eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Bioestatistika: Oinarrizko Ikastaroa

Inmaculada Arostegui

Arantza Urkaregi

ISBN: 978-84-9860-930-1

EUSKARAREN ARLOKO ERREKTOREORDETZAREN SARE ARGITALPENA

Liburu honek UPV/EHUko Euskararen Arloko Errektoreordetzaren dirulaguntza jaso du

Inmaculada Arostegui Madariaga e-mail: inmaculada.arostegui@ehu.es

Arantza Urkaregi Etxepare e-mail: arantza.urkaregi@ehu.es

Matematika Aplikatua eta Estatistika eta Ikerkuntza Operatiboa Saila

Zientzia eta Teknologia Fakultatea

Euskal Herriko Unibertsitatea (UPV/EHU)

Sarriena auzoa z.g., 48940 Leioa

Egileen aipamena:

Ikasmaterial hau UPV/EHUko Zientzia eta Teknologia Fakultatean eskaintzen den Biozientzietako hiru graduetako lehen mailako *Bioestatistika* irakasgairako garatuak izan dira. Egileok, geure ikuspegi pertsonala eta urte askotako esperientzia erabiliz, eta oinarritzat erreferentzietan aipatzen diren hainbat liburu hartuta, aipatutako irakasgaiaren ulermenerako laguntza gisa erabiltzeko egin dugun lana da.

ISBN: 978-84-9860-930-1

Aurkibidea

1 ESTATISTIKA DESKRIBATZAILEA 1

1.1 SARRERA 1

1.2 TAULA ESTATISTIKOAK 2

1.3 ADIERAZPIDE GRAFIKOAK 5

1.4 ESTATISTIKOAK 14

1.5 KUTXA-DIAGRAMAK 31

1.6 DATU ELKARTUAK 32

1.7 ARIKETAK 43

2 PROBABILITATEA 57

2.1 SARRERA 57

2.2 ZORIZKO ESPERIMENTUARI ELKARTUTAKO GERTAERAK 57

2.3 GERTAEREN ARTEKO ERLAZIOAK 58

2.4 PROBABILITATE KALKULUA 60

2.5 EREDU PROBABILISTIKOAK 62

2.6 PROBABILITATE BALDINTZATUA ETA GERTAEREN ARTEKO INDEPENDENTZIA 65

2.7 BAYES-en TEOREMA 70

2.8 TEST DIAGNOSTIKOA 73

2.9 KONBINATORIA (hautazkoa) 78

2.10 ARIKETAK 81

3 ZORIZKO ALDAGAIAK 89

3.1 SARRERA 89

3.2 ZORIZKO ALDAGAI DISKRETUAK 90

3.3 ZORIZKO ALDAGAI JARRAITUAK 94

3.4 ZORIZKO ALDAGAIAREN ITXAROPENA ETA BARIANTZA 98

3.5 BANAKETA BINOMIALA 101

3.6 POISSON-EN BANAKETA 105

3.7 BANAKETA NORMALA 111

3.8 LIMITEAREN TEOREMA ZENTRALA 117

3.9 BANAKETA NORMALA BESTE BANAKETA BATZUEN LIMITE GISA…… 118

3.10 ARIKETAK 122

4 INFERENTZIA ESTATISTIKOA. LAGIN BAKAR BATEN AZTERKETA.... 129

4.1 SARRERA 129

4.2 POPULAZIOAREN ETA LAGINAREN ARTEKO ERLAZIOA..... 130

4.3 ESTIMAZIOA 133

4.4 BANAKETAREN BATEZBESTEKOAREN ESTIMAZIOA..... 134

4.5 BANAKETAREN BARIANTZAREN ESTIMAZIOA 146

4.6 BARIANTZA EZEZAGUNEKO POPULAZIO NORMALAREN LAGIN-BATEZBESTEKOAREN ESTIMAZIOA..... 151

4.7 ARIKETAK 157

5 HIPOTESI-KONTRASTEA. LAGIN BAKAR BATEN AZTERKETA 165

5.1 SARRERA 165

5.2 LEHENDABIZIKO ADIBIDEA: HIPOTESI-KONTRASTEAREN KONTZEPTU INTUITIBOAK 166

5.3 PLANTEAMENDU OROKORRA 166

5.4 HIPOTESI-KONTRASTEA EGITEKO URRATSAK 168

5.5 BARIANTZA EZAGUNEKO POPULAZIOAREN BATEZBESTEKORAKO HIPOTESI-KONTRASTEA..... 170

5.6 TESTAREN AHALMENA 174

5.7 LAGIN-TAMAINAREN KALKULUA..... 179

5.8 KONFIANTZA-TARTEAREN ETA HIPOTESI-KONTRASTEAREN ARTEKO ERLAZIOA 181

5.9 BARIANTZA EZEZAGUNEKO POPULAZIO NORMALAREN BATEZBESTEKORAKO HIPOTESI-KONTRASTEA 183

5.10 BANAKETA NORMALAREN BARIANTZARAKO HIPOTESI-KONTRASTEA..... 185

5.11 ARIKETAK 188

5.12 ERANSKINA 193

6 BI POPULAZIORAKO INFERENTZIA 199

6.1 BI LAGIN ASKEREN BATEZBESTEKOEN ARTEKO DIFERENTZIAREN BANAKETA..... 199

6.2 FISHER-SNEDECOR-EN F BANAKETA..... 202

6.3 BANAKETA NORMALARI DARRAIZKION BI POPULAZIO ASKEREN BARIANTZEN KONPARAKETA 205

6.4 BI BANAKETA ASKEREN BATEZBESTEKOEN ARTEKO DIFERENTZIARAKO KONFIANTZA-TARTEA..... 207

6.5 BI POPULAZIO ASKEREN BATEZBESTEKOEN BERDINTASUNERAKO KONTRASTEA..... 217

6.6 BATEZBESTEKOEN ARTEKO BERDINTASUNA: BINAKAKO DATUAK ……..234

6.7 ARIKETAK..... 244

7 POPULAZIO BINOMIALETARAKO INFERENTZIA..... 253

7.1 BANAKETA BINOMIALAREN p PARAMETROAREN ESTIMAZIOA .. 253

7.2 BANAKETA BINOMIALAREN p PARAMETRORAKO HIPOTESI-KONTRASTEA..... 257

7.3 BI PROPORTZIOREN KONPARAKETA 266

7.4 ARIKETAK..... 275

8 KHI KARRATU BANAKETAREN APLIKAZIOAK 281

8.1 SARRERA..... 281

8.2 INDEPENDENTZIA-PROBA..... 282

8.3 HOMOGENEOTASUN-PROBA..... 289

8.4 DOIKUNTZA-EGOKITASUNERAKO PROBA 296

8.5 ARIKETAK..... 302

9 BARIANTZA-ANALISIA 307

9.1 SARRERA..... 307

9.2 FAKTORE BAKARREKO BARIANTZA-ANALISIA 307

9.3 BARIANTZA-ANALISIA *SPSS*-REN BITARTEZ 314

9.4 KONPARAKETA ANITZAK 316
9.5 ARIKETAK 324
10 ERREGRESIO LINEALA 331
10.1 SARRERA 331
10.2 DATUEN DESKRIBAPENA 333
10.3 KORRELAZIOA 334
10.4 ERREGRESIO LINEAL SINPLEA 337
10.5 ERREGRESIO LINEAL SINPLEA *SPSS*-REN BITARTEZ 344
10.6 ERREGRESIO LINEAL SINPLEAREN EREDUAREN DIAGNOSIA 348
10.7 ARIKETAK 351
A. ERANSKINA: TAULAK 361
B. ERANSKINA: ERREFERENTZIAK 377

1 ESTATISTIKA DESKRIBATZAILEA

1.1 SARRERA

Estatistika Deskribatzailea da esperimentuaren datuen bilketaz, azterketaz, antolakuntzaz eta laburpenaz arduratzen den Estatistikaren arloa.

Elementu multzo batetik informazioa jaso nahi denean, multzo horri **populazio** deritzogu. Populazioaren elementu bakoitzari **unitate estatistiko** edo **ale** deritzo. Populazioak finituak edo infinituak izan daitezke, baina beti zehaztasunez definitu behar dira, zalantzagarritasunik gerta ez dadin.

1.1 adibidea

Euskal Autonomi Elkarteko (EAEko) pertsonen osasunari buruzko azterketak egitea nahi bada, EAEn erroldaturik dauden pertsona guztiek osatuko dute populazioa. Multzo horren elementu kopurua infinitua dela kontsidera dezakegu. EAEko pertsona bakoitza unitate estatistikoa da.

Populazio baten elementu guztiak banan-banan aztertu ezin direnean (edo, egin daitekeelarik ere, interesik ez duenean), populazioaren zati bat aztertzen da. Zati horri **lagin** deritzogu. **Laginaren tamaina** laginaren elementu kopurua da, eta n-ren bidez adierazten da. Populazio osoa aztertzea zaila edo ezinezkoa denean, lagin bat aukeratzen da. Ikasmaterial honen laugarren gaian ikusiko den bezala, fenomenoaren azterketa estatistikoa gauzatzeko, honako hau da laginek bete behar duten baldintza garrantzitsuenetariko bat: zorizkoak eta populazioaren adierazgarriak izatea.

1.2 adibidea

EAEn dauden pertsona guztien osasunari buruzko datuak lortzea eta aztertzea ezinezkoa denez, populazioaren azpimultzo finitu bat hartzen da azterketa estatistikorako. EAEko pertsonen osasunari buruzko azterketak egiteko, Osasun Inkestak egin izan dira hainbat urtean. Esate baterako, 1997. urtean 4.052 pertsonek erantzun zioten inkestari, eta haiek osatu zuten EAEko populazioaren lagina.

Izaera da aztertzen den populazioaren ezaugarria, eta **modalitatea,** berriz, izaerak aurkez dezakeen era edo egoera bakoitza. Izaerak mota honetakoak izan daitezke:

1. **Kualitatiboak**. Ezaugarri batek aurkezten dituen kategoriak, ordenarik gabe eta izen baten bidez adierazten direnean, izaera kualitatibo edo nominala dela esaten dugu; bere modalitateak ezin daitezke zenbakien bidez adierazi. Adibideak: sexua, lanbidea edo begien kolorea.
2. **Kuantitatiboak.** Ezaugarriaren modalitateak neurgarriak dira. Adibideak: tenperatura, pisua edo adina.

Izaera baten bidez elementu talde baten ordena baino definitzen ez dugunean (elementu horien artean dauden distantziak neurtu ezinik), **izaera ordinala** edo **quasi-kuantitatiboa** dela esaten dugu.

Aztertutako ale bakoitzak izaeraren modalitate bakar bat aurkeztu behar du; hau da, modalitateek baztertzaile eta exhaustiboak izan behar dute. Adibidez, hiri batean hilabetean zehar izandako tenperaturak aztertzen baditugu, tenperatura izango da izaera eta modalitateak neurturiko gradu kopuru guztiak.

Izaera kuantitatiboaren modalitateen neurketari dagokion zenbaki multzoari **aldagai estatistiko** deritzo. Aldagai estatistikoak modu honetan sailka daitezke:

1. **Aldagai estatistiko diskretuak**: modalitateek hartzen dituzten balioak zenbaki diskretuak dira (balio isolatuak bakarrik har ditzakeen aldagaia da); adibidea: leka baten ilar kopurua.
2. **Aldagai estatistiko jarraituak**: tarte baten barnean, modalitateak edozein balio har dezake (bi balioren arteko edozein balio har dezakeen aldagaia da); adibidea: persona baten pisua.

Aldagai estatistiko diskretu eta jarraituen arteko ezberdintasuna guztiz erlatiboa da, neurrien akatsak direla kausa, azkenean, aldagaien balioak zenbaki diskretuak izango baitira beti. Baina, teorikoki, ezberdintasuna argi dago. Praktikoki, aldagaiak balio asko hartzen dituenean, aldagai estatistikoa jarraitutzat hartuko dugu, nahiz eta diskretua izan.

1.2 TAULA ESTATISTIKOAK

Estatistika deskribatzaileak egiten duen lehenengoa hauxe da: aztertu nahi den fenomenoari buruzko datuak biltzea. Lan hori arazo delikatua da, fase honetako erroreek datuen azterketa faltsutzen baitute. Kasu bakoitzean, egiten den azterketaren eta ikertzailearen irizpideen menpean daude lana gauzatzeko erabide eta teknikak.

Demagun, populaziotik ateratako lagin batean, aldagai estatistiko bat aztertu nahi dela. Datuak bildu eta gero, $x_1,\ldots,x_n$ izanik datuak, **taula estatistikoa** edo **maiztasun-taula** izenekoan ordenatuta aurkeztea da bigarren lana. Taula estatistikoetan honako hauek adieraz daitezke:

1. m_i **izaeraren modalitateak.** Izaerak aurkez ditzakeen era edo egoera ezberdinak. Izaera kuantitatiboa bada, izaeraren modalitateak aldagai estatistikoaren balioak dira ($m_i = x_i$) eta ordenaturik jartzen dira. Izaeraren modalitate kopurua k-ren bidez adieraziko dugu.
2. f_i **maiztasun absolutuak**. Modalitate bakoitzari dagozkion laginaren ale kopuruak. Laginaren tamaina n bada, maiztasun absolutuek honako propietate hauek beteko dituzte:

i) $0 \leq f_i \leq n \qquad \forall\, i = 1, \ldots, k$

ii) $\sum_{i=1}^{k} f_i = f_1 + f_2 + \ldots + f_k = n$

3. Izaera kuantitatiboa denean, F_i **maiztasun metatuak** adieraziko dira, hots, x_i edo x_i baino txikiagoak diren balio guztiei dagozkien maiztasun absolutuen batura. Matematikoki, $F_i = \sum_{j=1}^{i} f_j$. Maiztasun metatuek honako propietate hau betetzen dute: $F_k = n$, non k aldagaiaren balio ezberdinen kopurua eta n laginaren tamaina baitira.

4. Aldagaiaren balio bakoitzari dagokion h_i **maiztasun erlatiboa** maiztasun absolutuaren eta laginaren tamainaren arteko zatidura da: $h_i = \frac{f_i}{n}$. Maiztasun erlatiboa aldagaiaren balio bakoitzak adierazten duen aleen batekotzat uler daiteke. Bigarren gaian ikusiko dugun bezala, maiztasun horiek lotuta daude probabilitatearekin. Maiztasun erlatiboek honako propietate hauek betetzen dituzte:

   i) $0 \leq h_i \leq 1 \qquad \forall\, i = 1, \ldots, k$

   ii) $\sum_{i=1}^{k} h_i = 1$, non k aldagaiaren balio ezberdinen kopurua baita.

5. Izaera kuantitatiboa denean, H_i **maiztasun metatu erlatiboa** da, hots, x_i edo bera baino txikiagoak diren balio guztiei dagozkien maiztasun erlatiboen batura. Matematikoki, $H_i = \sum_{j=1}^{i} h_j$. Maiztasun metatu erlatiboa maiztasun metatuaren eta laginaren tamainaren arteko zatidura gisa ere defini daiteke: $H_i = \frac{F_i}{n} = \frac{\sum_{j=1}^{i} f_j}{n} = \sum_{j=1}^{i} \frac{f_j}{n} = \sum_{j=1}^{i} h_j$. Maiztasun metatuek honako propietate hau betetzen dute: $H_k = 1$, non k aldagaiaren balio ezberdinen kopurua baita.

Halaber, taula estatistikoan, **portzentajea** eta **portzentaje metatua** adierazten dituzten beste bi zutabe ager daitezke. Bi zutabe horiek maiztasun erlatiboak eta maiztasun metatu erlatiboak 100ez biderkatuz lortzen dira, hurrenez hurren.

Aurreko guztia kontuan izanik, 1.1 taulan agertzen da taula estatistikoaren adierazpen orokorra.

Aldagaia x_i	Maiztasun absolutua f_i	Maiz. abs. metatua F_i	Maiztasun erlatiboa h_i	Maiz. erl. metatua H_i
x_1	f_1	F_1	h_1	H_1
x_2	f_2	F_2	h_2	H_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_k	f_k	$F_k = n$	h_k	$H_k = 1$
Totala	$\boldsymbol{n}$		**1**	

1.1 taula. Aldagai estatistikoaren taularen adierazpen orokorra.

Zenbait adibideren bidez, ikus dezagun nola gauzatzen den deskribatutako prozesua:

1. Izaera kualitatiboa denean, taula estatistikoan interesa daukaten zutabeak, modalitateak, maiztasun absolutuak eta erlatiboak (edo portzentajeak) dira.

1.3 adibidea

1.2 taulan sailkatu dira Osasun Inkestari erantzun zioten 16 urte baino zaharragoak ziren pertsona kopuruak, EAEko lurralde historikoka.

Lurralde historikoa	f_i	h_i
Araba	488	0,12
Bizkaia	1.940	0,48
Gipuzkoa	1.624	0,40
Totala	**4.052**	**1**

1.2 taula. EAEko 1997ko Osasun Inkestaren erantzuleak, lurralde historikoka sailkatuta.

2. Izaera kuantitatiboa. Aldagai estatistikoa diskretua denean, aurretik definituriko bost zutabeek osatzen dute taula estatistikoa.

1.4 adibidea

Osasun Inkestan honako galdera hau agertu da: Azken bi asteetan, zenbat aldiz joan zara mendira? Mendira joandako aldi kopurua aldagai kuantitatibo diskretua da. Inkestatuen artean, 270 pertsona joan dira mendira eta 3.782 pertsona ez dira joan. Mendira joan direnen taula estatistikoa erakusten da 1.3 taulan.

Zenbat aldiz	f_i	F_i	h_i	H_i
1	89	89	0,330	0,330
2	99	188	0,366	0,696
3	23	211	0,085	0,781
4	25	236	0,093	0,874
5	6	242	0,022	0,896
6	4	246	0,015	0,911
7	6	252	0,022	0,933
8	1	253	0,004	0,937
9	1	254	0,004	0,941
10	8	262	0,029	0,970
12	1	263	0,004	0,974
14	7	270	0,026	1
Totala	**270**		**1**	

1.3 taula. EAEko 1997ko Osasun Inkestaren erantzuleetatik inoiz mendira joandakoen sailkapena, joandako aldi kopuruaren arabera.

1.3 ADIERAZPIDE GRAFIKOAK

Nahiz eta taula estatistikoan datuak deskribatzeko behar den informazio guztia izan, komenigarria da informazio hori grafikoen bidez adieraztea, batez ere, argiagoa edota ageriagoa egiteko eta lorturiko datuen begi-laburpena edukitzeko. Adierazpide grafikoek informazio orokorra, arina eta ulertzeko erraza eskaintzen digute, eta, horretaz gainera, informazioa erlatibizatzeko konparaketak egitea baimentzen digute. Grafikoa ona izan dadin, testua irakurri gabe ulertu behar da. Horretarako, ondo etiketaturik egon behar du, ardatzetan aldagaien unitateak eduki behar ditu, eta abar. Adierazpide grafiko mota asko daude. Horietariko batzuk azalduko dira ondoren:

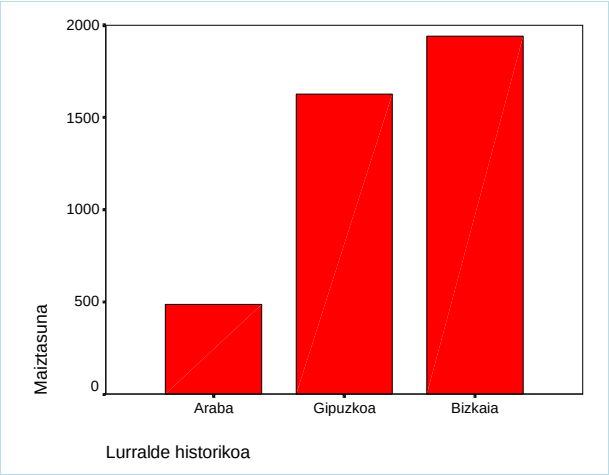
Izaera kualitatiboaren kasuan, bi adierazpide grafiko dira erabilgarrienak, hain zuzen, barra-diagrama eta sektore-diagrama.

1. **Barra-diagrama**: modalitateak (abzisa-ardatzean) eta maiztasunak (ordenatu-ardatzean) ardatz kartesiarren gainean aurkeztuko dira, batzuen eta besteen arteko egokitasuna barra-sistema baten bidez deskribatuz.
2. **Sektore-diagrama**: zirkulu baten azalera osoa zatituko da modalitate ezberdinen maiztasunen araberako zabaltasunak dauzkaten sektore zirkularretan.

Bi grafiko horiek maiztasun absolutuak zein erlatiboak erabiliz irudika daitezke.

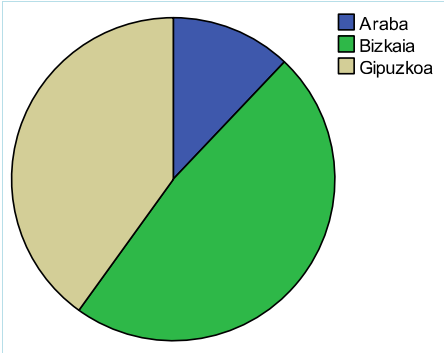
1.5 adibidea

Osasun Inkestari erantzun dioten 16 urtetik gorako biztanleen lurralde-historikoko 1.3 adibideko datuak oinarri hartuta, *SPSS* erabiliz eraiki dira honako barra-diagrama (1.1 irudia) eta sektore-diagrama (1.2 irudia) hauek.



1.1 irudia. Luralde historikoaren barra-diagrama.

$$n = 4052 \rightarrow 360^{o}$$
$$f_A = 488 \rightarrow 43{,}36^{o}$$
$$f_B = 1940 \rightarrow 172{,}36^{o}$$
$$f_G = 1624 \rightarrow 144{,}28^{o}$$



1.2 irudia. Lurralde historikoaren sektore-diagrama.

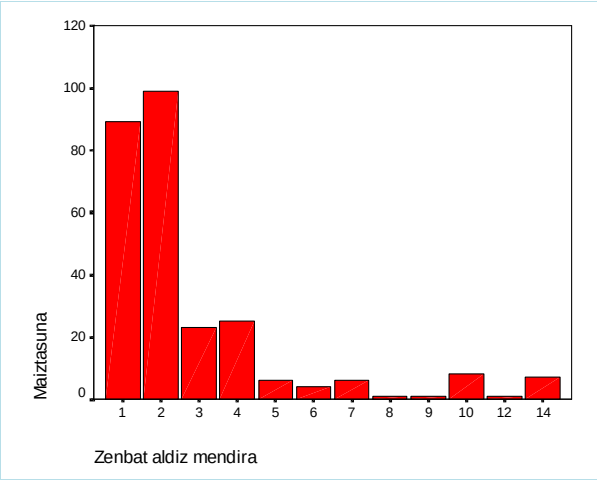
Izaera kuantitatiboaren kasuan, aldagai estatistikoa diskretua edo jarraitua izatearen menpe banatuko dugu grafikoen deskribapena. Aldagai diskretuaren kasuan, bi motako grafikoak ditugu: barra-grafikoa eta grafiko metakorra.

1. **Barra-grafikoa.** Aldagai diskretuen kasuan gehien erabiltzen denetariko bat da. Abzisa-ardatzean adierazten dira aldagaiaren balioak eta ordenatu-ardatzean maiztasun absolutu edo erlatiboak, eta barra-sistema baten bidez deskribatzen da batzuen eta besteen arteko egokitasuna. Barra guztiek zabalera berekoak izan behar dute eta elkarrekiko distantzia berean egon behar dute. Aurreko kasuko barra-diagramaren analogoa da, baina, kasu honetan modalitateak neurgarriak direnez, kontuan hartu behar da ordenak garrantzia duela.

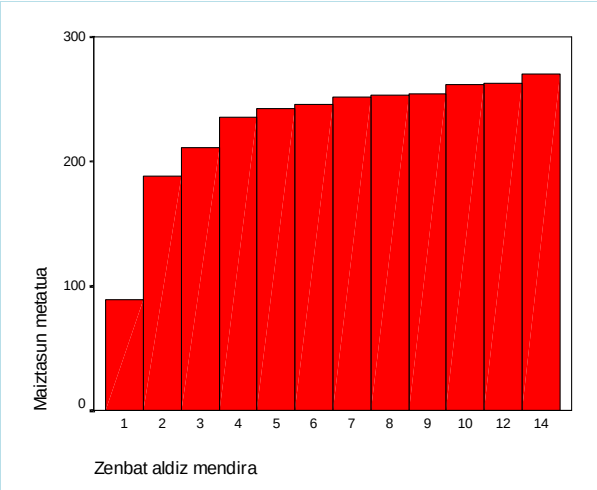
2. **Grafiko metakorra.** Abzisa-ardatzean aldagaiaren balioak eta ordenatu-ardatzean maiztasun metatuak adierazten dira, eta barra-sistema baten bidez deskribatzen da batzuen eta besteen arteko egokitasuna. Barra guztiek zabalera berekoak eta ukitzaileak izan behar dute. Barra bakoitzaren goi-segmentuak lotzen direnean, eskailera itxura duen grafiko metakorra lortuko da. Grafiko hori maiztasun metatu erlatiboak erabiliz ere egin daiteke.

1.6 adibidea

Osasun Inkestari erantzun dioten 16 urtetik gorako biztanleen artean mendira joan direnentzat (1.4 adibideko datuak), honako barra-grafiko hau (1.3 irudia) eta grafiko metakor hau (1.4 irudia) eraiki dira, mendira joandako aldi kopururako *SPSS* erabiliz.



1.3 irudia. Mendira joandako aldi kopururako barra-grafikoa.



1.4 irudia. Mendira joandako aldi kopururako grafiko metakorra

Izaera kuantitatiboaren eta aldagai estatistiko jarraituaren kasuan, oro har balio kopurua oso handia izaten denez, adierazpide grafikoak egiteko, datu gordinak erabili ordez, tartetan elkarturik erabiltzen dira. Datu elkartuak 1.6 atalean azaltzen diren arren, kasu honetarako adierazpide grafiko erabilgarrienen deskribapena egingo dugu hemen.

1. **Histograma.** Ardatz kartesiarren gainean, modu honetako errektangelu-sistema eraikitzen da: oinarriak klaseen muga errealek definituko dituzte, eta errektangelu bakoitzak adierazten duen klasearen maiztasun absolutuaren araberako azalera edukiko du. Klase guztiek luzera bera baldin badaukate, errektangeluaren azalera, eta baita altuera ere, maiztasun absolutuaren araberakoak izango dira. Klaseek luzera ezberdinak badauzkate, ordea, errektangelu bakoitzaren altuera maiztasunaren eta klasearen luzeraren arteko zatidura (edo zatiduraren araberakoa) izango da. Klaseek luzera ezberdinak dauzkatenean, histograma irudikatzerakoan, maiz egiten den akatsa da maiztasun absolutuen araberako altuerak eraikitzea. Horrek datuak oker interpretatzera garamatza. Maiztasun erlatiboak erabiliz ere eraiki daiteke histograma.

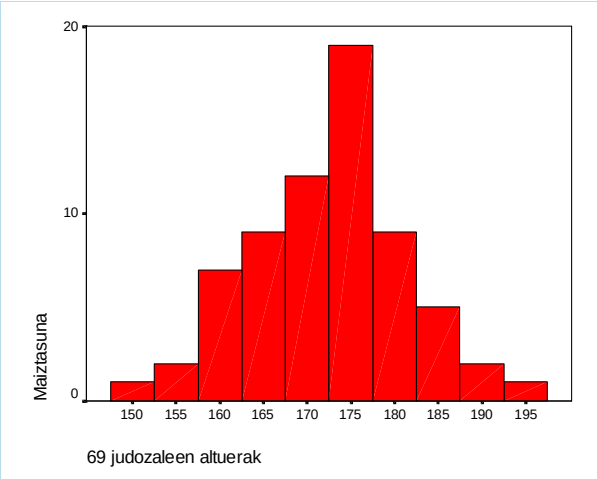
2. **Maiztasun absolutuen poligonoa.** Histogramako errektangeluen goiko oinarrien erdiko puntuak lotzen dituen poligonoa da. Maiztasun erlatiboen poligonoa ere eraiki daiteke. Maiztasun-poligonoak mugatzen duen azalera eta histogramak mugatzen duena berdinak dira (poligonoaren lehenengo klasea eta azkena bereziak dira).

3. **Maiztasun metatuen histograma.** Maiztasun metatuen araberako altuerak eta oinarritzat klase-tarteak dauzkan errektangelu-sistema da. Maiztasun erlatibo metatuen histograma ere egin daiteke.

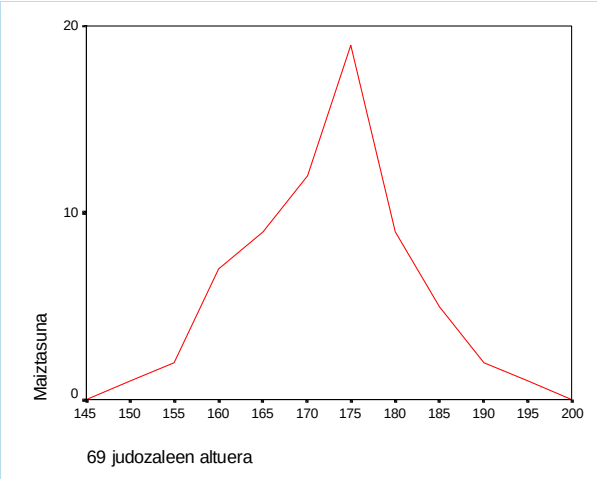
4. **Ojiba edo maiztasun-poligono metakorra.** Poligono hau abzisa-ardatzean klase bakoitzaren goi-muga eta ordenatu-ardatzean klasearen maiztasun metatua jarririk lortzen diren puntuak lotuz eraikitzen da. Hau da, (l_{i+1} , F_i) puntuak lotuz eraikitako lerro poligonala da, l_{i+1} = «T_i klasearen goiko muga erreala» eta F_i = «T_i klaseari dagokion maiztasun metatua» izanik. Ojiba izenaren arrazoia beronen itxura ojiba deritzonarena izatean datza. Maiztasun erlatibo metatuak erabiliz ere egin daiteke ojiba edo poligono metakorra.

1.7 adibidea

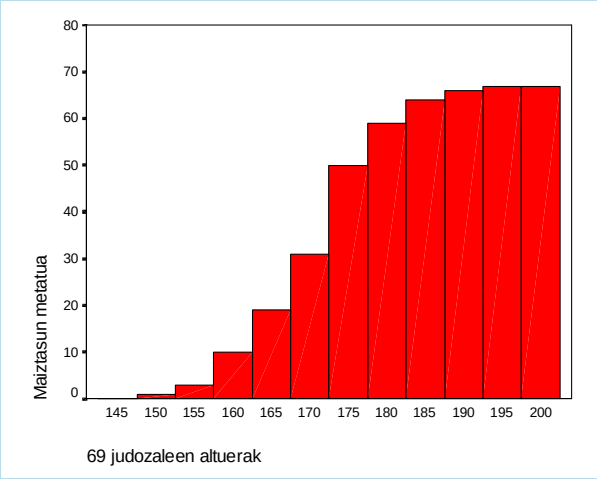
1997ko Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) irudikatzean, 1.5 irudian, 1.6 irudian, 1.7 irudian eta 1.8 irudian agertzen dira, hurrenez hurren, *SPSS*k ematen dizkigun histograma, maiztasun absolutuen poligonoa, maiztasun metatuen histograma eta maiztasun-poligono metakorra.



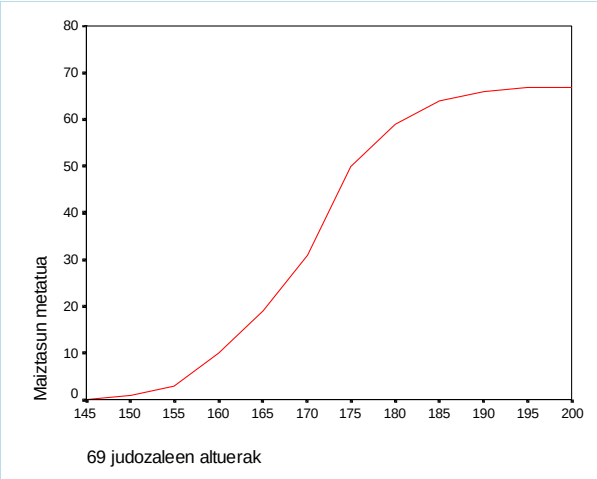
1.5 irudia. Judozaleen altuerarako histograma.



1.6 irudia. Judozaleen altuerarako maiztasun-poligonoa.



1.7 irudia. Judozaleen altuerarako histograma metakorra.



1.8 irudia. Judozaleen altuerarako maiztasun-poligono metakorra.

5. **Zurtoin eta hosto grafikoa (stem & leaf plot):** Adierazpide grafiko hau eraikitzeko honako urrats hauek eman behar dira:

   1. Lehendabizi, aldagaiaren balio bakoitza osagai bitan banatuko da: zurtoin-osagaia eta hosto-osagaia, non zurtoin-osagaia azken zenbakia izan ezik beste guztiek osatzen baitute eta hosto-osagaia azken zenbakiak osaturik baitago. Judozaleen altueraren adibidean, 159 balioaren zurtoin-osagaia 15 eta hosto-osagaia 9 dira. Zurtoin-osagai ezberdin guztiekin zurtoin-osagaien zerrenda prestatuko da.
   2. Balio txikienetik hasita ordenatuko da zurtoin-osagaien zerrenda.
   3. Zerrendan agertzen ez diren tarteko zurtoin-osagaiez osatuko da zerrenda.
   4. Irudia egiten hasteko, grafikoaren ezkerreko goi-muturrean jarriko da lortutako zurtoin-osagai txikiena. Idatzi dugun balioa zurtoin-osagaien zerrendatik kendu, zerrendan geratzen den zurtoin-osagai txikiena zerrendatik atera, eta grafikoan aurrekoaren azpian jarriko da, hurrengo lerroan. Azken pausoa errepikatu egingo da, zurtoin-osagairik handienera iritsi arte.
   5. Zurtoin-osagaien lerro kopuru egokiak datu kopuruaren araberakoa izan behar du. Zurtoin-osagaien lerro kopurua gehitzeko, zurtoin-osagaiak errepikatzen dira, normalean bi, hiru, lau edo bost aldiz. Esate baterako, hurrengo adibidean, bi aldiz errepikatzen dira.
   6. Zurtoin-osagaien eskuinaldean lerro (edo puntu ilada) bertikal bat jarriko da.

7. Aldagaiaren balio bakoitzerako, grafikoan bere zurtoin-osagaia aurkitu, eta horren parean —baina lerro bertikalaren eskuinaldean— bere hosto-osagaia idatziko da. Hurrengo adibidean, aldagaiaren balioen zurtoin-osagaiak bi aldiz errepikatuko direnez, hostoak bi lerrotan idatziko dira. Horrela, hosto multzoak datuen banaketaren itxura orokorra hartuko du.
8. Bukatzeko, errenkada bakoitzeko hosto-osagaiak txikienetik handienera ordenatu behar dira.

Sarritan, zurtoinei dagozkien maiztasunak adierazteko, ezkerreko zutabe bat gehitzen da. Batzuetan, beste zutabe bat ere gehitzen da, sakontasuna adierazteko. Bertan maiztasunak metatzen dira, lehenengo lerrotik behera eta azkenengo lerrotik gora. Lehendabizi, lehen maiztasunetik hasita, maiztasunak beherantz metatzen dira lagin tamainaren erdira iritsi arte. Lortutakoa lagin tamainaren erdia bera bada, prozesua behetik gora errepikatzen da, puntu berera iritsi arte. Lortutakoa lagin tamainaren erdia baino handiagoa bada, ez da idazten eta, haren ordez, klase horren maiztasun absolutua idazten da gakoen artean. Prozesua behetik gora errepikatzen da, zutabea bete arte.

Grafiko mota hori grafikoaren eta taularen nahasketa da: zenbakizko balioak taulan bezala agertzen dira, baina histograma itxura du.

1.8 adibidea

Aurreko adibidearekin jarraituz, honako hauek dira judoa praktikatzen duten 69 pertsonen altuerak (cm-tan):

168	168	168	172	174	170	175	153	178	159
178	183	158	160	180	185	174	181	172	166
194	170	160	175	170	186	180	166	187	175
177	167	177	176	192	177	180	164	175	176
171	173	163	177	172	161	165	160	169	180
160	176	181	154	185	175	180	165	177	181
175	152	165	176	164	178	176	170	182	

1.9 irudian erakusten da datu horien zurtoin eta hosto grafikoa. Ikusten denez, histogramaren itxura bera du.

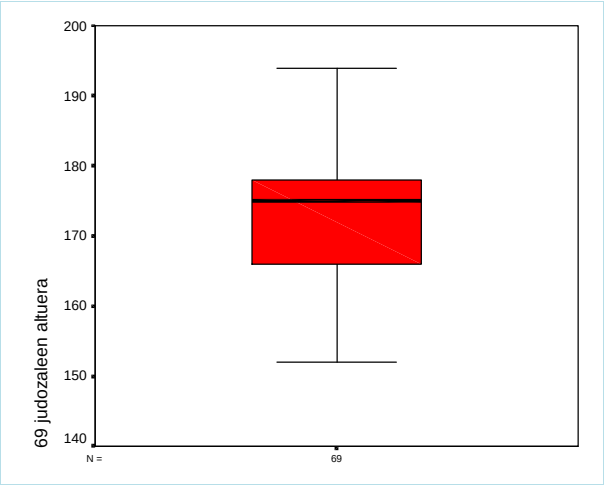
Sakontasuna	Maiztasuna	Zurtoina		Hostoa
3	3	15	*	2 3 4
5	2	15	·	8 9
13	8	16	*	0 0 0 0 1 3 4 4
23	10	16	·	5 5 5 6 6 7 8 8 8 9
34	11	17	*	0 0 0 0 1 2 2 2 3 4 4
(19)	19	17	·	5 5 5 5 5 5 6 6 6 6 6 7 7 7 7 7 8 8 8
16	10	18	*	0 0 0 0 0 1 1 1 2 3
6	4	18	·	5 5 6 7
2	2	19	*	2 4

1.9 irudia. Judozaleen altueren zurtoin eta hosto grafikoa.

6. **Kutxa-diagrama (box-plot):** Adierazpide grafiko hau gai honen amaieran azalduko da, berau eraikitzeko tresna guztiak azaldu ondoren. Edozein kasutan, bere itxura erakutsiko dugu adibide baten bidez.

1.9 adibidea

Aurreko adibidearekin jarraituz, 69 judozaleen altuerak erabiliz, *SPSS*k ematen digun kutxa-diagrama agertzen da 1.10 irudian.



1.10 irudia. Judozaleen altueraren kutxa-diagrama.

1.4 ESTATISTIKOAK

Osasun- edo biologia-azterketa estatistikoaren tratamendurako aztertzen dugun etapa deskribatzailean, laburpen-eredua behar izaten dugu. Eredu horrek datuen azterketa baimenduko digu modu maneiagarrian, eta, bide batez, ondorio posibleak ateratzeko gaitasuna emango digu. Lehenago ikasitako grafikoak laburpen mota bat dira. Hala ere, beharrezkoa da zenbakizko laburpena helburu gehienetan. Aztertutako aldagaiaren ezaugarriak laburki eta zehazki deskribatzen dituzten zenbakizko balioei **estatistiko deskribatzaileak** deritze. Estatistikoak, aldagairen ezaugarrien arabera, era honetan sailkatzen dira: joera zentraleko estatistikoak, posizio-estatistikoak, sakabanatze-estatistikoak eta forma-estatistikoak.

1.4.1 Joera zentraleko estatistikoak

Joera zentraleko estatistikoak, edo **zentralizazio-neurriak**, datuak beren inguruan biltzen direla adierazten duten zenbakizko balioak dira. Sarritan, lagin baten datu ordenatuen zentro inguruan kokatzen dira, baina ez beti. Jarraian azalduko ditugu estatistiko horietariko batzuk.

1.4.1.1 Batezbestekoa

Datu indibidualen baturaren eta laginaren elementu kopuruaren arteko zatidurari **batezbesteko aritmetikoa**, edo **batezbesteko** soilik, deritzo, eta bere adierazpena $\bar{x}$ da. Maiztasun-banaketa baten batezbesteko aritmetikoa banaketa horren grabitate-zentroa da.

Hau da, n aleko lagin bat hartzen badugu eta, bere ezaugarri estatistiko bat azterturik, x_1 , x_2 , ... , x_n balioak lortzen baditugu, honako hau izango da batezbesteko aritmetikoaren adierazpena:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Balioak errepikatuko balira, maiztasunak sortuko lirateke. Demagun k balio ezberdin daudela, hots x_1 , x_2 , ... , x_k, eta horiek f_1 , f_2 , ... , f_k maiztasunez banatzen direla. Orduan, honako hau izango litzateke batezbesteko aritmetikoaren adierazpena:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

1.10 adibidea

Osasun Inkestaren arabera, azken bi asteetan mendira joan diren erantzuleen aldi kopururako (1.4 adibideko datuak) 1.4. taula prestatu da, batezbestekoa

kalkulatzeko behar den zutabea gehituz.

Mendira joandako aldi kopurua (x_i)	f_i	$x_i f_i$
1	89	89
2	99	198
3	23	69
4	25	100
5	6	30
6	4	24
7	6	42
8	1	8
9	1	9
10	8	80
12	1	12
14	7	98
Totala	**270**	**759**

1.4 taula. Mendira joandako aldi kopurua

Datu horietatik abiatuz, azken bi asteetan mendira joan direnen batez besteko aldi kopurua kalkulatuko dugu, modu honetan:

$$\bar{x} = \frac{759}{270} = 2{,}81$$

Horren interpretazioa hau da: EAEn azken bi asteetan, batez beste, 2,81 aldiz (hots, ia 3 aldiz) joan dira mendira.

Batezbesteko aritmetikoaren abantailarik handiena kalkulatzeko erraztasuna da. Oztopoa, ordea, lagin-fluktuazioekiko sentikorra izan daitekeela da, muturretan jartzen diren balio arraroek eragina daukatelako batezbestekoan.

Batezbesteko aritmetikoa da, dudarik gabe, zenbaki multzoaren batez besteko joera adierazteko maizen erabiltzen den eta gehien ezagutzen den estatistikoa, baina ez da bakarra. Beste batezbesteko mota batzuk ere erabilgarriak dira, arlo batzuetan bereziki. Horregatik, batezbestekoaren beste bi aldaera azalduko ditugu atal honetan: batezbesteko haztatua eta batezbesteko geometrikoa, hain zuzen.

Elementu multzo baten batezbesteko aritmetikoa kalkulatzeko elementuen garrantzia kontuan hartzen bada, horrela lortutako batezbesteko aritmetikoari **batezbesteko aritmetiko haztatua** esaten zaio, eta $\bar{x}_{HAZ}$ adierazten da. Lagin baten elementu

bakoitzari ematen diogun garrantziaren neurria **pisua** da, eta p_i-z adierazten da. Batezbesteko haztatua kalkulatzeko honako formula hau erabiliko dugu:

$$\bar{x}_{HAZ} = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i}$$

1.11 adibidea

Demagun ikasgela bateko bi ikaslek bai hiruhilekoetan bai ekaineko azterketetan lortu dituzten puntuazioak 1.5 taulan agertzen direla.

	***A* ikaslearen puntuazioak**	***B* ikaslearen puntuazioak**
1. hiruhil.	5	5
2. hiruhil.	2	7
3. hiruhil.	4	6
ekaina	7	1

1.5 taula. Ikasleen puntuazioak. Iturria: A. Mtz. Lizardui. *Estatistika Deskribatzailea*. Elhuyar Matematika Taldea, 1983.

Ikasgai horretan lortu duten batez besteko puntuazioa ezagutu nahi dugu, ikasgaia gainditu duten ala ez jakiteko. Batez besteko puntuazioak honako hauek dira:

$$\bar{x}_A = \frac{5+2+4+7}{4} = 4.5 \qquad \bar{x}_B = \frac{5+7+6+1}{4} = 5$$

Beraz, emaitza horien arabera, lehenak ez du ikasgaia gainditu eta bigarrenak bai. Baina berehalaxe konturatuko gara emaitza horiek ez direla bidezkoak, hiruhilero egindako azterketek eta ekainean egindakoak ez baitute garrantzi bera. Garrantzia pisuaren bidez emango diegu eta, horretarako, 1.6 taula eratuko dugu.

	***A* ikaslearen puntuazioak**	***B* ikaslearen puntuazioak**	**Pisuak**
1. hiruhil.	5	5	1
2. hiruhil.	2	7	1
3. hiruhil.	4	6	1
ekaina	7	1	3
			6

1.6 taula. Ikasleen puntuazioak, pisuak barne.

Orain, ikasle bakoitzari dagokion batezbesteko haztatua kalkulatuko dugu:

$$\bar{x}_A = \frac{5\cdot 1 + 2\cdot 1 + 4\cdot 1 + 7\cdot 3}{6} = 5{,}33 \qquad \bar{x}_B = \frac{5\cdot 1 + 7\cdot 1 + 6\cdot 1 + 1\cdot 3}{6} = 3{,}67$$

Orduan, nahiz eta *B* ikasleak *A* ikasleak baino puntuazio hobeak atera dituela dirudien, azkenean, *A* ikasleak batezbesteko haztatu hobea du, zeren eta garrantzi handiena (pisu handiena) duen ekaineko azterketan *B* ikasleak baino askoz puntuazio hobea lortu baitu.

Batezbesteko geometrikoa da arlo batzuetan maiz erabiltzen den batezbestekoaren aldaera bat. Horrela definitzen da: *n* zenbakiz osaturiko multzo baten **batezbesteko geometrikoa** zenbaki horien biderkaduraren *n* erroa da, eta $\bar{x}_G$ adierazten da. Hots,

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Aldagaiaren balioen logaritmoen batezbesteko aritmetikoa da batezbesteko geometrikoaren logaritmoa. Beste era batera esanda, batezbesteko geometrikoaren logaritmoa da aldagaiaren balioen logaritmoen batezbesteko aritmetikoa. Hau da,

$$\log(\bar{x}_G) = \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$$

Horregatik, logaritmotan adierazten diren aldagaien deskribapena egiteko aproposa da, esate baterako laborategiko datuetan. Arrazoi edo ehunekoen batezbestekoa ateratzeko ere batezbesteko aritmetikoa baino egokiagoa da; horregatik, esate baterako ekonomian, erabilgarria da.

1.4.1.2 Mediana

Aldagai estatistikoaren balioak txikienetik handienera ordenaturik daudenean, **mediana** da alde bakoitzean balio kopuru erdia uzten duen balioa, eta *Me* adierazten da.

Aldagai estatistikoa diskretua denean, edo jarraitua izanik datu gordinak erabiltzen direnean, honako kasu hauek azter daitezke:

1. Balio kopuru bakoitia. (n+1)/2 tokia betetzen duen balioa da mediana.
2. Balio kopuru bikoitia. Kasu honetan, ez dago erdiko tokia betetzen duen baliorik, eta n/2 eta (n/2)+1 toki zentralak betetzen dituzten balioen arteko batezbesteko aritmetiko gisa kalkulatuko da mediana.

1.12 adibidea

Osasun Inkestaren erantzuleen artean, 69 judozaleen altueren (1.8 adibideko

datuak) bi azpitalde aukeratu ditugu:

a) Lehenengo zazpi elementuak, ordenaturik: 168 – 168 – 168 – 170 – 172 – 174 – 175

$Me = 170$ cm

b) Lehenengo zortzi elementuak, ordenaturik: 153 – 168 – 168 – 168 – 170 – 172 – 174 – 175

$$Me = \frac{168+170}{2} = 169 \text{ cm}$$

Datu kopurua handia denean, mediana kalkulatzeko era erosoa zurtoin eta hosto grafikoan agertzen den sakontasuna erabiltzea da. Azal dezagun adibide baten bidez.

1.13 adibidea

Judoa praktikatzen duten 69 pertsonen altueren mediana kalkulatu nahi dugu. Horretarako, 1.8 adibidean eratutako zurtoin eta hosto grafikoa begiratuko dugu.

Medianaren sakontasuna $\frac{n+1}{2} = \frac{69+1}{2} = 35$ da.

Beraz, 35. datua izango da mediana. Bosgarren klasearen azken balioa 34.a denez, mediana 6. klasearen lehenengo balioa izango da, hots, 175.

