteak. Beraz hau ikusirik, elementu ezberdinek nahikoa eritzi ezberdinak dauzkatela esan dezakegu, eta orokorki hartuta, telebista, irrati eta egunkarietan agertu diren iragarkien ondorioz lortu ditugun emaitzak ez direla signifikatiboki aldatzen. Hau Significanceren balioan ikus dezakegu, zeren ikusi ditugun ezberdintasunak zoriz lortzeko dagoen probabilitatea .1994 bait da.

4. OPZIOAK ETA ESTATISTIKOAK

Aipatu ditugun froga ez-parametriko guztiak, aukera eta estatistiko ezberdinekin osatzeko posibilitatea edukiko dugu. Guzti hauek hemen azalduko ditugu:

OPZIOAK

1 Datu galduak analisian sartzen ditu.

2 Datu galduak ez dira analisian sartzen.

4 Lagin handiak ditugunean, aukera honek zorizko azpi-laginketa batekin lana egiteko posibilitatea eskaintzen digu.

ESTATISTIKOAK

1 Estatistika deskribatzaileari dagozkion estatistikoak agertzen ditu: batezbestekoa, desbiderapen standarda...

2 Koartilak eta elementu-kopurua agertzen ditu.