Medianaren abantaila nagusiak dira kalkulatzeko oso erraza izatea eta lagin-fluktuazioekiko, edo balio arraroekiko, sentikorra ez izatea. Medianaren kalkuluan, datuak beren ordena finkatzeko soilik erabiltzen dira; beraz, datuen balioek ez dute medianan eraginik. Banaketa asimetrikoak ditugunean, mediana da joera zentraleko estatistikorik adierazgarriena, eta kasu horretan batezbestekoaren ordez erabiltzea gomendatzen da.

1.4.1.3 Moda

Aldagai estatistikoaren balioen artean gehien errepikatzen dena **moda** da, eta *Mo* adierazten da. Hots, maiztasun handiena duen aldagaiaren balioa da. Moda ez da zertan bakarra izan. Bi moda dauzkan banaketari bimodala deritzo. Halaber, $\cup$ forma daukan banaketaren beheko puntuari **antimoda** deritzo.

1.14 adibidea

Judoa praktikatzen duten 69 pertsonen altueren moda kalkulatu nahi dugu. Horretarako, 1.8 adibidean eratutako zurtoin eta hosto grafikoa aztertuko dugu. Maiztasun handiena daukan aldagaiaren balioa 175 da, sei aldiz errepikatzen delako. Beraz, moda 175 cm-ko altuera da.

Modaren kalkuluan lagineko datu bakarra erabiltzen denez, informazio asko galtzen da. Nahiz eta joera zentraleko neurria izan, modak ez du zertan erdialdean egon; banaketaren muturretan egon daiteke. Moda da literatura zientifikoan gutxien agertzen den joera zentraleko estatistikoa, eta ia inoiz ez da hari buruzko inferentzia estatistikorik egiten, bere tratamendu matematikoa zaila izaten baita.

1.4.2 Posizio-estatistikoak

Mediana da alde bakoitzean balioen % 50 uzten duen aldagaiaren balioa, hots, balioen posizioarekiko balio zentrala. Medianaz gainera, zenbait posizio adierazten dizkiguten beste estatistiko batzuk ere defini ditzakegu. Lagin ordenatua 100 zati berdinetan banatzen duten aldagaiaren balioei **pertzentil** deritze. P_i edo ***i.* pertzentila** da bere ezkerrean balioen % i uzten duen aldagaiaren balioa. Pertzentil batzuek izen bereziak dituzte. Honela, hamargarren pertzentila $P_{10} = D_1$ lehenengo **dezila** da, eta $P_{20} = D_2$ (bigarren dezila), $P_{30} = D_3$, $P_{40} = D_4$, ... , $P_{90} = D_9$, $P_{100} = D_{10}$. Beste pertzentilen artean, erabilgarrienak P_{25}, P_{50}, eta P_{75} dira, eta beren izen bereziak **koartilak** dira: lehenengo koartila Q_1, bigarren koartila Q_2, eta hirugarren koartila Q_3. Hain zuzen, $Me = Q_2 = D_5 = P_{50}$.

Pertzentilen kalkulua medianaren kalkuluaren antzera egiten da. Maiztasun metatuen zutabeari begiratuz, $(i \cdot n)/100$ balioa edo hori baino handiagoa den lehenengo maiztasunari dagokion balioa i. pertzentila da.

Gero eta hedatuago dauden beste posizio-estatistiko batzuk **Tukey-ren posizio-estatistikoak** dira, eta biren berreduretarako kalkulatzen dira. Hots, erdia (mediana), laurdenak, zortzirenak, etab. Horien kalkulua zurtoin eta hosto grafikoan definitu dugun sakontasunean dago oinarrituta.

1.15 adibidea

Judoa praktikatzen duten 69 pertsonen altueren 1. eta 3. koartilak, 2. dezila eta 35. pertzentila kalkulatuko ditugu.

Q_1: $\frac{i \cdot n}{100} = \frac{25 \cdot 69}{100} = 17{,}25$. Balio hori bera edo handiagoa den lehenengo maiztasun metatua 18 da, eta 166 balioari dagokio. Beraz, $Q_1 = 166$ da.

Q_3: $\frac{i \cdot n}{100} = \frac{75 \cdot 69}{100} = 51{,}75$. Balio hori bera edo handiagoa den lehenengo maiztasun metatua 53 da, eta 178 balioari dagokio. Beraz, $Q_3 = 178$ da.

D_2: $\frac{i \cdot n}{100} = \frac{20 \cdot 69}{100} = 13{,}8$. Balio hori bera edo handiagoa den lehenengo maiztasun metatua 16 da, eta 165 balioari dagokio. Beraz, $D_2 = 165$ da.

P_{35}: $\frac{i \cdot n}{100} = \frac{35 \cdot 69}{100} = 24{,}15$. Balio hori bera edo handiagoa den lehenengo maiztasun metatua 27 da, eta 170 balioari dagokio. Beraz, $P_{35} = 170$ da.

Halaber, Tukey-ren laurdenak era honetan kalkulatuko ditugu. Laurdenaren sakontasuna $\frac{([35]+1)}{2} = 18$ da, non [35] medianaren sakontasunaren zati osoa baita. Orduan, 18. datua, goitik eta behetik hasita, laurdenak 166 eta 178 cm dira.

1.4.3 Sakabanatze-estatistikoak

Posizioko eta joera zentraleko estatistikoek ez digute ematen laginaren kontzentrazioari buruzko informaziorik. Gerta daiteke joera zentraleko estatistiko berdinak dituzten bi lagin aukeratzea, non baten balioak bestearenak baino bilduagoak baitira.

1.16 adibidea

Medikuntza Fakultatean, 2. mailako zoriz aukeratu diren bost ikasleko laginaren adinak aztertzean, 17, 18, 18, 19 eta 23 balioak lortu ditugu. Beren batezbestekoa 19 da. Medikuntza Fakultateko 2. mailan zoriz aukeratu diren beste bost ikasleko laginaren adinak, berriz, 18, 19, 19, 19 eta 20 izan dira. Horien batezbestekoa ere 19 da, baina bilduagoak dira balioak. Bi laginen arteko ezberdintasun hori nabarmentzeko, sakabanatzea adierazten duten neurriak erabili behar dira.

Aldagaiak duen balio zentralarekiko kontzentrazioa adierazten duten neurriei **sakabanatze-estatistiko** deritze. Datuek aldakuntzarik ez daukatenean, hau da, zenbaki bakar batean kontzentraturik daudenean, zero izango da edozein sakabanatze-estatistikoren balioa, eta, beste kasuetan, balio positiboak bakarrik hartzen dituzte. Zenbat eta kontzentrazio txikiagoa izan, orduan eta handiagoa izango da sakabanatze-estatistikoen balioa.

1.4.3.1 Heina

Heina, anplitudea edo **aldakuntza-tartea** da sakabanatze-neurririk sinpleena eta intuitiboena. Laginaren balio handienaren eta txikienaren arteko diferentziatzat definitzen da, eta haren adierazpena R da. Hots, $R = X_{max} - X_{min}$.

Heinak laginaren % 100aren aldakuntza adierazten du. Heinaren eragozpenik garrantzizkoenak hauexek dira: bi balioren menpean bakarrik dagoela, eta, laginaren tamaina handitzerakoan, handitu egiten dela. Interes gutxi dauka, balio arraroek edo akastunek eragin handia baitaukate, baina, kalkulatzeko erraza denez gero, batzuetan erabiltzen da.

1.17 adibidea

Honako hauek dira aurreko 1.16 adibideko laginen heinak:

$R_1 = 23 - 17 = 6$

$R_2 = 20 - 18 = 2$

1.4.3.2 Koartilarteko heina

Koartilarteko heina, aldakuntza edo **anplitudea** hirugarren eta lehen koartilen arteko kendura gisa definitzen da eta *RI* adierazten da. Hots, $RI = Q_3 - Q_1$.

Koartilarteko heinak laginaren % 50aren aldakuntza eskaintzen digu. Hori dela eta, muturretako balio arraroek ez dute eraginik izango bere balioan, eta estatistiko hori heina baino sendoagoa izango da.

1.4.3.3 Batezbestekoarekiko desbideratzeak

Aldagaiaren balioen eta batezbesteko aritmetikoaren arteko diferentziei **batezbestekoarekiko desbideratze** deritze. Hots, aldagaiaren balioak $x_1, x_2, \ldots, x_n$ eta batezbestekoa $\bar{x}$ baldin badira, batezbestekoarekiko desbideratzeak $(x_1 - \bar{x})$, $(x_2 - \bar{x})$, . . . , $(x_n - \bar{x})$ dira. Batezbestekoarekiko desbideratzeen batura zero da. Hots, $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

1.4.3.4 Bariantza, desbideratze estandarra eta kuasibariantza

Laginaren **bariantza** batezbestekoarekiko desbideratze karratuen batezbestekoa da, eta s_n^2 adierazten da. Matematikoki, datuak banaka edo maiztasunetan bilduta egotearen menpe dago bere adierazpena. Hots,

<u>Datuak banaka</u> erabiliz: $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ edo $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$

<u>Maiztasunak</u> erabiliz: $s_n^2 = \frac{1}{n}\sum_{i=1}^{k}f_i(x_i - \bar{x})^2$ edo $s_n^2 = \frac{1}{n}\sum_{i=1}^{k}f_i x_i^2 - \bar{x}^2$

Bariantzaren erabilera matematikoa oso erraza denez, horixe da gehien erabiltzen denetariko sakabanatze-neurria. Bere eragozpenik garrantzizkoena aldagaien unitate karratuetan adieraztean datza.

1.18 adibidea

Hauexek dira Medikuntza Fakultatean 2. mailan zoriz aukeratu diren bost ikasleko bi laginen (1.16 adibideko datuak) bariantzak:

Lehen lagina:

$$s_n^2 = \frac{(17-19)^2 + (18-19)^2 + (18-19)^2 + (19-19)^2 + (23-19)^2}{5} = 4{,}40$$

$$s_n^2 = \frac{17^2 + 18^2 + 18^2 + 19^2 + 23^2}{5} - 19^2 = 4{,}40$$

Bigarren lagina:

$$s_n^2 = \frac{(18-19)^2 + (19-19)^2 + (19-19)^2 + (19-19)^2 + (20-19)^2}{5} = 0{,}40$$

$$s_n^2 = \frac{18^2 + 19^2 + 19^2 + 19^2 + 20^2}{5} - 19^2 = 0{,}40$$

Laginaren **desbideratze estandarra** bariantzaren erro karratu positiboa da eta s_n adierazten da. Hots, $s_n = +\sqrt{s_n^2}$. Bariantzaren abantaila berak ditu, eta, gainera, kasu honetan, aldagaiaren unitate berdinetan adierazten da. Oro har, **sakabanatze absolutua** neurtzeko gehien erabiltzen den estatistikoa da.

1.19 adibidea

Hauexek dira Medikuntza Fakultatean 2. mailan zoriz aukeratu diren bost ikasleko bi laginen (1.16 adibideko datuak) desbideratze estandarrak:

Lehen lagina: $s_n = \sqrt{4{,}40} = 2{,}10$

Bigarren lagina: $s_n = \sqrt{0{,}40} = 0{,}63$

Laginaren bariantza definitzerakoan, datuen batezbestekoarekiko desbideratze karratuen batezbestekoa dela esan dugu, eta beronen kalkulurako lagineko n datuak erabili ditugu (izendatzaileko balioa n izan da). Praktikan, n balioak eta beren batezbestekoa erabiltzen direnez, bariantzaren kalkulurako zera gertatzen da: $(n + 1)$ balio horiek ez direla askeak. Nolabait, esan dezakegu, n balioen batezbestekoa ezagutzen badugu, n balioetatik bat ez ezagutu arren ez genukeela informaziorik galduko. Hori dela eta, maiz bariantzaren beste aldaera bat erabiltzen da, kuasibariantza deritzona, eta formularen izendatzailean $(n - 1)$ erabiltzen du n erabili beharrean. Bariantzaren eta kuasibariantzaren arteko diferentzia oso txikia izaten da, askotan mespretxagarria, batez ere laginaren tamaina handia bada.

Testuliburu askotan ez da egiten bariantzaren eta kuasibariantzaren arteko bereizketarik; are gehiago, bata zein bestea erabiltzen dira maiz bariantzaz hitz egiterakoan. Hori dela eta, testu honetan biak definitzea erabaki dugu, nahiz eta testu osoan zehar bataren eta bestearen arteko diferentzia bereiziko dugun.

Laginaren **kuasibariantza** s_{n-1}^2 adierazten da, eta modu honetara kalkulatzen da:

<u>Datuak banaka</u> erabiliz: $s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ edo $s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{n}{n-1}\bar{x}^2$

<u>Maiztasunak</u> erabiliz: $s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{k}f_i(x_i - \bar{x})^2$ edo $s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{k}f_i x_i^2 - \frac{n}{n-1}\bar{x}^2$

Definizioan ikus daitekeenez, bariantzaren eta kuasibariantzaren artean bereizteko adierazpenaren azpindizean n edo $(n - 1)$ erabiliko dugu testu osoan zehar. Honako hau da bariantzaren eta kuasibariantzaren arteko erlazioa: $(n-1)\cdot s_{n-1}^2 = n \cdot s_n^2$.

1.4.3.5 Aldakuntza-koefizientea

Laginaren **aldakuntza-koefizientea** desbideratze estandarraren eta batezbestekoaren arteko zatidura da (ehunekotan emanda), eta *AK* adierazten da. Hots,

$$AK = \frac{s_n}{\bar{x}} \cdot 100$$

Dimentsiorik gabeko kantitatea da, eta **sakabanatze erlatiboa** neurtzen du. Salbuespen bakarra batezbesteko aritmetikoak zero inguruko balioak hartzen dituenean da. Kasu horretan, formulan ikusten den bezala, aldakuntza-koefizientea infinitorantz doa; orduan, ez da sakabanatzearen adierazgarri eta ez da sakabanatzea neurtzeko estatistiko egokia.

Neurri hori sakabanatzeak konparatzeko erabiltzen da. Horren bidez, honako hauek konpara daitezke: lagin ezberdinetan neurturiko aldagai berberaren sakabanatzeak, eta baita lagin berbereko bi izaera ezberdinen sakabanatzeak ere. Hurrengo adibideetan aldagaien sakabanatzearen konparazioak aurkeztuko dira, lagin berdinetan eta ezberdinetan.

1.20 adibidea

Demagun animaliaz osatutako bi lagin, bata elefantez eta bestea txindurriz osatua. Bi kasuetan, aldagaia animaliaren pisua izango da. Elefante taldearen batezbestekoaren balioa txindurri taldearena baino askoz handiagoa izango da,

elefanteen pisuak, noski, txindurrienak baino askoz handiagoak direlako. Jo dezagun, baita ere, sakabanatze absolutu berdina dutela bi lagin horiek (desbideratze estandarraren bidez neurtua). Datu horiek 1.7 taulan agertzen dira.

(Fikziozko datuak)	Elefanteak (pisua)	Txindurriak (pisua)
Batezbestekoa	6000 Kg	2 g
Desbideratze estandarra	1 g	1 g
Aldakuntza-koefizientea	% 0,0000167	% 50

1.7 taula. 1.20 adibideko datuak. Iturria: A. Mtz. de Lizardui. *Estatistika Deskribatzailea.* Elhuyar Matematika Taldea, 1983.

Sakabanatzea neurtzeko desbideratze estandarrera jotzen badugu, bi lagin horiek sakabanatze berdina dute. Baina elefanteen laginean, aldagaiak hartzen dituen balioak oso handiak izanik, gramo bateko sakabanatzea edukitzea oso txikia da. Alderantziz, txindurrien laginean aldagaiek hartzen dituzten balioak oso txikiak izanik, gramo bateko sakabanatzea edukitzea handia izan liteke.

Kontraesan galanta sortzen da sakabanatzea kalkulatzean, ez bada kontuan hartzen aldagaiak neurtzen dituen kantitateen handitasuna. Hori ekiditeko aldakuntza-koefizientea erabiltzen da, desbideratze estandarra beroni dagokion batezbesteko aritmetikoari egokituta agertzen baita.

Lehenengo laginaren desbideratze estandarrak eta bigarrenak balio berdinak izan arren, sakabanatze zeharo ezberdinak azaltzen dizkigute. Lehenengo laginean sakabanatzerik ez dagoela esan daiteke, eta, bigarrenean, oso sakabanatze handia agertzen da.

1.21 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) laburbiltzen dira, beren sexua kontuan hartuz, 1.8 taulan.

Altuera	Gizonezkoak	Emakumezkoak
Batezbestekoa	176,14 cm	163,26 cm
Desbideratze estandarra	6,79 cm	7,10 cm
Aldakuntza-koefizientea	% 3,85	% 4,35

1.8 taula. 1.21 adibideko judozaleen altuerak.

Kasu horretan, bai sakabanatze absolutua bai erlatiboa handiagoak dira emakumezkoen kasuan.

1.22 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen pisuak (kg-tan) eta altuerak (cm-tan) laburbiltzen dira 1.9 taulan.

Judozaleak	Pisua	Altuera
Batezbestekoa	69,93 kg	172,59 cm
Desbideratze estandarra	11,86 kg	8,95 cm
Aldakuntza-koefizientea	% 16,96	% 5,19

1.9 taula. 1.22 adibideko judozaleen altuerak eta pisuak.

Altuera eta pisu izaeren sakabanatzeak konparatzea ez da zilegi, altuera zentimetrotan eta pisua kilogramotan ematen baitira. Arazo hori gainditzeko, sakabanatze-neurri adimentsionala den aldakuntza-koefizientea erabiltzen da. Beraz, judozaleen pisuak altuerak baino sakabanatze handiagoa, absolutua zein erlatiboa, duela esan dezakegu.

1.4.4 Forma-estatistikoak

Forma-estatistikoak banaketaren itxura adierazten duten zenbakizko balioak dira; alegia, banaketaren adierazpide grafikoa egin gabe, itxura (simetria edo zorroztasuna) ezagutzea baimentzen digutenak.

1.4.4.1 Alborapena edo simetria

Maiztasun-banaketa **alboragabea** dela esango dugu, neurri zentral batekiko distantzia berean dauden balioen maiztasunak berdinak baldin badira. Beste kasuetan banaketa alboratua izango da. Maiztasun-banaketak eskuinerantz isats luzeagoa erakusten baldin badu, **eskuinerantz alboratua** deritzo; ezkerrerantz isats luzeagoa erakusten baldin badu, ordea, **ezkerrerantz alboratua**.

Alborapena, edo Pearson-en lehen alborapen-koefizientea deritzona, banaketaren simetria neurtzen duen estatistiko bat da; υ letraren bidez adierazten da eta matematikoki modu honetara kalkulatzen da:

$$\upsilon = \frac{\bar{x} - Mo}{s_n}$$

Kontuan hartu behar dugu alborapena bakarrik kalkula daitekeela moda bakar bat eta kanpai-itxura duten banaketen kasuan. Alborapenaren interpretazioa hau da:

- $\upsilon < 0$ baldin bada, banaketa ezkerrerantz alboratua da.
- $\upsilon \approx 0$ baldin bada, banaketa simetrikoa edo alboragabea da.
- $\upsilon > 0$ baldin bada, banaketa eskuinerantz alboratua da.

Definitu dugun alborapena, berez, Pearson-en lehenengo alborapen-koefizientea da. Horretaz gainera, honako beste hauek ditugu Pearson-en alborapen-koefizienteen artean:

Bigarren alborapen koefizientea: $\upsilon_2 = \dfrac{3 \cdot (\bar{x} - Me)}{s_n}$

Koartilarteko alborapena: $\upsilon_K = \dfrac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \dfrac{Q_3 - 2 \cdot Q_2 + Q_1}{Q_3 - Q_1}$

10-90 pertzentilen arteko alborapena: $\upsilon_P = \dfrac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})} = \dfrac{P_{90} - 2 \cdot P_{50} + P_{10}}{P_{90} - P_{10}}$

Alborapen-momentua: $a_3 = \dfrac{m_3}{s_n^3} = \dfrac{m_3}{\left(\sqrt{m_2}\right)^3}$,

non $m_r = \dfrac{\sum_{j=1}^{n} (x_j - \bar{x})^r}{n}$ r. momentu zentrala baita ($m_1 = 0$).

Eredu gisa erabiltzeko, *SPSS*k kalkulatzen duen alborapen-neurria a_3 da. Alborapenarekin gertatzen den era berean, $a_3 \approx 0$ denean, banaketa simetrikoa edo alboragabea da.

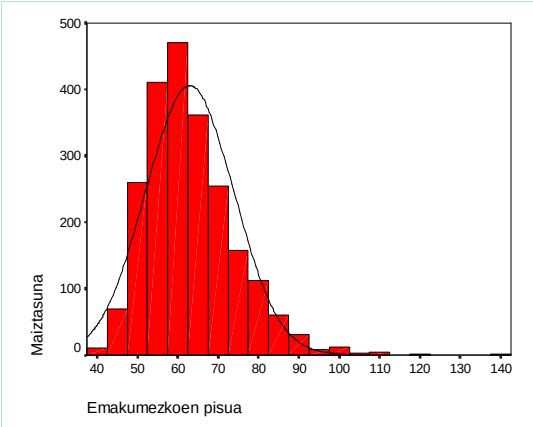
1.23 adibidea

Osasun Inkestako datuak erabiliz, kg-tan emandako emakumezkoen pisuen banaketak honako alborapen hau erakusten du:

$$\upsilon = \frac{\bar{x} - Mo}{s_n} = \frac{62{,}9739 - 60}{10{,}93} = 0{,}27 > 0$$

$a_3 = 1{,}033 > 0 \Rightarrow$ eskuinerantz alboratua (+)

1.11 irudian, datu horien histograma erakusten da, eta bertan, eskuin isatsari begiratuz, banaketa eskuinerantz alboratua dela susma daiteke.



1.11 irudia. Osasun Inkestako emakumezkoen pisuaren histograma, eskuinerantz alboratua.

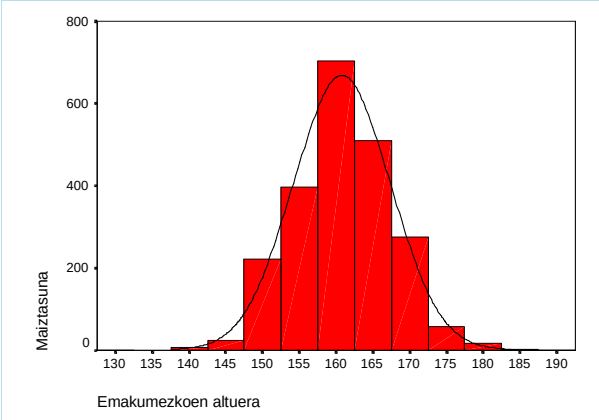
1.24 adibidea

Osasun Inkestako datuak erabiliz, cm-tan emandako emakumezkoen altueren banaketak honako alborapen hau erakusten du:

$$\upsilon = \frac{\bar{x} - Mo}{s_n} = \frac{160{,}7673 - 160}{6{,}6} \approx 0$$

$a_3 = 0{,}064 \approx 0 \Rightarrow$ alboragabea

1.12 irudian, datu horien histograma erakusten da, eta, isats bien luzeraren arabera, banaketa alboragabea dela susma daiteke.



1.12 irudia. Osasun Inkestako emakumezkoen altueraren histograma, alboragabea.

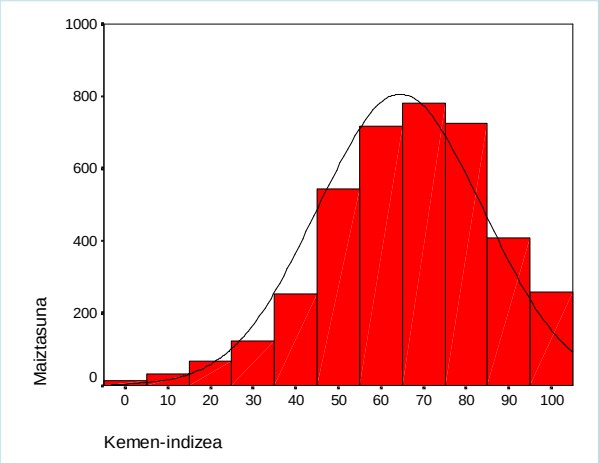
1.25 adibidea

Osasun Inkestako datuak erabiliz, bizi-kalitatea neurtzen duen kemen-indizeak honako alborapen hau erakusten du.

$$\upsilon = \frac{\bar{x} - Mo}{s_n} = \frac{65{,}5132 - 70}{19{,}3644} = -0{,}2833 < 0$$

$a_3 = -0{,}429 < 0 \Rightarrow$ ezkerrerantz alboratua (–)

1.13 irudian, datu horien histograma erakusten da, eta bertan, ezker isatsari begiratuz, banaketa ezkerrerantz alboratua dela susma daiteke.



1.13 irudia. Osasun Inkestako erantzuleen kemen-indizearen histograma, ezkerrerantz alboratua.

1.4.4.2 Kurtosia edo zorroztasuna

Kurtosiak kanpai-itxura duten banaketen zorroztasuna neurtzen du. Aztertzen den banaketak Gauss-en kanpai edo kurba normalak baino tontor altuagoa badu, zorrotza izango da eta **leptokurtikoa** dela esaten da (kurtosi normala baino handiagoa); normalaren antzekoa bada, **mesokurtiko** deritzo (kurtosi normala); eta banaketa normala baino leunagoa bada, **platikurtikoa** da (kurtosi normala baino txikiagoa). Kurtosi-neurrien artean, honako hauek ditugu:

Kurtosi-momentua: $a_4 = \frac{m_4}{s_n^4} = \frac{m_4}{m_2^2}$

Pertzentilen kurtosi-koefizientea: $\kappa = \frac{RI/2}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2 \cdot (P_{90} - P_{10})}$

Gauss-en kanpaiaren kurtosi-neurriak $a_4 = 3$ eta $\kappa = 0{,}263$ direnez, balio horiekin konparatzen dira banaketa baten kurtosi-neurriak, leptokurtikoa, mesokurtikoa edo platikurtikoa dela egiaztatzeko.

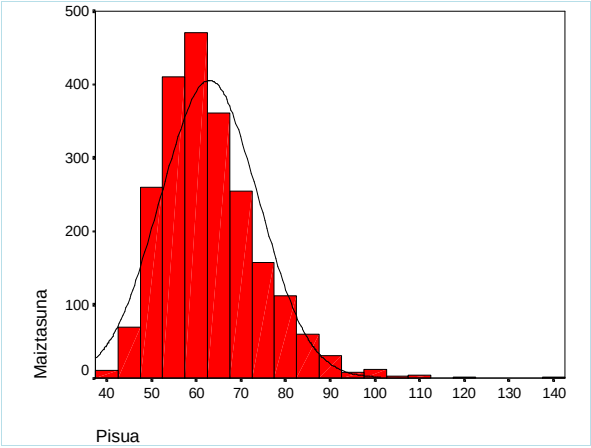
Berriro ere, eredu gisa erabiltzeko, *SPSS*k kalkulatzen duen kurtosi-koefizientea ($a_4 - 3$) da. Beraz, konparazioa, kasu horretan, 0 balioarekiko egin beharko dugu.

1.26 adibidea

Osasun Inkestako datuak erabiliz, kg-tan emandako emakumezkoen pisuen banaketak honako kurtosi-koefiziente hau erakusten du:

$a_4 - 3 = 2{,}169 > 0$

1.14 irudian, datu horien histograma erakusten da. Kurtosi normala baino handiagoa denez, histogramaren itxura banaketa normalaren kurba baino zorrotzagoa eta altuagoa da; beraz, banaketa leptokurtikoa da.



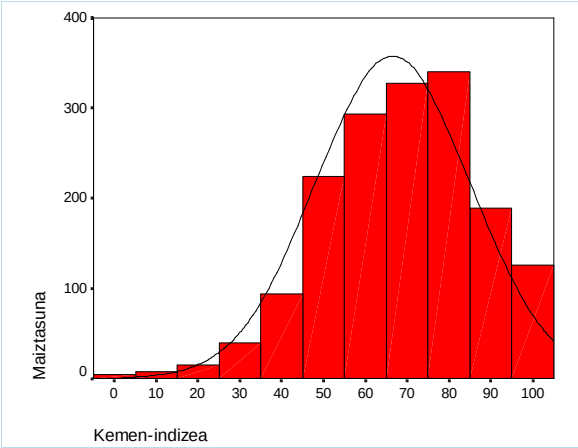
1.14 irudia. Osasun Inkestako emakumezkoen pisuaren histograma, banaketa leptokurtikoa.

1.27 adibidea

Osasun Inkestako datuak erabiliz, bizi-kalitatea neurtzen duen kemen-indizearen banaketak honako kurtosi hau erakusten du gizonengan:

$a_4 - 3 = 0{,}048 \approx 0$

1.15 irudian, datu horien histograma erakusten da. Kurtosia gutxi gorabehera normalarena denez, histogramaren eta kurba normalaren itxurak antzekoak dira; beraz, banaketa mesokurtikoa da.



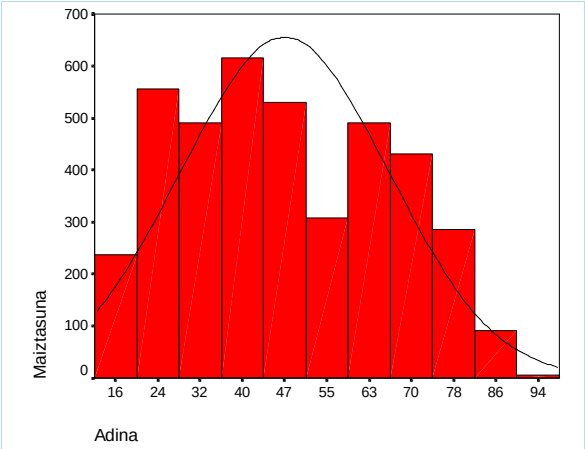
1.15 irudia. Osasun Inkestako gizonezkoen kemen-indizearen histograma, banaketa mesokurtikoa.

1.28 adibidea

Osasun Inkestako datuak erabiliz, gizabanakoen adinaren banaketak honako kurtosia hau erakusten du.

$a_4 - 3 = -1{,}055 < 0$

1.16 irudian, datu horien histograma erakusten da. Kurtosia normala baino txikiagoa denez, histogramaren itxura banaketa normalaren kurba baino leunagoa eta baxuagoa da, beraz, banaketa platikurtikoa da.



1.16 irudia. Osasun Inkestako erantzuleen adinaren histograma, banaketa platikurtikoa.

1.5 KUTXA-DIAGRAMAK

Aurreko 1.3. atalean adierazpide grafikoak deskribatu ditugunean, kutxa-diagrama aurkeztu dugu, nahiz eta nola egiten den ez dugun zehaztasunez azaldu. Estatistiko guztiak azaldu ondoren, kutxa-diagrama eraikitzeko behar diren tresna guztiak ezagutzen ditugu. Kutxa-diagrama da datu multzo baten posizioaren ikus-pertzepzioa errazten duen adierazpide grafikoa. Oso erabilgarria da hainbat datu multzoren banaketa konparatu nahi denean. *Outlier* deritzon balio arraro edo muturrekoak aurkitzea ere baimentzen digu.

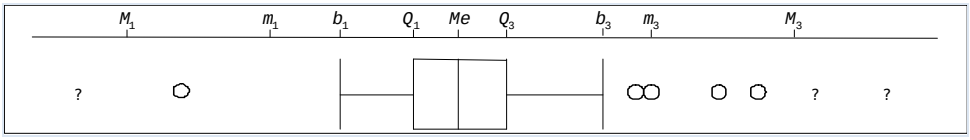
Kutxa-diagrama edozein software estatistikoren bidez eraiki daitekeen arren, pausoka azalduko dugu bere eraikuntza. Kutxa horizontala edo bertikala irudika daiteke, eta, horregatik, gure diagramaren norabidea aukeratu beharko dugu hasteko. Aukeratu eta gero, erreferentzia-eskala kutxaren luzeraren norabidearen arabera irudikatu behar da, paraleloan. Hauexek dira kutxa-diagrama eratzeko urratsak:

1. Koartilak eta koartilarteko heina kalkulatzen ditugu.
2. **Barne-hesi** deritzen m_1 eta m_3 puntuak kalkulatuko ditugu, era honetan:
   $m_1 = Q_1 - 1{,}5RI$
   $m_3 = Q_3 + 1{,}5RI$

   Puntu horiek *outlier*ak detektatzeko erabiliko ditugu.
3. **Albo-balio** deritzen b_1 eta b_3 balioak kalkulatuko ditugu: m_1 baino txikiagoak ez diren aldagaiaren balioen artean, m_1-etik hurbilena da b_1; eta m_3 baino handiagoak ez diren aldagaiaren balioen artean, m_3-tik hurbilena da b_3.
4. **Kanpo-hesi** deritzen M_1 eta M_3 puntuak kalkulatuko ditugu era honetara:
   $M_1 = Q_1 - 2(1{,}5)RI$
   $M_3 = Q_3 + 2(1{,}5)RI$
5. Orain arte kalkulaturiko balio guztiak eskalan kokatzen dira.
6. Q_1 eta Q_3 balioak muturretan ezarriz, kutxa bat irudikatuko dugu; mediana dagoen puntuan lerro bat jarriko dugu.
7. Albo-balioak besoen bidez adieraziko ditugu, eta kutxarekin lotuko ditugu.

Barne- eta kanpo-hesien bitartean dauden aldagaiaren balioak zirkuluen bidez adieraziko dira. Puntu horiek **outlier moderatu**tzat hartuko dira. Kanpo-hesien kanpoaldean dauden aldagaien balioak izartxoen bidez adieraziko dira. Puntu horiek **izugarrizko outlier**tzat hartuko dira.

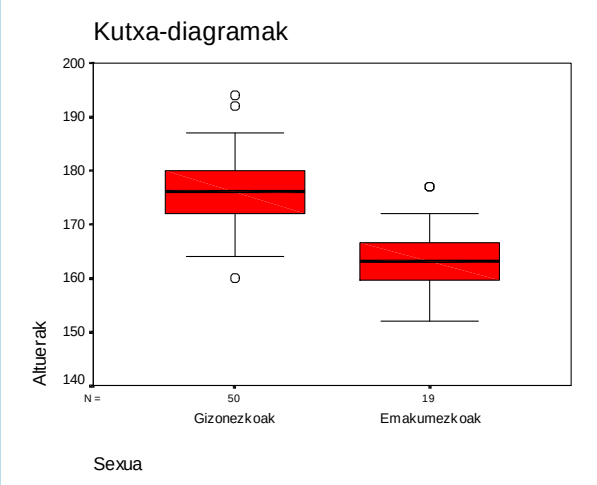
1.17 irudian kutxa-diagramaren adierazpide grafikoa erakusten da.



1.17 irudia. Kutxa-diagramaren adierazpide grafikoa

1.29 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) laburbiltzen dira, beren sexua kontuan hartuz, 1.18 irudiako kutxa-diagrametan. Bertan ikusten denez, gizonezkoak emakumezkoak baino altuagoak dira, baina antzeko sakabanatzea dute.



1.18 irudia. Osasun Inkestako judozaleen altueraren kutxa-diagramak, sexuaren arabera.

1.6 DATU ELKARTUAK

Aldagai estatistikoa jarraitua denean, hartzen duen balio kopurua handia izaten da (infinitua, alegia) eta, beraz, datu gordinen azterketa eskuz egiteko zaila izaten da. Nahiz eta informazioa eta zehaztasuna galdu, datuak elkartzea komeni da, haien azterketa errazteko (adibidez, txosten bat aurkeztu edo grafiko bat egiterakoan). Datuak klase edo azpitarteetan elkartu beharra dagoenean, honako teknika hau erabiltzen da:

1. Aldagai estatistikoaren ibiltartea, heina edo aldakuntza-tarte osoa determinatu. Horretarako, aldagaiak hartzen dituen baliorik handienaren eta txikienaren arteko diferentzia kalkulatu behar dugu.

2. Tartea maiztasun-banaketa deskribatu nahi den bezainbeste azpitartetan (**klase** deiturikoak) zatitzen da. Zenbaki egokirik ez dago, baina muturreko kasuak ekidingo ditugu ahal den neurrian. Klase kopuru txikiak ezin lezake neurrien arteko ezberdintasunik erakutsi eta, beraz, informazioa gal liteke. *Tukey*ren iritziz, hauexek dira klase kopuru egokia aukeratzeko laguntza gisa erabil daitezkeen metodo batzuk:
   - *Dixon eta Kronmal* ikerlariek (1965) proposatutako klase kopurua: $k = [10 \log_{10} n]$, non n laginaren tamaina baita, klase kopuruaren goiko muga egokia da.
   - *Sturges*en erregela (1926): $k = 1 + 3{,}322 \log_{10} n$ edo $k = 1 + \log_2 n$.
   - *Vellemanek* (1976) proposatutakoa: $k = \left[2\sqrt{n}\right]$

3. Klaseak beren muturrak seinalatuz mugatzen dira. Mutur horiei **beheko eta goiko klase-muga** deituko diegu, eta $T_i = (l_i, l_{i+1})$ adieraziko dugu. Klase-muga horiek inolako zalantzagarritasunik gabe hautatuko dira. Hau da, datu bakoitzari zer klase dagokion jakin behar dugu eta datu posible bakoitzari klase bakar bat egokitu behar diogu. Hori lortzeko bi teknika daude:
   - Datuek dituzten zifra adierazgarriak baino zifra gehiagorekin klase-mugak ematea.
   - Klase-mugak $[l_i, l_{i+1})$ moduan ematea. Horrela, klase horri dagozkion datuak l_i-ren berdinak edo l_i baino handiagoak eta l_{i+1} baino txikiagoak izango dira.

   Klaseak **muga-neurri** izenekoen bidez defini ditzakegu. Horiexek dira klaseari dagozkion neurririk handiena eta txikiena, beren zehaztasuna kontuan harturik. Kasu horretan, bi klase kontsekutiboren muga-neurriek ezin dezakete balio berbera hartu. Klase-muga errealak l_i eta l_{i+1} muga-neurri horien bidez kalkulatzen badira, klase horren behe-mugaren eta aurreko klasearen goi-mugaren arteko batez besteko balioa l_i izango da, eta klase horren goi-mugaren eta ondorengo klasearen behe-mugaren arteko batez besteko balioa l_{i+1} izango da.

   Klase bakoitzerako bi ezaugarri definituko ditugu:

   1. **Klasearen marka** tartearen erdiko puntu gisa definituko dugu. Hots,

      $$x_i = \frac{l_i + l_{i+1}}{2} \quad \forall\, i = 1, \ldots, k$$

      Balio horrek tarte bakoitzak daukan informazioa adierazten du eta estatistikoak kalkulatzeko erabiltzen da.
   2. **Klasearen luzera** tartearen bi mugen arteko diferentzia da. Hots, $d_i = l_{i+1} - l_i$ $\forall\, i = 1, \ldots, k$

4. Maiztasunak zenbatu eta, aurreko kasuetan egin dugun antzera, taula eratu. 1.10 taulan erakusten da datu elkartuetarako maiztasun-taula.

Klase-mugak $[l_i, l_{i+1})$	Klase-marka x_i	Klase-luzera d_i	Maiztasun absolutua f_i	Maiz. abs. metatua F_i	Maiztasun erlatiboa h_i	Maiz. erl. metatua H_i
$[l_1, l_2)$	x_1	d_1	f_1	F_1	h_1	H_1
$[l_2, l_3)$	x_2	d_2	f_2	F_2	h_2	H_2
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
$[l_k, l_{k+1})$	x_k	d_k	f_k	$F_k = n$	h_k	$H_k = 1$
Totala			n		1	

1.10 taula. Datu elkartuen taula estatistikoaren adierazpen orokorra

1.30 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) agertzen dira 1.8 adibidean. Aztertutako aldagaia —pertsonaren altuera— teorikoki jarraitua da, baina praktikoki, neurketa-prozesua dela kausa, balio zehatzak agertzen zaizkigu. Dena dela, neurri edo modalitate asko daudenez, klaseetan elkartutako maiztasun-taula estatistikoaren bidez adierazten da aldagai hori.

Datu elkartu moduan aurkezteko emandako aurreko urratsei jarraituz, baliorik handiena 194 cm da, eta txikiena 152 cm. Beraz, heina $R = 194 - 152 = 42$ da. Tarte kopurua kalkulatzeko, Sturges-en erregela erabiltzen badugu:

$k = 1 + 3{,}322\log_{10} n = 1 + 3{,}322 \cdot 1.839 = 7{,}109 \approx 8$

Luzera bereko tarteak eratzen baditugu, luzera hau izango da: $\frac{R}{k} = \frac{42}{8} = 5{,}25$, baina, lana errazteko, aukeratuko dugun luzera 5 izango da. Datu horien taula estatistikoa 1.11 taula da.

Klaseen muga-neurriak	Klaseen muga errealak (l_i, l_{i+1})	Klase-marka x_i	Maiztasun absolutua f_i
150 – 154	149,5 – 154,5	152	3
155 – 159	154,5 – 159,5	157	2
160 – 164	159,5 – 164,5	162	8
165 – 169	164,5 – 169,5	167	10
170 – 174	169,5 – 174,5	172	11
175 – 179	174,5 – 179,5	177	19
180 – 184	179,5 – 184,5	182	10
185 – 189	184,5 – 189,5	187	4
190 – 194	189,5 – 194,5	192	2
			$n = 69$

1.11 taula. Osasun Inkestako judozaleen altuerarako datu elkartuen taula estatistikoa.

Datu elkartuak lortzeko beste modu bat tarteko mugatzat aldagaiaren balio errealak hartzea izango litzateke. Azken hori erabiliz, 1.12 taula lortuko genuke (maiztasun metatuak eta erlatiboak gehitu dira taula honetan).

Klase-mugak $[l_i, l_{i+1})$	Klase-marka x_i	Klase-luzera d_i	Maiztasun absolutua f_i	Maiz. abs. metatua F_i	Maiztasun erlatiboa h_i	Maiz. erl. metatua H_i
[150 – 155)	152,5	5	3	3	0,04	0,04
[155 – 160)	157,5	5	2	5	0,03	0,07
[160 – 165)	162,5	5	8	13	0,12	0,19
[165 – 170)	167,5	5	10	23	0,14	0,33
[170 – 175)	172,5	5	11	34	0,16	0,49
[175 – 180)	177,5	5	19	53	0,28	0,77
[180 – 185)	182,5	5	10	63	0,14	0,91
[185 – 190)	187,5	5	4	67	0,06	0,97
[190 – 195)	192,5	5	2	69	0,03	1
Totala			$n = 69$		1	

1.12 taula. Osasun Inkestako judozaleen altuerarako datu elkartuen taula estatistikoaren beste adierazpen bat.

Bi aukerak dira egokiak, nahiz eta emaitzak ezberdinak izan. Ikertzaileak kasu bakoitzean gehien komeni zaiona aukeratuko du. Kontuan hartu beharrekoa da datuak elkartzeko eta taula eraikitzeko aukeratutako irizpideak estatistikoen kalkulua baldintzatuko duela.

1.6.1 Batezbestekoaren kalkulua

Datuak elkartuta daudenean, batezbestekoaren kalkulurako erabiliko dugun formulan agertzen diren x_i balioak klase-markak izango dira. Hots,

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

1.31 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altueraren (cm-tan) batezbestekoa kalkulatuko dugu. Horretarako, datuak

elkartzeko 1.11 taulan erabili dugun irizpidean oinarrituko gara. Datu horiei batezbestekoa kalkulatzeko behar dugun zutabea gehituko diogu, eta horrela 1.13 taula eraikiko dugu. Beraz, hauxe da batez besteko altuera:

$$\bar{x} = \frac{11943}{69} = 173{,}09$$

Klaseen muga errealak (l_i, l_{i+1})	Klase-markak x_i	Maiztasun absolutuak f_i	$x_i f_i$
149,5 – 154,5	152	3	456
154,5 – 159,5	157	2	314
159,5 – 164,5	162	8	1296
164,5 – 169,5	167	10	1670
169,5 – 174,5	172	11	1892
174,5 – 179,5	177	19	3363
179,5 – 184,5	182	10	1820
184,5 – 189,5	187	4	748
189,5 – 194,5	192	2	384
		$n = 69$	11943

1.13 taula. Judozaleen altueraren batezbestekoaren kalkulurako taula, datuak elkartuta.

1.6.2 Medianaren kalkulua

Aldagai estatistikoa jarraitua eta klasez bildua denean, mediana interpolazioz kalkulatzen da. Mediana jatorrizko datuak erabiliz kalkulatzea hobe da beti, noski. Baina, elkartu gabeko datuak ez ditugunean, propietate geometriko sinple batzuetan oinarritzen den formula aplikatuko dugu. 1.19 irudian erakusten da medianaren kalkulurako adierazpide grafikoa. Mediana barnean daukan tarteko datuak uniformeki banatuta daudela onartzen da. Baldintza horren pean, ABC eta ADE triangeluak baliokideak direla kontuan harturik, honako erlazio hau dugu:

$$\frac{\overline{AB}}{\overline{AD}} = \frac{\overline{BC}}{\overline{DE}} \Rightarrow \frac{Me - l_i}{d_i} = \frac{\frac{n}{2} - F_{i-1}}{f_i}$$

Beraz, *Me* askatuz, honako formula hau dugu mediana kalkulatzeko:

$$Me = l_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} d_i$$

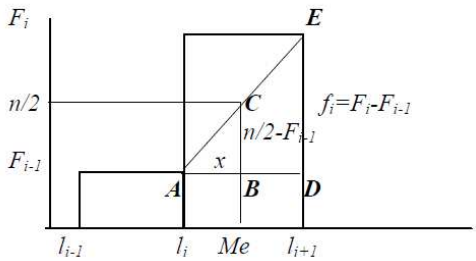
non l_i = mediana daukan klasearen behe-muga

n = laginaren tamaina

$F_{i-1} = \dfrac{n}{2}$ baino txikiagoa den maiztasun metaturik handiena

d_i = mediana daukan klasearen luzera

f_i = mediana daukan klasearen maiztasuna baitira.



1.19 irudia. Datu elkartuen medianaren kalkulurako adierazpide grafikoa.

Grafikoki ikusirik, honako propietate hau betetzen du medianak: aldagaiaren balio horretatik pasatzen den ordenatu-ardatzaren zuzen paraleloak histogramaren azalera erdibitzen du.

1.32 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) tartetan bilduta daudenean, mediana interpolazioz kalkulatzen da, 1.14 taulan oinarrituz.

Klaseen muga-neurriak	Klaseen muga errealak (l_i, l_{i+1})	Maiztasun absolutua f_i	Maiztasun metatua F_i	Maiztasun met. erlat. H_i
150 – 154	149,5 – 154,5	3	3	0,043
155 – 159	154,5 – 159,5	2	5	0,072
160 – 164	159,5 – 164,5	8	13	0,188
165 – 169	164,5 – 169,5	10	23	0,333
170 – 174	169,5 – 174,5	11	34	0,492
175 – 179	174,5 – 179,5	19	53	0,768
180 – 184	179,5 – 184,5	10	63	0,913
185 – 189	184,5 – 189,5	4	67	0,971
190 – 194	189,5 – 194,5	2	69	1,000
		n = 69		

1.14 taula. Judozaleen altueraren medianaren kalkulurako datu elkartuen taula.

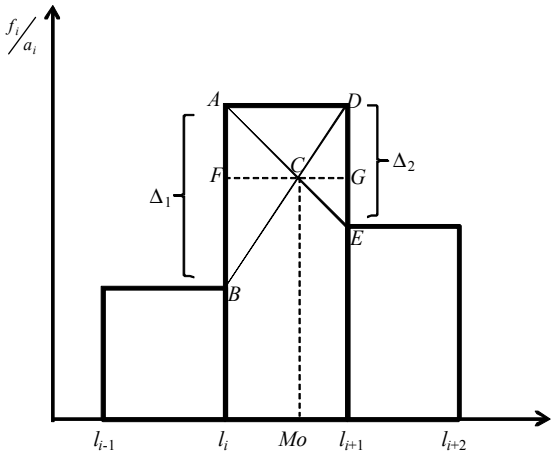
Bosgarren klasean banaketaren % 49,2 metatzen da eta 6. klasean % 76,8. Beraz, mediana 6. klasean dago eta l_i = 174,5, n = 69, F_{i-1} = 34, d_i = 5 eta f_i = 19 dira.

Hau da, $Me = l_i + \dfrac{\frac{n}{2} - F_{i-1}}{f_i} d_i = 174{,}5 + \dfrac{\frac{69}{2} - 34}{19} \cdot 5 = 174{,}63$ cm

1.6.3 Modaren kalkulua

Elkartutako datuak ditugunean, hots, jatorrizko daturik ez dugunean edo erabilgarriak ez direnean, moda kalkulatzeko propietate geometriko sinple batzuetan oinarritzen den beste formula bat aplikatuko dugu. Modaren kalkulurako adierazpide grafikoa 1.20 irudian erakusten da. Maiztasun-kontzentraziorik handiena daukan klaseari **modaren klasea** deritzogu, eta, kasu horretan, interpolazioz kalkulatzen da modaren balioa. ABC eta CDE triangeluak baliokideak direla kontuan harturik, honako erlazio hau dugu:

$$\frac{\overline{AB}}{\overline{DE}} = \frac{\overline{FC}}{\overline{GC}} \Rightarrow \frac{\Delta_1}{\Delta_2} = \frac{Mo - l_i}{l_{i+1} - Mo}$$



1.20 irudia. Datu elkartuen modaren kalkulurako adierazpide grafikoa.

Beraz, *Mo* askatuz, $Mo = l_i + \dfrac{\Delta_1}{\Delta_1 + \Delta_2} d_i$

non l_i = modaren klasearen behe-muga

$$\Delta_1 = \frac{f_i}{d_i} - \frac{f_{i-1}}{d_{i-1}}$$

$$\Delta_2 = \frac{f_i}{d_i} - \frac{f_{i+1}}{d_{i+1}}$$

d_i = modaren klasearen luzera
f_i = modaren klasearen maiztasuna baitira.

1.33 adibidea

Osasun Inkestako datuak erabiliz, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) biltzen dituen aurreko adibidearen taulan oinarrituz, moda kalkulatu nahi dugu. Maiztasun-kontzentraziorik handiena (19) 6. klasean dagoenez, klase horretan egongo da moda. Kasu honetan, klase guztien tamaina berdina denez, sinpletu egin daiteke moda kalkulatzeko formula:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} d_i = l_i + \frac{f_i - f_{i-1}}{f_i - f_{i-1} + f_i - f_{i+1}} d_i$$

non l_i = 174,5, d_i = 5, f_{i-1} = 11, f_i = 19 eta f_{i+1} = 10 baitira.

Beraz, $Mo = 174{,}5 + \frac{19-11}{19-11+19-10} \cdot 5$ = 174,5 + 2,3529 = 176,85 cm

1.6.4 Posizio-estatistikoen kalkulua

Pertzentilen kalkulua medianaren kalkuluaren antzera egiten da. Pertzentila dagoen tartea aurkitu ondoren, honako formula hau erabiltzen da kalkulatzeko:

$$P_j = l_i + \frac{\frac{j \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i$$

non sinboloen adierazpena medianaren formularen antzekoa baita:

$l_i = j$ portzentajea daukan klasearen behe-muga

n = laginaren tamaina

$F_{i-1} = \frac{j \cdot n}{100}$ baino txikiagoa den maiztasun metaturik handiena

d_i = esandako klasearen luzera

$f_i = j$ portzentajea daukan klasearen maiztasun absolutua.

1.34 adibidea

Osasun Inkestako datuak elkarturik badaude, judoa praktikatzen duten 69 pertsonen altuerak (cm-tan) klasetan banatuta agertzen dira 1.32 adibidearen taulan. Kasu honetan, kalkula ditzagun koartilak eta 9. dezila edo 90. pertzentila. Zenbat cm-tik gora egongo da judozaleen % 75aren altuera?

$$Q_1 = l_i + \frac{\frac{25n}{100} - F_{i-1}}{f_i} \cdot d_i = 164{,}5 + \frac{\frac{69}{4} - 13}{10} \cdot 5 = 166{,}63 \text{ cm}$$

$$Q_3 = l_i + \frac{\frac{75n}{100} - F_{i-1}}{f_i} \cdot d_i = 174{,}5 + \frac{\frac{3 \cdot 69}{4} - 34}{19} \cdot 5 = 179{,}17 \text{ cm}$$

$$D_9 = P_{90} = l_i + \frac{\frac{90n}{100} - F_{i-1}}{f_i} \cdot d_i = 179{,}5 + \frac{\frac{90 \cdot 69}{100} - 53}{10} \cdot 5 = 184{,}05 \text{ cm}$$

Kirolarien % 75aren altuera 166,63 cm (Q_1) baino handiagoa izango da.

1.6.5 Sakabanatze-estatistikoen kalkulua

Batezbestekoarekin gertatzen den gisan, bariantzaren formulan agertzen diren x_i balioak klase-markak dira. Hots,

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$$

1.35 adibidea

Osasun Inkestako datuak erabiliz, 1.15 taulan laburbiltzen dira judoa praktikatzen duten 69 pertsonen altuerak (cm-tan). Altueraren bariantza eta desbideratze estandarra kalkulatuko ditugu.

Klaseen muga errealak (l_i, l_{i+1})	Klase-markak x_i	Maiztasun absolutuak f_i	$x_i f_i$	$x_i^2 f_i$
149,5 – 154,5	152	3	456	69.312
154,5 – 159,5	157	2	314	49.298
159,5 – 164,5	162	8	1.296	209.952
164,5 – 169,5	167	10	1.670	278.890
169,5 – 174,5	172	11	1.892	325.424
174,5 – 179,5	177	19	3.363	595.251
179,5 – 184,5	182	10	1.820	331.240
184,5 – 189,5	187	4	748	139.876
189,5 – 194,5	192	2	384	73.728
		$n = 69$	11.943	2.072.971

1.15 taula. Judozaleen altueraren bariantzaren kalkulurako taula.

Beraz, $s_n^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2 = \frac{2072971}{69} - 173{,}09^2 = 82{,}9 \text{ cm}^2 \Rightarrow s_n = 9{,}11 \text{ cm}$

1.6.6 Forma-estatistikoaren kalkulua

Aurrekoetan bezala, kasu honetan ere, klase-markak dira x_i balioak.

1.36 adibidea

Aurreko adibideari jarraituz, judoa praktikatzen duten 69 pertsonen altueraren forma-estatistikoak kalkulatuko ditugu. Honako hau izango da alborapen-koefizientea, aurreko adibideetan kalkulaturiko estatistikoak erabiliz:

$$\upsilon = \frac{\bar{x} - Mo}{s_n} = \frac{173{,}09 - 176{,}85}{9{,}11} = -0{,}413$$

Asimetria- eta kurtosi-koefizienteak *SPSS*ren bidez kalkulatu ditugu, eta 1.16 taulan erakusten dira emaitzak.

Estatistiko deskribatzaileak

	N	Asimetria		Kurtosia	
	Estatistikoa	Estatistikoa	Errore estandarra	Estatistikoa	Errore estandarra
ALTUERA	69	-,329	,289	-,183	,570

1.16 taula. Judozaleen altueraren forma estatistikoak.

1.6.7 Ordenagailuaren eragina

Dena dela, ordenagailuak datu gordinak erabiltzen ditu estatistikoak kalkulatzeko, eta beraz, datu elkartuen metodoa nahiko zaharturik dago. Lehen aipatu dugunez, datuak elkartzean informazioa galtzen da, eta eskuz kalkulaturiko estatistikoak eta ordenagailuak ematen dizkigunak antzekoak dira, baina ez berdinak. Esate baterako, 69 judozaleen altueren kasuan, datu gordinak erabiliz lortutako estatistikoak eta datu elkartuak ditugunean lortutakoak laburbiltzen dira 1.17 taulan.

Estatistikoa	Datu gordinak	Datu elkartuak
$\bar{x}$	172,59	173,09
Me	175	174,63
Mo	175	176,85
Q_1	166	166,63
Q_3	178	179,17
s_n	8.96	9,11
υ	−0.269	−0,413
a_3	−0.194	−0,329
a_4	−0.114	−0,183

1.17 taula. Osasun Inkestaren arabera, EAEko judozaleen altueraren estatistikoak, datu gordinak eta elkartuak erabiliz.

1.7 ARIKETAK

1. 1. Geriatria-klinika bateko pazienteen ikerketa egin da. Helburua da klinika horretako pazienteen ezaugarriak eta klinika uztean duten helbidea ezagutzea. Azken urtean klinikan egon diren pazienteen honako ezaugarri hauek ezagutzen dira: sexua (E ≡ emakumezkoa, G ≡ gizonezkoa); diagnostikoa (AM ≡ atzerapen mentala, GM ≡ gaixotasun mentala, GF ≡ gaixotasun fisikoa); adina (urteak) eta helbidea (1 ≡ hilda, 2 ≡ familiaren etxea, 3 ≡ ospitalea, 4 ≡ kalea, 5 ≡ beste geriatriko bat, 6 ≡ klinika pribatua, 7 ≡ ez du geriatrikoa utzi). Datuak 1.18 taulan agertzen dira.

Sexua	Diagnostikoa	Adina	Helbidea	Sexua	Diagnostikoa	Adina	Helbidea
E	GM	67	6	G	AM	80	7
G	GF	71	7	G	GM	83	2
E	GF	54	1	G	GF	49	3
E	GM	63	7	E	GF	78	6
E	GM	48	7	G	GM	57	7
G	GM	56	7	G	AM	69	3
G	GF	62	3	E	GM	83	7
E	AM	57	2	E	GF	92	1
E	GF	81	7	E	GF	55	3
E	GF	36	7	E	GF	63	6
E	GF	72	3	E	GF	64	4
E	GF	65	3	G	GF	89	7

1.18 taula. 1.1 ariketako datuak. Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994.

a) Eraiki itzazu lau ezaugarrien maiztasun-taulak.

b) Irudika itzazu maiztasun erlatiboen barra-grafikoak, sexua, diagnostikoa eta helbidea adierazteko.

c) Irudika ezazu adinaren zurtoin eta hosto grafikoa. Jar itzazu maiztasunari eta sakontasunari dagozkien zutabeak ere.

d) Eraiki itzazu, bananduta, gizonen eta emakumeen helbidearen maiztasun-taulak eta adinaren zurtoin eta hosto grafikoak.

1. 2. Osakidetzako bost ospitaletako Aldakako Protesi Osoari (APO) buruzko datuak agertzen dira 1.19 taulan. APO izan duten 849 pazienteri egindako kirurgiari buruzko datuak erakusten dira, beren diagnosiaren eta erabilitako protesi motaren arabera ($Z \equiv$ zementuzkoa, $G \equiv$ zementu gabea eta $N \equiv$ nahastea).

	Ospitalea														
	1			2			3			4			5		
Protesi-metodoa	*Z*	*G*	*N*	*Z*	*G*	*N*	*Z*	*G*	*N*	*Z*	*G*	*N*	*Z*	*G*	*N*
Diagnosia															
-Osteoartrosia	15	94	9	80	29	5	4	112	-	-	82	8	7	151	8
-Nekrosia	-	4	-	2	2	-	-	5	-	-	1	3	1	13	-
-Berriz aldatzea	6	20	3	42	11	7	6	4	-	-	35	2	7	58	3
Guztira	21	118	12	124	42	12	10	121	-	-	118	13	15	222	11

1.19 taula. 1.2 ariketako datuak.

a) Egin itzazu ospitale, protesi-metodo eta diagnosia aldagaiei banaka dagozkien maiztasun-taulak eta sektore-diagramak.

b) Egin ezazu diagnosirako maiztasun-taula, ospitalearen arabera. Irudika ezazu ospitale bakoitzari dagokion sektore-diagrama.

c) Egin ezazu diagnosirako maiztasun-taula, protesi-metodoaren arabera. Irudika ezazu protesi-metodo bakoitzari dagokion sektore-diagrama.

d) Kalkula ezazu, ebakuntza guztiak kontuan hartuta, protesi mota bakoitza erabiliz egiten den ebakuntza-portzentajea bost ospitaleetan eta ospitale bakoitzean.

1. 3. Haurra oinez hasten denean, hilabetean zehar, erori gabe eta zuzenean, haur horrek ibiltzen dituen metroak neurtzen dira pediatria-klinika batean. Datuak 1.20 taulan agertzen dira.

Ume kopurua	2	6	10	5	10	3	2	2
Metroak	1	2	3	4	5	6	7	8

1.20 taula. 1.3 ariketako datuak. Iturria: V. Quesada, A Isidoro, L.A. López. *Curso y Ejercicios de Estadística.* Alhambra Universidad, 1982.

a) Eraiki ezazu maiztasun-taula.

b) Irudika ezazu maiztasun absolutuen barra-grafikoa.

c) Kalkula itzazu mediana, moda, batezbestekoa eta koartilak.

d) Kalkula itzazu heina, koartilarteko heina eta desbideratze estandarra.

1. 4. Hogeita hamabost egunean zehar lurralde batean bota den zuhaitz kopurua adierazten da segidan:

200	300	100	100	0	200	0	0	0	0	0	0
0	200	0	0	400	0	400	0	1.000	300	300	300
0	0	0	100	0	0	200	200	0	0	100	

Iturria: A. Amparan, I. Aróstegui, B. del Hoyo, S. Marcaida, A. Urkaregi, A. Valle. *Bioestatistikako apunteak*. EHU, 1997 (argitaratu gabe).

a) Eraiki ezazu botatako zuhaitz kopuruaren maiztasun-taula.

b) Irudika itzazu banaketa horren barra-grafikoa eta zurtoin- eta hosto-grafikoa.

c) Kalkula itzazu banaketa horrentzat ezagutzen dituzun joera zentraleko estatistikoak.

d) Kalkula itzazu banaketa horrentzat ezagutzen dituzun sakabanatze-estatistikoak.

1. 5. Honako 1.21 taula hau emanik:

x_i	Maiztasun absolutuak	Maiztasun absolutu metatuak	Portzentajea
3		6	
6		11	12,5
9	9		
12		27	
13	10		25
15			7,5

1.21 taula. 1.5 ariketako datuak. Iturria: V. Quesada, A Isidoro, L.A. López. *Curso y Ejercicios de Estadística*. Alhambra Universidad, 1982.

a) Osatu taula.

b) Kalkula itzazu batezbestekoa eta mediana.

c) Kalkula itzazu koartilak.

d) Kalkula ezazu aldakuntza-koefizientea.

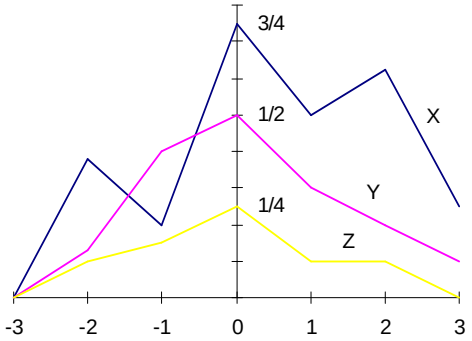
1. 6. EAEn egindako 1997ko Osasun Inkestan, gaixoek azken urtean ospitale batean igaro duten gau kopurua zenbatu da: batezbestekoa eta desbideratze estandarra

8,7 eta 10,7 gau izan dira, hurrenez hurren. Ospitalean igarotako batez besteko gau kopurua banaketa horren adierazgarri al da? Azal ezazu erantzuna.

1. 7. Enpresa bateko 80 langileen soldata honela banatzen da: 60 langilek orduko 8 euroko soldata daukate, eta gainerakoek orduko 6 eurokoa. Kalkula ezazu enpresa horretan duten orduko batez besteko soldata.

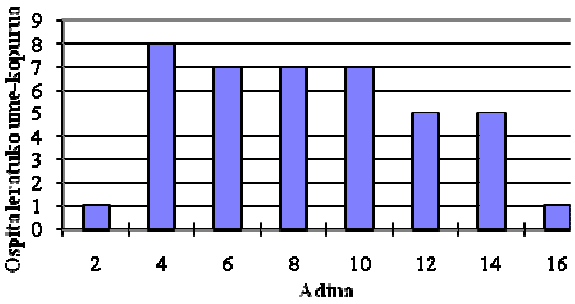
1. 8. Pertsona normalen talde bateko albumina kantitate ibilkorraren batezbestekoa 130 gramo da. Gaixotasunen bat duten pertsona talde batean, batezbesteko hori 100 gramo da. Osotasunean harturik, 125 gramo da bi pertsona taldeen albumina kantitate ibilkorraren batezbesteko haztatua. Zein da pertsona gaixoen portzentajea, osotasunean hartutako talde horretan?

1. 9. 1.21 irudi honetan agertzen den grafikoak *X, Y* eta *Z* aldagaien maiztasun-banaketak adierazten ditu aldi berean. Zeinek du modarik handiena? Zein da moda horren balioa?



1.21 irudia. 1.9 ariketako datuen adierazpide grafikoa. Iturria: V. Quesada, A. Isidoro, L.A. López. *Curso y Ejercicios de Estadística*. Alhambra Universidad, 1982.

1. 10. Ospitale batean dauden umeen adinen adierazpide grafikoa 1.22 irudian agertzen da (urteetan).



1.22 irudia. 1.10 ariketako datuen adierazpide grafikoa. Iturria: A. Amparan, I. Aróstegui, B. del Hoyo, S. Marcaida, A. Urkaregi, A. Valle. Bioestatistikako apunteak. EHU, 1997 (argitaratu gabe).

a) Eraiki ezazu ospitaleraturiko umeen adinaren maiztasun-taula.

b) Kalkula itzazu ospitaleraturiko umeen adinaren batezbestekoa, mediana eta moda.

c) Zein estatistikok neurtzen du aldagai horren sakabanatze absolutua? Zer unitatetan adieraziko da sakabanatze hori? Kalkula ezazu horren balioa.

d) Eta sakabanatze erlatiboa? Zein unitatetan adieraziko da? Kalkula ezazu horren balioa.

e) Zein izango da adin maximoa, umeen % 25ak hori baino urte gutxiago dituela ziurtatzeko?

1. 11. EAEn egindako 1997ko Osasun Inkestaren datuen arabera, puntuazio hauek lortu dira osasunarekin lotutako bizi-kalitatea neurtzen duen galdeketaren honako bi arlo hauetan: funtzionamendu sozialean, batezbestekoa 89,47 eta desbideratze estandarra 19,02; eta osasun mentalean, batezbestekoa 71,18 eta desbideratze estandarra 18,01. Konpara itzazu aurreko bi arloetako honako hauek:

a) Sakabanatze absolutua.

b) Sakabanatze erlatiboa.

1. 12. 1.22 taulan bildu dira EAEn dauden ibaien adar kopuruak.

Ibaiak	Adar kopurua
Bidasoa	2
Oiartzun	2
Urumea	2
Oria	12
Urola	4
Deba	9
Artibai	1
Lea	1
Ea	0
Oka	3
Butron	3
Ibaizabal	6
Nerbioi	4
Ibaizabal-Nerbioi	4
Barbadun	3
Agüera	0
Karrantza	0
Calera del Prado	0
Ebro	6

1.22 taula. 1.12 ariketako datuak. Iturria: EUSTAT. *EUE 97/98.*

a) Osa ezazu EAEko ibai-adarren kopuruaren maiztasun-taula eta irudika ezazu barra-grafikoa.

b) Kalkula ezazu EAEko ibaien batez besteko adar kopurua.

c) Kalkula itzazu EAEko ibaien adar kopuruaren mediana eta koartilak.

d) Kalkula ezazu EAEko ibaien adar kopuruaren desbideratze estandarra.

1. 13. 1997ko Osasun Inkestako datuak erabiliz, erretzaileen artean egunero erretzen duten puru kopuruaren maiztasun-taula (1.23 taula) eratu da.

Puru kopurua	Pertsona kopurua
0	960
1	54
2	25
3	11
4	5
5	2
6	1
7	1
8	3
9	0
10	2
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	1
Guztira	**1.065**

1.23 taula. 1.13 ariketako datuak.

a) Osa ezazu banaketa horren maiztasun-taula.

b) Kalkula itzazu egunero erretako puru kopuruaren batezbestekoa, mediana, moda eta koartilak.

c) Kalkula itzazu egunero erretako puru kopuruaren heina, koartilarteko heina, desbideratze estandarra eta aldakuntza-koefizientea.

d) Zer puru kopurutik gora egongo da gehien erretzen dutenen % 25a?

e) Azter ezazu banaketaren simetria, alborapena eta alborapen-momentua kalkulatuz.

f) Azter ezazu banaketaren zorroztasuna, kurtosi-momentua kalkulatuz. Nolakoa da banaketa horren zorroztasuna, normalarekin konparatuz?

g) Pururen bat erretzen dutenak soilik hartzen baditugu, eraiki ezazu berriro maiztasun-taula. Zer puru kopurutik behera egongo da gutxien erretzen dutenen % 25a?

1. 14. EAEn 1996. urtean lehorreratutako baxurako arrain kantitatea (toneladatan) agertzen da 1.24 taulan, itsas portuei jarraiki sailkatuta eta espezieen arabera.

	Antxoa	Atuna	Hegaluzea	Txitxarroa	Legatza	Bisigua	Berdela
Armintza	–	–	1,6	0,1	1,2	–	–
Bermeo	2.111,9	122,4	3.250,9	205,7	25,4	1,9	5.456,5
Donostia	306,9	27,2	96,2	76,1	22,9	–	434,9
Elantxobe	–	–	0,2	–	13,1	–	0,5
Getaria	2.780,1	989,0	2.476,4	337,7	–	–	1.400,8
Hondarribia	1.847,5	3.369,2	687,6	141,6	14,9	0,1	1.539,5
Lekeitio	449,4	9,2	241,9	4,5	10,5	–	115,3
Mundaka	–	–	–	0,3	–	–	–
Mutriku	871,8	54,4	553,5	112,3	7,3	–	390,9
Ondarroa	4.707,6	142,9	1.540,4	4.849,1	114,5	0,2	4.875,2
Pasaia	2.080,4	20,9	187,0	1.869,0	1.521,1	8,5	937,6
Santurtzi	6,0	–	–	0,3	2,4	–	–

1.24 taula. 1.14 ariketako datuak. Erabilitako ikurrak: – (batere ez). Iturria: EUSTAT. EUE 97/98.

a) Irudika ezazu espezie bakoitzerako EAEn lehorreratutako kantitatearen kutxa-diagrama.

b) Kalkula ezazu espezie bakoitzerako EAEn lehorreratutako batez besteko kantitatea.

c) Irudika itzazu espezie bakoitzerako Bizkaian eta Gipuzkoan lehorreratutako kutxa-diagramak grafiko berean.

d) Kalkula ezazu espezie bakoitzerako Bizkaian eta Gipuzkoan lehorreratutako batez besteko kantitatea.

e) Kalkula ezazu itsas portu bakoitzean lehorreratutako batez besteko arrain kantitatea.

f) Kalkula ezazu espezie bakoitzerako EAEn lehorreratutako arrain kantitatearen desbideratze estandarra. Konpara ezazu espezie bakoitzerako lehorreratutako arrain kantitatearen sakabanatzea.

1. 15. EAEn dauden behatoki plubiometrikoetan, 1996. urtean batutako zenbait aldagai klimatologiko nagusi bildu dira 1.25 taulan.

Behatokiak	Lurralde historikoa	Egun kopurua				
		Euria	Txingorra	Ekaitza	0°ren azpitik	25°tik gora
Abadiño	B	169	0	0	33	32
Agurain	A	122	9	7	29	58
Aia	G	189	8	28	8	13
Amurrio	A	126	0	13	17	72
Arkaute	A	75	–	6	41	18
Aretxabaleta	G	187	4	14	37	39
Arrasate	G	119	1	16	18	14
Arteaga	B	149	3	1	29	37
Balmaseda	B	20	4	20	21	61
Beizama	G	194	3	12	–	–
Bergara	G	195	5	27	20	92
Bernedo	A	83	0	11	–	–
Derio	B	155	4	13	9	57
Donostia (Ategorrieta)	G	180	8	21	9	18
Donostia (Igeldo)	G	184	11	22	8	16
Elduain	G	151	0	0	–	–
Elgeta	G	183	5	24	29	43
Ermua	B	126	4	13	30	40
Errenteria	G	197	0	1	20	43
Eskalmendi	A	156	2	18	24	17
Etxebarria	B	192	1	20	32	55
Gasteiz	A	173	3	14	38	53
Gaubea	A	82	1	19	38	91
Izarra	A	143	0	19	25	19
Kanpezu	A	94	2	8	–	–
Lagran	A	132	0	3	–	–
Laguardia	A	78	0	0	29	44
Lasarte	G	140	4	8	8	51
Lareo	G	141	0	21	41	15
Legazpi	G	184	3	13	27	46

Behatokiak	Lurralde historikoa	Egun kopurua				
		Euria	Txingorra	Ekaitza	0°ren azpitik	25°tik gora
Legazpi (Barrendiola)	G	203	4	8	31	48
Leza	A	120	0	3	44	32
Martioda	A	105	0	0	–	–
Ollerieta	A	172	2	12	18	26
Pasaia	G	171	6	12	28	47
Sondika	B	207	1	19	10	51
Urizaharra	A	95	2	13	–	–
Urnieta	G	–	–	–	8	26

1.25 taula. 1.15 ariketako datuak. Erabilitako ikurrak: - (ez dugu daturik eskuartean). Lurralde historikoak: A (Araba), B (Bizkaia) eta G (Gipuzkoa). Iturria: EUSTAT. EUE 97/98.

a) Irudika ezazu aldagai klimatologiko bakoitzeko zurtoin eta hosto grafikoa.

b) Kalkula itzazu aldagai klimatologiko bakoitzeko egun kopuruaren batezbestekoa eta mediana.

c) Kalkula ezazu aldagai klimatologiko bakoitzeko egun kopuruaren desbideratze estandarra. Zeinek dauka sakabanatzerik handiena?

d) Azter itzazu aldagai klimatologiko bakoitzaren simetria eta zorroztasuna, alborapen- eta kurtosi-momentuak erabiliz.

e) Behatoki klimatologikoak lurraldeka banandu ondoren, errepika ezazu aurreko ataletan egindako azterketa estatistikoa. Konpara itzazu hiru lurralde historikoetan lortutako emaitzak.

1. 16. Bost txanpon 1.000 aldiz jaurti dira eta 0, 1, 2, 3, 4 eta 5 aurpegi agertzen diren jaurtiketa kopuruak zenbatu dira. Emaitzak 1.26 taulan adierazi dira.

Aurpegi kopurua	Jaurtiketa kopurua
0	38
1	144
2	342
3	287
4	164
5	25
Totala	1.000

1.26 taula. 1.16 ariketako datuak.

a) Adieraz itzazu datu horiek, barra-grafikoaren bidez.

b) Osa ezazu taula, maiztasun erlatiboak eta metatuak kalkulatuz.

c) Irudika ezazu datu horien grafiko metakorra.

1. 17. Honako datu hauek Euskal Herriko Unibertsitateko Medikuntza eta Odontologia Fakultateko lehen mailako 45 ikasleren pisuak adierazten dituzte kilotan:

55	56	51	54	74	49	50	55	55
63	70	65	56	55	59	59	53	56
54	60	52	82	55	82	65	60	53
59	60	52	78	63	58	78	52	55
60	72	58	62	45	50	60	55	53

a) Kalkula itzazu ikasleen pisuaren batezbestekoa, mediana, koartilak, desbideratze estandarra, koartilarteko heina eta aldakuntza-koefizientea, datu gordinak erabiliz.

b) Irudika ezazu zurtoin- eta hosto-grafikoa.

c) Datuak tartetan elkarturik, eraiki itzazu maiztasun-taula eta histograma.

d) Irudika ezazu ojiba, c) ataleko taula erabiliz.

e) Kalkula itzazu a) atalean kalkulaturiko estatistiko guztiak, datu elkartuak erabiliz.

1. 18. Animalia mota bateko 65 jaioberriren luzeraren (mm-tan) maiztasun-banaketa adierazi da 1.27 taulan.

Luzera	Jaioberrien kopurua
250 – 260	8
260 – 270	10
270 – 280	16
280 – 285	14
285 – 290	10
290 – 300	5
300 – 310	2
Totala	65

1.27 taula. 1.18. ariketako datuak. Iturria: A. Amparan, I. Aróstegui, B. del Hoyo, S. Marcaida, A. Urkaregi, A. Valle. Bioestatistikako apunteak. EHU, 1997 (argitaratu gabe).

Kalkula itzazu:

a) Seigarren klasearen behe-muga.

b) Laugarren klasearen goi-muga.

c) Hirugarren klasearen marka.

d) Bosgarren klasearen luzera.

e) Hirugarren klasearen maiztasun absolutua.

f) Hirugarren klasearen maiztasun erlatiboa.

g) Modaren klasea.

h) 280 mm baino luzera laburragoa daukaten jaioberrien portzentajea.

i) Joera zentraleko estatistiko guztiak.

Adieraz itzazu:

j) Histograma.

k) Maiztasun-poligonoa.

1. 19. Anestesiko baten azterketa egiten da, arratoi biziak erabiliz. Sendagaia eman ondoren, odol arterialeko CO_2-ren presioaren aldakuntza-portzentajea neurtu da eta 1.28 taulan agertzen diren datuak lortu dira.

Aldakuntza-portzentajea, mm Hg														
27,2	30,1	30,5	28,4	30,7	31,3	30,5	30,1	29,6	30,2	33,0	32,1	31,7	33,0	32,1

1.28 taula. 1.19 ariketako datuak. Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994.

Kalkula itzazu honako estatistiko hauek, datuak elkarturik eta elkartu gabe, eta konpara itzazu emaitzak:

a) Laginaren batezbesteko aritmetikoa. Zein da maiztasun-banaketa bi zati berdinetan banatzen duen aldakuntza-portzentajearen balioa? Nola deritzo balio horri?

b) Laginaren bariantza eta desbideratze estandarra. Zein unitatetan neurtzen dira bi horiek?

c) Laginaren heina eta koartilarteko heina.

1. 20. Airearen kutsadura aztertzeko, 57 hiritan aztertu zen esekidura-materiaren partikula-edukia (mikrogramo/m^3-tan), eta honako emaitza hauek lortu ziren:

68	63	42	27	30	36	28	32	79	27	22	23	24	25	44	51	23	12	32
65	43	25	74	51	36	42	28	31	28	25	45	12	57	31	50	38	21	16
43	27	49	28	23	19	46	30	43	49	12	42	27	47	22	38	69	49	24

Iturria: A. Urkaregi, *Bioestatistika*. UEU, 1991.

a) Histogramaren eta maiztasun-poligonoaren bidez, irudika ezazu aireko esekidura-materiaren partikula-edukiaren banaketa.

b) Kalkula itzazu batezbestekoa, mediana, moda, heina, desbideratze estandarra, aldakuntza-koefizientea, koartilak eta koartilarteko heina.

c) Zein balioren artean egongo da hiri kutsatuenetariko % 25?

2 PROBABILITATEA

2.1 SARRERA

Lehenengo gaian egindako azterketa emandako datuen multzoaren azterketara mugatu da. Metodo hori ez da nahikoa zientzia esperimental gehienek eskatzen dituzten baldintzak betetzeko. Gehienetan, laginaren edo azpipopulazioaren azterketa eginez, beharrezkoa da bertatik ateratako ondorioak populazio osora hedatzea. Alegia, estatistika deskribatzailetik estatistika induktibora pasatzea. Urrats hori emateko, halabeharrezko bihurtzen da oinarrizko tresnaren erabilera garatzea; tresna matematiko hori **probabilitate-teoria** da.

Esperimentua egin aurretik emaitza ezagutzea ezinezkoa denean, **zorizko esperimentua** dugu. Teorikoki, baldintza berdinen pean etengabe errepika daiteke, eta ezaguna da emaitza posible guztien multzoa. Zorizkoak ez diren esperimentuei determinista deritze. Esate baterako, Fisikan aztertzen diren ezaugarri gehienetan (higidura, elektrizitatea…) emaitza esperimentua egin aurretik ezagutzen da. Osasun eta Natur Zientzietan aztertzen diren ezaugarrietan (heredentziaren transmisioa, odolaren konposaketa, uraren kutsadura-maila…), ordea, ezinezkoa da emaitza aurresatea. Beraz, Osasun eta Natur Zientzietan aztertzen diren arazo gehienak zorizkoak dira.

Orain, honako hau da galdera: esperimentuaren emaitza aurrezagutzea ez bada posible, nola finkatu fenomenoa menperatzen duten legeak?

Zorizko esperimentuaren emaitza finkoa ezezaguna da, baina esperimentua baldintza berdinen pean behin eta berriro errepikatzen bada, lortzen diren neurriak, nahiz eta ezberdinak izan, «antzekoak» izango dira. Hots, balio «normalen» artean egongo dira. Erregulartasun hori zorizko esperimentuek betetzen duten propietatea da, eta fenomeno horrek probabilitate-kalkuluaren bidez aztertzera bideratzen gaitu.

2.2 ZORIZKO ESPERIMENTUARI ELKARTUTAKO GERTAERAK

Zorizko esperimentuaren emaitza posible guztien multzoari **lagin-espazio** edo **multzo unibertsal** deritzo, eta Ω-ren bidez adierazten da. Zorizko esperimentu baten lagin-espazioa multzo finitua edo infinitua izan daiteke.

Lagin-espazioaren azpimultzoei **gertaera** deritze, eta letra maiuskulen bidez adierazten dira; horien artean, kontuan izan behar dira multzo hutsa –$\varnothing$ **ezinezko gertaera**– eta lagin-espazioa bera azpimultzo gisa –Ω **gertaera segurua**–. Elementu bakar batek osaturiko gertaerei **oinarrizko gertaera** deritze, eta elementu bat baino gehiagoz osaturikoei **gertaera konposatuak**.

2.1 adibidea

Demagun zorizko esperimentua pertsona baten A gripearen birusa (N1H1) detektatzean datzala, gripea detektatzeko testa erabiliz. Honako hau da lagin-espazioa: Ω = {positibo, negatibo}. Honako gertaera hauek hartzen ditugu: $P \equiv$ Positibo eta $N \equiv$ Negatibo. Beraz, testa positibo ateratzen baldin bada, P gertaera jazoko da. Testa negatibo ateratzen baldin bada, N gertaera jazoko da. Alegia, gertaera bat zorizko esperimentuaren proba egin ondoren beha daitekeen jazoera da. P eta N oinarrizko gertaerak dira.

Zorizko esperimentu bati loturiko lagin-espazioa Ω izanik, Ω-ren azpimultzo guztiek osaturiko multzoari **Ω-ren parteen multzo** deritzo, eta $\wp(\Omega)$ adierazten da. Zorizko esperimentu bati loturiko gertaera guztiek osaturiko multzoa da.

Gripea detektatzeko testaren esperimentuan, $\Omega = \{P,N\}$ da eta, beraz, $\wp(\Omega) = \{\varnothing, \{P\}, \{N\}, \{P, N\}\}$ da.

2.2 adibidea

Demagun zorizko esperimentua dado bat botatzean datzala. Kasu horretan, Ω lagin-espazioa, $\Omega = \{1, 2, 3, 4, 5, 6\}$ da. Demagun honako gertaera hauek hartzen ditugula:

$A = \{1, 2, 4\}$, $B = \{1, 3, 5\}$, $C = \{2, 4, 6\}$, $D = \{1, 2, 3\}$,

$E = \{3, 4, 5, 6\}$, $F = \{5\}$

Beraz, dadoa bota eta emaitza bakoitia ateratzen baldin bada, B gertaera jazoko da. Dadoa bota eta emaitza bikoitia ateratzen baldin bada, C gertaera jazoko da. Dadoa bota eta 4 baino emaitza txikiagoa ateratzen baldin bada, D gertaera jazoko da. Dadoa bota eta 2 baino emaitza handiagoa ateratzen baldin bada, E gertaera jazoko da. Dadoa bota eta 5 ateratzen baldin bada, F gertaera jazoko da, etab.

Adibide horretan F da oinarrizko gertaera bakarra; A, B, C, D eta E gertaerak konposatuak dira.

2.3 GERTAEREN ARTEKO ERLAZIOAK

Esperimentu bati lotutako gertaerak definitzen direnean, gertaera horien arteko erlazioak ere defini ditzakegu. Atal honetan, gertaeren artean gauza daitezkeen erlazio sinpleekin arituko gara.

A gertaera ez jazotzea beste gertaera bat da, eta A-ren **kontrako gertaera** deritzo; $\overline{A}$ adierazten da. Alegia, A-n ez dauden elementuek osaturiko lagin-espazioaren azpimultzoa da. Esate baterako, A «positibo» baldin bada, $\overline{A}$ «negatibo» izango da.

A eta B gertaeren arteko bat gutxienez jazotzea beste gertaera bat da. Gertaera horri A-ren eta B-ren arteko **bildura** deritzogu eta $A \cup B$ adierazten da. Alegia, A-n edo B-n dauden elementuek osaturiko lagin-espazioaren azpimultzoa da. Esate baterako, 2.2 adibideko A eta B gertaerak hartzen baditugu, $A \cup B = \{1, 2, 3, 4, 5\}$ izango da.

A eta B bi gertaerak batera gertatzen direnean, jazotzen den gertaerari A-ren eta B-ren **ebakidura** deritzo, eta $A \cap B$ moduan adierazten da. Alegia, A-n eta B-n dauden elementuek osaturiko lagin-espazioaren azpimultzoa da. Esate baterako, aurreko adibideari jarraituz, $A \cap B = \{1\}$ izango da.

Bi gertaera ezin badira batera jazo, **gertaera bateraezin** deritze. Esate baterako, 2.2 adibideko B eta C gertaerak bateraezinak dira, $B \cap C = \varnothing$ betetzen baita.

A gertaera jazo eta B gertaera ez jazotzeari A eta B gertaeren arteko **kendura** edo **diferentzia** deritzo, eta $A - B$ adierazten da. Alegia, A-n egonik, B-n ez dauden elementuek osaturiko lagin-espazioaren azpimultzoa da. Esate baterako, 2.2 adibideko A eta B gertaerak hartzen baditugu, $A - B = \{2, 4\}$ izango da. $\overline{B} = \{2, 4, 6\}$ denez, $A - B = A \cap \overline{B}$ betetzen da.

A gertaera jazotzeak B gertaera ere jazotzea ondorioztatzen badu, ***A* parte *B*** dela esaten da eta $A \subset B$ adierazten da. Esate baterako, 2.2 adibideko F eta E gertaerak hartzen baditugu, F gertaera Eren parte da, $F \subset E$ betetzen baita.

De Morgan-en legeek azpimultzoen arteko honako erlazio hauek adierazten dituzte: A eta B multzo baten azpimultzoak badira, $\overline{A \cap B} = \overline{A} \cup \overline{B}$ eta $\overline{A \cup B} = \overline{A} \cap \overline{B}$ dira.

2.3 adibidea

Demagun hiru seme-alabako familien azterketa egiten dugula, umeen sexuaren arabera. Zortzi konposaketa posible ditugu, hots, $\Omega = \{aaa, saa, asa, aas, ass, sas, ssa, sss\}$.

Honako gertaera hauek hartzen ditugu: $A \equiv$ «Alaba bakarra dago», $B \equiv$ «Lehenengoa alaba da» eta $C \equiv$ «Hirurak alabak dira».

A eta B gertaerak konposatuak dira, $A = \{ass, sas, ssa\}$ eta $B = \{ass, asa, aas, aaa\}$ baitira. C, aldiz, oinarrizkoa da, $C = \{aaa\}$ baita.

$\overline{A} \equiv$ «Alaba bakarra ez egotea» = {alabarik ez, edo bi alaba edo hiru egotea} $= \{sss\} \cup \{saa, asa, aas\} \cup \{aaa\} = \{sss, saa, asa, aas, aaa\}$

$A \cap B =$ {alaba bakarra dago} $\cap$ {lehenengoa alaba da} $= \{ass, sas, ssa\} \cap \{ass, asa, aas, aaa\} = \{ass\}$. Hau da, lehena alaba bakarra da.

A eta C bateraezinak dira, $A \cap C$ = {alaba bakarra} $\cap$ {hirurak alabak} = $\{ass, sas, ssa\} \cap \{aaa\} = \varnothing$ baita.

$A \cup C = \{ass, sas, ssa, aaa\}$. Hau da, alaba bakarra dago edo hirurak alabak dira.

$A - B = \{ass, sas, ssa\} - \{ass, asa, aas, aaa\} = \{sas, ssa\}$. Hau da, alaba bakarra dago, baina ez da lehena.

2.4 PROBABILITATE KALKULUA

Zorizko esperimentu bati lotutako gertaeren ziurtasunaren edo ziurgabetasunaren neurria adierazten duen ezaugarri edo funtzio bat definitu nahi dugu. Horretarako, zorizko esperimentuan behaturiko gertaeraren maiztasun erlatiboa erabiliko dugu.

Demagun zorizko esperimentu baten n proba egiten direla, eta A gertaera f aldiz jazo dela. Orduan, A-ren maiztasun absolutua $f(A) = f$ da eta maiztasun erlatiboa $\frac{f}{n}$ da; $h(A)$ adierazten da. Maiztasun erlatiboak honako propietate hauek betetzen ditu:

1. $h(A) \in [0,1]$, edozein A gertaeratarako, zeren eta $f \leq n$ baitira ($f, n \geq 0$).
2. Izan bitez B eta C gertaera bateraezinak, non f_1 eta f_2 B eta C gertaeren maiztasun absolutuak baitira, hurrenez hurren. $B \cap C = \varnothing$ denez, $B \cup C$-ren maiztasun absolutua ($f_1 + f_2$) da. Beraz: $h(B \cup C) = \frac{f_1 + f_2}{n} = h(B) + h(C)$.
3. $h(\Omega) = 1$. Gertaera segurua beti jazotzen denez, bere maiztasun absolutua n da eta, ondorioz, bere maiztasun erlatiboa $\frac{n}{n} = 1$.

Intuitiboki, maiztasun erlatiboen ezaugarritzat hartzen ditugun hiru propietate horiek erlatiboki neurtzen digute gertaeren jazoera-maila. Erabil ditzagun, orduan, edozein gertaeraren probalitatea formalki definitzeko.

Izan bedi zorizko esperimentu bati lotutako Ω lagin-espazioa. ***P* probabilitate** deituko diogu $\wp(\Omega)$ definizio-barrutitzat duen eta bere balioak [0,1] tarte itxian hartzen dituen zenbakizko funtzioari:

$$\begin{aligned} P\colon \wp(\Omega) &\longrightarrow [0,1] \\ A &\longrightarrow P(A) \end{aligned}$$

honako **axioma** edo oinarrizko baldintza hauek betetzen baditu:

1. $0 \leq P(A) \leq 1$ edozein $A \in \wp(\Omega)$
2. $P(A \cup B) = P(A) + P(B)$ edozein $A, B \in \wp(\Omega)$ non $A \cap B = \varnothing$ baita.
3. $P(\Omega) = 1$

Lehen axiomak edo baldintzak gertaera bat jazotzearen probabilitatearen mugak adierazten ditu, alegia, 0 eta 1 bitartean. Bigarren axiomak edo baldintzak lagin-espazioko edozein bi gertaera bateraezinen arteko bilduraren probabilitatea beren probabilitateen batura izatera behartzen du. Hirugarren axiomak edo baldintzak gertaera segurua lortzearen ziurtasuna adierazten du.

Bigarren axioma gertaera bateraezinen edozein kopurutara heda daiteke. Hots, edozein $A_1, \ldots, A_n, \ldots$ gertaeratarako, non $A_i \cap A_j = \varnothing$ baita $\forall\ i \neq j$ honako hau betetzen da:

$$P(A_1 \cup \ldots \cup A_n \cup \ldots) = P(A_1) + \ldots + P(A_n) + \ldots$$

Batzuetan, ez da erraza izaten gertaera baten probabilitatea nola kalkulatu. Kasu horietan, balio enpiriko edo esperimental bat lor daiteke atal honen hasieran aipatu dugun bezala, gertaeraren maiztasun erlatiboa erabiliz. Demagun esperimentua hainbat aldiz errepikatzen dela, eta gertaeraren jazoera kopurua zenbatzen dela. Gertaeraren maiztasun erlatiboa honako zatidura honen bidez kalkulatzen da:

$$\frac{\text{Gertaeraren jazoera kopurua}}{\text{Proba kopuru osoa}}$$

Orduan, A gertaera jazotzeko probabilitatea $P(A) = \lim\limits_{n \to \infty} \left(\frac{f}{n} \right)$ da, non esperimentuaren n errepikapenetan A gertaeraren jazoera kopurua f baita, eta $n \to \infty$ adierazpenak errepikapen kopuru aski handia hartzen dela adierazten baitu. Beraz, n errepikapen kopurua nahiko handia baldin bada, $\frac{f}{n}$ kantitateak gutxi ezberdintzen dira elkarren artean; hots, zenbaki finko baterantz jotzen dute, eta zenbaki hori A gertaeraren probabilitatea da.

<u>Probabilitatearen oinarrizko propietateak</u>

1. $P(\overline{A}) = 1 - P(A) \qquad \forall\ A \in \wp(\Omega)$

   Bereziki: $P(\varnothing) = 1 - P(\Omega) = 1 - 1 = 0$

2. $P(B) \leq P(A)$ eta $P(A - B) = P(A) - P(B) \qquad \forall\ A, B \in \wp(\Omega)$ non $B \subset A$ baita.

Bereziki, honako bi hauek betetzen dira:

$P(B \cap A) \leq P(B) \leq P(B \cup A)$ eta $P(B \cap A) \leq P(A) \leq P(B \cup A)$

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $\forall\, A, B \in \wp(\Omega)$

   Beraz, $\forall\, A, B \in \wp(\Omega) \Rightarrow P(A \cup B) \leq P(A) + P(B)$

<u>Frogak</u>:

1. $\forall\, A \in \wp(\Omega) \Rightarrow A \cup \overline{A} = \Omega$

   $A \cap \overline{A} = \varnothing$ denez, $P(A \cup \overline{A}) = P(A) + P(\overline{A}) \Rightarrow P(\Omega) = P(A) + P(\overline{A}) \Rightarrow$
   $\Rightarrow 1 = P(A) + P(\overline{A}) \Rightarrow P(\overline{A}) = 1 - P(A)$.

2. $B \subset A$ denean, $A = B \cup (A - B)$ adieraz daiteke.

   $B \cap (A - B) = B \cap (A \cap \overline{B}) = \varnothing$ denez, $P(A) = P(B) + P(A - B) \geq P(B) \Rightarrow$
   $P(B) \leq P(A)$ eta $P(A) - P(B) = P(A - B)$.

3. $A \cup B = A \cup (B - A)$ eta $A \cap (B - A) = \varnothing$ denez, $P(A \cup B) = P(A) + P(B - A)$.

   $B = (B - A) \cup (B \cap A)$ eta $(B - A) \cap (B \cap A) = \varnothing$ denez,
   $P(B) = P(B - A) + P(B \cap A) \Rightarrow P(B - A) = P(B) - P(B \cap A)$.
   Beraz, $P(A \cup B) = P(A) + P(B) - P(B \cap A)$.

Formula hori gertaera kopuru handiagorako orokortu daiteke. Esate baterako, hiru gertaeraren kasuan, honako formula hau dugu:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Beste aldetik, $P(A \cup B) = P(A) + P(B) - P(B \cap A)$ berdintzatik abiatuz, $P(B \cap A) \geq 0$ denez, $P(A \cup B) \leq P(A) + P(B)$ dela ondorioztatzen da. Indukzioz, formula hori gertaera kopuru handiagorako orokortu daiteke. Hots,

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n)$$

2.5 EREDU PROBABILISTIKOAK

Izan bedi lagin-espazio finitua duen zorizko esperimentua: $\Omega = \{g_1, g_2, \ldots, g_{n-1}, g_n\}$. Orduan, zorizko esperimentuari loturiko gertaera guztiek osatzen duten multzoa, $\wp(\Omega)$, era honetara lortzen da: gertaera segurua (Ω), ezinezko gertaera ($\varnothing$), $\forall\, i \leq n$ oinarrizko gertaerak $\{g_i\}$, eta bildurak (oinarrizko gertaerak i-naka osatzen dituzten bildurak). Gure helburua $\wp(\Omega)$-n probabilitate bat definitzea da; horretarako, oinarrizko gertaeretan

definitzea nahikoa dela ikusiko dugu. Honako kasu hauek bereiziko ditugu: zorizko eredu klasikoa eta eredu probabilistiko finitua.

2.5.1 Zorizko eredu klasikoa

Probabilitatearen definizio klasikoa XVII. mendearen inguruan sortu zen, zorizko jokoei lotuta. Laplace-k honako definizio hau eman zuen: gertaeraren probabilitatea da gertaera jazotzearen aldeko kasuen eta kasu posibleen kopuruen arteko zatidura. Definizio hori oso eztabaidatua izan zen bere garaian; izan ere, kasu gehienetan betetzen zen, baina baziren salbuespen batzuk. Salbuespenak aztertzerakoan, definizio klasikoa egiazkoa izateko baldintza ezarri zen: kasu posible guztiak ekiprobableak izatearen beharra, hain zuzen.

Izan bedi oinarrizko gertaera kopuru finitua daukan zorizko esperimentua, lagin-espazioa $\Omega = \{g_1, g_2, \dots, g_{n-1}, g_n\}$ izanik. Gertaera horiek guztiak ekiprobableak dira, honako hau betetzen bada:

$$p = P(\{g_1\}) = P(\{g_2\}) = \dots = P(\{g_{n-1}\}) = P(\{g_n\})$$

Horren arabera, $\Omega = \{g_1\} \cup \{g_2\} \cup \dots \cup \{g_{n-1}\} \cup \{g_n\}$ denez eta oinarrizko gertaerak bateraezinak direnez, $P(\Omega) = P(\{g_1\}) + P(\{g_2\}) + \dots + P(\{g_{n-1}\}) + P(\{g_n\})$ da. Ondorioz:

$$1 = p + p + \dots + p + p \quad \Rightarrow \quad 1 = n \cdot p \quad \Rightarrow \quad p = \frac{1}{n}$$

Hots: $P(\{g_1\}) = P(\{g_2\}) = \dots = P(\{g_{n-1}\}) = P(\{g_n\}) = \dfrac{1}{n}$

Behin oinarrizko gertaeren probabilitatea ezagututa, oinarrizko gertaeren bilduraren bidez kalkulatuko da edozein gertaera konposaturen probabilitatea.

2.4 adibidea

Hiru seme-alabako familiekin jarraituz, gertaera hauen probabilitateak kalkulatuko ditugu: «alaba bakarra ez egotea» eta «alaba bakarra edo lehenengoa alaba izatea».

Aurreko 2.3 adibidean ikusi dugunez, $A \equiv$ «Alaba bakarra dago» = $\{ass, sas, ssa\}$ da eta $B \equiv$ «Lehenengoa alaba da» = $\{ass, asa, aas, aaa\}$.

$\Omega = \{aaa, saa, asa, aas, ass, sas, ssa, sss\}$ denez, Laplace-ren formula aplikatuz,

$$P(A) = \frac{A \text{ izatearen aldeko gertaera kopurua}}{\text{Gertaera kopuru osoa}} = \frac{3}{8}$$

$$P(B) = \frac{B \text{ izatearen aldeko gertaera kopurua}}{\text{Gertaera kopuru osoa}} = \frac{4}{8} = \frac{1}{2}$$

Alaba bakarra ez egotea $\overline{A}$ denez, $P(\overline{A}) = 1 - P(A) = \frac{5}{8}$.

Beste aldetik, alaba bakarra edo lehenengoa alaba izatea $A \cup B$ denez,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ da.

Baina, $A \cap B = \{ass, sas, ssa\} \cap \{ass, asa, aas, aaa\} = \{ass\}$ da $\Rightarrow$

$P(A \cap B) = \frac{1}{8}$.

Beraz, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{8} + \frac{1}{2} - \frac{1}{8} = \frac{3}{4}$.

Nahitaezkoa da gertaera baten probabilitatea kalkulatzeko gertaera eta lagin-espazioa osatzen duten oinarrizko gertaera kopurua ezagutzea. Horretarako, 2.9 atalean laburtzen den konbinatoriak eskaintzen dizkigun zenbatze-teknikak erabiliko ditugu.

2.5.2 Eredu probabilistiko finitua

Izan bedi oinarrizko gertaera kopuru finitua daukan zorizko esperimentua, alegia, $\Omega = \{g_1, g_2, \ldots, g_{n-1}, g_n\}$, baina gertaera horiek guztiak ekiprobableak ez izanik. Hots,

$P(\{g_1\}) = p_1, P(\{g_2\}) = p_2, \ldots, P(\{g_{n-1}\}) = p_{n-1}, P(\{g_n\}) = p_n$

Demagun honako bi propietate hauek betetzen direla:

1. $\forall\ 1 \le i \le n \qquad 0 \le p_i \le 1$
2. $P(\Omega) = 1 \quad \Rightarrow \quad p_1 + p_2 + \ldots + p_{n-1} + p_n = 1$

Oinarrizko gertaerak bateraezinak direnez, edozein gertaera konposatu oinarrizko gertaeren bilduraren modura adieraz daiteke.

2.5 adibidea

Hereda daitezkeen karaktereak geneen menpekoak dira. Horiek binaka agertzen dira eta, kasurik sinpleenean, bikotearen gene bakoitzak *A* eta *a* forma edo alelo har ditzake. Hots, hiru bikote era daitezke (zeren eta *Aa* = *aA* baita): *AA*, *aa* eta *Aa*. Ugalketa-prozesuan, umearen gene-pare bakoitzean,

aitarengandiko gene batek eta amarengandiko beste batek eragiten dute, eta horiek transmititzeko probabilitate berdina daukate.

Adibidez: Aita: *Aa* Ama: *Aa*

Umea

Aukerak: *AA* *Aa* *aa*

(baina ez dira ekiprobableak)

$$P(AA) = \frac{1}{4} \quad P(Aa) = \frac{1}{2} \quad P(aa) = \frac{1}{4} \qquad \text{(esperimentalki frogatua).}$$

Dena dela, benetako aukerak: *AA, Aa, aA* eta *aa*, ekiprobableak dira, baina lehen aipatu dugunez, *Aa* = *aA* da eta, orduan: $P(Aa)= \frac{1}{4}$ gehi $P(aA)= \frac{1}{4}$ eginez, $P(Aa$ edo $aA)= \frac{1}{2}$ dela lortzen dugu.

Demagun *B* gertaera konposatuak umea homozigotikoa izatea adierazten duela. Alegia, $B = \{AA, aa\}$. Orduan, $P(B) = P(AA) + P(aa) = \frac{1}{4}+\frac{1}{4}=\frac{1}{2}$.

<u>Oharra.</u> Arreta handia izan behar dugu hiru oinarrizko gertaera horiek ekiprobableak ez direla gogoratzeko. Esate baterako, umea homozigotikoa izateko probabilitatea ezin izango genuke honako era honetan kalkulatu:

$P(B) = \dfrac{B\text{ - ren aldeko gertaera kopurua}}{\text{Gertaera kopuru osoa}} = \dfrac{2}{3}$, emaitza posibleak (*AA*, *Aa*, *aa*) ez baitira ekiprobableak.

2.6 PROBABILITATE BALDINTZATUA ETA GERTAEREN ARTEKO INDEPENDENTZIA

Batzuetan, zera gertatzen da: zorizko esperimentuari buruzko informazio zehatza lor dezakegu, eta horrek berari buruzko ikuspen orokorra aldarazten digu. Horrela, adibidez, *A* gertaera gauzatu dela ezagutzen badugu, beste zenbait gauzen artean, $\overline{A}$ gertaera ez

dela gauzatu esan dezakegu, edo $B \subset A$ gertaerari gertatzeko probabilitate handiagoa eman diezaiokegu. Aurreko nozio intuitiboaren informazioa gure eredu probabilistikoari eransteko arazoa agertzen zaigu, beraz, segidan.

2.6 adibidea

Demagun ikastetxe bateko 6 urtetik beherako haurren populazioa 100 haurrek osatzen dutela. Horietatik 50 haurtzaindegian daude (4 urtetik beherakoak) eta beste 50 eskolaurrean (4 edo 5 urte). Haurren artean gaixotasun bat agertu da. Gaixotasun horren prozesu epidemiologikoa aztertu nahi dugu, eta haurren adinarekin zerikusia daukan baieztatu. 50 haur gaixo daude guztira; horietatik 40 haurtzaindegikoak dira eta 10 eskolaurrekoak.

Beraz, honako informazio hau daukagu: $P(\text{Gaixo}) = \dfrac{\text{Aldeko kasuak}}{\text{Kasu posibleak}}$

Guztira: $P(\text{Gaixo}) = \dfrac{50}{100} = 0{,}5$

Haurtzaindegian: $P(\text{Gaixo}) = \dfrac{40}{50} = 0{,}8$

Eskolaurrean: $P(\text{Gaixo}) = \dfrac{10}{50} = 0{,}2$

Beraz, ikastetxe horretan, ume bat haurtzaindegian gaixotzeko probabilitatea eskolaurrean gaixotzekoa baino handiagoa da. Alegia, umearen adina ezagutzeak bere gaixotzeko probabilitatea aldarazten du.

2.6.1 Probabilitate baldintzatua

Aurreko ideia abstraditu, honako definizio formal hau emango dugu. Izan bitez zorizko esperimentua eta $P(A) > 0$ baldintza betetzen duen hari elkartutako A gertaera. Demagun B beste edozein gertaera dela. Orduan, ***A* gertaerak baldintzaturiko *B* gertaeraren probabilitatea** $P(B \mid A)$ adierazten da, eta honako era honetara definitzen da:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

Kontuan hartu behar dugu $P(B \mid A)$ balioa edozein B gertaeratarako definitzen dela, eta probabilitate baldintzatua honako funtzio honek definitzen duela:

$$P(\cdot \mid A): \quad \wp(\Omega) \longrightarrow [0, 1]$$

$$B \longrightarrow P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

Funtzio hori probabilitate-neurria da, aurretik aipaturiko hiru axiomak betetzen baititu:

1. $0 \leq P(B \mid A) \leq 1 \qquad \forall B \in \wp(\Omega)$
2. $P(B \cup C \mid A) = P(B \mid A) + P(C \mid A) \qquad \forall B, C \in \wp(\Omega)$ non $B \cap C = \varnothing$ baita
3. $P(\Omega \mid \mathrm{A}) = 1$

Beraz, honako hiru propietate hauek ere beteko ditu:

1. $P(\overline{B} \mid A) = 1 - P(B \mid A) \qquad \forall B \in \wp(\Omega)$
2. $P(B \mid A) \leq P(C \mid A)$ eta $P(C - B \mid A) = P(C \mid A) - P(B \mid A)$

   $\forall B, C \in \wp(\Omega)$ non $B \subset C$ baita
3. $P(B \cup C \mid A) = P(B \mid A) + P(C \mid A) - P(B \cap C \mid A) \qquad \forall B, C \in \wp(\Omega)$

   Eta, beraz, $P(B \cup C \mid A) \leq P(B \mid A) + P(C \mid A)$

Aurreko adibidera bueltatzen bagara, $G \equiv$ Gaixoa, $H \equiv$ Haurtzaindegiko haurrak eta $E \equiv$ Eskolaurreko haurrak adierazten dituzten gertaerak erabiliz, hauxe izango litzateke ezarritako probabilitateen idazkera: $P(G) = 0{,}5$, $P(G \mid H) = 0{,}8$ eta $P(G \mid E) = 0{,}2$.

2.6.2 Gertaeren arteko independentzia

Intuitiboki, **gertaeren arteko independentzia**ren nozioak erakusten du ez dagoela eraginik gertaera horien artean. Esate baterako, 2.6 adibidean, gaixotasuna pairatzea eta umearen adina menpekoak direla dirudi. Nahiz eta gertaera askeen esanahiari buruzko ideia hartzea erraza izan, oso problematikoa gerta daiteke bi gertaera askeak izatearen arazo praktikoa. Gizakiaren pisua eta altuera, edo pultsua eta tenperatura, zenbateraino dira askeak? Zenbait aldagai biologikoren arteko menpekotasun-maila baliogabeko mailetatik elkarrekiko eragin handietara oszila daiteke. Alegia, ezin da absolutuki hitz egin menpekotasunaz ala independentziaz, menpekotasun-maila ezberdinez baizik. Dena den, bi gertaeraren arteko independentziaren definizio formala emango dugu.

Izan bitez A eta B gertaerak, $P(A) > 0$ eta $P(B) > 0$ izanik. A eta B **gertaera askeak** dira, honako hau betetzen baldin bada: $P(A \cap B) = P(A) \cdot P(B)$.

Horrek, praktikan, esan nahi du A gertaerak B gertaerarekiko informaziorik ez duela ematen. Edo, era berean, B gertaerak A-rekiko informaziorik ez duela ematen.

Honako baieztapen hauek baliokideak dira:

1. A eta B gertaerak askeak dira.

2. $P(A \cap B) = P(A) \cdot P(B)$
3. $P(A \mid B) = P(A)$, hots, A gertaera B-rekiko askea da.
4. $P(B \mid A) = P(B)$, hots, B gertaera A-rekiko askea da.

Azken bi atalek erakusten dute independentziaren nozioa simetrikoa dela, eta, beraz, «gertaera askeez» hitz egin dezakegula justifikatzen dute. Hiru horietatik edozein har daiteke independentziaren definiziotzat.

2.7 adibidea

Hiru seme-alabako familien kasuan, askeak al dira $B \equiv$ «Lehenengoa alaba da» eta $D \equiv$ «Bi sexuetako seme-alabak daude» gertaerak?

$\Omega = \{aaa, saa, asa, aas, ass, sas, ssa, sss\}$, $B = \{ass, asa, aas, aaa\}$ eta $D = \{ass, sas, ssa, aas, asa, saa\}$ direnez, Laplace-ren formula aplikatuz:

$$P(B) = \frac{4}{8} = \frac{1}{2} \text{ eta } P(D) = \frac{6}{8} = \frac{3}{4}$$

Beste aldetik, $B \cap D = \{ass, asa, aas\}$ denez, honako hau betetzen da:

$$P(B \cap D) = \frac{3}{8} = \frac{1}{2} \cdot \frac{3}{4} = P(B) \cdot P(D).$$

Beraz, B eta D gertaerak askeak dira.

Oharra. Bateraezintasuna eta independentzia guztiz kontzeptu ezberdinak dira, nahiz eta sarritan nahasten diren. Bateraezintasuna batera jazo ezin izatean datza: hau jazotzen baldin bada, beste hori ezin da jazo. Independentziaren esanahia da bataren jazoerak bestearengan eraginik ez izatea. Aurreko adibidean azterturiko gertaerak askeak ziren, eta bateragarriak. Posible da bi gertaera bateraezin eta menpeko aurkitzea ere.

2.8 adibidea

Hiru seme-alabako familien kasuan, bateraezinak dira $C = \{$hirurak alabak dira$\}$ eta $E = \{$hirurak semeak dira$\}$ gertaerak, baina ez dira askeak.

$$P(C) = P(E) = \frac{1}{8} \text{ dira eta } P(C \cap E) = 0 \text{ da, baina } 0 \neq \frac{1}{8} \cdot \frac{1}{8}.$$

2.6.3 Probabilitate Konposatuaren Teorema

Jadanik, posible dugu $P(A \cap B)$ kalkulatzea, A eta B gertaerak askeak direnean. Antzeko formula bat nahi genuke gertaerak askeak ez direnean erabili ahal izateko ere. Hori probabilitate baldintzatuaren definiziotik abiatuz lortuko dugu. Hots, $P(B \mid A) =$

$\frac{P(B \cap A)}{P(A)}$ erabiliz, $P(A \cap B) = P(A) \cdot P(B \mid A)$ dela lortuko dugu. Formula hori **probabilitate konposatuaren teorema** edo **biderkadura-legea** izenez ezagutzen da.

Era berean, erraz froga daiteke honako hau, baldin eta $P(A \cap B \cap C) > 0$ bada:

$P(A \cap B \cap C) = P(A \cap B) \cdot P(C \mid A \cap B) = P(A) \cdot P(B \mid A) \cdot P(C \mid A \cap B)$

Indukzioz, n gertaeratarako orokor daiteke aurreko formula, eta horrela probabilitate konposatuaren teoremaren adierazpen orokorra lor daiteke. Izan bitez $A_1, \ldots, A_n$ gertaerak, non $P(A_1 \cap A_2 \cap \ldots \cap A_n) > 0$ baita. Orduan, honako hau betetzen da:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot \cdot \cdot P(A_n \mid A_1 \cap \ldots \cap A_{n-1})$$

Probabilitate baldintzatu ezberdinak ezaguturik, n gertaera aldi berean gertatzeko probabilitatea kalkulatzeko metodoa ematen digu teorema horrek.

2.9 adibidea

Populazioen genetikako ikerketen arabera, Rh faktorearen informazioa duten geneen % 39k Rh negatiboa (Rh–) izatea ondorioztatu dute. Pertsona bat Rh– izango da, baldin eta soilik baldin Rh faktorearen informazioa duten bi geneak negatiboak badira, hau da, Rh– (––). Ama Rh– bada eta semea Rh positiboa (Rh+), odol-bateraezintasuna dago eta gerta daiteke amak fetuaren globulu gorriak deuseztatzen dituen antigorputzak sortzea. Zer probabilitate dago arazo hori gertatzeko? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

Arazoa sortzeko era hauxe da: Ama Rh– izanik, aita Rh+ da eta gene positibo bat ematen dio semeari. Beraz, arazo hori gertatzeko probabilitatea $P(A_1 \cap A_2 \cap A_3)$ izango da, non A_1 = {aita Rh+}, A_2 = {aitak gene positibo bat semeari ematea} eta A_3 = {ama Rh–}baitira.

Biderkadura Legearen arabera, kalkulua hauxe izango litzateke:

$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2)$

Baina A_3 gertaera A_1 eta A_2 gertaerekiko askea da, amaren genotipoak ez baitu eraginik aitarengan edo hark semeari genea ematean. Hau da, $P(A_3 \mid A_1 \cap A_2) = P(A_3)$. Beraz, $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3)$.

Gene bakoitza guraso batengandik jasotzen denez, bi geneak askeak direla onar daiteke. Horren ondorioz:

$P(A_3) = P(––) = 0{,}39 \cdot 0{,}39 = 0{,}1521$

$P(A_1) = P(++ \cup +- \cup -+) = 0{,}61 \cdot 0{,}61 + 0{,}61 \cdot 0{,}39 + 0{,}39 \cdot 0{,}61 = 0{,}8479$

Beste aldetik, aitak gene positibo bat semeari emateko probabilitatea kalkulatzeko, aita homozigotikoa ($O = ++$) ala heterozigotikoa ($E = +- \cup -+$) den bereizi behar dugu.

$P(O) = P(++) = 0{,}61 \cdot 0{,}61 = 0{,}3721$

$P(E) = P(+- \cup -+) = 0{,}61 \cdot 0{,}39 + 0{,}39 \cdot 0{,}61 = 0{,}4758$

Orduan, $P(A_2|A_1) = P(A_2 \mid O \cup E) = \dfrac{P[A_2 \cap (O \cup E)]}{P(O \cup E)} = \dfrac{P[(A_2 \cap O) \cup (A_2 \cap E)]}{P(O \cup E)} =$

$= \dfrac{P(A_2 \cap O) + P(A_2 \cap E)}{P(O) + P(E)} = \dfrac{P(O) \cdot P(A_2 \mid O) + P(E) \cdot P(A_2 \mid E)}{P(O) + P(E)} =$

$\dfrac{0{,}3721 \cdot 1 + 0{,}4758 \cdot 0{,}5}{0{,}3721 + 0{,}4758} = \dfrac{0{,}61}{0{,}8479} = 0{,}7194$

Beraz, $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3) = 0{,}8479 \cdot 0{,}7194 \cdot 0{,}1521 = 0{,}093$.

2.7 BAYES-en TEOREMA

Batzuetan, gertaera baten probabilitatearen kalkuluan, komenigarria izaten da beste gertaera batzuek daukaten eragina erabiltzea. Arazo horren ebazpena bideratzen duen probabilitate osoaren izeneko teorema frogatzen hasiko gara, laster tresna gisara erabiltzeko.

2.7.1 Probabilitate Osoaren Teorema

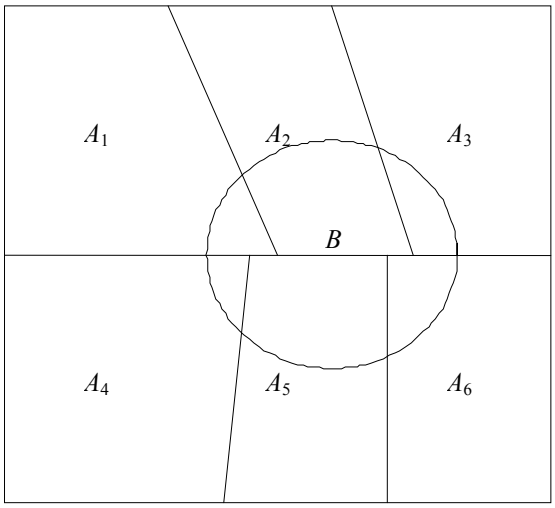
Izan bitez $A_1, \ldots, A_n$ **gertaera-sistema osoa**, hau da, $\bigcup_{i=1}^{n} A_i = \Omega$ eta $A_i \cap A_j = \varnothing$ $\forall\ i \neq j$ betetzen dituen n gertaerako familia. Demagun $P(A_i) > 0$ $i = 1, \ldots, n$ eta izan bedi B edozein gertaera. Orduan, honako hau beteko da:

$$P(B) = P(B \mid A_1) \cdot P(A_1) + \ldots + P(B \mid A_n) \cdot P(A_n)$$

<u>Froga</u>:

Kontuan har dezagun B gertaeren arteko bildura gisa adieraz daitekeela, alegia:

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{n} A_i \right) = (B \cap A_1) \cup \ldots \cup (B \cap A_n)$$



Lagin-espazioa $\Omega = \{A_1,A_2,A_3,A_4,A_5,A_6\}$

2.1 irudia. B gertaera eta $\{A_1, A_2, \dots, A_6\}$ gertaera-sistema osoa.

Gertaera horiek disjuntuak dira: $(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = \varnothing \quad \forall\, i \neq j$

Beraz, $P(B) = P(B \cap A_1) + \dots + P(B \cap A_n)$

Azkenik, $P(B \cap A_i) = P(B \mid A_i) \cdot P(A_i)$ dela gogoratuz, hauxe dugu:

$P(B) = P(B \mid A_1) \cdot P(A_1) + \dots + P(B \mid A_n) \cdot P(A_n)$

<u>Kasu berezia:</u> $\{A, \overline{A}\}$ gertaera-sistema osoa da edozein A gertaeratarako, eta, beraz, $P(B) = P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot P(\overline{A})$ betetzen da, beti, edozein B gertaerarentzat.

2.10 adibidea

Demagun tuberkulina testean negatiboa eman duten 10.000 pertsonen arteko batek tuberkulosia garatu duela, eta positiboa eman duten 100 pertsonen arteko batek garatu duela tuberkulosia. Beste aldetik, badakigu populazioaren % 1ak tuberkulina testean positibo eman duela. Populazioko pertsona bat zoriz aukeratzen badugu, zein da pertsona horrek tuberkulosia pairatzeko probabilitatea? (<u>Iturria</u>: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press,1995).

Testean positibo edo negatibo ematea + edo − adierazten baditugu, hurrenez hurren, {+, −} gertaera-sistema osoa da. T = {Tuberkulosia pairatzea} baldin bada,

$$P(T) = P(T\,|+) \cdot P(+) + P(T\,|-) \cdot P(-) = \frac{1}{100} \cdot 0{,}01 + \frac{1}{10.000} \cdot 0{,}99 =$$
$$= 0{,}000199$$

2.7.2 Bayes-en teorema

Izan bitez $A_1, \ldots, A_n$ gertaera-sistema osoa eta B beste edozein gertaera; guztien probabilitateak, berriz, hertsiki positiboak. B gertaerak baldintzaturiko familiaren edozein A_j gertaeraren probabilitatea modu honetan kalkula daiteke:

$$P(A_j \mid B) = \frac{P(B \mid A_j) \cdot P(A_j)}{P(B \mid A_1) \cdot P(A_1) + \ldots + P(B \mid A_n) \cdot P(A_n)}$$

<u>Froga</u>:

Definizioz, $P(A_j \mid B) = \dfrac{P(A_j \cap B)}{P(B)}$ da.

Zenbakitzailea $P(A_j \cap B) = P(B \mid A_j) \cdot P(A_j)$ erara ere jar daiteke eta, $P(B)$ adierazteko probabilitate osoaren teorema erabiltzen badugu, teoremaren adierazpena lortuko dugu berehala.

Interesgarria da Bayes-en teoremaren interpretazioaren azalpena egitea. Jo dezagun B gertaera $A_1, \ldots, A_n$ «kausa» ezberdinek (elkar ukatzaileek) sorrarazia izan daitekeela. Halaber, jo dezagun A_i kausaren presentzian B gertatzeko posibilitatea $P(B \mid A_i)$ dela (i = 1, … , n) eta A_i kausa bakoitzaren probabilitatea $P(A_i)$ dela (i = 1, … , n). B gertaera jazo dela behatu baldin badugu, A_i «kausak» sorrarazia izateko probabilitateaz galdetuko diogu geure buruari. Probabilitate hori Bayes-en teoremak ematen du. $P(A_i)$ zenbakiei «a priori» probabilitate esaten zaie, $P(B \mid A_i)$ zenbakiei egiantzak, eta $P(A_i \mid B)$ zenbakiei «a posteriori» probabilitateak.

2.11 adibidea

1997ko Osasun Inkestan lortutako datuen arabera, 3.782 pertsona ez dira mendira joan azken bi asteetan. Gertaera hori herrialdeka aztertuz, honako datu hauek lortu ditugu:

a) Arabarren % 88 ez da mendira joan azken bi asteetan.

b) Bizkaitarren % 96 ez da mendira joan.

c) Gipuzkoarren % 92 ez da mendira joan.

Beste aldetik, lehenengo gaiko 1.3 adibidean adierazten denez, 1997ko Osasun Inkesta bete dutenen % 12 arabarra da, % 48 bizkaitarra, eta % 40 gipuzkoarra. Badakigu zoriz aukeraturiko EAEko pertsona bat azken bi asteetan mendira joan dela. Zein da arabarra izateko probabilitatea?

Probabilitateak maiztasun erlatiboen bidez hurbiltzen baditugu, honako hauek ditugu:

$P(A) = 0{,}12$ $\quad P(B) = 0{,}48$ $\quad P(G) = 0{,}4$

$P(M|A) = 1 - P(\overline{M}|A) = 1 - 0{,}88 = 0{,}12$

$P(M|B) = 1 - P(\overline{M}|B) = 1 - 0{,}96 = 0{,}04$

$P(M|G) = 1 - P(\overline{M}|G) = 1 - 0{,}92 = 0{,}08$

Beraz, $P(A|M) = \dfrac{P(M|A)\cdot P(A)}{P(M|A)\cdot P(A) + P(M|B)\cdot P(B) + P(M|G)\cdot P(G)} =$

$$= \frac{0{,}12 \cdot 0{,}12}{0{,}12 \cdot 0{,}12 + 0{,}04 \cdot 0{,}48 + 0{,}08 \cdot 0{,}4} = \frac{0{,}0144}{0{,}0656} = 0{,}2195$$

2.8 TEST DIAGNOSTIKOA

Historia klinikoak, azterketa fisikoak, laborategiko analisiak eta erradiologia-ebaluazioak emandako informazioak gaixoaren egoera diagnostikatzen lagunduko digu. Gaixo bakoitzarentzat, sendagileak zenbait diagnosi posible hartzen ditu kontuan. Sendagileak, diagnosi bakoitzari buruzko azken informazioa erabiliz, aukera horien probabilitate erlatiboak doitzen ditu. Baina, azterketen emaitzek zenbateraino alda dezakete gaixotasunaren probabilitatea? Tuberkulosia detektatzeko testak positibo emateak ziurtatzen al digu gaixoak tuberkulosi aktiboa duenik? Argi dago ezetz, ez dagoelako test zeharo zehatzik; dena den, test horien emaitza positiboak gaixotasuna egotearen probabilitatea handitzen dute.

Gure ikuspuntu idealizatuaren arabera, daukagun informazioa inperfektua da, zeren eta, gaixotasunari probabilitatea ematerakoan, aldi berean eta pertsona bakoitzarengan gaixotasuna badagoela eta ez dagoela jotzen baitugu. Gaixotasunaren probabilitateak haren presentziari buruz daukagun sinesmen-maila neurtzen du. Gaixotasuna gaixoarengan dago; haren probabilitatea, ordea, gure buruan.

Sentikortasuna eta espezifikotasuna bi modu dira froga baten ziurtasuna deskribatzeko:

- Gaixotasuna pairatzen duten guztiek test «positiboa» badute, testa gaixotasunaren presentziarekiko **sentikorra** dela esango dugu.
- Gaixotasuna ez daukaten guztiek test «negatiboa» badute, testa gaixotasunaren ausentziarekiko **espezifikoa** dela esango dugu.

Testa espezifikoa eta sentikorra baldin bada, ez dago inolako zehaztasun faltarik haren interpretazioan. Baldintza ideal horiek oso gutxitan gertatzen dira. Dena den, test-diagnostikoak maiz ematen ditu gaixoaren benetako egoerarekin bat datozen emaitzak.

Atal honetan, testa zeharo zehatza ez denean, test diagnostikoak emandako informazioa nola erabili azaltzen da gaixotasunaren probabilitatea aldatzeko. Testaren emaitzak laburtzeko modu bat datuak, gaixotasuna dutenen eta ez dutenen arabera, 2×2 taulan azaltzean datza. Bi zutabek eta bi lerrok osatzen duten lau gelaxkako taula eraiki behar dugu. Bazter-totalak zutabe eta lerro bakoitzaren bukaeran azaltzen dira, eta total osoa eskuin-behe muturrean. Gaixotasunaren presentziaren ($G+$) edo ausentziaren ($G-$) arabera sailkatzen ditu pertsonak zutabe-faktoreak edo gaixotasunak. Lerro-faktoreak, testaren emaitzak, positibo ($T+$) edo negatibo ($T-$) ematearen arabera sailkatzen ditu pertsonak. 2.1 taulan agertzen da taula honen adierazpen orokorra.

		Gaixotasuna		
		Bai ($G+$)	Ez ($G-$)	Totala
Testa	$T+$	a	b	$a + b$
	$T-$	c	d	$c + d$
	Totala	$a + c$	$b + d$	n

2.1 taula. Gaixotasunaren eta testaren emaitzaren arabera antolaturiko 2 x 2 taula.

Definituriko 2×2 taulak era egokia eskaintzen digu, besteak beste, diagnosia daukaten pertsonen proportzioa edo testaren emaitza positiboa daukaten pertsonen proportzioa kalkulatzeko. Esate baterako, n pertsonak osaturiko taldean, gaixo daudenen proportzioa $\frac{a+c}{n}$ da, testa positiboa izan dutenen proportzioa $\frac{a+b}{n}$ da, edo testa positiboa izan eta gaixo daudenen proportzioa $\frac{a}{n}$ da.

Oro har, proportzioak ez baizik gaixotasuna pairatzeko probabilitatea edo testak positibo emateko probabilitatea ezagutu nahi ditugu, besteak beste. Horretarako, probabilitateak proportzioen bidez zenbatestea besterik ez dugu. Demagun n pertsonak osaturiko taldea populaziotik zoriz aukeratzen dela, eta gaixotasuna pairatzearen eta testaren emaitzaren arabera 2×2 taula eraikitzen dela. Orduan, honako gertaera hauen probabilitateak zenbatets ditzakegu, besteak beste:

P(Gaixotasuna) = $P(G+) = \dfrac{a+c}{n}$ eta P(Osasuna) = $P(G-) = \dfrac{b+d}{n}$

P(Test positiboa) = $P(T+) = \dfrac{a+b}{n}$ eta P(Test negatiboa) = $P(T-) = \dfrac{c+d}{n}$

P(Gaixotasuna eta test positiboa) = $P(G+ \cap T+) = \dfrac{a}{n}$

Honako bi hauek dira kontuan hartzeko erlazioak:

$$P(T+) + P(T-) = \frac{a+b}{n} + \frac{c+d}{n} = 1$$

$$P(G+) + P(G-) = \frac{a+c}{n} + \frac{b+d}{n} = 1$$

Alegia, pertsona baten testaren emaitzak positiboa edo negatiboa izan behar du nahitaez, eta gaixotasun-egoerak gaixoa edo osasuntsua.

Oharra. Emaitza horien interpretazioa kontu handiz egin behar da. Demagun, n pertsona zoriz aukeratu beharrean, n_1 gaixo eta n_2 osasuntsu aukeratu direla, $n_1 + n_2 = n$ izanik. Pertsona horiei testa egin ondoren, aurretik aipaturiko erara antolatu dira datuak 2×2 taulan. Alegia, zutabeetako bazter-totalak, zoriaren menpe egon beharrean, aurrefinkaturik daude, $a + c = n_1$ eta $b + d = n_2$. Egoera horretan, ezin da inolaz ere hitz egin bazter-total horiek parte hartzen duten gertaerei buruzko probabilitatearen estimazioaz. Esate baterako, gaixotasuna pairatzeko probabilitatea $\dfrac{a+c}{n} = \dfrac{n_1}{n}$ kantitateak estimatzen duela esaterakoan, akats larria egiten ari gara, bi kopuru horiek ikertzaileak aurrefinkaturikoak baitira. Praktikan arazo hori maiz agertzen da, eta konponbidea da uko egitea gaixotasuna pairatzearen edo ez pairatzearen probabilitatea 2×2 taulan oinarrituz estimatzeari. Egoera horrek ez du bermatuko atal honetan definituko ditugun gaineratiko kontzeptuak estimatzeko aukera.

Eritasuna detektatzerakoan, testaren zehaztasunaren neurri bat eritasuna pairatzen duten gizabanakoek test positiboa emateko probabilitatea da. $P(T+|G+)$ probabilitate baldintzatu horri **testaren sentikortasuna** deritzogu. Aurreko 2×2 taulan oinarrituz, testaren sentikortasunaren estimazioa $\dfrac{a}{a+c}$ da.

Testaren zehaztasunaren beste neurri bat eritasuna pairatzen ez duten gizabanakoek test negatiboa emateko probabilitatea izango litzateke. $P(T-|G-)$ probabilitate baldintzatu

horri **testaren espezifikotasuna** deritzogu. Aurreko 2×2 taulan oinarrituz, testaren espezifikotasunaren estimazioa $\frac{d}{b+d}$ da.

Testaren sentikortasuna eta espezifikotasuna ez daude eritasunaren probabilitatearen menpe. Nahiz eta ospitale bakoitzean, edo talde bakoitzean, eritasunaren probabilitatea ezberdina izan, testaren sentikortasunak eta espezifikotasunak konstante diraute. Horren ondorioz, eritasuna maizago edo nekezago izaten den egoera kliniko bakoitzean, beste ikertzaile batek zehazturiko testaren sentikortasuna eta espezifikotasuna erabil daitezke.

Aurretik aipatu dugunez, test idealik ez da existitzen. Sentikortasuna eta espezifikotasuna elkarri lotuta dauden kontzeptuak dira, eta bata txikitu nahiak bestea handitzera garamatza. Praktikan, testak ebaluatzeko, balio aurresankor positibo eta negatibo kontzeptuak erabiltzen dira. Nahiz eta sentikortasuna eta espezifikotasuna ez egon gaixotasunaren prebalentziaren edo gaixoen proportzioaren menpe, balio aurresankorrak bai testaren zehaztasunaren eta bai eritasunaren prebalentziaren menpekoak dira.

Testean positibo eman duen gizabanakoak gaixotasuna pairatzeko probabilitateari **testaren balio aurresankor positiboa** deritzogu, eta $P(G+|T+)$ probabilitate baldintzatuak ematen du. Aurreko 2×2 taulan oinarrituz, testaren balio aurresankor positiboaren estimazioa $\frac{a}{a+b}$ da.

Testean negatibo eman duen gizabanakoak gaixotasuna ez pairatzeko probabilitateari **testaren balio aurresankor negatiboa** deritzogu, eta $P(G-|T-)$ probabilitate baldintzatuak ematen du. Aurreko 2×2 taulan oinarrituz, testaren balio aurresankor positiboaren estimazioa $\frac{d}{c+d}$ da.

2.12 adibidea

Beller-ek eta laguntzaileek (1971) suero digitalikoaren kontzentrazioaren eta toxizitate digitalikoaren arteko erlazioa aztertu zuten. Boston City ospitalean, ikerketa-epearen barnean onartutako 931 gaixoen artean digoxina eta digitoxina hartzen ari ziren 135 gaixo ikertu zituzten. Ospitaleko gaixo guztien artean, digitalis prestakina jasotzen zuten 135 gaixo horiek osatu zuten beren populazioa. Digitalis intoxikazioa definitu ondoren, sueroko digoxina-maila erabili zuten test gisa, eta 2.2 taulan agertzen dira berek lortutako datuak.

Honako hauek izango dira sueroko digoxina-mailak digitalis intoxikazioa aurresateko daukan sentikortasuna eta espezifikotasuna:

Sentikortasuna = $P(T+ \mid G+) = \frac{25}{43} = 0{,}58$

$$\text{Espezifikotasuna} = P(T- \mid G-) = \frac{78}{92} = 0{,}85$$

		Toxizitatea		
		Bai ($G+$)	Ez ($G-$)	Totala
Testa	$T+$	25	14	39
	$T-$	18	78	96
	Totala	43	92	135

2.2 taula. 2.12 adibideko datuak. Iturria: Ingelfinger J.A., Mosteller, F., Thibodean L.A., Ware J.H., *Biostatistics in Clinical Medicine*. MacMillan Publishimg Co., Inc. New York, 1987.

Beller-ek eta laguntzaileek ikertutako 135 gaixoekin lortutako sentikortasuna eta espezifikotasuna kontuan hartuz, test horren balio aurresankor positiboa eta negatiboa:

$$\text{Balio aurresankor positiboa} = P(G+ \mid T+) = \frac{25}{39} = 0{,}64$$

$$\text{Balio aurresankor negatiboa} = P(G- \mid T-) = \frac{78}{96} = 0{,}81$$

Beraz, ospitalera ailegatzen den pertsona batek gaixotasuna pairatzeko duen probabilitatea $P(G+) = \frac{43}{135} = 0{,}32$ da. Bere sueroko digoxina-maila positiboa dela aztertu ondoren, ordea, gaixotasuna pairatzeko probabilitatea 0,64 da.

Oharra. 2.12 adibidean aztertu ditugun datuetan, Beller-ek eta laguntzaileek, ospitaleko gaixo guztien artean, digitalis prestakina jasotzen zuten 135 gaixoek osaturiko populazioa erabili zuten sueroko digoxina-mailaren arabera definituriko testaren sentikortasuna eta espezifikotasuna kalkulatzeko. Praktikan, sentikortasun eta espezifikotasun ezagunak dituzten testak erabiltzen dira, gure populazioan eritasunak daukan prebalentziaren arabera, testaren balio aurresankorrak estimatzeko. Prozedura hori gauzatzeko, testaren bi zehaztasun-neurriak egiantza gisa eta gaixo/osasun probabilitateak «a priori» probabilitate gisa erabiliz, Bayes-en teoremaren bidez kalkulatzen dira balio aurresankorrak. Ikus dezagun metodo hori praktikan nola gauzatzen den adibide baten bidez.

2.13 adibidea

Demagun odol-presioa automatikoki neurtzeko tresna batek hipertentsotzat sailkatzen dituela hipertentsoen % 84 eta normotentsoen % 23. Zein dira

tresnaren balio aurresankor positiboa eta balio aurresankor negatiboa, pertsona helduen % 20 hipertentsoak badira?

Sentikortasuna = $P(T+ \mid G+) = 0{,}84$

Espezifikotasuna = $P(T- \mid G-) = 1 - P(T+ \mid G-) = 1 - 0{,}23 = 0{,}77$

Gaixotasunaren prebalentzia = $P(G+) = 0{,}2$

Bayes-en teorema erabiliz:

Balio aurresankor positiboa = $P(G+ \mid T+) =$

$$\frac{P(T+|G+)\cdot P(G+)}{P(T+|G+)\cdot P(G+)+P(T+|G-)\cdot P(G-)} =$$

$$= \frac{0{,}84\cdot 0{,}2}{0{,}84\cdot 0{,}2+0{,}23\cdot 0{,}8} = \frac{0{,}168}{0{,}352} = 0{,}477$$

Balio aurresankor negatiboa = $P(G- \mid T-) =$

$$\frac{P(T-|G-)\cdot P(G-)}{P(T-|G+)\cdot P(G+)+P(T-|G-)\cdot P(G-)}$$

$$= \frac{0{,}77\cdot 0{,}8}{0{,}16\cdot 0{,}2+0{,}77\cdot 0{,}8} = \frac{0{,}616}{0{,}648} = 0{,}95$$

Tresnaren emaitza negatiboa oso aurresankorra da, % 95eko ziurtasun-mailaz pertsona hori normotentsoa dela baiezta baitezakegu. Tresnaren emaitza positiboa, aldiz, ez da oso aurresankorra, % 48 ziurtasun-mailaz soilik esan baitezakegu pertsona hori hipertentsoa dela.

2.9 KONBINATORIA (hautazkoa)

Aurreko 2.5. atalean gertaera baten probabilitatea kalkulatzeko formulak aurkeztu ditugu. Horretarako, gertaera eta lagin-espazioa osatzen duten oinarrizko gertaera kopurua ezagutzea nahitaezkoa dela aipatu dugu. Hori egiteko aukera eskaintzen digu konbinatoriak. Adibide gisa m elementuko multzoa emanik, zenbat eratara ordena daiteke? m elementuren arteko n elementuko azpimultzoak ($n \leq m$) eratu nahi badira, zenbat aukera daude? Konbinatorian, horiek eta antzeko egoerak aztertzen dira. Multzo finitu bateko elementuek osaturiko azpimultzo kopurua zenbatzeko erabiltzen den teknika-sistemari **konbinatoria** deritzo, eta honako hauek dira azpimultzoak eratzeko irizpideak: aukeratzen diren elementuak, beren ordena, edo biak.

Aldez aurretik, behar ditugun zenbait nozio gogoraraziko ditugu. Lehena ***m* faktoriala** ($m!$) da. Edozein m zenbaki arrunterako, m lehen zenbaki arrunten arteko biderkadura da, hots, $m! = m\,(m-1)\,(m-2)\cdots(2)\,(1)$. Horregatik, honako hau betetzen da:

$$m! = m\,(m-1)! = m\,(m-1)\,(m-2)! = \ldots$$

Komenientziazko arrazoiak direla eta, $0! = 1$ definitzen da.

Halaber, m, n zenbaki arruntak emanik, $n \leq m$ izanik ($m = n = 0$ izan daitezke), horrela definituko dugu **«*m* gain *n*» konbinazio-zenbakia**: $\binom{m}{n} = \frac{m!}{n!(m-n)!}$

Hauxek dira frogatzeko errazak diren propietateak:

- $\binom{m}{1} = m$
- $\binom{m}{0} = \binom{m}{m} = 1$
- $\binom{m}{n} = \binom{m}{m-n}$
- $\binom{m}{n} + \binom{m}{n+1} = \binom{m+1}{n+1}$

Egingo ditugun baieztapenak frogatu nahi gabe, hel diezaiogun gaiari. Edozein m elementuko multzotik atera daitezkeen m elementuko taldeen ezberdintasuna elementuen ordenean soilik datza, eta ***m* elementuren permutazio (arrunt)** deritze; P_m adierazten dira.

Baina, multzoko elementuen arteko n ($n \leq m$) elementuko taldeak eraiki nahi baditugu, bi aukera ditugu. Taldeak ezberdintzat hartzen ditugu, elementuren bat desberdina denean. Kasu horretan, talde ezberdinei ***n*-naka hartutako *m* elementuren konbinazio (arrunt)** deritze, eta C_m^n adierazten dira. Beren kopurua $C_m^n = \binom{m}{n} = \frac{m!}{n!(m-n)!}$ da. Taldeak ezberdintzat hartzen ditugu, elementuren bat desberdina denean, edo, elementu berberak izanik, beren ezarrera-ordena desberdina denean. Kasu horretan, talde ezberdinei ***n*-naka hartutako *m* elementuren aldakuntza (arrunt)** deritze, eta V_m^n adierazten dira. Beren kopurua $V_m^n = \frac{m!}{(m-n)!}$ da.

Kontuan har dezagun honako propietate hau: $C_m^n \cdot P_n = V_m^n$.

Elementuak errepikatzea posible denean, haietatik atera daitezkeen n elementuko (n hautazkoa) taldeak era berean definitzen dira, eta **errepikatuzko permutazio, aldakuntza eta konbinazio** deritze.

*V' e***rrepikatuzko aldakuntzak**: bi talde bereizteko irizpidea aldakuntza arruntena da, eta haien kopurua $V'^{n}_{m} = m^{n}$ da.

*C' e***rrepikatuzko konbinazioak**: bi talde bereizteko irizpidea konbinazio arruntena da, eta haien kopurua $C'^{n}_{m} = \binom{m+n-1}{n}$ da.

$P_{n}^{n_1,n_2,\ldots,n_{k-1},n_k}$ **errepikatuzko permutazioak**: n elementuen ordenazio ezberdinak dira, non elementu bat n_1 aldiz, beste bat n_2 aldiz… azkenaurrekoa n_{k-1} aldiz, azkena n_k aldiz, eta $n_1 + n_2 + \ldots + n_{k-1} + n_k = n$ baitira. Haien kopurua hauxe da: $P_{n}^{n_1,n_2,\ldots,n_{k-1},n_k} = \dfrac{n!}{n_1!\,n_2!\ldots n_{k-1}!\,n_k!}$.

2.10 ARIKETAK

2.1. Rh negatiboa daukan odol-emaile bat behar dugu, transfusioa egiteko. Eskaria egin ondoren, sei pertsonako familia aurkeztu da odola emateko prest. Familia horretan bi Rh negatibo daudela dakigu, baina ez dakigu zein diren. Familiako kideen odola banan-banan aztertu da, Rh negatibodun bat aurkitu arte emaile izateko. Deskriba ezazu esperimentuaren lagin-espazioa.

2.2. Demagun Euskal Autonomi Elkarteko eta Nafarroako mendiek osaturiko lagin-espazioa aztertzen dugula. Honako gertaera hauek definitzen dira:

$H \equiv$ 1.200 metro baino altitude handiagoa.

$E \equiv$ 1.000 eta 1.200 metro arteko altitudea.

$T \equiv$ 600 eta 1.000 metro arteko altitudea.

$B \equiv$ Bizkaiko lurraldean.

$G \equiv$ Gipuzkoako lurraldean.

$A \equiv$ Arabako lurraldean.

$N \equiv$ Nafarroako lurraldean.

$O \equiv$ Euskal Autonomi Elkartean.

$P \equiv$ Pirinio mendikatean.

a) Zer adierazten dute honako gertaera hauek: $H \cup E \cup T$, $B \cup G \cup A$, $H \cap N$, $T \cap A$, $H \cap \overline{N}$, $H \cap \overline{P}$, $H \cap E \cap T$, $N \cap P$, $O - B$, $\overline{O}$?

b) B eta A bateraezinak dira? Eta B eta N? Eta G eta T?

c) Adieraz ezazu H, *E-ren* eta T-ren arabera.

d) Adieraz ezazu O, *B-ren*, *G-ren* eta A-ren arabera.

e) Adieraz ezazu $\overline{N}$, *B-ren*, *G-ren* eta A-ren arabera.

2.3. Pazientea ikusterakoan, honako sintoma hauek dituen ala ez neurtzen du sendagileak: buruko mina, sukarra, azaleko negala edo muskuluetako mina. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Eraiki ezazu sintomen 16 konbinazio posible adierazten dituen zuhaitz-diagrama.

b) Janariei alergia sistemikoaren diagnosia edukitzeko, azaleko negala izan behar da. Seinala itzazu diagnosi horretara eramaten gaituzten bideak.

c) Gripearen diagnosia edukitzeko, sukarra eta muskuluetako mina izan behar dira. Seinala itzazu diagnosi horretara eramaten gaituzten bideak.

d) Ba al dago aurreko bi diagnosietara eramaten gaituen bideren bat? Horrela bada, zein dira bide horiek?

e) Seinala itzazu aurreko bi diagnosietatik bat ere ondorioztatzen ez duten bide guztiak.

2.4. EAEko populazioren % 54,9 Bizkaian bizi da, eta EAEko populazioaren % 29 EAEtik kanpo jaioa da. Halaber, EAEko populazioaren % 16,6 EAEtik kanpo jaioa da eta Bizkaian bizi da (Iturria: EUSTAT. *EUE* 97/98.). Datu horien arabera, kalkula itzazu honako gertaera hauen probabilitateak EAEko biztanle batentzat:

a) EAEn jaioa izatea.

b) Bizkaian ez bizitzea.

c) EAEtik kanpo jaioa izatea edo Bizkaian bizitzea.

d) EAEn jaioa izatea edo Bizkaian ez bizitzea.

e) EAEn jaioa izatea eta Bizkaian ez bizitzea.

f) EAEtik kanpo jaioa izatea eta Bizkaian ez bizitzea.

g) EAEn jaioa izatea eta Bizkaian bizitzea.

2.5. Familia batek lau seme-alaba ditu; sexuaren eta adinaren arabera ezberdintzen dira beren artean. Adieraz itzazu dauden 16 posibilitateak zuhaitz-diagrama baten bidez (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994). Seme bat edo alaba bat izateko probabilitatea berdina baldin bada, kalkula itzazu honako gertaera hauen probabilitateak:

a) $A \equiv$ «Lehena semea da».

b) $B \equiv$ «Zehazki bi seme dira».

c) $C \equiv$ «Nagusia eta txikiena semeak dira».

d) $D \equiv$ «Alaba bat eta hiru seme dauzkate».

2.6. Galsoro batean, A, B eta C hiru ongarri mota erabiltzen dira. Galsoroa tamaina bereko hiru zatitan banatzen da, eta bakoitzean ongarri mota ezberdina

erabiltzen da (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Eraiki ezazu hiru lur zatietan ongarrien esleipenerako dauden sei era ezberdinak adierazten dituen zuhaitz-diagrama.

b) Esleipena zorizkoa izan baldin bada, zuhaitz-diagramaren bide bakoitza ekiprobablea izango da. Zein da lehen zatian *A* ongarria erabiltzeko probabilitatea?

2.7. Fruta-euliaren gorputza bi koloretakoa izan daiteke: grisa (*E*) eta ebanoa (*e*); grisa da menperatzailea. Hegoak ere bi motatakoak izan daitezke: normalak (*V*), eta laburrak edo bestigialak (*v*); normalak dira menperatzaileak. Euli homozigotikoak (*EEVV* eta *eevv*) elkarrekin gurutzatzean, ondorengo guztiak heterozigotiko bikoitiak izaten dira (*EeVv*). (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994). Euli horietako bi gurutzatzen badira:

a) Adieraz itzazu hirugarren belaunaldiko ondorengoaren emaitza posibleak, zuhaitz-diagrama erabiliz.

b) Zenbat bidetatik lortuko da ebano koloreko euli bat?

c) Zenbat bidetatik lortuko da hego normaleko euli bat?

d) Euli bat zoriz aukeratzen baldin bada, zein da hego normalak edukitzeko probabilitatea? Zein da ebano kolorekoa eta hego laburduna izateko probabilitatea? Zein da hego laburduna eta grisa izateko probabilitatea?

2.8. Emakume bat hemofilia deritzon gaixotasunaren eramailea da. Alegia, nahiz eta emakumeak gaixotasuna ez pairatu, seme-alabei kutsa diezaieke. Emakumeak hiru ume ditu. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Zein da esperimentuaren lagin-espazioa?

b) Zein da ume bakar batek ere hemofilia ez pairatzeko probabilitatea?

c) Zein da, zehazki, bi umek hemofilia pairatzeko probabilitatea?

2.9. Kimikari batek papergintza-enpresa bateko hondakin-uretatik hartutako hamar ur-lagin ditu. Kimikariak jakin gabe, horietatik lauk azido gehiegi daukate. Zoriz hiru lagin aukeratzerakoan, zein da zehazki bi azido gehiegi dituztenetarikoak izateko probabilitatea? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

2.10. Zazpi sendagaik aknearen kontra duten eraginkortasuna neurtu da. Ikertzaileak sendagaiak ordenan jarri ditu eraginkortasunaren arabera: 1 balioa eraginkorrenari eta 7 balioa eraginkortasun baxuena duenari. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Zenbat eratara eman dakizkieke zazpi balioak zazpi sendagaiei?

b) A eta B sendagaiak proba honetan interesaturik dagoen enpresa batek fabrikatu ditu, baina hori ez daki ikertzaileak. Demagun ikertzaileak ezin duela zazpi sendagaien eraginkortasuna bereizi, eta balioak zoriz banatu dituela. Zein da A edo B sendagaia lehen postuan (1 balioa) gelditzeko probabilitatea?

2.11. Demagun aitak, amak eta bi seme-alabak osaturiko familia daukagula. (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1995). Honako gertaera hauek definitu dira:

$A \equiv$ «Amak gripea dauka».

$B \equiv$ «Aitak gripea dauka».

$C \equiv$ «Seme-alaba nagusiak gripea dauka».

$D \equiv$ «Seme-alaba gazteak gripea dauka».

$E \equiv$ «Seme-alaba batek behintzat gripea dauka».

$F \equiv$ «Gurasoetako batek behintzat gripea dauka».

$G \equiv$ «Familiako pertsona batek behintzat gripea dauka».

a) Zer adierazten dute gertaera hauek: $A \cup B$, $A \cap B$, $C \cup E$, $C \cap E$, $\overline{A}$, $\overline{B}$?

b) C eta D bateraezinak al dira?

c) Adieraz ezazu E, *C-ren* eta D-ren arabera.

f) Adieraz ezazu G, *E-ren* eta F-ren arabera.

d) Adieraz ezazu $\overline{G}$, *E-ren* eta F-ren arabera.

Demagun hiri batean gripe-epidemiari buruz dauden datuak honako hauek direla: familien % 10ean amak gripea dauka; familien % 10ean aitak gripea dauka; familien % 2an aitak eta amak (biek) gripea daukate. Egoera horretan:

e) A eta B gertaerak askeak al dira?

Halaber, gripea edukitzeko ume bakoitzaren probabilitatea % 20koa da; baina bi umeko familien % 10ean bi umeek gripea daukate. Egoera horretan:

f) Zein da ume batek behintzat gripea edukitzeko probabilitatea?

2.12. Animalien ezaugarri batzuek sexuaren eragina daukatela esaten da. Adibidez, ardien adakera alelo bikote batek menperatzen du: *H* eta *h*. *H*, adakera presentziaren aleloa, arrengan menperatzailea eta emeengan errezesiboa da. *h*, adakera ezaren aleloa, arrengan errezesiboa eta emeengan menperatzailea da. Beraz, ar heterozigotiko bat eta eme heterozigotiko bat (*Hh*) izanik, arrak adakera izango du eta emeak ez (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994). Demagun bi animalia horiek parekatzen direla.

a) Adieraz ezazu, zuhaitz-diagrama baten bidez, adakera presentziaren arabera ager daitekeen ondorengo posiblea.

b) Ume bakoitzak arra zein emea izateko probabilitate bera du. Kalkula ezazu ume bat adakeradun arra izateko probabilitatea. Kalkula ezazu ume bat adakeradun emea izateko probabilitatea.

c) Kalkula ezazu ume batek adakera izateko probabilitatea. Froga ezazu arra izatea eta adakeraduna izatea ez direla askeak.

2.13. Bakteriorik egotekotan, testak *B* motako bakterioen presentzia detektatzen du uretan, eta 0,9 da probabilitatea. Bakteriorik ez badago, egoneza detektatzen du, eta 0,8 da probabilitatea. Ur-laginak benetan *B* motako bakterioak edukitzeko probabilitatea 0.2 dela jakinik, kalkula itzazu honako hauek (Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991):

a) Testa positibo ateratzeko probabilitatea.

b) Testa positiboa izan denean, ur-laginak *B* motako bakterioak edukitzeko probabilitatea.

c) Testa negatiboa izan denean, ur-laginak *B* motako bakterioak edukitzeko probabilitatea.

d) Kalkula itzazu test horrek positibo eta negatibo faltsuak emateko probabilitateak.

e) Ur-laginak *B* motako bakterioak edukitzeko, eta, gainera, testa positibo izateko probabilitatea.

2.14. EAEn, elikadura-desoreka dela kausa, honako hau da psikiatrarengana joaten diren pazienteen banaketa, beren diagnosiaren arabera: anorexia murriztailea (AM) % 28, anorexia bulimikoa (AB) % 30, bulimia (B) % 33 eta betekada-desoreka (BD) % 9. Lehenengo kontsultara sintoma larriekin heltzeko probabilitatea diagnosiaren menpekoa da: anorexia murriztailerako 0.3, anorexia bulimikorako 0.57, bulimiarako 0.47 eta betekada-desorekarako 0.18 .

a) Kalkula ezazu elikadura-desorekadun edozein pazientek bere lehenengo kontsultan sintoma larriak edukitzeko probabilitatea.

b) Kalkula ezazu lehenengo kontsultan sintoma larririk erakusten ez duen pazientearen diagnosia bulimia izateko probabilitatea.

c) 1998. urtean Osakidetzako ospitale batera 67 paziente joan ziren elikadura-desoreka zela kausa. Kalkula ezazu horietatik gutxienez bosten diagnosia betekada-desoreka izateko probabilitatea.

2.15. Gizabanakoaren *Rh* ezaugarria alelo bikote batek menperatzen du: presentziak (+) eta ausentziak (–), eta positiboa da menperatzailea. Izan bedi honako zuhaitz genealogiko hau:

$$\underbrace{\underbrace{+-}_{X} \quad +-}_{} \quad +-$$

$$\text{Ondorengoa} = \{+-,\ --,\ --,\ --\}$$

Aurki itzazu *X* gizabanakoa:

a) Homozigotiko positiboa izateko probabilitatea.

b) Heterozigotiko positiboa izateko probabilitatea.

c) Negatiboa izateko probabilitatea.

2.16. Euskal Herriko odol-taldeen banaketa, azken mendean zehar, honako hau izan da: % 43 A taldekoak, % 8 B-koak, % 3 AB-koak eta % 46 O-koak. Gerra garaian, A taldeko % 88 ongi sailkatu zela jotzen da, baina O, B eta AB taldeetako % 4, % 4 eta % 10, hurrenez hurren, A taldekotzat sailkatu zirela. Soldadu bat zaurituz gero, erizaindegira eramaten zuten, eta A taldeko gisa sailkatzen zuten. Zein da soldadu hori benetan talde horretakoa izateko probabilitatea?

2.17. Badakigu familia batek seme bat edo alaba bat izateko probabilitatea ½ dela. (Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991).

a) Bi umeko familia bat aukeratu dugu, eta haien arteko bat gizonezkoa da; zein da biak gizonezkoak izateko probabilitatea?

b) Izan bedi lau umeko familia bat. Askeak al dira $H \equiv$ «Familiak bi sexuetako umeak dauzka» gertaera eta $A \equiv$ «Ez dago emakumezko bat baino gehiago» gertaerak?

2.18. Jakina da pertsona osasuntsuen kolesterol-maila adinaren menpekoa dela. Hiri txiki batean behaturiko datuen arabera, honako hau da adinaren banaketa: % 30, 21 urte baino gazteagoa; % 20, 21 eta 29 urte bitartekoa (biak barne); eta

gainerakoa, 29 urte baino gehiagokoa. Halaber, pertsona baten kolesterol-maila altua izateko probabilitatea 0,05, 0,1 edo 0,3 da, hurrenez hurren, 21 urtetik beherakoa, 21 eta 29 urte bitartekoa, edo 29 urte baino gehiagokoa izatearen arabera.

a) Kalkula ezazu zoriz aukeraturiko pertsonaren kolesterol-maila altua ez izateko probabilitatea.

b) Aurreko pertsonak kolesterol-maila altua ez duela jakinik, zein da bere adina 21 eta 29 urte bitartekoa izateko probabilitatea?

c) Askeak al dira «29 urte baino gehiago edukitzea» eta «kolesterol-maila altua edukitzea» gertaerak?

2.19. Irakaskuntza ertaineko ikastetxe guztietako ikasleei tuberkulosiaren proba egiten zaie, tuberkulosia errotik ateratzea helburu duen kanpainan zehar. Proba horren sentikortasuna eta espezifikotasuna 0,98 eta 0,95 dira, hurrenez hurren. Halaber, jakina da proba egindako ikasle taldearen % 1ak tuberkulosia pairatzen duela.

a) Kalkula ezazu zoriz aukeraturiko ikasle baten proba negatiboa izateko probabilitatea.

b) Zoriz aukeraturiko ikasle baten proba positiboa izan dela jakinik, kalkula ezazu ikasle horrek tuberkulosia pairatzeko duen probabilitatea. Eta proba negatiboa izan bada?

c) Zenbat positibo faltsu eta negatibo faltsu esperoko dira 1.800 umek osaturiko taldean?

2.20. Tuberkulosiaren presentzia detektatzeko proba-diagnostikoa aplikatu da ehungintzako langile talde batean. Orain arte egindako ikerketen arabera, mota horretako langileengan gaixotasunaren prebalentzia (gaixorik dauden pertsonen proportzioa) % 14 dela onartu da. Ikerketa horietan, behatu da langile osasuntsuen % 17k positibo ($T+$) eta langile gaixoen % 5ak negatibo ($T-$) eman dutela. Kalkula ezazu osasuntsu (O) dauden eta positibo eman duten langileen proportzioa; eta gaixorik (G) dauden eta negatibo eman duten langileen proportzioa (Iturria: A. Amparan, I. Arostegi, B. Del Hoyo, S. Marcaida, A. Urkaregi, A. Valle. *Bioestatistikako Apunteak*. EHU. Argitaratu gabe).

3 ZORIZKO ALDAGAIAK

3.1 SARRERA

Aurreko gaian, probabilitate-teoria aztertu dugu. Azaldu dugu, baita ere, aski korapilatsuak diren arazoak aztertzeko nola erabil daitezkeen probabilitatearen axiomak, propietateak eta teoremak. Baina ez ditugu aipatu probabilitate-teoriak datu-analisian dituen ondorioak. Probabilitate-teoria tresna matematiko beharrezkoa da datuen azterketa estatistikoa egin eta ulertzeko. Eta probabilitate-teoria eta estatistika aplikatuaren arteko lotura aztertuko dugu, gai honetan, zorizko aldagaien bidez.

Zorizko aldagaien definizio matematikoa aurkeztu baino lehen, definizio intuitiboagoa sartuko dugu. Populazioaren ezaugarri batzuek gure interesa sortzen dute. Ezaugarri horien balioa aldakorra da populazioko aletik alera. Ezaugarri horiei zorizko aldagai deritze: aldagaiak, balioz aldatzen dutelako; zorizkoak, hartzen dituzten balioak zoriaren menpekoak eta, beraz, aurretik jakinezinak direlako. Zorizko aldagaiaren kontzeptua ez da ulertzeko zaila. Intuitiboki, zorizko aldagaia da bere balioa zoriaren menpekoa duen aldagaia. Esate baterako, 18 urte edo gehiago dituzten pertsonen artean, bihotzeko gaixotasun baten azterketan, gaixotasunaren garapenari lotutako lau ezaugarri edo faktore aurkitu ditu ikertzaileak: adina, pisua, egunero erretzen den zigarro kopurua eta bihotz-gaixotasunaren familiako aurrekariak. Zorizko esperimentuari —bihotzeko gaixotasun baten garapenari— lotutako ezaugarri bakoitzaren balioa aldatu egin daiteke populazioko ale bakoitzaren arabera (alea gaixotasuna pairatzen duen pertsona bakoitza da). Beraz, esperimentu horretan, ezaugarri bakoitzari lotutako lau zorizko aldagai defini daitezke.

Zorizko esperimentu bati lotutako ezaugarri bakoitzari **zorizko aldagai** deritzo. Horrela, X zorizko aldagaia, Ω lagin-espazioan definiturik egonik, balio errealak hartzen dituen funtzioa da. Beraz, zorizko esperimentu bati elkartutako zorizko aldagaia esperimentuaren emaitza posibleen unibertsoaren edo osotasunaren parteen gainean definitutako funtzio erreala da. Esperimentuaren unibertsoa zorizko aldagaiaren **eremua** da, eta aldagaiaren balioen multzoa da bere **heina**. Bihotzeko gaixotasunean eragina duten faktoreen azterketan, interesatzen zaizkigun zorizko aldagaiak dira gaixoaren adina, pisua, egunero erretzen duen zigarro kopurua eta familiako aurrekariak.

Zorizko aldagaiak letra maiuskulaz adierazten dira, eta ale bakoitzean hartzen dituzten balioak letra minuskulaz. Zorizko aldagaiak diskretuetan eta jarraituetan sailkatzen dira, hartzen duten balio kopuruaren arabera.

X zorizko aldagaia **diskretua** izango da, hartzen duen balio kopurua finitua edo infinitu zenbakigarria (zenbaki arruntena bezala) baldin bada. Egunero erretzen den zigarro kopurua diskretua da. Bere balio posibleak $\{0, 1, 2, 3, \ldots\}$ dira, hots, kopurua infinitu zenbakigarria da. Familiako aurrekariak bihotz-arazoren bat pairatu duten guraso eta aitona-amonen kopurua zenbatuz aztertzen baldin badira, aldagai hori ere diskretua da, eta bere balio posibleak $\{0, 1, 2, 3, 4, 5, 6\}$ dira, hots, kopurua finitua da.

X zorizko aldagaia **jarraitua** izango da, segmentu edo zuzenerdi baten balio guztiak (adibidez, zenbaki erreal positibo guztiak) har baldin baditzake. Bihotzeko gaixotasunaren azterketan, adina aldagai jarraitua da, eta baita pisua ere. Adibidez,

pertsona baten adinak 18 eta 100 urte bitartean (denbora-tarte jarraituan) edozein balio har dezake, eta pisua 40 eta 150 kg artean egon daiteke.

Ikus ditzagun zorizko aldagai diskretuen eta jarraituen hainbat adibide:

$X \equiv$ «Bizitzako lehen bi urteetan haur batek otitis pairatzen duen aldi kopurua» (diskretua).

$X \equiv$ «Zortzi seme-alaba dituzten familietako seme kopurua» (diskretua).

$X \equiv$ «Ebakuntza batean pazienteak galtzen duen odol kantitatea, litrotan» (jarraitua).

$X \equiv$ «Zirkulazio-istripu bat izan ondoren, lehen anbulantzia heldu arte pasatzen den denbora» (jarraitua).

3.2 ZORIZKO ALDAGAI DISKRETUAK

Zorizko aldagai diskretua balio kopuru finitua edo infinitu zenbakigarria hartzen duena dela esan dugu. Aldagaien tratamenduan, zorizkoa dela soilik esatea ez da nahikoa. Gai izan behar dugu aldagaiak edozein unetan hartuko duen balioa nolabait aurresateko. Baina, aldagaiaren portaera zoriak menperatzen duenez, beharrezkoa izango dugu ziurgabetasunaren azterketa sakona. Azken hori egiteko bi funtzio erabiliko ditugu: probabilitate-legea eta banaketa-funtzioa.

3.2.1 Zorizko aldagai diskretuaren probabilitate-legea

Demagun X zorizko aldagai diskretua dela; **probabilitate-legea** da balio bakoitzari bere gertatze-mailaren neurria ematen dion funtzioa: $P(X = x) = f(x)\ \forall x \in \mathbb{R}$.

X aldagaiaren heina $\{x_1, x_2, \ldots, x_n\}$ baldin bada, $f(x_1) = P(X = x_1), \ldots, f(x_n) = P(X = x_n)$ adieraziko dugu, eta aldagaiak hartzen ez dituen x balio guztientzat $f(x) = P(X = x) = 0$ izango da.

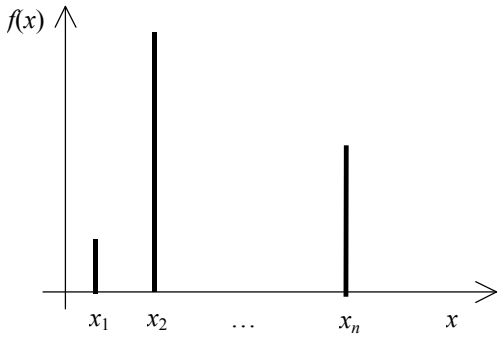
Probabilitate-legeak honako bi propietate hauek betetzen ditu:

1. $f(x) \geq 0 \qquad \forall x \in \mathbb{R}$
2. $\sum_{x \in \Re} f(x) = \sum_{x \in \Re} P(X = x) = 1$

Alegia, zorizko aldagairen baten probabilitate-legea da ez-negatiboa den eta hartzen dituen balio posible guztiak batzerakoan 1 lortzen den edozein funtzio erreal.

Probabilitate-legearen adierazpide grafikoa **probabilitatearen barra-grafikoa** da. Aldagaiaren balioak eskala horizontal baten gainean adieraztean datza hori, eta balio bakoitzean probabilitatearen altuera berdineko barra bat eraikiko da. 3.1 irudian erakusten da barra-grafikoaren itxura orokorra. Populazioaren ordez laginaren azterketa

egiten denean, grafiko horren parekoa da 1.3 atalean azaldu dugun maiztasun erlatiboen barra-grafikoa.



3.1 irudia. Probabilitate-legearen adierazpide grafikoa: probabilitatearen barra-grafikoa.

3.2.2 Zorizko aldagai diskretuaren banaketa-funtzioa

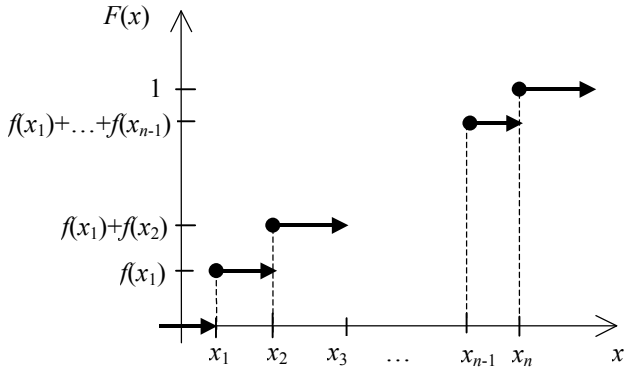
Demagun X zorizko aldagai diskretua dela X-ren **banaketa-funtzioa**, $F(x)$. Zenbaki errealen multzoan definituriko funtzioa da, non $F(x) = P(X \leq x)$ baita $\forall x \in \mathbb{R}$. Hau da, zorizko aldagaiak x-ren balio bera edo x baino balio txikiagoa hartzeko probabilitatea da $F(x)$. Hots,

$$F: \mathbb{R} \longrightarrow [0, 1]$$
$$x_0 \longrightarrow F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} f(x)$$

Zorizko aldagai diskretuaren $F(x)$ banaketa-funtzioak honako propietate hauek betetzen ditu:

1. $F(+\infty) = 1$
2. $F(-\infty) = 0$
3. Ez-beherakorra: $a \leq b \Rightarrow F(a) \leq F(b)$
4. Eskuinetik jarraitua: $\lim_{x \to a^+} F(x) = F(a)$
5. X-ren heina $\{x_1, x_2, \ldots, x_n\}$ baldin bada, $F(x)$-ek $\{x_1, x_2, \ldots, x_n\}$ puntuetan, eta horietan soilik, etenak aurkeztuko ditu; eta jauziak $f(x_1), f(x_2), \ldots, f(x_n)$-koak izango dira, hurrenez hurren. Grafikoki, eskaileraren itxura hartzen du funtzio horrek.

Banaketa-funtzioaren adierazpide grafikoari **grafiko metakor** deritzo, eta 3.2 irudian erakusten da bere itxura orokorra. Populazioaren ordez laginaren azterketa egiten denean, grafiko horren parekoa da 1.3 atalean azaldu dugun maiztasun erlatiboen grafiko metakorra.



3.2 irudia. Zorizko aldagai diskretuaren banaketa-funtzioaren adierazpide grafikoa: grafiko metakorra.

Praktikan, nahikoa da aurretik definituriko bi funtzio horietatik bat ezagutzea, bestea kalkulatzeko edo interesatzen zaizkigun gertaeren probabilitateak kalkulatzeko.

3.1 adibidea

Hiru seme-alabako familietan, X alaba kopurua aztertu dugu. X zorizko aldagaiaren balioak 0, 1, 2 eta 3 dira. 2.3 adibidean ikusi dugunez, zortzi dira familia horien konposaketa posibleak: *sss*, *ass*, *sas*, *ssa*, *aas*, *asa*, *saa*, *aaa*.

$X = 0$ izateko, konposaketak *sss* izan behar du. Beraz, $P(X = 0) = \frac{1}{8}$.

$X = 1$ izateko, hiru konposaketa posible ditugu: *ass*, *sas*, *ssa*. Beraz, $P(X = 1) = \frac{3}{8}$.

$X = 2$ izango da, konposaketa *aas*, *asa* edo *saa* denean. Beraz, $P(X = 2) = \frac{3}{8}$.

Azkenik, $X = 3$ izango da, konposaketa *aaa* denean. Beraz, $P(X = 3) = \frac{1}{8}$.

Beraz, honako hau da hiru seme-alabako familien X alaba kopuruaren probabilitate-legea:

$$f(x) = \begin{cases} \frac{1}{8} & \text{baldin } x = 0 \text{ edo } x = 3 \\ \frac{3}{8} & \text{baldin } x = 1 \text{ edo } x = 2 \\ 0 & \text{beste kasuetan} \end{cases}$$

Kalkula dezagun X-ren banaketa-funtzioa:

$x < 0$: $F(x) = P(X \leq x) = 0$

$0 \le x < 1$: $F(x) = P(X \le x) = P(X = 0) = \frac{1}{8}$

$1 \le x < 2$: $F(x) = P(X \le x) = P(X = 0) + P(X = 1) = \frac{1}{2}$

$2 \le x < 3$: $F(x) = P(X \le x) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{7}{8}$

$x \ge 3$: $F(x) = P(X \le x) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$

Beraz, honako hau da hiru seme-alabako familien X alaba kopuruaren banaketa-funtzioa:

$$F(x) = \begin{cases} 0 \text{ baldin } x < 0 \\ \frac{1}{8} \text{ baldin } 0 \le x < 1 \\ \frac{1}{2} \text{ baldin } 1 \le x < 2 \\ \frac{7}{8} \text{ baldin } 2 \le x < 3 \\ 1 \text{ baldin } x \ge 3 \end{cases}$$

3.2 adibidea

Hegazti batek tximeletak jaten ditu, eta tximeleta horiek landare pozoitsuak jaten dituzte. Hegaztiak tximeleta pozoitu bat jaten duenean, tximeletak harrapatzeari uzten dio, egun horretan behintzat. Demagun tximeleta-populazioaren % 40ak landare pozoitsuak jan dituela; eta izan bedi $X \equiv$ «hegaztiak egun batean jaten duen tximeleta kopurua». (Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991).

X zorizko aldagaiaren balioak 1, 2, 3… izango dira, tximeleta pozoitua jaten duen lehena, bigarrena eta abar izatearen arabera.

Hau da, $X = 1$ izango da, baldin eta jaten duen tximeleta pozoitua bada. Beraz, $P(X = 1) = 0{,}4$.

$X = 2$ izateko, jandako lehenengo tximeleta ez da pozoitua izango, eta bigarrena bai. Beraz, $P(X = 2) = 0{,}6 \cdot 0{,}4$.

$X = 3$ izango da, jandako lehenengo bi tximeletak ez badira pozoituak, baina hirugarrena, bai. Beraz, $P(X = 3) = 0{,}6^2 \cdot 0{,}4$.

Oro har, $X = k$ izango da, jandako lehenengo $(k - 1)$ tximeletak ez badira pozoituak, baina azkena bai. Beraz, $P(X = k) = 0{,}6^{(k-1)} \cdot 0.4$, non $k = 1, 2, 3, \ldots$ baita.

Probabilitate horiek dituen zorizko aldagaiari geometriko deritzo, eta probabilitate-legearen propietateak betetzen direla froga daiteke:

1. $f(k)=P(X=k) = 0{,}6^{(k-1)} \cdot 0{,}4 \geq 0$ da.

2. $$\sum_{k=1}^{\infty} f(k) = \sum_{k=1}^{\infty} P(X=k) = \sum_{k=1}^{\infty} 0{,}6^{k-1} \cdot 0{,}4 = 0{,}4 \cdot \sum_{k=1}^{\infty} 0{,}6^{k-1} = 0{,}4 \cdot \frac{1}{1-0{,}6} = 1$$

Oro har, X zorizko aldagai diskretua p parametroko **banaketa geometrikoa**ri darraio, baldin eta bere probabilitate-legeak honako adierazpen hau badu:

$P(X=k) = (1-p)^{k-1} \cdot p$ non $k = 1, 2, 3, \ldots$ baita.

3.3 ZORIZKO ALDAGAI JARRAITUAK

Zorizko aldagai jarraituen azterketan, diskretuaren kontzeptu paraleloak ezarri baino lehen, aldagai jarraituak balio zehatz bat hartzeko probabilitatea nulua dela ikusiko dugu. Hori egiteko, demagun olagarro heldu baten gorputzaren diametroa neurtzen duen zorizko aldagaia eta zentzuzko limiteak 10 eta 30 zentimetroen artean daudela. Alegia, [10, 30] tartean balioak hartzen dituen zorizko aldagai jarraitua da. Baina, aukerarik egin baino lehen, olagarro baten gorputzaren diametroa zehazki 12,981321069217031 zentimetrokoa izateko probabilitatea zein den galdetzen badugu, erantzuna 0 da. Izan ere, oszilaziorik gabe, justu-justu diametro hori duen olagarro bat aurkitzea ia-ia ezinezkoa da. Horrela, definizioa bete dadin, funtsezkoa da zorizko aldagaiak balio zehatz bat hartzeko probabilitate nulua daukala suposatzea. Dena den, propietate hori matematikoki froga daiteke. Zorizko aldagai diskretuetan hori egia ez denez, propietate horrek zorizko aldagai diskretuak eta jarraituak bereiztea baimentzen digu.

Zorizko aldagai jarraituetan ere, diskretuetan egin dugun era berean, bi funtzio erabiliko ditugu aldagaiaren ziurgabetasunaren azterketa egiteko: banaketa-funtzioa eta dentsitate-funtzioa.

3.3.1 Zorizko aldagai jarraituaren banaketa-funtzioa

X zorizko aldagaia jarraitua baldin bada, X-ren **banaketa-funtzioa**, $F(x)$, zenbaki errealen multzoan definituriko funtzioa da, non $F(x) = P(X \leq x)$ baita $\forall x \in \mathbb{R}$. Hau da, zorizko aldagaiak x-ren balio bera edo x baino txikiagoa hartzeko probabilitatea da $F(x)$. Hots,

$$F: \mathbb{R} \longrightarrow [0, 1]$$

$$x_0 \longrightarrow F(x_0) = P(X \leq x_0)$$

Zorizko aldagai jarraituaren $F(x)$ banaketa-funtzioak honako propietate hauek betetzen ditu:

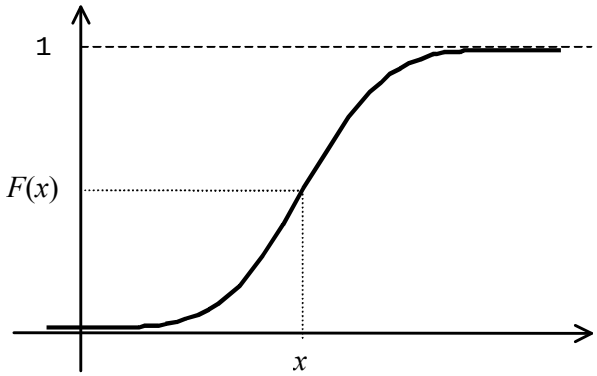
1. $F(+\infty) = 1$

2. $F(-\infty) = 0$
3. Ez-beherakorra: $a \leq b \Rightarrow F(a) \leq F(b)$
4. Jarraitua: $\forall\, x \in \mathbb{R},\ \lim_{x \to a} F(x) = F(a)$
5. $\forall\, x \in \mathbb{R} \quad P(X = x) = 0$ denez, honako hauek betetzen dira:

   $P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = F(b) - F(a)$

   $P(X \leq a) = P(X < a) = F(a)$

Banaketa-funtzioaren adierazpide grafikoari **grafiko metakorra** deritzo, eta ojiba forma hartzen du. 3.3 irudian agertzen da bere itxura. Populazioaren ordez laginaren azterketa egiten denean, 1.3 atalean azaldu dugun maiztasun erlatiboen ojiba edo maiztasun poligono metakorra da grafiko horren parekoa.



3.3 irudia. Zorizko aldagai jarraituaren banaketa-funtzioaren adierazpide grafikoa: grafiko metakorra.

3.3.2 Zorizko aldagai jarraituaren dentsitate-funtzioa

Kontzeptu hau zorizko aldagai diskretuen probabilitate-legearen paraleloa da, baina konplexuagoa. Ezinezkoa da zorizko aldagai jarraitu batek hartzen dituen balio guztiak zerrendatzea, balio kopuru infinitu ez-zenbakigarria hartzen baitu. Halaber, zorizko aldagai jarraituak x balio zehatz bat hartzeko probabilitatea zero denez, ez da hori ere dentsitate-funtzioaren bidez kalkulatu nahi genukeen emaitza. Hala ere, funtsezkoa da probabilitateak kalkulatzea baimentzen digun adierazpena edukitzea. X zorizko aldagaia jarraitua baldin bada, X tarte batean edo bi balioren artean egoteko probabilitatea da kalkulatu nahi duguna. Esate baterako, leuzemiaren tratamenduan, medikazio berri baten bidez, paziente baten biziraupen-denbora 6 hilabete baino gehiago izateko probabilitatea kalkulatu nahi badugu —X da biziraupen-denbora—, $P(X > 6)$ da gure intereseko balioa.

$F(x + h) - F(x)$ kendurak h luzerako $(x, x + h)$ tartean dagoen probabilitate-masa adierazten du, hots, $P(x < X < x + h)$. Hori dela eta, tartearen batez besteko probabilitate-

dentsitatetzat har daiteke $\frac{F(x+h)-F(x)}{h}$ zatidura. $F(x)$ banaketa-funtzioa deribagarria dela jotzen baldin badugu, $\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}=F'(x)$ izango da. Ideia hori erabiliko dugu zorizko aldagai jarraituaren dentsitate-funtzioa definitzeko.

Zorizko aldagai jarraituaren **dentsitate-funtzioa** x erreal guztietarako x-rekiko banaketa-funtzioaren deribatu gisa definiturik dagoen funtzioa da; hau da:

$$f(x) = F'(x) = \frac{d}{dx}F(x) = \lim_{h\to 0}\frac{F(x+h)-F(x)}{h}$$

Dentsitate-funtzioak honako propietate hauek betetzen ditu:

1. $f(x) \geq 0 \qquad \forall x \in \mathbb{R}$

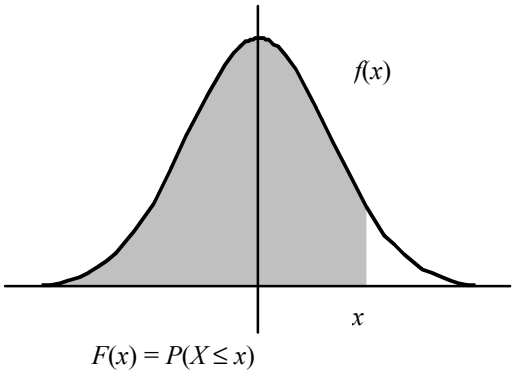
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

Azken propietate horrek adierazten digu f-ren grafikoaren eta abzisa-ardatzaren bitartean dagoen azalera 1 dela.

Populazioaren azterketaren ordez laginaren azterketa egiten denean, 1.3 atalean azaldu dugun maiztasun erlatiboen poligonoa da dentsitate-funtzioaren adierazpide grafikoaren parekoa.

<u>Oharra</u>: banaketa- eta dentsitate-funtzioek elkar determinatzen dute, zeren $F(x)$ ezaguturik, $f(x) = F'(x)$ baita; eta, alderantziz, $f(x)$ ezaguturik, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$ baita.

Banaketa-funtzioak x-ren ezkerrerantz f-ren grafikoaren eta abzisa-ardatzaren bitartean dagoen azalera adierazten du. 3.4 irudian erakusten da banaketa- eta dentsitate-funtzioen arteko erlazioa.



3.4 irudia. Banaketa- eta dentsitate-funtzioen arteko erlazioa.

3.3 adibidea

Demagun X zorizko aldagaiak bakterioa zatitu arte iragandako aldia (segundotan) neurtzen duela, eta bere dentsitate-funtzioa honako hau dela:

$$f(x) = \begin{cases} 0 & \text{baldin } x < 0 \\ 3e^{-3x} & \text{baldin } x \geq 0 \end{cases}$$

Kalkula dezagun bere banaketa-funtzioa:

$$x < 0 \text{ bada, } F(x) = \int_{-\infty}^{x} f(t)dt = 0$$

$$x \geq 0 \text{ bada, } F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} 3e^{-3t}dt = -e^{-3t}\Big]_{0}^{x} = 1 - e^{-3x}$$

Beraz, $F(x) = \begin{cases} 0 & \text{baldin } x < 0 \\ 1 - e^{-3x} & \text{baldin } x \geq 0 \end{cases}$

Hauxe da bakterioa 2.7 eta 3 segundo artean zatitzeko probabilitatea:

$P(2.7 \leq X \leq 3) = F(3) - F(2,7) = 1 - e^{-3\cdot 3} - (1 - e^{-3\cdot 2,7}) = 3,04\cdot 10^{-4} - 1,23\cdot 10^{-4} = 1,8\cdot 10^{-4}$

Zenbait egoeratan, «gertaera bat jazo arte iragandako aldia» zorizko aldagaitzat jotzea interesatzen zaigu. Egoera orokor horren kasu berezi bat izango litzateke aurreko adibidea. Zorizko aldagai hori, baldintza batzuen pean, banaketa esponentzialaren bidez deskriba daiteke egokiro. Aurreko adibidean, bakterioa zatitu arte iragandako aldia 3 parametroko banaketa esponentzialari darraiola esaten da.

Oro har, X zorizko aldagai jarraitua $\lambda \geq 0$ parametroko **banaketa esponentziala**ri darraio, baldin eta bere dentsitate-funtzioak honako adierazpen hau baldin badu:

$$f(x) = \begin{cases} 0 & \text{baldin } x < 0 \\ \lambda e^{-\lambda x} & \text{baldin } x \geq 0 \end{cases}$$

Hauxe da bere banaketa-funtzioa: $F(x) = \begin{cases} 0 & \text{baldin } x < 0 \\ 1 - e^{-\lambda x} & \text{baldin } x \geq 0 \end{cases}$

3.4 ZORIZKO ALDAGAIAREN ITXAROPENA ETA BARIANTZA

Zorizko aldagaiak hartzen duen balio kopurua handia edo infinitua denean, probabilitate- edo dentsitate-funtzioak ez dira laburpen-neurri egokiak izango. Izatez, lehen gaian lagin-balioak laburbiltzeko izan dugun arazo bera izango dugu kasu honetan ere. Zorizko aldagaientzat ere, laginean definitu diren era antzekoan definituko ditugu joera zentraleko neurriak eta sakabanatze-neurriak. Populazioaren ezaugarri bat adierazten duen zorizko aldagai baten propietateak laburki deskribatzen dituen neurriari **parametro** deritzo. Parametroak letra grekoen bidez adierazten dira.

Ezaugarri batentzat sarritan dugun parametroa populazioaren **batezbestekoa** edo **itxaropen matematikoa** da. Parametro hori μ letra grekoaren bidez adierazten da: $\mu \equiv E(X)$. Itxarotako balioak zorizko aldagai baten «batez besteko balioa» adierazten du nolabait. Adibidez, bihotzeko gaixotasunaren azterketan, ikertzaileak populazioko aleek egunero erretzen duten batez besteko zigarro kopurua kalkulatu nahi izango du. Parametro horren balio zehatza kalkulatzea ez da posible izango, ez bada populazio osoa aztertzen.

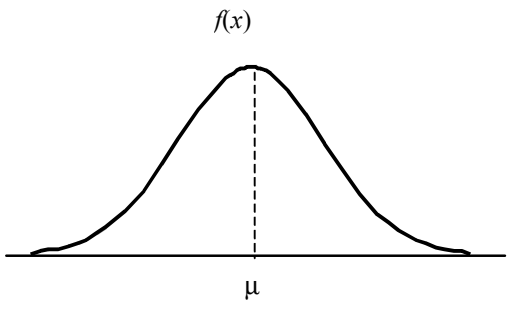
Demagun X zorizko aldagaia diskretua dela, eta $f(x)$ dela bere probabilitate-legea. Honako era honetara kalkulatzen da X-ren batezbestekoa edo itxaropen matematikoa:

$$\mu \equiv E(X) = \sum_x x \cdot f(x)$$

non x aldagaiak hartzen dituen balioak baitira.

Demagun X zorizko aldagaia jarraitua dela, eta $f(x)$ dela bere dentsitate-funtzioa. Honako modu honetan kalkulatzen da X-ren batezbestekoa edo itxaropen matematikoa:

$$\mu \equiv E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$



3.5 irudia. μ= E(X) dentsitate-funtzioaren oreka-puntua.

Geometrikoki irudika daiteke aldagai jarraituaren $E(X)$ batezbestekoa. 3.5 irudian ikus daitekeenez, X-ren dentsitate-funtzioaren grafikoan, $E(X)$-k oreka-puntua adierazten du. Probabilitate-masa da bere inguruan kontzentratzen den balioa adierazten duen zentralizazio-neurria.

Demagun X eta Y zorizko aldagaiak diskretuak zein jarraituak direla. Orduan, itxaropenak honako propietate hauek betetzen ditu:

1. $E(k) = k \quad \forall k \in \mathbb{R}$ (Balio bakar bat hartzen duen zorizko aldagai diskretuaren itxaropena balio bera da)
2. $E(X + k) = E(X) + k \quad \forall k \in \mathbb{R}$
3. $E(aX) = a \cdot E(X) \quad \forall a \in \mathbb{R}$
4. $E(X + Y) = E(X) + E(Y)$
5. X eta Y zorizko aldagaiak askeak badira, $E(XY) = E(X) \cdot E(Y)$

<u>Oharra</u>. Zorizko aldagai askeak honela definitzen dira:

- X eta Y bi aldagai diskretuak askeak dira, baldin $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ bada, x eta y, X-k eta Y-k, hurrenez hurren, hartzen dituzten edozein balio izanik.
- X eta Y bi aldagai jarraituak askeak dira, baldin $P\{(X \leq a) \cap (Y \leq b)\} = P(X \leq a) \cdot P(Y \leq b) \; \forall a, b \in \mathbb{R}$ bada.

<u>Oharra</u>. Itxaropen matematikoa aldagaiaren unitate berdinetan adierazten da. Hots, bakterioa zatitu arte iragandako aldia segundotan adierazten bada, orduan, segundotan adieraziko da $E(X)$ batez besteko aldia.

Zorizko aldagaiaren itxaropen matematikoak edo batezbestekoak aldagaia fluktuatzen den inguruaren balioa adierazten du, baina ez du fluktuazio horren tamainari buruzko informaziorik ematen. Zorizko aldagaiaren fluktuazioa edo sakabanatzea adierazteko, lagin-bariantzarekin bat datorren parametroa definituko dugu. Parametro horrek aldagaiaren batezbestekoarekiko sakabanatzea neurtzen du, bariantza deritzo eta

Var(X) ≡ σ^2 adierazten da. Demagun X zorizko aldagaia diskretua zein jarraitua dela, non $\mu \equiv E(X)$ baita. Honela definituko da ***X*-ren bariantza**: Var(X) ≡ $\sigma^2 = E[(X - \mu)^2]$ edo karratua garatuz, Var(X) = $E(X^2) - \mu^2$. ***X*-ren desbideratze estandarra** bariantzaren erro karratu positibotzat definitzen da: DE(X) ≡ $\sigma = \sqrt{Var\ X}$

Demagun X eta Y zorizko aldagaiak diskretuak zein jarraituak direla. Orduan, bariantzak honako propietate hauek beteko ditu:

1. Bariantza beti positiboa edo nulua da.
2. Var(X) = 0 da, baldin eta soilik baldin X zorizko aldagaia konstantea bada.
3. Var(aX) = a^2Var(X) $\forall a \in \mathbb{R}$
4. Var($X + a$) = Var(X) $\forall a \in \mathbb{R}$
5. X eta Y zorizko aldagaiak askeak baldin badira, honako bi hauek betetzen dira:
   Var($X + Y$) = Var(X) + Var(Y)
   Var($X - Y$) = Var(X) + Var(Y)

<u>Oharra</u>. Bariantzaren unitateak aldagaiaren unitate karratuak dira, eta desbideratze estandarrarenak aldagaiaren berdinak. Hots, bakterioa zatitu arte iragandako aldia segundotan adierazten bada, orduan, bariantza segundo^2-tan adierazten da, eta desbideratze estandarra segundotan.

3.4 adibidea

3.1 adibidean oinarriturik, kalkula ditzagun hiru seme-alabako familien X alaba kopurua adierazten duen aldagai diskretuaren itxaropena eta bariantza.

$$E(X) = \sum_x x\, P(X = x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1{,}5$$

Beraz, hiru seme-alabako familia batean itxarotako alaba kopurua 1,5 da.

$$E(X^2) = \sum_x x^2\, P(X = x) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = \frac{24}{8} = 3$$

Var(X) = $E(X^2) - \mu^2 = 3 - 1{,}5^2 = 0{,}75$

X-ren desbideratze estandarra $\sigma = \sqrt{0{,}75} = 0{,}866$ da.

3.5 adibidea

3.3 adibidean oinarriturik, kalkula ditzagun X bakterioa zatitu arte iragandako aldiaren itxaropena eta bariantza.

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_0^{\infty} x 3e^{-3x} dx = -xe^{-3x}\Big]_0^{\infty} + \int_0^{\infty} e^{-3x} dx = -\frac{1}{3}e^{-3x}\Big]_0^{\infty} = \frac{1}{3}$$

Zatika: $x = u \rightarrow dx = du$

$$3\,e^{-3x}dx = dv \rightarrow v = \int 3\,e^{-3x} = -e^{-3x}$$

Bakterioa zatitu arte iragandako denbora 1/3 segundo izatea espero da.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{\infty} x^2 3e^{-3x} = -x^2\,e^{-3x}\Big]_0^{\infty} + 2\int_0^{\infty} xe^{-3x}dx$$

$$= 2\int_0^{\infty} x\,e^{-3x}dx = \frac{2}{3^2}$$

Zatika: $x^2 = u \rightarrow 2xdx = du$

$$3e^{-3x}dx = dv \rightarrow v = -e^{-3x}$$

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = \frac{2}{3^2} - \frac{1}{3^2} = \frac{1}{3^2} = \frac{1}{9}$$

X-ren desbideratze estandarra $\sigma = \sqrt{\frac{1}{3^2}} = \frac{1}{3}$ da.

Oro har, λ parametroko banaketa esponentzialaren itxaropena eta desbideratze estandarra $\frac{1}{\lambda}$ dira.

3.5 BANAKETA BINOMIALA

Probabilitate-kalkuluaren helburuetariko bat da zientzia esperimentaletan —beren artean Biologian eta Medikuntzan— agertzen diren zorizko fenomeno ugariren eredutzat har daitezkeen banaketak aurkitzea. Horietariko bat da banaketa binomiala deritzona. Azter dezagun berriro hiru seme-alabako familien X alaba kopurua, baina beste era batera. Seme-alaba bat izatea zorizko esperimentutzat hartuko dugu. Horri elkartutako gertaerak bi besterik ez dira: alaba ala semea. Bata zein bestea izateko probabilitateak berdinak direla suposatuko dugu, hau da, $P(a) = P(s) = \frac{1}{2}$.

Hiru seme-alaba izatea zorizko esperimentua hiru aldiz gauzatzean datza, eta, gainera, hirurak askeak dira. Askeak direla kontuan hartuz, honako era honetara kalkula ditzakegu probabilitateak:

$$P(X=0) = P(sss) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$$P(X=1) = P(ass \cup sas \cup ssa) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = 3\cdot\frac{1}{2}\cdot\left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

$$P(X=2) = P(aas \cup asa \cup saa) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = 3\cdot\left(\frac{1}{2}\right)^2\cdot\frac{1}{2} = \frac{3}{8}$$

$$P(X=3) = P(aaa) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

Beste era batera azaldurik, $P(X = r) = \binom{3}{r}\cdot\left(\frac{1}{2}\right)^r\cdot\left(\frac{1}{2}\right)^{3-r}$, $r = 0, 1, 2, 3$ izanik. Hau da, hain zuzen, $\mathrm{Bin}(3, \frac{1}{2})$ banaketa binomialaren probabilitate-legea.

Oro har, izan bedi zorizko esperimentua, non hari elkarturiko gertaerak bi besterik ez baitira: A gertaera (arrakasta) eta $\overline{A}$ kontrakoa (porrota). Jo dezagun $P(A) = p$ eta $P(\overline{A}) = 1 - p = q$ direla ($p + q = 1$). Horrelako esperimentuari **Bernouilli-ren proba** deritzogu.

Demagun n proba egiten ditugula, honako baldintza hauen pean:

1. Proba bakoitzean, elkarren ukitzaileak diren emaitza posibleetariko bat lortzen da. Emaitza posibleetako bati arrakasta eta besteari porrot deritzegu.
2. Arrakastaren probabilitateak (p), probaz proba, konstante dirau. Porrotaren $1 - p$ probabilitateari q deritzogu.
3. Probak askeak dira, hots, edozein probaren emaitzak ez dauka eraginik beste edozein probaren emaitzan.

$(x_1, x_2, \ldots, x_n)$ n-kotearen x_i osagaiei A eta $\overline{A}$ balioak era posible guztietara egokituz lortzen dira lagin-espazioaren elementuak:

$$\Omega = \{(x_1, x_2, \ldots, x_n) \,/\, x_i \in \{A, \overline{A}\} \ \forall\, i = 1, 2, \ldots, n\}$$

Posible izango dira A eta $\overline{A}$ elementuen n ordenako errepikatuzko aldakuntzak diren $VR_2^n = 2^n$ gertaera.

Lagin-espazio horren gainean definitzen dugu $X \equiv$ «Kontuan harturiko gertaeraren arrakasta kopurua» **zorizko aldagai binomiala**.

$P(X = r)$ probabilitate-legea aurkitzeko, kontuan hartuko ditugu ordenazio konkretu bat, r arrakasta eta $(n - r)$ porroteko gertaera bat: $\overbrace{AA,\ldots,AA}^{r}\ \overbrace{\overline{A}\,\overline{A},\ldots,\overline{A}}^{n-r}$

Honako hau da ordenazio horren probabilitatea: $\overbrace{p \cdot p \cdot \ldots \cdot p}^{r}\ \overbrace{q \cdot q \cdot \ldots \cdot q}^{n-r} = p^r q^{n-r}$

Baina r arrakasta eta $(n - r)$ porroten ordenamendu posibleak A eta $\overline{A}$ bi letren n ordenako konbinazioak dira, non A letra kopurua r eta $\overline{A}$ letrarena $(n - r)$ baitira; hots, $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

Orduan, $P(X = r) = \frac{n!}{r!(n-r!)} p^r q^{n-r} = \binom{n}{r} p^r q^{n-r}$

$n = 1, 2, 3, \ldots$
$r = 0, 1, 2, \ldots, n$
$0 < p < 1$

Hori da banaketa binomiala izenekoaren probabilitate-legea. Binomiala izena, hain zuzen, probabilitate hauek honako binomio honen garapeneko gaiak izatetik dator:

$$1^n = (p+q)^n = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} = \sum_{r=0}^{n} P(X = r)$$

3.6 irudian erakusten da banaketa binomialaren probabilitatearen barra-grafikoaren itxura, parametroen balio batzuetarako.

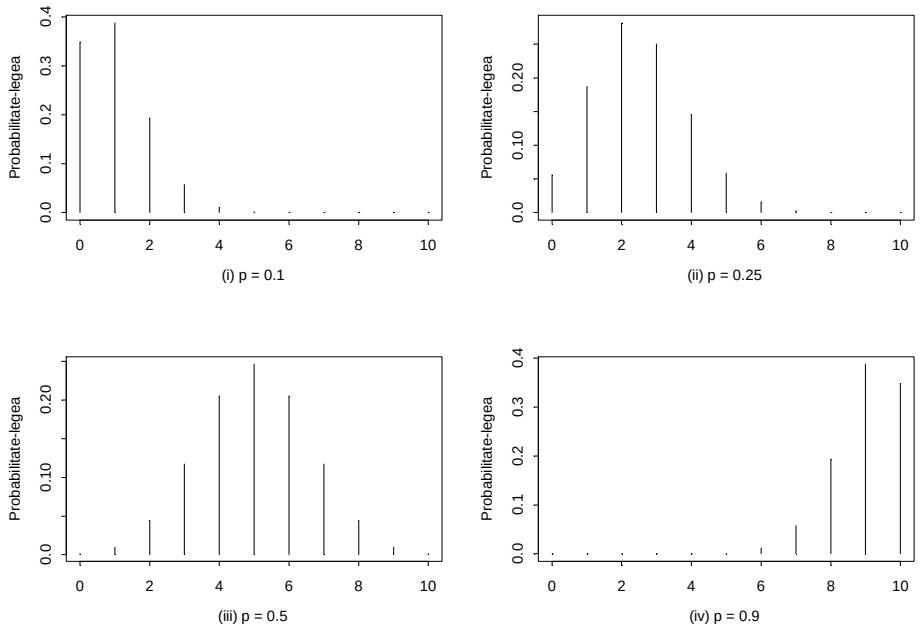
Banaketa-funtzioa honako hau izango da: $P(X \le x) = F(x) = \sum_{r \le x} \binom{n}{r} p^r q^{n-r}$

Itxaropen matematikoa $E(X) = np$ da.

Hauek dira bariantza eta desbideratze estandarra:

$\mathrm{Var}(X) \equiv \sigma^2 = npq \Rightarrow \mathrm{DE}(X) = \sigma = \sqrt{npq}$

Ikusten dugunez, banaketa binomiala n proba kopuruak eta proba bakoitzaren p arrakasta-probabilitateak definiturik dago. Kantitate horiek banaketa binomialaren parametroak dira, eta n eta p parametroetako X zorizko aldagai binomiala X: $\mathrm{Bin}(n, p)$ adierazten da.



3.6 irudia. Banaketa binomialaren probabilitate-legearen adierazpide grafikoa, n = 10 eta p = 0,1 0,25 0,5 0,9 parametroen balioetarako.

A eranskineko 1. taulan agertzen dira parametro ezberdinetako banaketa binomialen probabilitate-legeak, n = 2tik 10era eta p = 0,1etik 0,5era.

Oharra: X: Bin(n, p) zorizko aldagaiak arrakasta kopurua adierazten badu, $Y = n - X$ zorizko aldagaiak porrot kopurua adieraziko du, eta bere banaketa Y: Bin(n, q) izango da.

3.6 adibidea

Tuberkulosia izateko joera duten hamar pertsona gaixotasunaren eramaile batekin erlazionaturik daude. Eramaileak gaixotasuna beste edozeini kutsatzeko probabilitatea 0,1 da. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Kalkula ezazu bi pertsona, edo bi pertsona baino gutxiago kutsatzeko probabilitatea.

b) Zenbat pertsona kutsatzea espero dugu?

Emaitza:

Honako zorizko aldagai binomial hau definituko dugu:

$X \equiv$ «Kutsatzen den pertsona kopurua»: Bin(10, 0,1)

Beraz, honako hau da eskatutako probabilitatea:

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}3487 + 0{,}3874 + 0{,}1937 =$
$= 0{,}9298$

$E(X) = np = 10 \cdot 0{,}1 = 1$; hots, pertsona bakarra kutsatzea espero dugu.

3.6 POISSON-EN BANAKETA

*Poisson*en aldagai diskretuen familia definituko dugu jarraian. Izen hori *Simeón Denis Poisson* (1781-1840) Frantziako matematikariari zor zaio. *Poisson*en zorizko aldagaiak *Poisson*en prozesuekin loturik daude. Prozesu horiek, denbora, luzera edo espazio tarte jarraitu batean, gertaera multzo diskretu baten behaketan dautza. Esate baterako, odol tanta batean dagoen globulu gorrien kopurua. Interesatzen zaigun gertaera globulu gorrien presentzia da eta, tarte jarraitua, odol tanta. Beste adibide bat litzateke hiru hilabetean energia nuklearreko planta batek gas erradiaktiboak igortzen dituen aldi kopurua. Kasu horretan, gas erradiaktiboak igortzea da gertaera eta, denbora-tarte jarraitua, hiru hilabeteak. Osasunarekin edo biologiarekin zerikusia daukaten egoera askori deskribapen egokia eskaintzen dio *Poisson*en ereduak, eta banaketa horri darraizkion beste adibide asko aurki daitezke: mm^3 odoleko agente patogenoen kopurua; substantzia erradiaktiboak igorritako partikula kopurua denboraldi batean; ur-bolumenean dagoen bakterio kopurua; urtebetean sukar tifoidez hildakoen kopurua; urteroko buruhiltze-tasa; eta abar. *Poisson*en banaketa da, beharbada, binomialaren ostean maizen erabiltzen den banaketa diskretua.

*Poisson*en eredua gertaeren behaketa neurtzeko egokia izan dadin, honako hiru baldintza hauek bete behar dira:

1. Demagun Δs s unitatek osaturiko tartearen edozein azpitarte txiki dela. Orduan:
   - Gertaera bat behatzeko probabilitatea da Δs azpitartearen luzeraren proportzio zuzena. Hau da, P(gertaera bat) $\approx \lambda \cdot \Delta s$, λ konstantea izanik.
   - Δs azpitartean gertaerarik ez behatzeko probabilitatea, gutxi gorabehera, $1 - \lambda \cdot \Delta s$ da.
   - Azpitarte horretan gertaera bat baino gehiago behatzeko probabilitatea 0 da.
2. s unitatek osaturiko tarte osoan zehar, unitateko gertaera kopurua berdina da.

3. Azpitarte batean gertaera behin gertatzeak ez du eraginik hurrengo azpitarte batean berriro gertatzeko probabilitatean.

Aurreko baldintzetan oinarrituta, honako probabilitate-lege honek ematen du s unitatek osaturiko tartean k gertaera izateko probabilitatea:

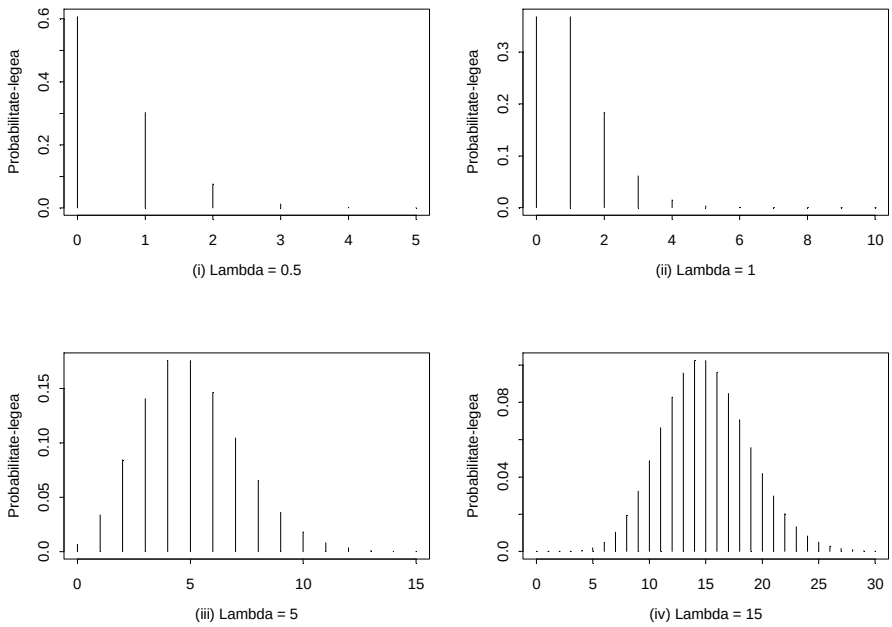
$$P(X = k) = \frac{e^{-\lambda s} \cdot (\lambda s)^k}{k!} \quad , \qquad k = 0, 1, 2, \ldots\ldots$$

Probabilitate-lege horri darraion zorizko aldagaiari λs parametroko X ***Poisson*en zorizko aldagai** deritzo, eta bere adierazpena X: $\mathcal{P}(\lambda s)$ da. A eranskineko 2. taulan agertzen dira parametro ezberdineko *Poisson*en probabilitate-legeak.

λs parametroko *Poisson*en zorizko aldagaiaren itxaropena $\mu = \lambda s$ da. Hori dela eta, *Poisson*en zorizko aldagaiaren probabilitate-legea beste era honetara adierazten da sarritan:

$$P(X = k) = \frac{e^{-\mu} \cdot \mu^k}{k!} \quad , \qquad k = 0, 1, 2, \ldots\ldots$$

λ parametroak unitateko itxarotako gertaera kopurua adierazten du, eta $\mu = \lambda s$ parametroak s unitatek osaturiko tartean zehar itxarotako gertaera kopurua.



3.7 irudia. *Poisson*en banaketaren adierazpide grafikoa $\lambda = 0.5, 1, 5, 15$ parametroen balioetarako.

λs parametroko *Poisson*en zorizko aldagaiaren bariantza ere λs da. *Poisson*en zorizko aldagaiaren ezaugarri bat da itxaropena eta bariantza berdinak edukitzea. 3.7 irudian erakusten da *Poisson*en banaketaren probabilitatearen barra-grafikoaren itxura, parametroen balio batzuetarako.

3.7 adibidea

Kontsidera dezagun sukar tifoidez hildakoen kopuruaren banaketa epe luze batean, adibidez, urtebetean. Onartuko dugu edozein egunetan sukar tifoidez heriotza bat gertatzeko probabilitatea oso txikia dela, eta edozein bi egunetan deklaratutako kasuen kopuruak zorizko aldagai askeak direla. Demagun sukar tifoideari dagokion hildakoen kopurua urtebetean λ = 4,6 parametrotako *Poisson*en zorizko aldagaia dela. Zein da, sei hilabetean zehar, 6 edo 6 baino hildako gehiago egoteko probabilitatea? Eta bi urtean zehar? (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1994).

Izan bitez $X \equiv$ «sei hilabetean zehar hildakoen kopurua». $\lambda = 4{,}6$ eta $t = 0{,}5$ direnez, $\mu = \lambda t = 2{,}3$ da. Orduan,

$$P(X=0) = e^{-2.3} = 0.100$$

$$P(X=1) = \frac{e^{-2,3} \cdot (2,3)^1}{1!} = 0{,}231$$

$$P(X=2) = \frac{e^{-2,3} \cdot (2,3)^2}{2!} = 0{,}265$$

$$P(X=3) = \frac{e^{-2,3} \cdot (2,3)^3}{3!} = 0{,}203$$

$$P(X=4) = \frac{e^{-2,3} \cdot (2,3)^4}{4!} = 0{,}117$$

$$P(X=5) = \frac{e^{-2,3} \cdot (2,3)^5}{5} = 0{,}054$$

$$P(X \geq 6) = 1 - (0{,}100 + 0{,}231 + 0{,}265 + 0{,}203 + 0{,}117 + 0{,}054) = 0{,}030$$

Izan bitez $Y \equiv$ «bi urtean zehar hildakoen kopurua». $\lambda = 4{,}6$ eta $t = 2$ direnez, $\mu = \lambda t = 9{,}2$ da. Orduan,

$$P(Y=0) = e^{-9,2} = 0{,}00010$$

$$P(Y=1)=\frac{e^{-9,2}\cdot(9,2)^1}{1!}=0,00093$$

$$P(Y=2)=\frac{e^{-9,2}\cdot(9,2)^2}{2!}=0,00428$$

$$P(Y=3)=\frac{e^{-9,2}\cdot(9,2)^3}{3!}=0,01311$$

$$P(Y=4)=\frac{e^{-9,2}\cdot(9,2)^4}{4!}=0,03016$$

$$P(Y=5)=\frac{e^{-9,2}\cdot(9,2)^5}{5}=0,05549$$

$P(Y \geq 6) = 1 - (0,0001 + 0,00093 + 0,00428 + 0,01311 + 0,03016 + 0,0549) = 1 - 0,10348 = 0,8965$

Adibide hori denboran zehar izaten den gertaera arraro bati buruzkoa da. Baina, gertaera arraroak ez dira denboran zehar soilik gertatzen; azalera batean ere gerta daitezke.

3.8 adibidea

Demagun 100 cm^2-ko agar-plaka daukagula, eta oso txikia dela bakterio-kolonia bat a puntu batean (edo zehazkiago a-ren inguruko azalera txiki batean) aurkitzeko probabilitatea, eta askea dela a_1 eta a_2 edozein bi puntutan bakterio-koloniak aurkitzea. Agar-plaka osoaren gainean, bakterio kolonien kopurua *Poisson*en banaketari darraio. Onartzen da plakaren edozein puntutan Δa neurriko azalera batean bakterio-kolonia bat aurkitzeko probabilitatea $\lambda \cdot \Delta a$ dela, eta plakaren bi puntutan aurkitutako bakterio-kolonien kopuruak zorizko aldagai askeak direla. Orduan, *Poisson*en probabilitate-legeak emango digu a neurriko azalera batean k bakterio-kolonia topatzeko probabilitatea, $\mu = \lambda a$ izanik. cm^2-ko bakterio-kolonia bat aurkitzeko probabilitatea $\lambda = 0,02$ dela jakinik, zein da 100 cm^2-ko agar-plaka batean bost edo bost baino bakterio-kolonia gehiago aurkitzeko probabilitatea? (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1994).

a = 100 cm^2 eta $\lambda = 0,02$ direnez, $\mu = \lambda a = (0,02) \cdot 100 = 2$ daukagu. Beraz, $\mathcal{P}(2)$ izango da $X \equiv$ «100 cm^2-tan aurkitutako bakterio kolonien kopurua» aldagaiaren banaketa:

$$P(X=0)=e^{-2}=0,135$$

$$P(X=1)=\frac{e^{-2}\cdot(2)^1}{1!}=0{,}271$$

$$P(X=2)=\frac{\mathrm{e}^{-2}\cdot(2)^2}{2!}=0{,}271$$

$$P(\mathrm{X}=3)=\frac{\mathrm{e}^{-2}\cdot(2)^3}{3!}=0{,}180$$

$$\mathrm{P}(\mathrm{X}=4)=\frac{\mathrm{e}^{-2}\cdot(2)^4}{4!}=0{,}090$$

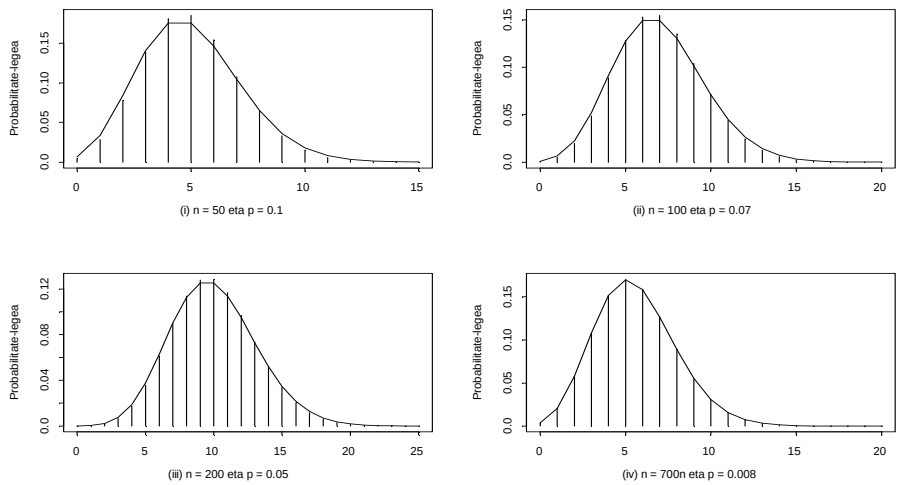
$$\mathrm{P}(\mathrm{X}\geq 5)=1-(0{,}135+0{,}271+0{,}271+0{,}180+0{,}090)=0{,}053$$

Argiro, zenbat eta λ handiagoa izan, orduan eta probabilitate handiagoa izango dugu bakterio-kolonien kopuru edo sukar tifoidez hildakoen kopuru edo beste edozeren kopuru zehatz bat lortzeko.

Bai zorizko aldagai binomialak eta bai *Poisson*enak gertaera kopuruaren gertatze-maila neurtzen dute. Bien arteko ezberdintasun garrantzitsu bat saio eta gertaera kopuruei dagokie. Banaketa binomial batean, *n* saio kopurua finitua da, eta gertaera kopurua ezin da *n* baino handiagoa izan. *Poisson*en banaketaren kasuan, saio kopurua infinitua da, eta gertaera kopurua mugagabea izan daiteke; baina *k* oso handia bada, *k* gertaera agertzeko probabilitatea oso txikia izango da. Hala ere, gertaera kopuruaren gertatze-maila neurtzerakoan *n* handia eta *p* txikia direnean, banaketa binomialaren hurbilketan datza *Poisson*en banaketaren beste erabilera garrantzitsua. Kontsidera dezagun *n* handia eta *p* txikia dituen banaketa binomiala. *np* eta *npq* banaketaren batezbestekoa eta bariantza dira. Ohar zaitez *p* txikia bada $q \approx 1$ dela, eta horrela $npq \approx np$ dela. Beraz, kasu horretan, ia berdinak dira banaketa binomialaren batezbestekoa eta bariantza, *Poisson*en zorizko aldagaian gertatzen den moduan.

Horretaz gainera, $\lim\limits_{\substack{n\to\infty \\ p\to 0}}\binom{n}{k}p^k q^{n-k}=\frac{e^{-\mu}\mu^k}{k!}$ dela froga daiteke. Hau da, *n* handia eta *p* txikia dituen banaketa binomiala zehazki hurbil egin daiteke $\mu = np$ parametroa duen *Poisson*en banaketaren bidez. Praktikan, $p \leq 0{,}05$ eta $n \geq 20$ direnean onartzen da hurbilketa hori. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

Grafikoki, 3.8 irudian ikus daiteke hurbilketa horren egokitasuna, parametroen balio ezberdinetarako. Barra bertikalek banaketa binomialaren probabilitate-legea adierazten dute, eta poligonalek *Poisson*en banaketaren probabilitate-legea.



3.8 irudia: Banaketa binomialaren (marra bertikaletan) eta *Poisson*en hurbilketaren (poligonalean) probabilitate-legeen adierazpide grafikoak, parametroen balio ezberdinetarako, itxaropenaren eta bariantzaren balioak kontserbatuz ($\lambda = np$).

Hurbilketa hori p oso txikia denean erabiltzen denez, *Poisson*en banaketari gertaera arraroen lege deritzo maiz.

3.9 adibidea

«Escherichia coli» gizakiaren digestio-aparatuan sarritan agertzen den bakterioa da. «Escherichia coli»ren kasuan, batez beste, 10^9 zeluletatik batek mutatzen du estreptomizinari sentikor izatetik estreptomizinari sentikaitz izatera. Mutazio horrek inplikatutako gizabanakoa estreptomizinari sentikaitz izatera bihur dezake. Halako zeluletatik $2 \cdot 10^9$ zelula behatuz, zein da bat ere ez mutatzeko probabilitatea? Eta zein gutxienez bat mutatzeko probabilitatea? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

Kasu hori benetan binomiala da, $n = 2 \cdot 10^9$ eta $p = \frac{1}{10^9}$ izanik. Hots, $\frac{1}{10^9}$ oso txikia denez, zelula baten mutazioa oso kasu arraroa da. Horregatik, $X \equiv$ «Mutatzen duen zelula kopurua», gutxi gorabehera $\mu = \lambda s = np = 2 \cdot 10^9 \cdot \frac{1}{10^9}$ = 2 parametroa duen *Poisson*en banaketari darraiola kontsidera daiteke.

*Poisson*en taula begiratuta, mutaziorik ez egoteko probabilitatea $P(X = 0) = 0{,}135$ dela ikusten dugu. Diferentziaz kalkulatzen da gutxienez mutazio bat egoteko probabilitatea: $P(X \geq 1) = 1 - P(X = 0) = 1 - 0{,}135 = 0{,}865$.

3.7 BANAKETA NORMALA

Familia normala zorizko aldagai jarraituen familia bat da. *De Moivre*k definitu zuen lehenengoz 1773. urtean, proba kopurua emendatzerakoan banaketa binomialaren limite gisa. Aurkikuntza horrek ez zuen arrakasta handirik izan, eta banaketa normala *Laplace*k eta *Gauss*ek ediren zuten berriro mende erdi geroago, behaketen akats-banaketaren azterketatik abiatuz. Banaketa normala da erabilera teorikorik errazena duen banaketa eta, hurrengo gaietan ikusiko dugunez, banaketa normalean oinarritzen dira teknika estatistiko gehienak. Beraz, probabilitate-banaketa jarraiturik garrantzitsuena da, eta bere aplikazio praktiko anitz daude datu esperimentalen azterketan. Halaber, interes handiko beste zenbait banaketaren sortzaile ere bada.

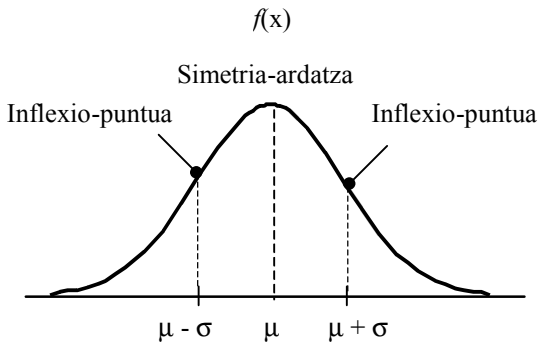
n eta p parametroen balioen arabera banaketa binomial asko dauden era berean, banaketa normal asko daude μ batezbestekoaren eta σ desbideratze estandarraren balioen arabera.

Izan bedi X zorizko aldagai jarraitua, non bere heina ardatz erreal osoa baita eta bere dentsitate-funtzioa honako adierazpen honek emanik baitago:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty \qquad \mu \in \mathbb{R} \text{ eta } \sigma \in \mathbb{R}^+$$

Orduan, X **banaketa normala**ri darraio eta X: $N(\mu, \sigma)$ adierazten da. Banaketa normala μ (zenbaki erreala) eta σ (zenbaki erreal positiboa) bi parametroen menpekoa da. Banaketa normalaren dentsitate-funtzioak honako propietate hauek betetzen ditu:

1. $f(x)$ dentsitate-funtzioa da; hots,
   - $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
   - $\int_{-\infty}^{\infty} f(x)dx = 1$
2. $E(X) = \mu$ eta $Var(X) = \sigma^2$.
3. Banaketa normalaren $f(x)$ dentsitate-funtzioaren grafikoa kanpai-itxurako kurba simetrikoa da eta zentroa μ batezbestekoan dauka, zeina maximoa baita.
4. $x = \mu - \sigma$ eta $x = \mu + \sigma$ abzisako puntuak $f(x)$ kurbaren inflexio-puntuak dira. $(\mu - \sigma, \mu + \sigma)$ tartean ganbila eta bertatik kanpo ahurra da kurba hori. Zenbat eta handiagoa izan σ, orduan eta leunagoa izango da kurba.



3.9 irudia. Banaketa normalaren dentsitate-funtzioaren adierazpide grafikoa.

3.10 adibidea

Kutsadura atmosferikoan, autoen ihes-tutuetatik sortutako hidrokarburoek dute ekarpenik handienetarikoa. Auto batek kilometroko sortzen duen hidrokarburo gramoak adierazten dituen zorizko aldagaiari *X* deritzogu. Demagun *X*-ren banaketa normala dela, eta bere batezbestekoa gramo bat eta desbideratze estandarra 0,25 gramo direla. Honako hau da *X*-ren dentsitate-funtzioa:

$$f(x) = \frac{1}{0{,}25\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-1}{0{,}25}\right)^2}$$

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

Horren grafikoa $x = 1$ zuzenarekiko simetrikoa da, eta 0,75 eta 1,25 puntuetan daude bere inflexio-puntuak.

Teorikoki, zorizko aldagai normalak edozein balio erreal har dezake. Kasu horretan, baldintza hori nahiko irreala da, auto batek ezin baitu hidrokarburo kantitate negatiborik sortu. Beraz, X-k banaketa normalari darraiola esaten denean, X-k har ditzakeen balio fisikoekin kurba normalak probabilitate onargarriak ematen dituela esan nahi dugu. Ikuspegi horrekin, zoriz aukeraturiko auto batek sortzen duen hidrokarburo kantitatea 0,9 eta 1,5 gramo bitartean egoteko probabilitatea kalkula dezakegu, f dentsitate-funtzioaren azpitik geratzen den azalera kalkulatuz, x = 0,9 eta x = 1,5 zuzenen artean. Hots, honako hau da kalkulatu nahi dugun probabilitatea:

$$P(0{,}9 \le X \le 1{,}5) = \int_{0.9}^{1.5} \frac{1}{0{,}25\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-1}{0{,}25}\right)^2} dx$$

Integral horren kalkulua ez da berehalakoa, baina ikus dezagun kalkulu hori errazteko metodoa.

Banaketa normalen kopurua infinitua da, eta bakoitza μ eta σ parametroen menpekoa da. Kanpai guztien artean, μ = 0 eta σ = 1 balioetakoari «kanpai estandar» edo «*Gauss*en kanpai» deritzogu, eta ordenatu-ardatzarekiko simetrikoa da. Parametroak μ = 0 eta σ = 1 dituen aldagai normalari **aldagai normal estandarizatu** deritzo, eta Z hizkiaren bidez adierazten da, Z: $N(0, 1)$. Banaketa horrek aparteko garrantzia du, zeren eta beste edozein banaketa normali dagozkion probabilitateak kalkulatzeko erreferentzia gisa erabiliko baita.

Demagun X: $N(\mu, \sigma)$ dela. Nola kalkula daitezke $P(x_1 < X < x_2)$ probabilitateak, non x_1 eta x_2 ($x_1 < x_2$) edozein zenbaki erreal baitira? Teorikoki, hauxe da:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Aurreko adibidean aipatu dugunez, integral horren kalkulua ez da erraza. Horretaz gainera, μ eta σ-ren balio bakoitzerako kalkulu berriak egin beharko genituzke.

<u>Propietatea</u>. Demagun X zorizko aldagaia dela, non X: $N(\mu, \sigma)$ baita. Orduan, $Z = \dfrac{X-\mu}{\sigma}$ aldagai normal estandarizatua Z: $N(0, 1)$ da.

Prozesu horren izena **estandarizazioa** da, eta modu honetara kalkula daiteke ondorioz planteaturiko edozein probabilitate:

$$P(x_1 < X < x_2) = P(x_1 - \mu < X - \mu < x_2 - \mu) = P\left(\frac{x_1-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right) =$$

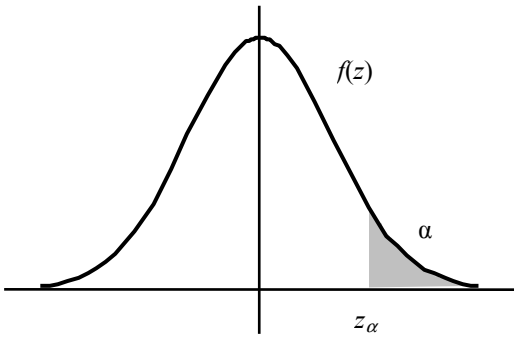
$$= P(z_1 < Z < z_2)$$

non $z_1 = \dfrac{x_1-\mu}{\sigma}$ eta $z_2 = \dfrac{x_2-\mu}{\sigma}$ baitira.

Horrela, X: $N(\mu, \sigma)$ dentsitate-kurbaren pean, x_1 eta x_2 bitarteko azalera aurkitzeko arazoa $N(0, 1)$ kurbaren pean z_1 eta z_2 bitarteko azalera aurkitzeko arazo bihurtzen da. Azken hori A eranskinean agertzen den banaketa normal estandarizatuaren taula (3. taula) erabiliz ebazten da.

Banaketa normalaren balio kritiko deritzo banaketa normalaren dentsitate-funtzioaren eta abzisa-ardatzaren bitartean eskuinerantz aurrefinkaturiko azalera uzten duen balioari (banaketa-funtzioaren balioaren osagarria da). Aurrefinkaturiko azalera hori α baldin bada, balio kritikoa z_α adieraziko da, hots, $P(Z > z_\alpha) = \alpha$. Erlazio hori 3.10 irudian

erakusten da grafikoki. *A* eranskineko 3. taulan agertzen dira α azalerak eta z_α banaketa normal estandarizatuaren balio kritikoak.



3.10 irudia. Banaketa normalaren z_α balio kritikoaren adierazpide grafikoa.

3.11 adibidea

Banaketa normal estandarizatuaren taularen erabilera. Dentsitate-kurba normal estandarizatuaren simetria eta taula erabiliz, honako probabilitate hauek kalkulatuko ditugu:

a) $P(Z > 2) = 0{,}0228$

b) $P(Z \leq -1{,}5) = P(Z \geq 1{,}5) = 0{,}0668$

c) $P(Z < 2{,}36) = 1 - P(Z \geq 2{,}36) = 1 - 0{,}00914 = 0{,}9909$

d) $P(0{,}8 \leq Z < 1{,}24) = P(Z \geq 0{,}8) - P(Z \geq 1{,}24) = 0{,}2119 - 0{,}1075 = 0{,}1044$

e) $P(Z > a) = 0{,}0301 \Rightarrow a = 1{,}88$

f) $P(Z \leq b) = 0{,}025 \Rightarrow P(Z \geq -b) = 0{,}025 \Rightarrow -b = 1{,}96 \Rightarrow b = -1{,}96$.

3.12 adibidea

3.10 adibideari jarraituz, kalkula dezagun zoriz aukeraturiko auto batek sortzen duen hidrokarburo kantitatea 0,9 eta 1,5 gramo bitartean egoteko probabilitatea:

$$P(0{,}9 \leq X \leq 1{,}5) = P\left(\frac{0{,}9-1}{0{,}25} \leq \frac{X-1}{0{,}25} \leq \frac{1{,}5-1}{0{,}25}\right) = P(-0{,}4 \leq Z \leq 2) =$$

$$= 1 - P(Z \leq -0{,}4) - P(Z \geq 2) = 1 - P(Z \geq 0{,}4) - P(Z \geq 2) =$$

$$= 1 - 0{,}3446 - 0{,}0228 = 0{,}6326$$

3.13 adibidea

Demagun, aurreko esperientziaren arabera, bitik hamalau urte bitarteko

haurren kolesterol-maila 175mg/dL batezbestekoa eta 30mg/dL desbideratze estandarra dituen banaketa normalari darraiola. Kolesterol altua heredatzen den aztertu nahi dugu. Horretarako, hiperkolesterolemia duten gurasoen 2 eta 14 urte bitarteko haurren kolesterol-maila neurtuko dugu. Horietariko haur baten kolesterol-mailak 200mg/dL gainditzen badu, medikuaren kontrolpean jarriko da. (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1994).

a) Kalkula ezazu hiperkolesterolemia duten gurasoen 2 eta 14 urte bitarteko haurra kontrolpean jartzeko probabilitatea.

b) Hamar haurrek osaturiko laginean, zein da horietariko bat gutxienez kontrolpean egoteko probabilitatea?

Emaitza:

a) $X \equiv$ «Haurraren kolesterol-maila»: $N(175, 30)$

$$P(X > 200) = P\left(\frac{X-175}{30} > \frac{200-175}{30}\right) = P(Z > 0{,}83) = 0{,}2$$

Hiperkolesterolemiadun gurasoen 2 eta 14 urte bitarteko haur bat medikuaren kontrolpean egoteko probabilitatea 0,2 da.

b) Aldagai berri bat definituko dugu:

$Y \equiv$ «Hamar haur artean, medikuaren kontrolpean daudenen kopurua»: Bin(10, 0,2).

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1 - 0{,}1074 = 0{,}8926$$

Beraz, hamar haurrek osaturiko taldean, bat gutxienez medikuaren kontrolpean egoteko probabilitatea 0,9 da.

Propietatea. Demagun X_1 eta X_2 zorizko aldagai askeak direla, non X_1: $N(\mu_1, \sigma_1)$ eta X_2: $N(\mu_2, \sigma_2)$ baitira. Bi horien batura eta diferentzia definitzen dira: $B = X_1 + X_2$ eta $D = X_1 - X_2$. Orduan, B eta D aldagaien banaketa normala da, eta beren parametroak honako hauek dira:

B: $N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

D: $N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Propietate hori aldagai normal eta askeen edozein konbinazio linealetarako orokor daiteke.

3.14 adibidea

Hiru seme-alabako familietan, X alaba kopurua aztertzen dugu. X zorizko aldagaiaren balioak 0, 1, 2 eta 3 dira. 2.3 adibidean ikusi dugunez, zortzi dira familia horien konposaketa posibleak: *sss*, *ass*, *sas*, *ssa*, *aas*, *asa*, *saa*, *aaa*.

Demagun X: $N(2, 2)$ eta Y: $N(5, 1)$ zorizko aldagai askeak direla. Orduan, aurreko propietatea erabiliz, $T = \frac{1}{2}X + 3Y$ aldagaia normala izango da. Kalkula ditzagun bere parametroak:

$$E(T) = E(\frac{1}{2}X + 3Y) = \frac{1}{2}E(X) + 3E(Y) = \frac{1}{2}\cdot 2 + 3\cdot 5 = 16$$

$\mathrm{Var}(T) = \mathrm{Var}(\frac{1}{2}X + 3Y) = \frac{1}{4}\mathrm{Var}(X) + 9\cdot\mathrm{Var}(Y) = \frac{1}{4}\cdot 4 + 9\cdot 1 = 10$ (independentzia erabiliz).

Beraz, T: $N(16, \sqrt{10})$.

3.15 adibidea

3.10 adibideari jarraituz, demagun ordubetean hogei auto pasatzen direla kilometro bateko errepide batek zeharkatzen duen herrian. Kalkula dezagun herri horretan ordubetean 22 gramo hidrokarburo baino gehiago sortzeko probabilitatea.

<u>Emaitza</u>: X: $N(1, 0{,}25)$ auto batek kilometroko sortzen duen hidrokarburo kantitatea da. Hogei autok sortzen duten hidrokarburo kantitatea adierazteko, beste aldagai bat definituko da:

$$B = X_1 + X_2 + \ldots + X_{20}$$

Jo dezagun auto bakoitzak sortzen duen hidrokarburo kantitatea besteak sortzen duenarekiko askea dela. Orduan, B normala izango da eta parametro hauek izango ditu:

$$E(B) = E(X_1 + X_2 + \ldots + X_{20}) = 20\cdot E(X) = 20\cdot 1 = 20$$

$$\mathrm{Var}(B) = \mathrm{Var}(X_1 + X_2 + \ldots + X_{20}) = 20\cdot\mathrm{Var}(X) = 20\cdot 0{,}25^2 = 1{,}25$$

$$\Rightarrow \mathrm{DE}(B) = \sqrt{1{,}25} = 1{,}12$$

Hain zuzen, B: $N(20, 1{,}12)$.

$$P(B > 22) = P\left(\frac{B-20}{1{,}12} > \frac{22-20}{1{,}12}\right) = P(Z > 1{,}79) = 0{,}0367$$

Beraz, herrian ordubetean 22 gramo hidrokarburo baino gehiago sortzeko probabilitatea 0,04 da.

3.8 LIMITEAREN TEOREMA ZENTRALA

Banaketa normalari (doi-doi ala gutxi gorabehera) jarraitzen dion zorizko aldagai jarraituen kopurua izugarri handia da. Horren azalpen matematikoa teoria estatistikoaren garrantzitsuenetakoa den Limitearen Teorema Zentrala delakoan datza. Nahiz eta hemen ezingo dugun hura frogatu, haren esanguraren deskribapen intuitiboa azalduko dugu, inferentzia estatistikoan daukan aplikazio anitz dela kausa. Batzen diren oso antzeko bariantza finituko faktore txiki askeen konkurrentziaz, zorizko aldagai ugariak sortzen dira. Adibidez, uzta bateko laranja baten pisua, beste faktore batzuen artean, lurraren ezaugarri, erabilitako ongarri, hezetasun-maila, tenperatura eta abarren pean dagoen zorizko aldagaia da. Eta bariantza txiki eta ezberdin samarra duten aldagai txiki askeen kopuru handia batuz eratzen diren zorizko aldagai guztiak *Gauss*en lege normalari darraizkio. Aldagai biologiko anitz horrela sortzen direnez gero, ez gaitu harritu behar banaketa normala hain maiztasun handiz agertzeak.

Limitearen Teorema Zentrala. Demagun X_1, X_2, …, X_n zorizko aldagai askeak eta berdinki banatuak direla, eta μ batezbestekoa eta $\sigma^2 \neq 0$ bariantza finitua direla. Orduan, $S_n = X_1 + X_2 + \ldots + X_n$ zorizko aldagaia gutxi gorabehera normala da, eta bere parametroak $n\mu$ eta $\sigma\sqrt{n}$ dira. Hain zuzen, $S_n \approx N(n\mu, \sigma\sqrt{n})$.

Emaitza baliokidea beste hau da: $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \approx N(0, 1)$.

Banaketa gutxi gorabehera edo asintotikoki normala izateak limitearen esanahia dauka. Hots, n aldagai kopuruak infiniturantz jotzen duenean, S_n aldagaiaren banaketak normalerantz joko du. Beraz, zenbat eta handiagoa izan n, orduan eta hobea da hurbilketa. Praktikan, $n \geq 30$ denean onartuko dugu hurbilketa hori.

Teorema horretan, emaitza berezia da gutxi gorabehera normala izatea berdinki banatuak eta askeak diren edozein aldagai segidaren batura, S_n. Emaitza hori ez da ikasmaterial honetan frogatuko, horretarako behar den probabilitate-teoriako ezagutza sakona delako. Behin normala dela jakin ondoren, bere parametroak kalkulatzeko itxaropenaren eta bariantzaren propietateak erabiltzea besterik ez dugu egin behar. Hots,

$$E(S_n) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) = n\mu$$

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n) = n\sigma^2$$

Baina, teorema horren garrantzia, nahiz eta X_i aldagaiak normalak ez izan, S_n aldagaien batura ia-ia normala izatean datza.

3.16 adibidea

Hiru farmazia-laborategik sendagai bera ekoizten dute. Sendagaia 100 paketeko kutxatan gordetzen da, eta pakete bakoitzak akastuna izateko 0,01 probabilitatea dauka. Laborategiko 90 kutxa ireki dira eta sendagai-pakete

akastunen kopuru osoa zenbatu da. Zein da pakete akastunen kopuru oso hori 100 baino txikiagoa izateko probabilitatea? (Iturria: A. Amparan, I. Aróstegui, B. del Hoyo, S. Marcaida, A. Urkaregi, A. Valle. *Bioestatistikako apunteak*. EHU, 1997, argitaratu gabe).

Emaitza: honako aldagai hauek definituko ditugu:

$Y_1 \equiv$ «1. kutxako sendagai-pakete akastunen kopurua»: Bin(100, 0,01)

$Y_2 \equiv$ «2. kutxako sendagai-pakete akastunen kopurua»: Bin(100, 0,01)

……

$Y_{90} \equiv$ «90. kutxako sendagai-pakete akastunen kopurua»: Bin(100, 0,01)

Izan bedi $S_n = Y_1 + Y_2 + \ldots + Y_{90} \equiv$ «sendagai-pakete akastunen kopuru osoa». Honako hauek dira bere itxaropena eta bariantza:

$E(S_n) = E(Y_1 + Y_2 + \ldots + Y_{90}) = 90E(Yi) = 90\cdot(100\cdot 0{,}01) = 90\cdot 1 = 90$

$\mathrm{Var}(S_n) = \mathrm{Var}(Y_1 + Y_2 + \ldots + Y_{90}) = 90\cdot\mathrm{Var}(Yi) = 90\cdot(100\cdot 0{,}01\cdot 0{,}99) = 90\cdot 0{,}99 = 89{,}1$

Limitearen Teorema Zentralean proposatzen den hurbilketaren ondorioz, $S_n \approx N(90, \sqrt{89{,}1})$ da. Beraz,

$$P(S_n < 100) = P\left(\frac{S_n - 90}{\sqrt{89{,}1}} < \frac{100-90}{\sqrt{89{,}1}}\right) = P(Z < 1{,}05) = 1 - P(Z > 1{,}05) =$$

$= 1 - 0{,}1469 = 0{,}8531$

Hots, 0,85 da pakete akastunen kopuru osoa 100 baino txikiagoa izateko probabilitatea.

Adibide horretan ikusten denez, parametro berdindun banaketa binomialari darraioten zorizko aldagai askeak batzen direnean, gutxi gorabehera normala den aldagaia lortzen da.

3.9 BANAKETA NORMALA BESTE BANAKETA BATZUEN LIMITE GISA

3.9.1 Banaketa binomialerako banaketa normalaren hurbilketa

Gai honetako 3.5 atalean, banaketa binomiala *n* saiotan agertutako arrakasta kopuru gisa definitu dugu, eta *p* arrakastaren probabilitatea probaz proba konstantea izan da. Proba

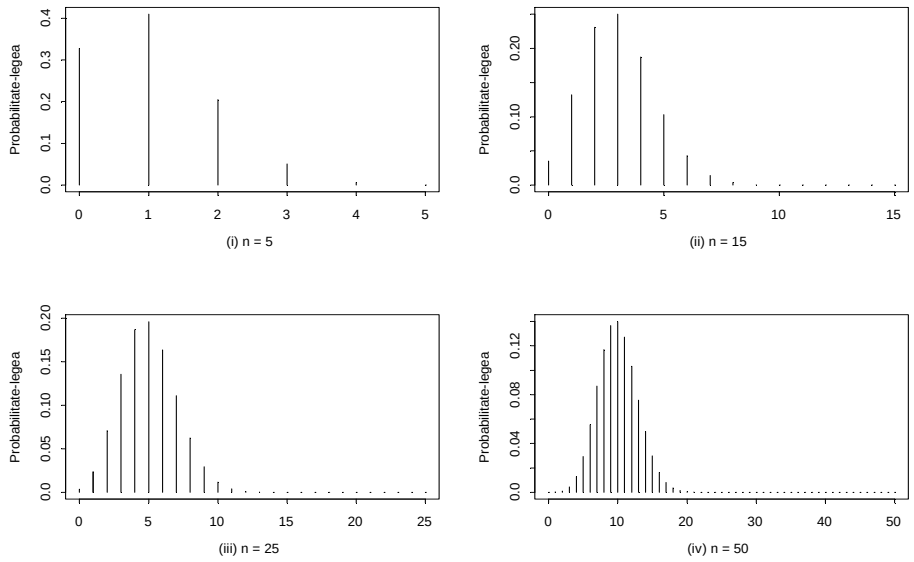
kopurua, n, handia denean, nahiko nahasi bihurtzen da banaketa binomialari dagozkion probabilitateak kalkulatzea. A eranskinean agertzen den banaketa binomialaren taula eraiki da arazo hori konpontzeko, baina, hala ere, taula horretan ez dira n-ren balio handiak agertzen. Kasu horretan, probabilitate zehatza kalkulatu beharrean, hobe da probabilitatearen hurbilketa bat edo gutxi gorabeherako balioa kalkulatzea. Gutxi gorabeherako probabilitatea kalkulatzeko, Limitearen Teorema Zentralaren ondorio bat erabiliko dugu.

*Moivre*ren teorema. Izan bedi X banaketa binomialari darraion zorizko aldagaia, X: Bin(n, p), non $p\in$ (0, 1) eta n handia baitira. Orduan, X gutxi gorabehera normala da eta bere parametroak np eta $\sqrt{npq}$ dira. Hots, $X \approx N(np, \sqrt{npq})$.

Honako hau da emaitza baliokidea: $\dfrac{X-np}{\sqrt{npq}} \approx N(0,1)$.

Oharra. Praktikan, $n \geq 30$ eta $0{,}1 < p < 0{,}9$ direnean onartzen da hurbilketa hori; hots, Bin(n, p) $\approx N(np, \sqrt{npq})$. (Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991). Arrakastaren probabilitatea txikia zein handia denean ($p \leq 0{,}1$ edo $p \geq 0{,}9$), hobe da banaketa binomialerako Poissonen hurbilketa erabiltzea normalarena baino.

Grafikoki, 3.11 irudian erakusten da hurbilketa horren egokitasuna p = 0,2 baliorako. Zenbat eta n handiagoa izan, orduan eta hurbilketa hobea dela ikusten da.



3.11 irudia. Banaketa binomialaren probabilitate-legearen adierazpide grafikoa, p = 0,2 eta n = 5, 15, 25, 50 parametroen balioetarako.

3.17 adibidea

Herri bateko biztanleen % 40 A odol-taldekoa da. Lor ezazu, 200 pertsonako komunitatean, horietariko 90 baino gehiago A taldekoak izateko probabilitatea. (Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991).

Emaitza: honako zorizko aldagai hau definitzen dugu:

$X \equiv$ «200 pertsonako komunitatean, A odol-taldekoa den pertsona kopurua»: Bin(200, 0,4)

Normalaren hurbilketa onartzen da, zeren eta $n \geq 30$ eta $0{,}1 < p < 0{,}9$ betetzen baitira. Orduan, X: $Bin(200, 0{,}4) \approx N(80, \sqrt{48})$. Beraz,

$$P(X > 90) = P\left(\frac{X-80}{\sqrt{48}} > \frac{90-80}{\sqrt{48}}\right) = P(Z > 1{,}44) = 0{,}0749$$

Komunitate horretan, 0,07 da A odol-taldekoen kopurua 90 baino handiagoa izateko probabilitatea.

3.9.2 *Poisson*en banaketarako banaketa normalaren bidezko hurbilketa

Banaketa normala, binomialaz gainera, beste banaketa diskretu batzuk hurbiltzeko ere erabil daiteke; esate baterako, *Poisson*en banaketa. Horren arrazoia da, λ handia denean, *Poisson*en probabilitateen kalkulua korapilatsu bihurtzen dela. Banaketa binomialaren hurbilketarako erabili den teknika bera erabiltzen da. Hain zuzen, *Poisson*en banaketaren itxaropena eta bariantza gutxi gorabeherako normalaren parametroen berdinak egiten dira.

Teorema. Izan bedi X *Poisson*en banaketari darraion zorizko aldagaia, X: $\mathcal{P}(\lambda)$, non λ handia baita. Orduan, X gutxi gorabehera normala da eta bere parametroak λ eta $\sqrt{\lambda}$ dira. Hots, $X \approx N(\lambda, \sqrt{\lambda})$.

Honako hau da emaitza baliokidea: $\dfrac{X-\lambda}{\sqrt{\lambda}} \approx N(0, 1)$.

Oharra. Praktikan, $\lambda \geq 10$ denean onartzen da hurbilketa hori; hots, $\mathcal{P}(\lambda) \approx N(\lambda, \sqrt{\lambda})$. (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1994). Grafikoki, 3.12 irudian erakusten da hurbilketa horren egokitasuna. Zenbat eta handiagoa izan λ, orduan eta hurbilketa hobea dela ikusten da.

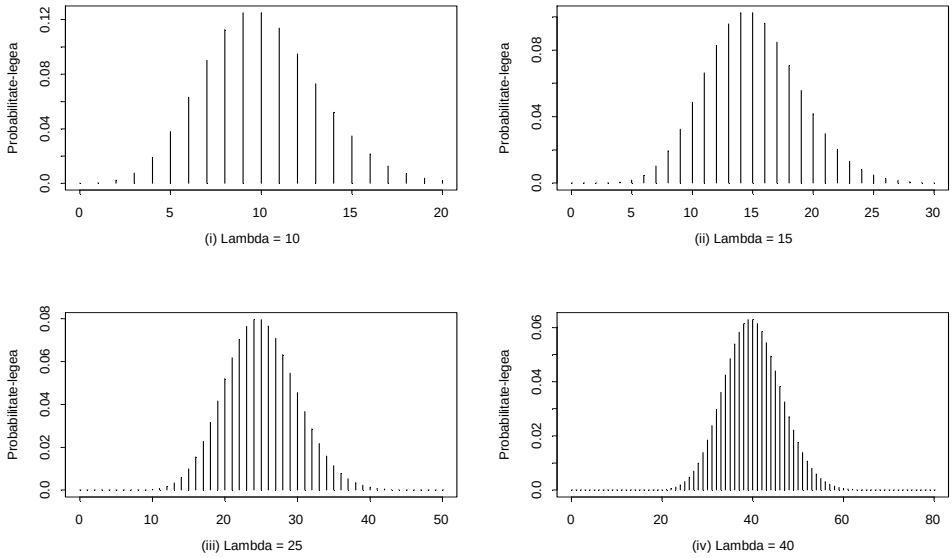
3.18 adibidea

3.7 adibidearekin jarraituko dugu. Sukar tifoidea errotik aterata dagoela esateko, bost urteko epean, gaixotasun horrekin hildakoen kopuruak 10 baino gutxiago izan behar du. Kalkula ezazu populazio horretan sukar tifoidea errotik aterata egoteko probabilitatea.

$\lambda = 4{,}6$ eta $t = 5$ direnez, $\mu = \lambda t = 23$ da. Orduan, $X \equiv$ «5 urtean zehar hildakoen kopurua»: $\mathcal{P}(23) \approx N(23, 4{,}79)$ da.

$$P(X < 10) = P\left(\frac{X-23}{\sqrt{23}} < \frac{10-23}{\sqrt{23}}\right) = P(Z < -2{,}71) = 0{,}0034$$

Beraz, populazio horretan sukar tifoidea errotik desagertuta egoteko probabilitatea oso txikia da; hain zuzen, 0.003.



3.12 irudia. *Poisson*en banaketaren probabilitate-legearen adierazpide grafikoa, λ = 10, 15, 25, 40 parametroen balioetarako.

3.10 ARIKETAK

3. 1. Birusaren bizitza-aldia (ordutan) honako dentsitate-funtzio hau daukan zorizko aldagaia da:

$$f(x) = \begin{cases} 0 & \text{baldin } x < 1 \\ k/x^4 & \text{baldin } x \geq 1 \end{cases}$$

(Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991).

a) Kalkula ezazu $f(x)$ dentsitate-funtzioa izateko k-ren balioa.

b) Zein da birusaren batez besteko bizitza-aldia?

c) Aurki ezazu birusaren bizitza-aldiaren banaketa-funtzioa.

d) Kalkula ezazu birusaren bizitza-aldiaren desbideratze estandarra.

e) Aurki ezazu zoriz aukeraturiko birusa 5 ordu baino gehiago bizitzeko probabilitatea.

3. 2. Ospitale batean, umeen jaiotza-pisua (kilogramotan) honako dentsitate-funtzio hau duen zorizko aldagaia dela behatu da:

$$f(x) = \begin{cases} kx & \text{baldin } 2 \leq x \leq 4 \\ 0 & \text{beste kasuetan} \end{cases}$$

(Iturria: V. Quesada, A Isidoro, L. A. López. *Curso y Ejercicios de Estadística*. Alhambra Universidad, 1982).

a) Aurki ezazu k konstantearen balioa, $f(x)$ dentsitate-funtzioa izan dadin.

b) Aurki ezazu umeen jaiotza-pisuaren banaketa-funtzioa.

c) Kalkula itzazu umeen jaiotza-pisuaren batezbestekoa, bariantza eta desbideratze estandarra.

d) Kalkula ezazu zoriz aukeraturiko ume baten pisua 3 kilogramo baino handiagoa izateko probabilitatea.

e) Zenbat kilogramo izan beharko ditu ume batek, umeen % 90aren pisua balio horretatik beherakoa izan dadin?

3. 3. Demagun $X \equiv$ «Hiri batean, familiako ume kopurua» neurtzen duen zorizko aldagaiak 3.1 taulan agertzen den probabilitate-legeari jarraitzen diola.

X	$P(X = x)$
0	0,47
1	0,30
2	0,10
3	0,06
4	0,04
5	0,02
6	0,01

3.1 taula. 3.3 ariketako aldagaiaren probabilitate-legea. Iturria: V. Quesada, A Isidoro, L. A. López. *Curso y Ejercicios de Estadística*. Alhambra Universidad, 1982.

a) Kalkula ezazu familiako ume kopuruaren banaketa-funtzioa.

b) Kalkula ezazu familiako ume kopuruaren itxaropen matematikoa. Zein da zenbaki horren esanahia?

c) Kalkula itzazu familiako ume kopuruaren bariantza eta desbideratze estandarra.

d) Jo dezagun hiri horretako udalak ume bakoitzeko 2.000 pezeta ordaintzen diola familiari. $Y = 2000 \cdot X$ zorizko aldagaia definitzen badugu, zer adierazten du Y-k? Zein da Y-ren probabilitate-legea?

e) Kalkula itzazu Y-ren batezbestekoa, bariantza eta desbideratze estandarra.

3. 4. Demagun $X \equiv$ «Animalia mikroorganismoaren bizitza-aldia (ordutan)» zorizko aldagaia honako dentsitate-funtzio honek zehazten duela:

$$f(x) = \begin{cases} 0 & \text{baldin } x < 0 \\ e^{-2x} & \text{baldin } x \geq 0 \end{cases}$$

(Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991).

Honako hauek eskatzen dira:

a) Kalkula ezazu X-ren banaketa-funtzioa eta aurki ezazu mikroorganismoa ordubete baino gehiago bizitzeko probabilitatea.

b) Zein da mikroorganismoaren itxarotako bizitza-aldia?

c) Bost mikroorganismo isolatuz gero, zein da bostek ordubete baino gehiago bizirik irauteko probabilitatea?

3. 5. Leuzemia akutuaren diagnostikoa izan ondoren, honako dentsitate-funtzio hau du biziraupen-denbora urtetan adierazten duen zorizko aldagai jarraituak:

$$f(x)=\begin{cases} -\dfrac{x}{2}+1 & \text{baldin } 0<x<2 \\ 0 & \text{beste kasuetan} \end{cases}$$

(Iturria: A. Amparan, I. Aróstegui, B. del Hoyo, S. Marcaida, A. Urkaregi, A. Valle. *Bioestatistikako apunteak*. EHU, 1997, argitaratu gabe).

a) Kalkula ezazu X-ren banaketa-funtzioa.

b) Kalkula itzazu $P(X \geq 1)$, $P(X > 1)$ eta $P(X = 1)$.

c) Kalkula ezazu leuzemia akutua pairatzen duen pazientearen itxarotako biziraupen-denbora.

d) Kalkula ezazu leuzemia akutua pairatzen duen pertsona batek sei hilabete baino gutxiago bizirik irauteko probabilitatea.

e) Leuzemia akutua pairatzen duten ehun paziente zoriz aukeratu dira. Zein da horietatik hamar baino gehiagok gutxienez sei hilabete bizirik irauteko probabilitatea?

3. 6. Gaur egun, badakigu 15 eta 35 urte bitarteko emakumeen artean anorexia nerbiosoaren intzidentzia % 15ekoa dela. Populazio horretatik zortzi emakume zoriz aukeratuta, kalkula itzazu honako hauek:

a) Anorexia nerbiosoa batek ere ez pairatzeko probabilitatea.

b) Bi gaixo soilik aurkitzeko probabilitatea.

c) Bi gaixo baino gehiago aurkitzeko probabilitatea.

d) Gehienera bost gaixo aurkitzeko probabilitatea.

e) Zein da gaixotasun horrekin itxaroten dugun emakume kopurua, talde horretan?

3. 7. Gizonezko osasuntsu batek eta hemofiliaren eramaile den emakumezko batek osatzen duten bikotearen ondorengoa emakumezko eramailea, gizonezko eria (baita eramailea ere) eta osasuntsua (ez eria, ez eramailea) izateko probabilitateak 1/4, 1/4 eta 1/2 dira, hurrenez hurren. Bikoteak zortzi seme-alaba ditu.

a) Zein da guztiak osasuntsuak izateko probabilitatea?

b) Zein da osasuntsuen kopurua lau baino handiagoa izateko probabilitatea?

c) Zein da itxaroten den eramaile kopurua? Eta eri kopurua?

d) Zein da eramaile kopuruaren bariantza?

e) Zein da bi alaba eramaile, bi seme eri, eta gainerakoak osasuntsuak izateko probabilitatea?

3. 8. *Migraina* gaixotasuna (buruko mina) sendatzeko konposatu bat garatu da. Fabrikatzaileak baieztatu du % 99an eraginkorra dela. Lau gaixorengan probatu da konposatu hori. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Zenbat pertsona sendatzea itxaro dugu?

b) Gutxienez gaixo bat sendatu dela jakinik, zein da sendaturiko gaixo kopurua gehienera hiru izateko probabilitatea?

c) Konposatuak gaixo bat ere ez badu sendatu, ba al duzu arrazoirik konpainiak baieztatzen duen % 99ko eraginkortasun-maila zalantzan jartzeko? Arrazoitu ezazu erantzuna.

d) Hiru ospitale aukeratu ditugu; bakoitzean ehun pertsona daude migrainarekin, eta horiei guztiei konposatua eman diete. Fabrikatzailearen esana egiatzat hartzen badugu, kalkula ezazu gutxienez ospitale batean sendatu ez den gaixo kopurua gehienera hiru izateko probabilitatea.

3. 9. Genetistek bi sexu-kromosoma (R eta Y) identifikatu dituzte gizakiengan. Gizaki guztiok R kromosoma dugu, eta Y kromosoma izatea da arrak bereizten dituena. Beraz, emea RR izango da, eta arra RY. Daltonismoa R kromosoman errezesiboa den r aleloak sorrarazten du. Y kromosoma ez dago daltonismoari lotuta. Beraz, daltonismoaren arabera, emakumeentzat hiru genotipo daude: RR (normala), Rr (eramailea) eta rr (daltonikoa); eta bi gizonentzat: RY (normala) eta rY (daltonikoa). Seme-alaba bakoitzak guraso bakoitzaren kromosoma bat heredatzen du zoriz. Daltonismorik ez duen emakumezko eramaile batek eta gizonezko batek ume bat dute. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Zuhaitz-diagrama baten bidez, adieraz itzazu umearen genotipo posibleak.

b) Zer probabilitate dago ume hori seme daltonikoa izateko?

c) Bikoteak hiru seme-alaba baldin baditu, zer probabilitate dago bi seme daltoniko izateko?

d) Bikoteak bost seme-alaba baldin baditu, zein da itxarotako seme daltonikoen kopurua? Zer probabilitate dago gehienera jota bi seme daltoniko izateko? Zer probabilitate dago hiru seme daltoniko edo hiru baino gehiago izateko?

3. 10. Gripearen aurka txertatutako 1.000 pertsonen artean birengan bigarren mailako efekturen bat agertzen da. Zer probabilitate dago txertatuen artean bigarren mailako efekturen bat daukatenen tasa 100.000tik 400 baino handiagoa izateko?

3. 11. Sueroa xiringatu eta gero, gizabanakoak positiboki erreakzionatzeko probabilitatea 0,991 dela jakinik, zehaztu ezazu, txertatuta dauden 200

pertsonako taldean, gehienera jota negatiboki erreakzionatzen duten hiru egoteko probabilitatea. (Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991).

3. 12. Fase epidemiologikoan, 2 urte baino ume gazteagoengan, % 0,2ko balioa dauka sepsi meningokozikoaren eraginak. Bi urte baino gazteagoak diren 200 umeen artean, kalkula itzazu:

a) Gaixotasuna lau umek pairatzeko probabilitatea.

b) Gaixotasuna lau umek edo lau baino gutxiagok pairatzeko probabilitatea.

3. 13. Ikertzaile batzuek kolesterol serikoa Z balioen bidez adieraztea nahi izango lukete, zeren kolesterol serikoa adinarekin eta sexuarekin erlazionaturik baitago. Kolesterol seriko gordina adierazten duen aldagaia X izanik, $Z = \dfrac{X - \mu}{\sigma}$ definituko dugu, non μ eta σ baitira aurrefinkaturiko adin eta sexu talde bateko kolesterol serikoaren batezbestekoa eta desbideratze estandarra. Z banaketa normal estandarizatua dela suposatzen da. (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1994).

a) Zein da $P(Z < 0{,}5)$?

b) Zein da $P(Z > 0{,}5)$?

c) Zein da $P(-1{,}0 < Z < 1{,}5)$?

Demagun, $Z > 2{,}0$ denean, pertsonak kolesterol altua duela eta, $1{,}5 < Z < 2{,}0$ denean, pertsonak kolesterola mugan daukala esaten dela.

d) Populazioren zer proportziok du kolesterol altua?

e) Populazioren zer proportziok du kolesterola mugan?

f) Zein dira banaketa normal estandarizatuaren dezilak, hau da, 10, 20, 30 . . . . 90 pertzentilak?

g) Zein dira banaketa normal estandarizatuaren koartilak?

3. 14. Diabetikoen artean, baraurik, odoleko glukosa-maila gutxi gorabehera banaketa normalari darraiola suposa daiteke; batezbestekoa 106 mg/dL da eta desbideratze estandarra 8 mg/dL. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Aurki ezazu $P(X \leq 120)$.

b) Diabetikoen ehuneko zenbatek dauka 90 eta 120 mg/dL arteko glukosa-maila?

c) Aurki ezazu $P(106 \leq X \leq 110)$.

d) Aurki ezazu $P(X \geq 121)$.

e) Aurki ezazu x_0 puntua, diabetikoen % 25ak baraurik x_0 baino glukosa-maila baxuagoa eduki dezan.

3. 15. EAEko ibai eta ibaiadarren emaria (ur kantitatea) 8,1 m^3/s-ko batezbestekoa eta 7,4 m^3/s-ko desbideratze estandarra dituen banaketa normalari darraio (Iturria: EUSTAT. *EUE* 97/98).

a) Kalkula ezazu EAEko ibai batek daraman ur kantitatea 4 m^3/s-tik beherakoa izateko probabilitatea.

b) Kalkula ezazu EAEko ibai batek daraman ur kantitatea 20 m^3/s baino handiagoa izateko probabilitatea.

c) Kalkula ezazu EAEko ibai batek daraman ur kantitatea 6 eta 10 m^3/s bitartean egoteko probabilitatea.

d) EAEn 19 ibai daude. Kalkula ezazu horietatik 20 m^3/s baino ur kantitate handiagoa daramaten ibaien kopurua gehienera hiru izateko probabilitatea.

3. 16. EAEn, 600 m-tik gorako mendien altitudea 965 metroko batezbestekoa eta 49.450 metro^2-ko bariantza dituen banaketa normalari darraio (Iturria: EUSTAT. *EUE* 97/98).

a) Kalkula ezazu EAEko mendi baten altitudea 800 metrotik beherakoa izateko probabilitatea.

b) Kalkula ezazu EAEko mendi bat 1.200 metro baino altuagoa izateko probabilitatea.

c) Kalkula ezazu EAEko mendi baten altitudea 900 eta 1.000 metro bitartean egoteko probabilitatea.

d) EAEn, 600 m-tik gorako 283 mendi daude. Kalkula ezazu horietatik 1.200 metroak gainditzen ez dituen kopurua % 75ekoa izateko probabilitatea.

3. 17. EAEko hiriburuetan, egun batean 25 gradutik gorako tenperatura egoteko probabilitateak honako hauek dira: Bilbon 0,137, Gasteizen 0,145 eta Donostian 0,044. (Iturria: EUSTAT. *EUE* 97/98).

a) Kalkula ezazu EAEko hiriburu bakoitzean urtebetean espero dugun 25 gradutik gorako egun kopurua.

b) Zer probabilitate dago Gasteizen hilabetean zehar 25 gradutik gorako tenperatura dagoen egun kopurua 10 baino gutxiago izateko? Eta Bilbon? Eta Donostian?

c) Zer probabilitate dago Donostian urtebetean 25 gradutik gorako tenperatura dagoen egun kopurua 50etik gorakoa izateko? Eta Bilbon? Eta Gasteizen?

3. 18. Etxeko animalia gisa erabiliak izateko Amazonako arroan harrapaturiko 50 papagairen artean, aldaketaren ostean bakar batek irauten duela bizirik jotzen da. Egun batean 700 papagai harrapatu dira. Zenbat papagaik egun bat bizirik irautea espero izatekoa da? Zer probabilitate dago egun batean gehienera hamar papagaik bizirik irauteko?

Hiru egunean zehar 700 papagai harrapatzen badira egunero, zer probabilitate dago hiru egun horietako bakoitzean gehienera hamar papagaik bizirik irauteko? Zer probabilitate dago hiru egunean zehar bizirik iraun duen papagai kopuru osoa hogeita hamar baino handiagoa izateko? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

3. 19. Lantegiek ibaira botatzen duten hondakinen azterketan, eguneroko isurketa kutsakorren kontrola ezarri da. Kontrola daraman erakundeak ibai bakoitzean egunero izandako isurketa kutsakorren kopurua zenbatzen du. Ibaizabal ibaiari dagokion azterketan oinarrituz, eguneroko isurketa kutsakorren kopurua 10eko itxaropena eta bariantza dauzkan zorizko aldagaia da. Egun ezberdinetako isurketa kopuruak askeak direla jotzen da. Zehaztu ezazu 150 egunean Ibaizabal ibaian egon den isurketa kutsakorren kopuruak 1.480 gainditzeko probabilitatea.

3. 20. Azken bolada honetan, ospitale bateko larrialdi-zerbitzua lanez gainezka dago, arreta hori behar duen paziente kopurua izugarri gehitu baita. Egunero erabiltzen dituzten batez besteko gasa kopurua 200 da, eta desbideratze estandarra 20. Kalkula ezazu, urte erdiko denboraldiko beharrizanak % 99ko probabilitatearekin betetzeko, larrialdietako materialaren arduradunak agindu behar duen gasa kopuru minimoa.

4 INFERENTZIA ESTATISTIKOA. LAGIN BAKAR BATEN AZTERKETA

4.1 SARRERA

Aurrefinkaturiko populazioko aleetan agertzen den ezaugarri (aldagai) baten azterketa da edozein ikerketa estatistikoren helburua. Populazioa osatzen duten ale guztien informazioa ezagutzen denean, **zentsua** edo **errolda** egiten dela esaten da. Baina hori ez da beti posiblea edo egokia izaten. Bai populazioa infinitua delako; bai proba hondagarriak direlako; bai elementu potentzialek osaturiko populazioa aztertzen ari garelako (esate baterako, sendagaiz tratatu ondoren, bigarren mailako ondorioak pairatzen dituen gaixoen portzentajea); edo, gehienean, ale bakoitza «kostu» bati lotuta dagoelako.

Horrelako egoerek **lagin** deritzon populazioaren zati baten informazioa hartzera behartzen dute ikertzailea.

Honako arazo hauek sortzen dira berehala:

1. Nola atera populazioari buruzko ondorioak?
2. Lagineko datuetatik abiatuz, zenbateko konfiantza-maila daukate ondorio horiek?
3. Lagin-balioak zein puntutaraino dira populazio osoaren adierazgarriak?

Estatistikaren oinarrizko metodologia, beraz, indukzioa da. Indukzioa edo inferentzia egiteko probabilitate-teoria erabiltzen da. Emaitzak interpretatzerakoan, kontuan hartu behar da «seguruak» ez direla, baizik eta nolabaiteko «konfiantza-maila» daukatela.

Inferentzia estatistikoa, probabilitate-teorian oinarriturik, honako helburu hauek betetzen dituen metodo multzoa da:

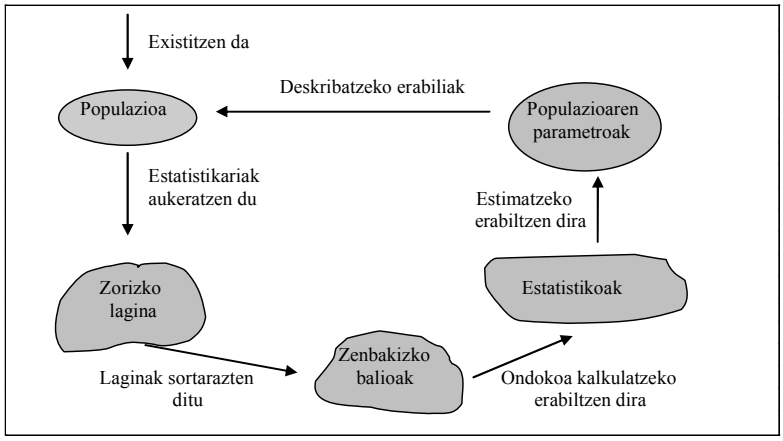
1. Emaitzen inferentziak edo orokortzeak gauzatzea, eta horien konfiantza-maila neurtzea.
2. Aurreko etapan lortutako emaitzak interpretatzea eta erabakiak hartzea.

Inferentzia estatistikoa bi arlotan sailka daiteke: Estimazioa eta hipotesi-kontrastea. Laginean lorturiko estatistikoen bidez populazioko parametro ezezagunak hurbiltzean datza **estimazioa**. **Hipotesi-kontrastea**, berriz, populazioko parametro bati edo batzuei buruzko hipotesia egin ondoren, lagineko estatistikoen balioen arabera hipotesia onartzeko edo errefusatzeko metodoa eratzean datza.

4.2 POPULAZIOAREN ETA LAGINAREN ARTEKO ERLAZIOA

Populazioa da azterketa estatistikoaren helburu den elementu multzoa. **Zorizko lagina** populazioaren elementuen azpimultzoa izango da, non elementu bakoitzaren aukera askea baita, eta aukeratua izateko probabilitatea ezaguna.

Esan dugunez, populazioarekiko inferentzia egingo da lagin deritzon populaziotik ateratako elementu sortatik lortutako informazioan oinarrituz. Lagina populazio txikiagotutzat uler daiteke. Zorizko aldagaiak laginean duen portaera populazioan duenaren deskribapen zehatza izatea espero dugu. Arrazoi horregatik, populazioko parametroen hurbilketa gisa erabiltzen dira laginean kalkulaturiko estatistikoen balioak. 4.1 irudian azaltzen da ideia hori.



4.1 irudia. Azterketa estasitikoa nola gauzatzen den adierazten duen eskema. Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994.

Nola aukeratzen da zorizko lagina? Galdera horren erantzuna ez da berehalakoa. Azterketa bakoitzean, faktore askoren menpekoa da aukeratzeko era. Populazioaren tamaina, populazio mota, erantzun behar diren galderak, erabil daitezkeen denbora eta diru kantitatea, eta nahi den zehaztasuna dira laginketa mota aukeratzen laguntzen duten faktoreetariko batzuk. Laginketa motarik egokiena azterketaren diseinuan finkatu behar da, ikertzailearen eta estatistikariaren artean. Ondoren, zorizko lagin bakunaren kontzeptua ezarriko dugu. Laginketa mota hori maiz erabiltzen da, ondo funtzionatzen du gehienetan, eta erraza da bai ulertzen bai erabiltzen.

Populazioaren elementu guztiek aukeratuak izateko probabilitate berdina daukatenean, zorizko laginari **zorizko lagin bakun** deritzo. Baina ikertzaileak ezin du ale konkretu bat laginean sartu, ez eta laginetik atera ere. Zorizko lagin bakunak lortzeko tresna kutxa

batetik izenak ateratzea bezain sinplea edo zorizko zenbakien sortzaile elektronikoa erabiltzea bezain konplikatua izan daiteke.

Populazio finitutik zorizko lagin bakuna aukeratzeko erarik sinpleenetarikoa da zorizko zenbakien taularen bidez egitea. Halaber, laginketa baliagarria izateko, laginak populazio osoaren adierazgarri izan behar du, eta zorizko lagin adierazgarria lortzeko metodo seguruenetarikoa zorizko zenbakien taulak erabiltzea da.

Zorizko zenbakia (edo **zorizko digitua**) 0, 1... 9 balioak probabilitate berdinarekin hartzen dituen X zorizko aldagaia da. Alegia, $P(X = 0) = \ldots = P(X = 9) = pr(X=0) = pr(X=1) = \ldots = pr(X=9) = \frac{1}{10}$.

Zorizko zenbakien taula da honako propietate hauek betetzen dituen digitu sorta:

1. Digitu bakoitzak (0, 1... 9) agertzeko probabilitate bera du.
2. Digitu baten balioa askea da beste taulako edozein digiturekiko.

A eranskinean, zorizko zenbakien taula agertzen da (4. taula).

4. 1 adibidea

Demagun pisua galtzeko tratamendu baten eraginkortasuna neurtu nahi dugula. Zorizko zenbakien taula erabiliz, tratamendupean dauden 1.000 pertsonen artean, nola aukera daitezke 25 pertsona zoriz?

Parte-hartzaileak ordena alfabetikoan ordenatu eta 000tik 999ra zenbatuko ditugu. Zorizko zenbakien taulan, edozein puntutatik hasten dela ere, aukera ditzagun hiru digituko hogeita bost talde. Adibide gisa, zorizko zenbakien taulako lehen errenkadatik hasten baldin bagara, 4.1 taulan adierazten diren zenbakiak lortuko ditugu.

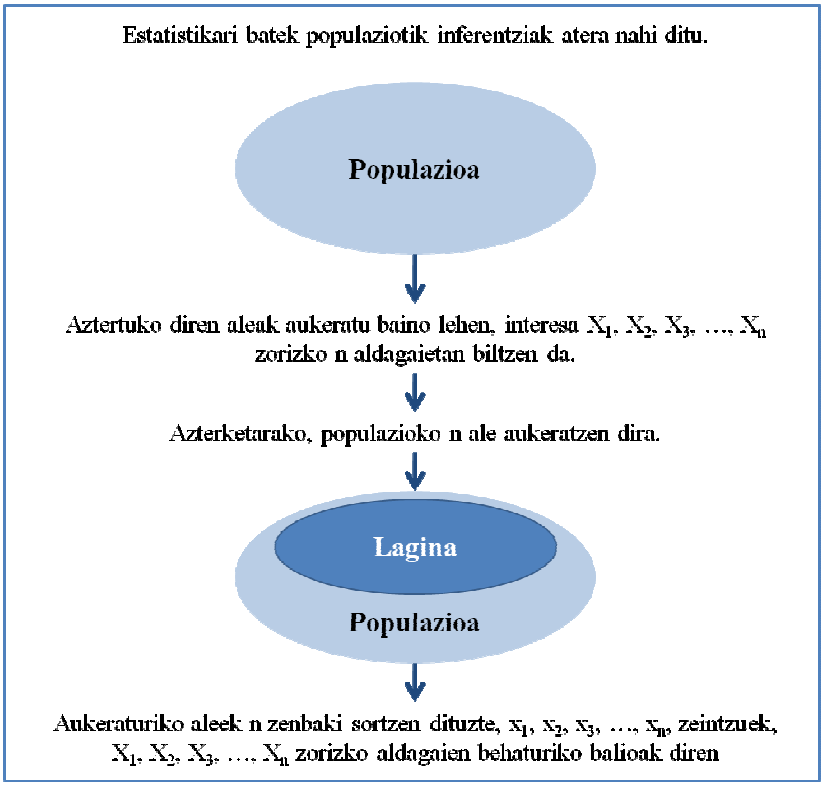
Zorizko zenbakien taulako lehengo bi errenkadak				Aukeraturiko zorizko zenbakiak						
57785	73283	69160	81349	577	857	328	369	160	813	497
79421	47737	42582	02361	942	147	737	425	820	236	188
88277	69862	95341	38066	277	698	629	534	138	066	823
82338	06389	24003	70989	380	638	924	003			

4.1 taula. Pisua galtzeko tratamendupean dauden 1.000 pertsonen artetik zoriz aukeraturiko hogeita bost pertsonak.

«Zorizko lagin bakuna» terminoak hiru erabilera ditu. Adibidez, 4. 1 adibidearekin jarraituz, demagun tratamendupean dauden pertsonak galtzen duten pisua adierazten duen X zorizko aldagaia aztertu nahi dugula, eta, horretarako, aurreko hogeita bost tamainako zorizko lagin bakuna atera dugula.

Azterketarako pertsonak aukeratu baino lehen, hogeita bost zorizko aldagai ditugu, X_1, X_2 … X_{25}, non X_i-k azterketarako aukeraturiko i. pertsonak tratamendupean galdu duen pisua adierazten baitu. Hogeita bost zorizko aldagai horiek X-ren banaketaren zorizko lagin bakuna adierazten dute. Letra maiuskulak erabiltzen dira, zorizko aldagaiak adierazteko erabili diren bezala, laginaren elementu bakoitza oraindik zorizko aldagai dela adierazteko. Kontura zaitez X_1, X_2, …, X_{25} aldagaiak askeak direla, eta X-ren banaketa bera daukatela.

Ondoren, azterketarako hogeita bost pertsona aukeratuko dira. Pertsona horiek pisua galtzeko tratamendupean dauden populaziotik ateratako zorizko lagin bakuna osatzen dute. Orain, laginaren elementu bakoitza pertsona bat da.



4.2 irudia. «Zorizko lagin bakuna» terminoa ulertzeko hiru era. Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994.

Pertsonak identifikatu eta tratamendupean galdu duten pisua erregistratu ondoren, hogeita bost zenbaki ditugu: x_1, x_2, …, x_{25}. Zenbaki horiek X_1, X_2, …, X_{25} zorizko aldagaien behaketak adierazten dituzte, eta zorizko lagin bakuna ere osatzen dute. Ideia hori 4.2 irudian adierazten da grafikoki laburbilduz.

4.3 ESTIMAZIOA

Estimazioa, laginean lorturiko estatistikoen bidez, populazioko parametroak hurbiltzean datza. Estimazioa parametroari balio bat ematean baldin badatza, **puntu-estimazioa** gauzatzen dugu; populazioaren parametroaren balioa tarte baten balioren bat dela estimatzen badugu, **konfiantza-tartea** ematen dugu.

4.3.1 Puntu-estimazioa

Populazioko parametroa hurbiltzeko erabiltzen den estatistikoa **estimatzailea** da. Estimatzaileak lagin zehatz batean hartzen duen zenbakizko balioari parametroaren **estimazio** deritzo. Parametro baten estimatzailea kapela baten bidez adieraziko dugu. Esate baterako, λ parametroaren estimatzailea $\hat{\lambda}$ adieraziko dugu.

Bi kontzeptu horiek, estimatzailea eta estimazioa, ondo bereiztea garrantzizkoa denez, horiek sartzeko adibide erraz bat erabiliko dugu.

4. 2 adibidea

Jo dezagun animalia ugaztunen espezie baterako umaldiko X ume kopurua ezagutu nahi dugula. Aski arrazoizkoa da Poissonen eredura jotzea. Baina ezezaguna den λ parametroaren menpean dago. Beraz, hari buruzko informazioa lortu behar dugu. Horretarako, urtebetean, 35 animalia behatuko ditugu, eta umaldiko ume kopurua zenbatuko dugu. Horrela, 4.2 taula eratuko dugu.

Umaldiko ume kopurua	Animalia kopurua
0	2
1	3
2	10
3	10
4	5
5	0
6	5
7	0

4.2 taula. 4. 2 adibideko datuen maiztasun-taula. Iturria: A. Urkaregi. *Bioestatistika*. UEU, 1991.

λ estimatzeko nola erabili informazio hori? Zenbait aukera dauzkagu:

1. $E(X) = \lambda$ dela badakigu, eta, laginaren batezbestekoa itxaropen teorikoaren balio «analogoa» denez, $\hat{\lambda} = \bar{x} = 2{,}94$ estima dezakegu. λ-ren estimatzailea $\bar{x}$ da, eta estimazioa 2,94.

2. X: $\mathcal{P}(\lambda)$ baldin bada, $P(X=0) = e^{-\lambda} \cdot \frac{\lambda^0}{0!} = e^{-\lambda}$ da. Beraz, $\lambda = -\ln[P(X=0)]$. Gertaeraren probabilitatea eta maiztasun erlatiboa «analogoak» direnez, aurreko berdintza maiztasunen bidez interpretatuz, hauxe da λ-ri dagokion balioa: $\hat{\lambda} = -\ln[h(0)] = -\ln(\frac{2}{35}) = 2{,}86$.
λ-ren estimatzailea $-\ln[h(0)]$ da, eta estimazioa 2,86.

3. $\mathrm{Var}(X) = \lambda$ denez, honako lagin-balio «analogoa» kalkulatuko dugu:
$$\hat{\lambda} = s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 2{,}33.$$
λ-ren estimatzailea s_{n-1}^2 eta estimazioa 2,33 dira.

Arrazoibide kopurua aski luza daiteke. Eta bakoitzak λ-ren balio batera garamatzanez, «hoberena» aukeratu behar dugu. Baina, zer ulertzen dugu «hoberenaz»?

Galdera horri orokorrean edozein parametrotarako erantzun beharrean, kasu zehatz batzuk aztertuko ditugu. Hain zuzen, gai honetan populazioaren itxaropena edo batezbestekoa (μ) eta bariantza (σ^2) estimatzen arituko gara.

4.4 BANAKETAREN BATEZBESTEKOAREN ESTIMAZIOA

4.4.1 Batezbestekoaren puntu-estimazioa

Behaturiko laginaren batezbestekoa erabiltzen da lagina atera dugun populazioaren batez besteko balioa hurbiltzeko. Laginaren batezbestekorako lortutako zenbakizko balioak aldatu egiten dira lagin batetik bestera. Estatistiko hori zorizko aldagaia da, eta horregatik, laginaren batezbestekoa adierazteko, $\bar{X}$ letra maiuskula erabiliko da. Lagina aukeratu ondoren, laginaren batezbestekoaren zenbakizko balioa $\bar{x}$ letra minuzkulaz adieraziko dugu.

4. 3 adibidea

Airearen kalitatea neurtzen duen adierazleetariko bat esekidura-partikulen

batez besteko edukia (mikrogramo/mm^3-tan) da. Hau da, X aldagaiaren μ batezbestekoa interesatzen zaigu, non X m^3 aireko esekidura-partikulen mikrogramo kopurua baita. Egoera hori kontrolatzeko, sei egunetik behin irakurketa bat egiten da, metro kubiko bat aire hartu eta filtro baten bidez, eta bertako esekidura-partikulen edukia neurtzen da. Hogeita hamar eguneko tarte baten ostean, honako bost tamainako zorizko lagin bakun hau sortzen da: X_1, X_2, X_3, X_4, X_5. Demagun 30 egunean aldagai horientzat behaturiko balioak honako hauek direla: x_1= 58; x_2 = 70; x_3 =57; x_4 = 61; x_5 = 59. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

$\overline{X}$ estatistikoaren behaturiko balioa bost balio horien batezbestekoa da.

Alegia, $\bar{x} = \dfrac{58+70+57+61+59}{5} = 61$

Izan bedi X_1, X_2, …, X_n zorizko aldagai askeak eta berdinki banatuak direla, eta bakoitzaren itxaropena μ dela. Beraz, X_1, X_2, …, X_n zorizko aldagaiek μ itxaropena duen zorizko aldagai baten zorizko lagin bakuna osatzen dute. Orduan, $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$, laginaren batezbestekoa **μ-ren puntu-estimatzailea** da. Puntuzkoa esaten diogu, zeren eta lagin bakoitzean balio bakar bat esleitzen baitio μ-ri.

Esate baterako, 4. 3 adibidean, $\overline{X}$ μ-ren estimatzailea da; emandako laginean oinarriturik, 61 μ-ren estimazioa da.

Demagun X_1, X_2… X_n zorizko aldagai askeak eta berdinki banatuak direla, eta bakoitzaren itxaropena μ dela. $\hat{\mu}$ μ-ren **estimatzaile zentratua** edo **alboragabea** da, honako hau betetzen baldin bada: $E(\hat{\mu}) = \mu$.

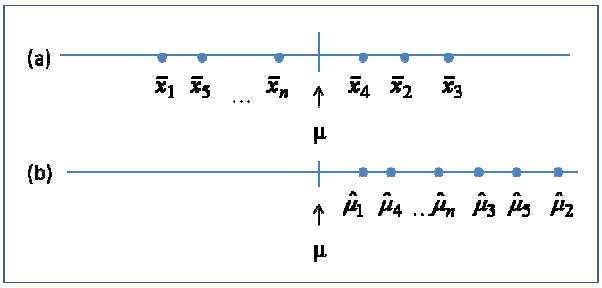
Propietatea. Demagun μ batezbestekoa eta σ^2 bariantza dituen populaziotik ateratako zorizko lagin bakuna $X_1, X_2, ..., X_n$ dela. $\overline{X}$ μ-ren estimatzaile zentratua da.

Froga:

$$E(\overline{X}) = E\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{E(X_1) + E(X_2) + ... + E(X_n)}{n} =$$

$$= \frac{\mu + \mu + ... + \mu}{n} = \frac{n \cdot \mu}{n} = \mu$$

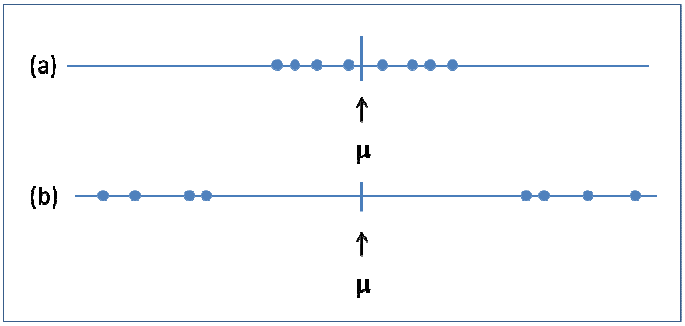
Alborapenaren esanahia erakusteko, 4.3 irudian μ-ren bi estimatzaile agertzen dira: bata alboragabea, $\overline{X}$, eta bestea alboraduna, $\hat{\mu}$.



4.3 irudia. (a), $\overline{X}$ μ-ren estimatzaile alboragabea da. Behaturiko laginen batezbestekoak, $\bar{x}_1, \bar{x}_2, ..., \bar{x}_n$, μ populazioko batezbestekoaren inguruan daude. (b) μ-ren estimatzaile alboratu baten behaturiko balioak dira. Balioak ez daude μ-n zentraturik, ez daude estimatzen ari garen parametroaren inguruan.

$\overline{X}$ estimatzaile alboragabea izatea ez da arrazoi nahikoa μ-ren estimatzaile hoberentzat hartzeko. Estimatzen ari garen parametroaren baliotik hurbil dauden estimazioak lortzea interesatzen zaigu, eta alboragabea izateak ez du hori bermatzen. Konfiantza-maila on batekin hurbiltasun hori ziurtatzeko, kontuan izan behar dugu estimatzailearen bariantza. Esandakoaren garrantzia azpimarratzeko, μ-ren bi estimatzaile erakusten dira 4.4 irudian; bariantza oso ezberdinak dituzte.

Lehen egoera bigarrena baino askoz hobea da, noski. Alegia, parametro baten puntu-estimazio bakar batek parametrotik hurbil dagoen balioa duela ziur egoteko, **estimatzaile alboragabea eta bariantza txikienekoa** erabili beharko genituzke, eta hori da, hain zuzen ere, $\overline{X}$.



4.4 irudia. (a) Bariantza txikidun μ-ren estimatzaile alboragabean behaturiko balioak. Estimazioak, oro har, μ-tik hurbil daude. (b) Bariantza handidun μ-ren estimatzaile alboragabean behaturiko balioak. Estimazio gehienen balioak ez daude, inolaz ere, μ-tik hurbil.

4.4.2 Lagin-batezbestekoaren errore estandarra

Lagin-batezbestekoaren itxaropena μ dela frogatu ondoren, haren bariantzaren eta hasierako aldagaien bariantzaren arteko erlazioa zein den aztertuko dugu.

Propietatea. Demagun X_1, X_2, …, X_n μ batezbestekoa eta σ^2 bariantza dituen populaziotik ateratako zorizko lagin bakuna dela. Orduan, $\overline{X}$ lagin-batezbestekoaren bariantza $\frac{\sigma^2}{n}$ da.

Froga:

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n)}{n^2}$$

$$= \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\overline{X}$ zorizko aldagaiaren desbideratze estandarrari **batezbestekoaren errore estandarra** edo **errore estandarra** deritzo: DE ($\overline{X}$) = $\sqrt{\mathrm{Var}(\overline{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

Beraz, errore estandarra $\frac{\sigma}{\sqrt{n}}$ da, eta, $\frac{s}{\sqrt{n}}$ -ren bidez estimatzen da, non s kuasibariantzaren erro karratua baita: $s_{n-1}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$.

Populazio beretik ateratako n tamainako laginetatik lortutako batezbestekoen aldakortasunaren neurri kualitatiboa da errore estandarra. Errore estandarra zuzenki proportzionala da, bai $\frac{1}{\sqrt{n}}$ -rekiko, eta baita σ behaketa baten populazioaren desbideratze estandarrarekiko ere. Horrek, μ populazioaren batez besteko ezezaguna estimatzerakoan, $\overline{X}$ -ren zehaztasun neurri-gisa laginaren tamaina erabiltzea justifikatzen du. Zenbat eta lagin handiagoa izan, orduan eta estimazio zehatzagoa izango dugu. 4.5 irudian erakutsi da ideia horren adierazpide grafikoa.

4. 4 adibidea

1997ko Osasun Inkestako datuen arabera, EAEko 1.658 pertsona ibili dira paseoan azken bi asteetan. Paseoan zenbat denbora ibiltzen diren aztertu dugu;

4.3 taulan agertu ditugu zoriz aukeraturiko probintzia bakoitzeko hamar pertsonak paseoan emandako denbora (minututan).

	Araba	Bizkaia	Gipuzkoa
1	30	120	120
2	90	120	30
3	60	120	90
4	60	120	60
5	120	60	60
6	180	60	60
7	30	60	120
8	40	30	60
9	30	20	90
10	60	10	45
$\bar{x}$	70	72	73.5
S	48.3	44.67	30.37

4.3 taula. Probintzia bakoitzeko hamar pertsonak paseo batean emandako denbora.

Hogeita hamar pertsonak paseo batean emandako denboraren heina 10 minututik 180 minuturakoa da. Batezbestekoa 71.83 minutukoa da, eta desbideratze estandarra 39,72 minutukoa. Hiru laginetako batez besteko denboren heina 70 minututik 73,5 minuturakoa da, eta desbideratze estandarra 1.43 minutukoa. Hau da, hamar behaketetan oinarritutako batezbestekoak, 1 tamainako laginetako batezbesteko gisa har daitezkeen banakako behaketak baino gutxiago sakabanaturik daude.

Hauxe da Gipuzkoako hamar pertsonak paseo batean emandako denboraren errore estandarraren estimazioa: $\frac{s}{\sqrt{n}} = \frac{30.37}{\sqrt{10}} = 9{,}6$

4.4.3 Lagin-batezbestekoaren banaketa

Lagin-batezbestekoaren itxaropena eta bariantza μ eta $\frac{\sigma^2}{n}$ direla frogatu ondoren, haren itxura zein den aztertuko dugu banaketaren bidez.

Populaziotik zoriz ateratako n tamainako laginetatik abiatuta, $\overline{X}$ zorizko aldagaiak har ditzakeen balio guztien banaketari **$\overline{X}$ -ren lagin-banaketa** deritzo.

Aurreko gaian ikusi dugunez, banaketa normaleko zorizko aldagaien konbinazio linealak banaketa normala du. Demagun μ batezbestekoa eta σ^2 bariantza dituen banaketa normalari darraion populaziotik ateratako zorizko lagin bakuna dela $X_1, X_2, \ldots, X_n$. Hain zuzen, $X_i : N(\mu, \sigma)\ \forall\ i = 1, 2, \ldots, n$. Aipatutakoaren arabera, $\overline{X}$ hasierako aldagaien konbinazio lineala denez, haren banaketa normala izango da. Hots, $\overline{X} : N(\mu, \frac{\sigma}{\sqrt{n}})$ da.

4. 5 adibidea

Demagun paseoan ibilitako denbora banaketa normalari darraiola. 1997ko Osasun Inkestako datuetan oinarrituta, paseoan ibiltzen diren EAEko biztanleek osaturiko populazioaren batez besteko denbora 75 minutu eta desbideratze estandarra 45 minutu direla jotzen dugu. Zer probabilitate dago zoriz aukeraturiko hamar pertsonek paseo batean ibilitako batez besteko denbora ordubete eta ordu eta erdi bitartean egoteko?

Hauxe da $\overline{X}$ -ren banaketa: $N(75, \frac{45}{\sqrt{10}}) = N(75, 14{,}23)$.

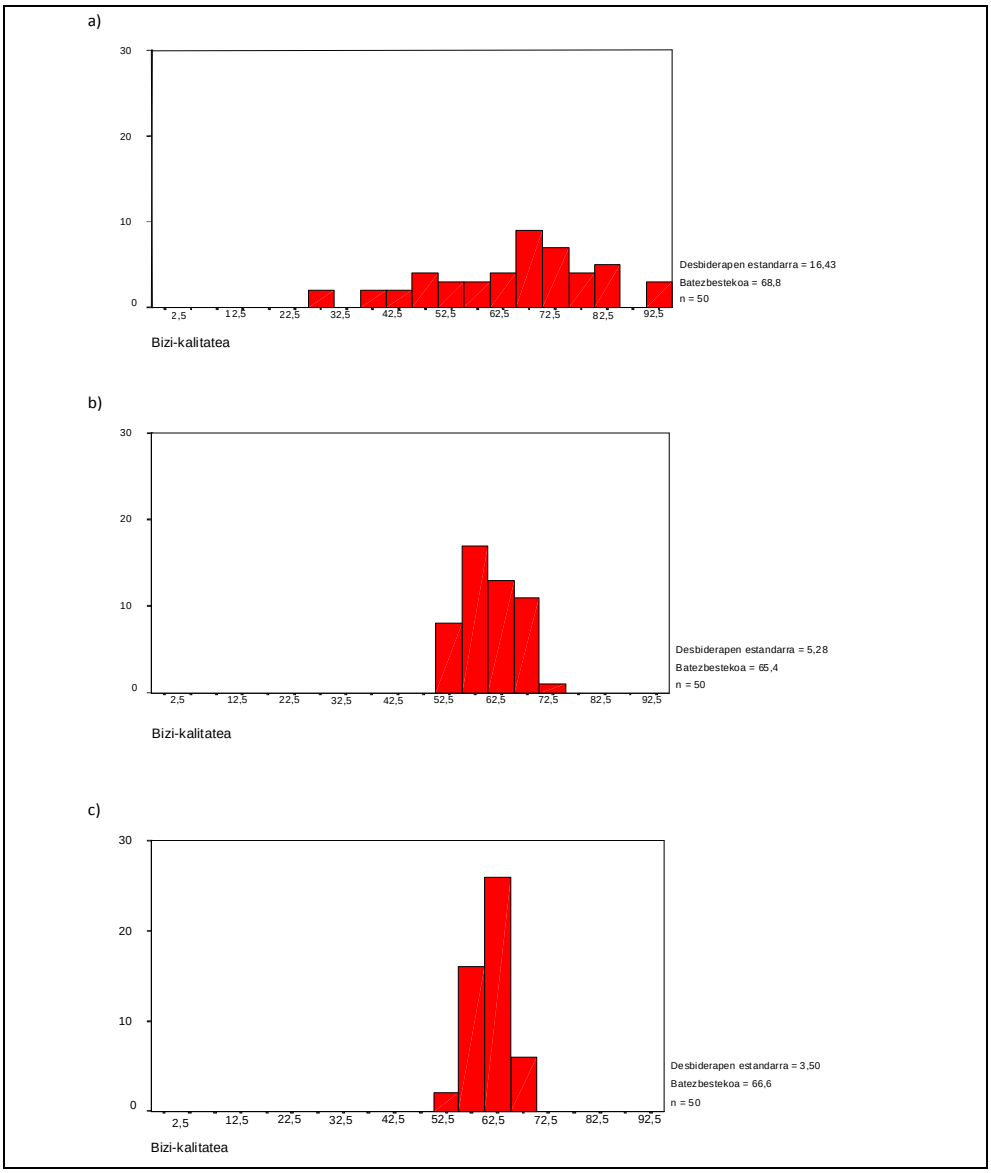
Beraz, $P(60 \le \overline{X} \le 90) = P\left(\frac{60-75}{14{,}23} \le \frac{\overline{X}-75}{14{,}23} \le \frac{90-75}{14{,}23}\right) = P(-1{,}05 \le Z \le 1{,}05)$

$= 1 - 2 \cdot P(Z > 1{,}05) = 1 - 2 \cdot 0{,}1469 = 0{,}7062$

<u>Propietatea</u>. Demagun μ batezbestekoa eta σ^2 bariantza dituen populaziotik ateratako zorizko lagin bakuna dela $X_1, X_2, \ldots, X_n$. Orduan, n handia ($n \ge 30$) denean, $\overline{X}$ -ren banaketa normala da gutxi gorabehera, nahiz eta X-ren banaketa normala ez izan. Hain zuzen, $\overline{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$ da.

<u>Froga</u>: $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{X_1}{n} + \frac{X_2}{n} + \ldots + \frac{X_n}{n}$

Hau da, berdinki banatuak eta askeak diren zorizko aldagaien arteko batura gisa adieraz daiteke $\overline{X}$. Limitearen Teorema Zentrala aplikatuz, $n \ge 30$ denez, $\overline{X}$ -ren banaketa gutxi gorabehera normala dela ziurta daiteke; alegia, $\overline{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$.



4.5 irudia. Batezbestekoaren errore estandarraren ilustrazioa, laginaren tamainaren arabera. 1997ko Osasun Inkestaren arabera, EAEko pertsonen bizi-kalitatearen osasun orokorra adierazten duen indizea agertzen da: a) 1 tamainako 50 laginetan; b) 10 tamainako 50 laginetan; c) 50 tamainako 50 laginetan.

4. 6 adibidea

1997ko Osasun Inkestan, azken bi asteetan pertsona bakoitza ibili den aldi kopurua aztertu da. Aldagai horren batezbestekoa 8 da, eta desbideratze estandarra 5. EAEko berrogei pertsona zoriz aukeratzen baditugu, zer probabilitate dago haiek azken bi asteetan ibili diren batez besteko aldi kopurua hamar baino handiagoa izateko?

X aldagaiaren banaketa ezezaguna da, baina n = 40 denez, Limitearen Teorema Zentrala aplikatuz, badakigu $\overline{X}$-ren banaketa $N(8, \frac{5}{\sqrt{40}}) = N(8, 0{,}79)$ dela.

Beraz, $P(\overline{X} > 10) = P\left(\frac{\overline{X}-8}{0{,}79} > \frac{10-8}{0{,}79}\right) = P(Z > 2{,}53) = 0{,}0057$

4.4.4 Batezbestekorako konfiantza-tartea (bariantza ezaguna)

Populazioaren batezbestekoaren gutxi gorabeherako balioaren informazioa ez ezik haren zehaztasunari buruzkoa ere lortzeko, konfiantza-tartearen metodoa azalduko dugu.

% 95eko konfiantza-mailan (L_1, L_2), μ-ren **tarte estimatzailea** zorizko tartea da, non L_1 eta L_2 honako baldintza hau betetzen duten estatistikoak baitira: $P(L_1 < \mu < L_2) = 0{,}95$

Lagin batean ateratako balioetatik abiatu, eta zenbakizko tartea lortzen da.

Demagun μ batezbestekoa eta σ^2 bariantza ezaguna dituen populaziotik ateratako zorizko lagin bakuna dela $X_1, X_2, \ldots, X_n$. Orduan, **% 95eko μ-ren konfiantza-tartea (KT)** $I_\mu^{0,95}$ adieraziko dugu, eta hauxe da: $I_\mu^{0,95} = \left(\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right)$.

Hain zuzen, $\overline{X}$ lagin-batezbestekoaren lagin-banaketa $\overline{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$ dela ikusi dugu.

Beraz, aldagaia tipifikatuz, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx N(0, 1)$ lortu da.

A eranskineko 3.taulan ikus daitekeenez, Z: $N(0, 1)$ den zorizko aldagaiarentzat ondokoa betetzen da, $P(-1{,}96 < Z < 1{,}96) = 0{,}95$. Orduan, Z-ren ordez, batezbesteko tipifikatua erabiliz, honako hau daukagu:

$$0{,}95 = P\left(-1{,}96 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < 1{,}96\right) = P\left(-1{,}96\cdot\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < 1{,}96\cdot\frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-\overline{X}-1{,}96\cdot\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X}+1{,}96\cdot\frac{\sigma}{\sqrt{n}}\right) = P\left(\overline{X}-1{,}96\cdot\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+1{,}96\cdot\frac{\sigma}{\sqrt{n}}\right)$$

Hau da, (L_1, L_2) tartea lortu dugu, non $P(L_1 < \mu < L_2) = 0{,}95$ baldintza betetzen baitu:

$$L_1 = \overline{X} - 1{,}96\cdot\frac{\sigma}{\sqrt{n}} \quad \text{eta} \quad L_2 = \overline{X} + 1{,}96\cdot\frac{\sigma}{\sqrt{n}}$$

Demagun x_1, x_2, …, x_n zorizko lagin bakuna dela; $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$ laginbatezbestekoa kalkulatuz eta aurreko adierazpeneko behe- eta goi-muturretan ordezkatuz, % 95eko μ-ren konfiantza-tartea lortuko dugu. Hau da, $I_\mu^{0,95} = \left(\bar{x} - 1{,}96\cdot\frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96\cdot\frac{\sigma}{\sqrt{n}}\right)$.

Konfiantza-tartearen interpretazioa. Ezin da esan μ parametroa % 95eko konfiantza-tarte zehatz batean egoteko probabilitatea 0,95ekoa denik. Aldiz, n tamainako laginketan oinarriturik atera daitezkeen % 95eko konfiantza-tarteen % 95ak μ bere barnean duela esan behar da.

4. 7 adibidea

EAEko biztanleak paseoan ibiltzen diren batez besteko denbora kalkulatu nahi dugu % 95eko konfiantza-mailaz, 4.3 taulan agertzen diren datuetan oinarrituz. Denbora horren desbideratze estandarra 45 minutukoa dela suposatzen dugu.

Taulako datuak lagin bakartzat hartuta, honako tarte hau lortuko dugu:

$$\left(\bar{x} - 1{,}96\cdot\frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96\cdot\frac{\sigma}{\sqrt{n}}\right) = \left(71{,}83 - 1{,}96\cdot\frac{45}{\sqrt{30}}, 71{,}83 + 1{,}96\cdot\frac{45}{\sqrt{30}}\right) =$$

$$= (71{,}83 - 16{,}1,\ 71{,}83 + 16{,}1) = (55{,}73,\ 87{,}93)$$

Datuak probintziaka hartzen baditugu, hauxe da Arabako biztanleentzat batez besteko denborari dagokion tartea:

$$\left(\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(70 - 1{,}96 \cdot \frac{45}{\sqrt{10}}, 70 + 1{,}96 \cdot \frac{45}{\sqrt{10}}\right) =$$

= (70 – 27,89, 70 + 27,89) = (42,11, 97,89)

Bizkaiko biztanleentzat:

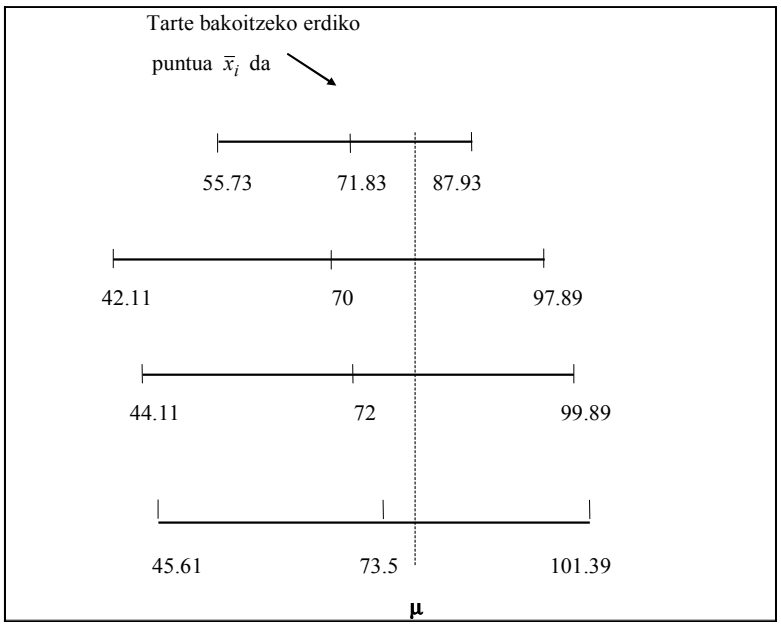
$$\left(\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(72 - 1{,}96 \cdot \frac{45}{\sqrt{10}}, 72 + 1{,}96 \cdot \frac{45}{\sqrt{10}}\right) =$$

= (72 – 27,89, 72 + 27,89) = (44,11, 99,89)

Eta Gipuzkoako biztanleentzat:

$$\left(\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(73{,}5 - 1{,}96 \cdot \frac{45}{\sqrt{10}}, 73{,}5 + 1{,}96 \cdot \frac{45}{\sqrt{10}}\right) =$$

= (73.5 – 27,89 , 73,5 + 27,89) = (45,61 , 101,39)



4.6 irudia. 4.3 taulako lau laginetan oinarriturik, μ populazioko batezbestekorako % 95eko konfiantza-tarteak.

4.6 irudian erakusten da konfiantza-tarte hori probintzia (lagin) bakoitzerako bat dela. μ-ren balioa adierazten duen irudizko lerroa jarri dugu. Hamar tamainako hainbat laginetan zehar, μ parametroa % 95eko konfiantza-tarteen % 95aren barnean egongo da.

Defini dezagun orain edozein konfiantza-mailako μ-ren konfiantza-tartea. Demagun μ batezbestekoa eta σ^2 bariantza ezaguna dituen populaziotik ateratako zorizko lagin bakuna dela X_1, X_2, …, X_n. Orduan, **%(1−α)×100eko μ-ren konfiantza-tartea (KT)** $I_\mu^{1-\alpha}$ adieraziko dugu, eta hauxe da: $I_\mu^{1-\alpha} = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$

Tarte hori populazioa normala denean, edo, Limitearen Teorema Zentrala aplikatuz, laginaren tamaina $n \geq 30$ denean da baliozkoa.

4. 8 adibidea

EAEko 1997ko Osasun Inkestako datuetan oinarrituta, paseoan ibiltzen diren Bizkaiko biztanleek osaturiko populazioa hartuko dugu. Badakigu populazio horrek paseo batean emandako denboraren desbideratze estandarra 46 minutukoa dela. 4.3 taulan agertzen diren datuetan oinarrituz, kalkula ezazu Bizkaiko biztanleek paseo batean emandako batez besteko denborarako % 99ko konfiantza-tartea.

$$I_\mu^{0,99} = \left(\bar{x} - z_{0,005} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0,005} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(72 - 2{,}576 \cdot \frac{46}{\sqrt{10}}, 72 + 2{,}576 \cdot \frac{46}{\sqrt{10}}\right)$$

= = (72 – 37,47 , 72 + 37,47) = (34,53 , 109,47)

<u>Konfiantza-tartearen luzeran eragina duten faktoreak</u>: **%**(1−α)×100eko μ-ren konfiantza-tartearen luzera $2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ da, eta α konfiantza-mailak eta n laginaren tamainak zehazten dute. Hain zuzen,

- Zenbat eta n lagin-tamaina handiagoa izan, orduan eta konfiantza-tarte laburragoa.
- Zenbat eta konfiantza-maila handiagoa (α txikiagoa) izan, orduan eta konfiantza-tarte luzeagoa.

4. 7 eta 4. 8 adibideetan, Bizkaiko biztanleek paseo batean emandako batez besteko denborarako % 95eko eta % 99ko konfiantza-tarteak kalkulatu ditugu. Lortutako bi tarteak $I_\mu^{0,95}$ = (44,11 , 99,89) eta $I_\mu^{0,99}$ = (34,53 , 109,47) dira. Zenbat eta konfiantza-maila handiagoa izan, orduan eta konfiantza-tarte luzeagoa dela ikus daiteke.

4. 9 adibidea

4. 5 adibidean azaldu dugunez, EAEko biztanleek paseo batean batez beste 75 minutu ematen dituztela jotzen dugu, eta desbideratze estandarra 45 minutu

dela. 1997ko Osasun Inkestako datuen arabera, 181 arabar ibili dira paseoan azken bi asteetan, eta paseo bakoitzean 71 minutu eman dituzte batez beste. Paseoari dagokionez, arabarren portaera EAEko biztanle guztien adierazgarri al da, % 95eko adierazgarritasun-mailaz?

$$I_{\mu}^{0,95} = \left(\bar{x} - z_{0,025} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0,025} \cdot \frac{\sigma}{\sqrt{n}} \right) = \left(71 - 1{,}96 \cdot \frac{45}{\sqrt{181}}, 71 + 1{,}96 \cdot \frac{45}{\sqrt{181}} \right) =$$

= (71 – 6,56 , 71 + 6,6) = (64,44 , 77,56)

$\mu = 75 \in (64{,}44 \ , \ 77{,}56)$ denez, arabarren portaera, paseoari dagokionez, EAEko biztanle guztien adierazgarri dela onar dezakegu, % 95eko adierazgarritasun-mailaz.

4.4.5 Laginaren tamainaren kalkulua, bariantza ezaguneko populazioaren batezbestekorako konfiantza-tartearen luzeran oinarriturik

Demagun σ^2 ezaguna duen populazioko batezbestekoa estimatu nahi dugula, **%**$(1-\alpha)\times 100$eko konfiantza-tartearen luzera L izateak baldintzatuz. Baldintza hori bete dezan, zer tamainatakoa izan behar du laginak?

$L = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ betetzea nahi baldin badugu, n isolatuz, $n = 4 \cdot z_{\alpha/2}^2 \cdot \frac{\sigma^2}{L^2}$ izan beharko da.

4. 10 adibidea

1997ko Osasun Inkestaren datuetan oinarrituta, EAEko biztanleen adinaren desbideratze estandarra hemeretzi urtekoa dela jotzen dugu. EAEko biztanleen batez besteko adinaren % 95eko konfiantza-tartearen luzera hamar urtekoa izan dadin, zenbat biztanle aukeratu behar ditugu?

$$n = 4 \cdot z_{\alpha/2}^2 \cdot \frac{19^2}{10^2} = 4 \cdot 1{,}96^2 \cdot \frac{19^2}{10^2} = 45{,}28 \approx 46 \text{ biztanle}$$

Orain arte, bariantza ezaguna daukan banaketaren batezbestekorako konfiantza-tartearen kalkulua aztertu dugu. Baina egoera hori artifizial samarra da, zeren eta populazioaren bariantza ez baita gehienean ezagutzen, eta estimatu behar izaten baita.

4.5 BANAKETAREN BARIANTZAREN ESTIMAZIOA

Batezbestekorako aztertu dugun era berean aztertuko dugu populazioko bariantza estimatzearen arazoa.

4.5.1 Bariantzaren puntu-estimazioa

Behaturiko kuasibariantza erabiltzen da lagina atera dugun populazioko bariantzaren balioa hurbiltzeko. Kuasibariantzarako lortutako zenbakizko balioak aldatu egiten dira lagin batetik bestera. Estatistiko hori zorizko aldagaia da, eta, horregatik, kuasibariantza adierazteko S^2 letra maiuskula erabiliko da. Lagina aukeratu ondoren, kuasibariantzaren zenbakizko balioa s^2 letra minuskulaz adieraziko dugu. Alegia, demagun $X_1, X_2, \ldots, X_n$ μ batezbestekoa eta σ^2 bariantza dituen populaziotik ateratako zorizko lagin bakuna dela; orduan, kuasibariantza $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ da.

Izan bitez $X_1, X_2, \ldots, X_n$ zorizko aldagai askeak eta berdinki banatuak, bakoitzaren itxaropena eta bariantza μ eta σ^2 izanik, hurrenez hurren. Beraz, $X_1, X_2, \ldots, X_n$ zorizko aldagaiek μ eta σ^2 itxaropen eta bariantza duten zorizko aldagai baten zorizko lagin bakuna osatzen dute. Orduan, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ kuasibariantza **σ^2-ren puntu estimatzailea** da.

Estimatzaile horren banaketa ezagutzeko, banaketa berri bat definitu behar dugu: khi karratu banaketa, hain zuzen.

4.5.2 Pearson-en khi karratu banaketa

Demagun $Z_1, Z_2, \ldots, Z_n$ $N(0,1)$ banaketari darraizkion n zorizko aldagai askeak direla. Orduan, $\chi_n^2 = Z_1^2 + Z_2^2 + \ldots + Z_n^2$ zorizko aldagaia ***n* askatasun-graduko khi karratuaren banaketari** darraio.

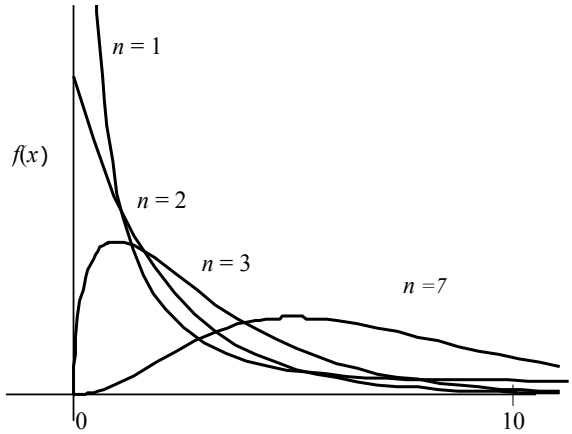
<u>Propietateak</u>:

1. Khi karratu banaketa $[0, +\infty)$ tartean definiturik dago.
2. Banaketa n askatasun-graduen kopuruaren pean dago. 4.7 irudian erakusten da askatasun-graduek banaketaren itxuran duten eragina.
3. Khi karratu banaketa jarraitua da definizio-tarte osoan.
4. Khi karratu banaketak maximoa dauka $x = n - 2$ puntuan, $n \geq 2$ suposaturik.

5. Khi karratu banaketa ez da simetrikoa.

6. $E(\chi^2_n) = n$ eta $Var(\chi^2_n) = 2n$.

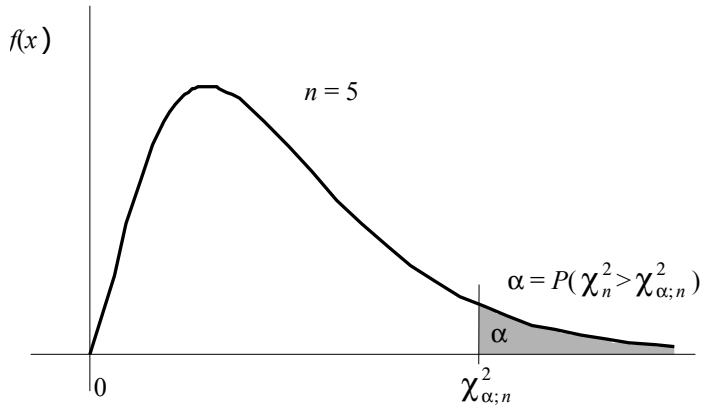
7. Baturaren teorema. Demagun $\chi^2_{n_1}$, $\chi^2_{n_2}$, …, $\chi^2_{n_k}$, hurrenez hurren, n_1, n_2, …, n_k askatasun-graduko khi karratuaren banaketari darraizkion k zorizko aldagai askeak direla. Orduan, beren batura $n_1 + n_2 + \ldots + n_k$ askatasun-graduko khi karratu zorizko aldagaia izango da. Hots, $\chi^2_{n_1} + \chi^2_{n_2} + \ldots + \chi^2_{n_k} = \chi^2_n$, non $n = n_1 + n_2 + \ldots + n_k$ baita.



4.7 irudia. n-ren balio ezberdinentzat χ^2_n hainbat banaketaren adierazpide grafikoak.

***n* askatasun-graduko khi karratuaren banaketaren balio kritikoa** deritzo n askatasun-graduko khi karratuaren banaketaren dentsitate-funtzioaren eta abzisa-ardatzaren bitartean eskuinerantz aurrefinkaturiko azalera uzten duen balioari (banaketa-funtzioaren balioaren osagarria da). Aurrefinkaturiko azalera hori α baldin bada, balio kritikoa $\chi^2_{\alpha,n}$ adieraziko da, hots, $P(\chi^2_n > \chi^2_{\alpha,n}) = \alpha$. Grafikoki 4.8 irudian erakusten da erlazio hori. *A* eranskineko 5. taulan agertzen dira α azalerak eta $\chi^2_{\alpha,n}$ balio kritikoak.

Oharra. n emendatzerakoan, $n \to \infty$, khi karratu banaketaren grafikoa normalarenera hurbilduz doa. Bereziki, $n > 30$ denean, honako hurbilketa hau erabiltzen da: $\chi^2_{\alpha,n} = \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$



4.8 irudia. n askatasun-graduko khi karratuaren banaketaren $\chi^2_{\alpha,\,n}$ balio kritikoaren adierazpide grafikoa.

4. 11 adibidea

Khi karratu banaketaren taulak erabiliz, honako balio hauek kalkulatuko ditugu.

a) $P(\chi^2_{10} \geq 3{,}25) = 0{,}975$

b) $P(\chi^2_{16} \leq 9{,}3) = 1 - P(\chi^2_{16} \geq 9{,}3) = 1 - 0{,}9 = 0{,}1$

c) $P(\chi^2_{21} \geq a) = 0{,}025 \Rightarrow a = \chi^2_{0{,}025\,,21} = 35{,}479$

d) $P(\chi^2_{13} \leq b) = 0{,}1 \Rightarrow P(\chi^2_{13} > b) = 1 - 0{,}1 = 0{,}9 \Rightarrow b = \chi^2_{0{,}9\,,\,13} = 7{,}042$

e) $P(c \leq \chi^2_{24} \leq d) = 0{,}95 \Rightarrow$ Aukera bat (ez bakarra) honako hau da:

$c = \chi^2_{0{,}975\,,\,24} = 12{,}401$ eta $d = \chi^2_{0{,}025\,,\,24} = 39{,}364$

f) $P(\chi^2_{50} \geq e) = 0{,}95 \Rightarrow e = \chi^2_{0{,}95\,,\,50} = \frac{1}{2}\left(z_{0{,}95} + \sqrt{2 \cdot 50 - 1}\right)^2 =$

$\frac{1}{2}\left(-1{,}645 + \sqrt{99}\right)^2 = 34{,}49$

g) $P(\chi^2_{50} \geq 32) = \alpha \Rightarrow \chi^2_{\alpha,50} = 32 \Rightarrow \frac{1}{2}\left(z_\alpha + \sqrt{2 \cdot 50 - 1}\right)^2 = 32 \Rightarrow z_\alpha =$

$\sqrt{2 \cdot 32} - \sqrt{99} = -1{,}95 \Rightarrow \alpha = P(Z \geq -1{,}95) = 1 - P(Z \geq 1{,}95) =$

$= 1 - 0{,}0256 = 0{,}97$

4.5.3 S^2 kuasibariantzaren banaketa

Populaziotik zoriz ateratako n tamainako laginetatik abiatuz, **S^2-ren lagin-banaketa** deritzo S^2 zorizko aldagaiak har ditzakeen balio guztien banaketari,.

Demagun $X_1, X_2, \ldots, X_n$ μ batezbestekoa eta σ^2 bariantza dituen eta banaketa normalari darraion populaziotik ateratako zorizko lagin bakuna dela. Orduan, S^2 kuasibariantzaren banaketa $\frac{(n-1)\cdot S^2}{\sigma^2} : \chi^2_{n-1}$ adierazpenak ematen du.

Propietate hori kontuan hartuz, S^2 kuasibariantza σ^2-ren estimatzaile alboragabea dela froga dezakegu. Hots, $E(S^2) = \sigma^2$ betetzen dela.

$$E(\chi^2_{n-1}) = n-1 \Rightarrow E\left(\frac{(n-1)\cdot S^2}{\sigma^2}\right) = n-1 \Rightarrow \frac{(n-1)\cdot E\left(S^2\right)}{\sigma^2} = n-1 \Rightarrow E(S^2) = \sigma^2$$

S^2 kuasibariantza σ^2-ren estimatzaile alboragabea izatean datza populazioaren σ^2 bariantza estimatzeko kuasibariantza erabiltzea, eta ez laginaren bariantza, estimatzaile alboratua baita.

4.5.4 Populazio normaleko bariantzarako konfiantza-tartea

Batezbestekorako egin dugun era berean, populazioaren bariantzaren gutxi gorabeherako balioaren informazioaz gainera, konfiantza-tartearen metodoa ere erabiliko dugu bariantza horren zehaztasunari buruzko informazioa lortzeko.

Demagun μ batezbestekoa eta σ^2 bariantza dituen populazio normaletik ateratako n tamainako zorizko lagin bakuna dela $X_1, X_2, \ldots, X_n$. Populazio horren **bariantzarako %(1−α)×100 mailako konfiantza-tartea (KT)** honako tarte honen bidez definitzen da:

$$I_{\sigma^2}^{1-\alpha} = \left(\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,(n-1)}}, \frac{(n-1)\cdot s^2}{\chi^2_{(1-\alpha/2),(n-1)}}\right).$$

Hain zuzen, χ^2_n banaketarentzat honako hau betetzen da: $P(\chi^2_{(1-\alpha/2),n} < \chi^2_n < \chi^2_{\alpha/2,n})$ $= 1-\alpha$. Beraz, kuasibariantzaren banaketa bertan ordezkatuz, hauxe dugu:

$$1-\alpha =$$

$$P\left(\chi^2_{(1-\alpha/2),(n-1)} < \frac{(n-1)\cdot S^2}{\sigma^2} < \chi^2_{\alpha/2,(n-1)}\right) = P\left(\frac{\chi^2_{(1-\alpha/2),(n-1)}}{(n-1)\cdot S^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{\alpha/2,(n-1)}}{(n-1)\cdot S^2}\right) =$$

$$= P\left(\frac{(n-1)\cdot S^2}{\chi^2_{\alpha/2,(n-1)}} < \sigma^2 < \frac{(n-1)\cdot S^2}{\chi^2_{(1-\alpha/2),(n-1)}}\right)$$

Demagun x_1, x_2, …, x_n zorizko lagin bakuna dela; bere kuasibariantza, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, kalkulatuz eta aurreko adierazpeneko behe- eta goi-muturretan ordezkatuz, %(1−α)×100eko σ^2-ren konfiantza-tartea lortuko dugu. Hau da,

$$I_{\sigma^2}^{1-\alpha} = \left(\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,(n-1)}}, \frac{(n-1)\cdot s^2}{\chi^2_{(1-\alpha/2),(n-1)}}\right)$$

Honako tarte honen bidez definitzen da populazio horren **desbideratze estandarrerako %(1−α)×100 mailako konfiantza-tartea (KT)**:

$$I_{\sigma}^{1-\alpha} = \left(\sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,(n-1)}}}, \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{(1-\alpha/2),(n-1)}}}\right).$$

4. 12 adibidea

1997ko Osasun Inkestaren datuetan oinarrituta, azken bi asteetan EAEn txirrinduz ibili diren hogei pertsona aukeratu ditugu zoriz. Pertsona horiek txirrinduz ibilitako denborak (minututan) honako hauek izan dira: 180 – 15 – 60 – 30 – 15 – 30 – 90 – 40 – 45 – 60 – 300 – 15 – 60 – 30 – 60 – 60 – 90 – 30 – 180 – 60.

Txirrinduz ibilitako denbora-banaketa normalari darraiola suposatuz, estima ezazu % 90eko konfiantza-mailaz EAEko biztanleek txirrinduz ibilitako denboraren desbideratze estandarra.

Laginaren tamaina hogei denez, askatasun-graduak hemeretzi dira. Khi karratu banaketaren taulan $P(10{,}117 \le \chi^2_{19} \le 30{,}144) = 0{,}90$ dela ikusiko dugu.

Datu horietarako $\sum x_i = 1.450$ eta $\sum x_i^2 = 200.500$ dira. Beraz, $s^2 = 5.019{,}74$ da.

Kasu horretan, honako hau da khi karratu zorizko aldagaia:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{19 \cdot S^2}{\sigma^2} \cdot \chi^2_{19}.$$

Beraz, dakiguna hauxe da: 0,90 = $P(10{,}117 \leq \frac{19 \cdot S^2}{\sigma^2} \leq 30{,}144)$ =

$$P\left(\frac{10{,}117}{19 \cdot S^2} \leq \frac{1}{\sigma^2} \leq \frac{30{,}144}{19 \cdot S^2}\right) = P\left(\frac{19 \cdot S^2}{30{,}144} \leq \sigma^2 \leq \frac{19 \cdot S^2}{10{,}117}\right)$$

Lagin honetan, s^2 = 5019,74 denez, orduan, honako hau da bariantzaren % 90eko konfiantza-tartea:

$$I^{0,90}_{\sigma^2} = \left(\frac{19 \cdot 5019{,}74}{30{,}144}, \frac{19 \cdot 5019{,}74}{10{,}117}\right) = (3.163{,}98,\ 9.427{,}2).$$

Beraz, desbideratze estandarraren % 90eko konfiantza-tartea (56,25 , 97,09) da.

4.6 BARIANTZA EZEZAGUNEKO POPULAZIO NORMALAREN LAGIN-BATEZBESTEKOAREN ESTIMAZIOA

Demagun μ batezbestekoa eta σ^2 bariantza dituen populaziotik ateratako zorizko lagin bakuna dela X_1, X_2, …, X_n. Orduan, σ ezezaguna denean, S, kuasibariantzaren erro karratuaren bidez hurbil dezakegu bere balioa. Arazoa, kasu horretan, $\frac{\overline{X} - \mu}{S / \sqrt{n}}$ kantitatearen banaketa normala ez izatean datza.

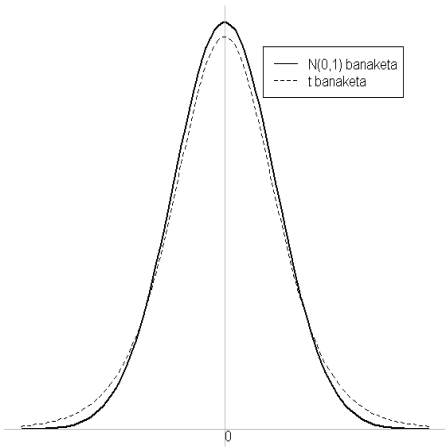
4.6.1 Student-en *t* banaketa

Arazo hori «ikasle» (*student*) deritzon William Gosset izeneko estatistikari batek konpondu zuen 1908. urtean. Hain zuzen, hori izan zen $\frac{\overline{X} - \mu}{S / \sqrt{n}}$ kantitatearen banaketari **Student-en *t*** deitzeko arrazoia.

Demagun Z: $N(0,1)$ eta χ^2_n n askatasun-graduko khi karratu zorizko aldagai askeak direla. Orduan, ***n* askatasun-graduko Student-en *t_n*** deritzo honako era honetara eraikitzen den zorizko aldagaiaren banaketari : $\dfrac{Z}{\sqrt{\chi^2_n/n}} = t_n$

<u>Propietateak</u>:

1. $E(t_n) = 0$
2. $\mathrm{Var}(t_n) = \dfrac{n}{n-2} > 1 \quad (n > 2)$ eta $\mathrm{Var}(t_n) \xrightarrow[n \to \infty]{} 1$
3. Student-en t banaketak $(-\infty, +\infty)$ tartean hartzen ditu balioak.
4. Student-en t banaketa batezbestekoarekiko ($t = 0$) simetrikoa da eta kanpai-itxura dauka.
5. Banaketa normalarekin konparatuz, t banaketa erdialdean normal tipifikatua baino baxuagoa da, eta isatsetan altuagoa. 4.9 irudian erakusten da banaketa normalaren eta t banaketaren arteko erlazioa.

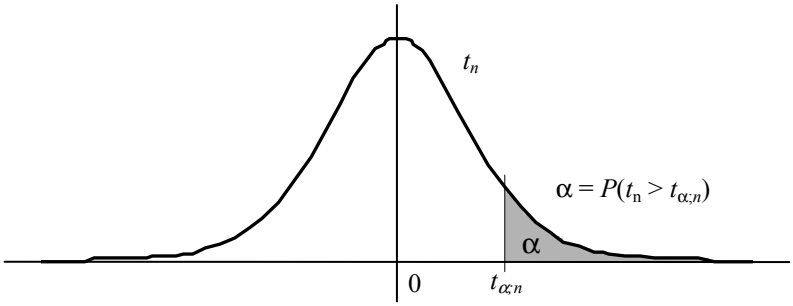


4.9 irudia. n askatasun-graduko Student-en tn banaketaren eta Z: $N(0,1)$ banaketa normalaren arteko konparaketa.

6. $n \geq 30$ denean, $t_n \approx Z$: $N(0,1)$ da.

n askatasun-graduko Student-en t_n banaketaren dentsitate-funtzioaren eta abzisa-ardatzaren artean eskuinerantz aurrefinkaturiko azalera uzten duen balioari ***n* askatasun-graduko Student-en *t* banaketaren balio kritikoa** deritzo (banaketa-funtzioaren balioaren osagarria da). Aurrefinkaturiko azalera hori α baldin bada, balio kritikoa $t_{\alpha,n}$

adierazten da, hots, $P(t_n > t_{\alpha n}) = \alpha$. Grafikoki 4.10 irudian erakusten da erlazio hori. *A* eranskineko 6. taulan, α azalerak eta $t_{\alpha,n}$ balio kritikoak agertzen dira.



4.10 irudia. *n* askatasun-graduko Student-en t_n banaketaren $t_{\alpha,n}$ *n* balio kritikoaren adierazpide grafikoa.

4. 13 adibidea

n askatasun-graduko t_n banaketaren taulak erabiliz, honako balio hauek kalkulatuko ditugu:

a) $P(t_{10} \leq 2{,}8) = 1 - P(t_{10} \geq 2{,}8) = 1 - 0{,}01 = 0{,}99$

b) $P(t_8 > -1{,}4) = 1 - P(t_8 \leq -1{,}4) = 1 - P(t_8 \geq 1{,}4) = 1 - 0{,}1 = 0{,}9$

c) $P(t_{10} \leq a) = 0{,}90 \Rightarrow P(t_{10} > a) = 0{,}10 \Rightarrow a = 1{,}372$

d) $P(t_{18} \leq b) = 0{,}20 \Rightarrow P(t_{18} > b) = 0{,}80 \Rightarrow P(t_{18} > -b) = 0{,}20 \Rightarrow -b = 0{,}862 \Rightarrow b = -0{,}862$

e) $P(-d \leq t_{10} \leq d) = 0{,}95 \Rightarrow P(t_{10} \geq d) = 0{,}025 \Rightarrow d = 2{,}228$

f) $P(t_{80} \leq 2{,}8) \approx P(Z \leq 2{,}8) = 1 - P(Z > 2{,}8) = 1 - 0{,}00256 = 0{,}9974$

4.6.2 Bariantza ezezaguneko populazio normalaren lagin-batezbestekoaren banaketa

Propietatea. Izan bitez $X_1, X_2, \ldots, X_n$ $N(\mu, \sigma)$ eta askeak diren zorizko aldagaiak. Orduan, $\dfrac{\overline{X} - \mu}{S / \sqrt{n}}$ adierazpenaren banaketa $(n - 1)$ askatasun-graduko Student-en *t* banaketa da.

Froga:

σ ezaguna izan bazen, $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} = Z$: $N(0,1)$.

Beste aldetik, $\dfrac{(n-1)\cdot S^2}{\sigma^2} = \chi^2_{n-1}$. Hots, $\dfrac{\chi^2_{n-1}}{n-1} = \dfrac{S^2}{\sigma^2} \Leftrightarrow \sqrt{\dfrac{\chi^2_{n-1}}{n-1}} = \dfrac{S}{\sigma}$.

S eta $\overline{X}$ zorizko aldagai askeak direla kontuan izanik, honako hau dugu:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} : \frac{S}{\sigma} = \frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}} = t_{n-1}$$

4.6.3 Bariantza ezezaguneko populazio normalaren batezbestekorako konfiantza-tartea

Demagun μ batezbestekoa eta σ^2 bariantza ezezaguna dituen populazio normaletik ateratako zorizko lagin bakuna dela X_1, X_2, …, X_n. Hauxe da populazio horren μ **batezbestekorako %(1−α)×100 mailako konfiantza-tartea (KT)**:

$$I_\mu^{1-\alpha} = \left(\bar{x} - t_{\alpha/2;(n-1)} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2;(n-1)} \cdot \frac{s}{\sqrt{n}}\right)$$

Hain zuzen, Student-en t banaketak honako hau betetzen du: $1 - \alpha = P(-t_{\alpha/2;(n-1)} < t_{n-1} < t_{\alpha/2;(n-1)})$.

Ekuazio horretan, hauxe dugu, bariantza ezezaguna daukan populazio normalaren $\overline{X}$ lagin-batezbestekoaren banaketaren adierazpena ordezkatuz:

$$1-\alpha = P\left(-t_{\alpha/2;(n-1)} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2;(n-1)}\right) =$$

$$= P\left(-t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}} < \overline{X}-\mu < t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}}\right) =$$

$$= P\left(-\overline{X} - t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}} < -\mu < -\overline{X} + t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}}\right) =$$

$$= P\left(\overline{X} - t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}}\right)$$

Hots, $P(L_1 < \mu < L_2) = 1 - \alpha$ adierazpena betetzen duen (L_1, L_2) tartea lortu dugu:

$$L_1 = \overline{X} - t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}} \quad \text{eta} \quad L_2 = \overline{X} + t_{\alpha/2;(n-1)} \cdot \frac{S}{\sqrt{n}}$$

Demagun $x_1, x_2, \ldots, x_n$ zorizko lagin bakuna dela; bere lagin-batezbestekoa, $\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$ eta kuasibariantza $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ kalkulatu ondoren, aurreko adierazpeneko behe- eta goi-muturretan ordezkatuz, % 95eko μ-ren konfiantza-tartea lortuko dugu. Alegia, $I_\mu^{1-\alpha} = \left(\overline{x} - t_{\alpha/2;(n-1)} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2;(n-1)} \cdot \frac{s}{\sqrt{n}}\right)$.

4. 14 adibidea

Era ezazu Gipuzkoako biztanleek paseo batean ematen duten batez besteko denborarako % 95eko konfiantza-tartea, 4.3 taularen datuetan oinarrituta.

$$I_\mu^{0.95} = \left(\overline{x} - t_{\alpha/2;(n-1)} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2;(n-1)} \cdot \frac{s}{\sqrt{n}}\right) =$$

$$\left(73{,}5 - t_{0{,}025;9}\frac{30{,}37}{\sqrt{10}}, 73{,}5 + t_{0{,}025;9}\frac{30{,}370}{\sqrt{10}}\right) =$$

$$= \left(73{,}5 - 2{,}262\frac{30{,}37}{\sqrt{10}}, 73{,}5 + 2{,}262\frac{30{,}370}{\sqrt{10}}\right) = =$$

$$= (73{,}5 - 46{,}97,\ 73{,}5 + 46{,}97) = (26{,}53,\ 120{,}47)$$

<u>Oharra</u>. Aurreko atalean esandakoaren arabera, $\frac{\overline{X} - \mu}{S/\sqrt{n}}$: t_{n-1} da. Baina Student-en t banaketaren 6. propietatea erabiltzen badugu: $n \geq 30 \Rightarrow t_n \approx N(0,\ 1)$. Beraz, kasu horretan, $\frac{\overline{X} - \mu}{S/\sqrt{n}} \approx N(0,\ 1)$ banaketa asintotikoa erabiliko dugu konfiantza-tartea eraikitzeko.

Demagun μ batezbestekoa eta σ^2 bariantza ezezaguna dituen populazio normaletik ateratako zorizko lagin bakuna dela $X_1, X_2, \ldots, X_n$, $n \geq 30$ izanik. Hauxe da populazio horren μ **batezbestekorako %(1−α)×100 mailako konfiantza-tartea (KT)**:

$$I_\mu^{1-\alpha} = \left(\overline{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \overline{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

4. 15 adibidea

Era ezazu EAEko biztanleek paseo batean ematen duten batez besteko denborarako % 90eko konfiantza-tartea, 4.3 taularen datuak lagin bakartzat hartuta.

4.3 taularen datuak batera hartuta, n = 30 tamainako lagina dugu. Kasu horretan, $\frac{\overline{X}-\mu}{S/\sqrt{n}} \approx N(0,\ 1)$ denez,

$$I_{\mu}^{0,90} = \left(\overline{x} - z_{0,05} \cdot \frac{s}{\sqrt{n}}, \overline{x} + z_{0,05} \cdot \frac{s}{\sqrt{n}}\right)$$

$$= \left(71{,}83 - 1{,}645 \cdot \frac{39.,174}{\sqrt{30}}, 71{,}83 + 1{,}645 \cdot \frac{39{,}7174}{\sqrt{30}}\right)$$

$$= (71{,}83\ -\ 11{,}93,\ 71{,}83\ +\ 11{,}93) = (59{,}9\ ,\ 83{,}76)$$

4.7 ARIKETAK

4. 1. Konifero espezie gehienek polen-pinaburuak eta hazi-pinaburuak dituzte. Pinaburu arrak askatutako polena pinaburu emeraino darama haizeak, eta han arrautzak ernaltzen dira. Demagun X dela polinizazio eta ernalketa bitartean iragandako denbora adierazten duen aldagaia. Eta, pinuen kasuan, aldagai hori sei hilabeteko batezbestekoa eta bi hilabeteko desbideratze estandarra dituen banaketa normalari darraiola. Izan bedi hogeita bost pinaburu emek osaturiko zorizko laginean oinarrituriko $\overline{X}$ estatistikoa. Zenbatekoa da $E\left[\overline{X}\right]$? Eta Var($\overline{X}$)? Eta batezbestekoaren errore estandarra? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

4. 2. EAEko 1997ko Osasun Inkestako datuen arabera, badakigu pertsonen osasun orokorraren maila adierazten duen bizi-kalitate indizea $N(66,\ 18)$ banaketari darraiola. Hamar pertsona aukeratzen dira zoriz, eta 74,9 da haien bizi-kalitate indizearen batezbestekoa.

a) Beste hamar pertsonako lagina zoriz aukeratzen badugu, zer probabilitate dago lagin berriaren batez besteko indizea emandako laginarena baino altuagoa izateko?

b) 10 tamainako laginaren batez besteko indizea 50 baino altuagoa eta A balioa baino txikiagoa izateko probabilitatea % 5ekoa izatea nahi dugu. Kalkula ezazu A balioa.

4. 3. Osakidetzako ospitale batean jaiotzen den ume-populazioaren pisua aztertu nahi dugu. Jaiotza-pisuaren banaketa normala da. Azken hamar urteetako datuetan oinarrituz, badakigu jaiotza-pisuaren batezbestekoa 3.225 gramo eta desbideratze estandarra 750 gramo direla.

a) Populazio horretatik 100 tamainako zorizko laginak ateratzen badira, kalkula ezazu lagin horietako umeen batez besteko jaiotza-pisua 3.100 eta 3.400 gramo bitartean egoteko probabilitatea.

b) Populazio horretatik ateratako 10 tamainako zorizko laginean, 3.115 gramo izan da batez besteko pisua. Lagin hori erabiliz, era ezazu populazioaren batez besteko jaiotza-pisurako % 95eko konfiantza-tartea. Egiazta ezazu populazioaren batezbestekoa kalkulatu duzun tartean dagoen ala ez.

4. 4. Bigarren ariketan aipatu dugunez, pertsonen osasun orokorraren maila adierazten duen bizi-kalitate indizea banaketa normalari darraio. EAEko

1997ko Osasun Inkestari erantzun diotenen artean, 31 emakume eta 38 gizon aukeratu ditugu zoriz. 31 emakumeen bizi-kalitate indizearen batezbestekoa 63,35 izan da, eta bariantza 521. 38 gizonen bizi-kalitate indizearen batezbestekoa 62,74 eta bariantza 490 izan dira. % 95eko konfiantza-mailaz, estima itzazu EAEko emakumezkoen batez besteko bizi-kalitate indizea eta desbideratze estandarra, eta EAEko gizonezkoen batez besteko bizi-kalitate indizea eta desbideratze estandarra.

4. 5. Badakigu Ulibarri urtegian 1997. urtean zehar bildutako ur-bolumena (Hm^3-tan) banaketa normalari darraiola. Hauexek izan ziren 1997ko hamabi hilabeteetan bildutako ur-bolumenak: 106,5 – 116,8 – 107 – 99,2 – 91,7 – 88,6 – 90,6 – 91,3 – 84,5 – 76,8 – 67,4 – 69,3. Datu horietan oinarrituta, % 95eko konfiantza-mailaz, zein balioren artean egongo da urtebetean urtegian bildutako batez besteko ur-bolumena? (Iturria: EUSTAT. *EUE* 97/98).

4. 6. Bigarren ariketan aipatu dugunez, pertsonen osasun orokorraren maila adierazten duen bizi-kalitate indizearen banaketa $N(66,18)$ da. Pertsona talde baten batez besteko bizi-kalitate indizea, 0,95eko probabilitateaz, 66tik 4 baino gehiago ez aldentzea nahi dugu (edozein norantzatan). Zer tamainatakoa izan beharko du pertsona talde horrek?

4. 7. EAEko 1997ko Osasun Inkestari erantzun dioten pertsonen altuerak 4.4 taulan agertzen dira.

Altuerak	Pertsonak
[130-140)	3
[140-150)	53
[150-160)	864
[160-170)	1.639
[170-180)	1.010
[180-190)	273
[190-200)	10

4.4 taula. EAEko1997ko Osasun Inkestari erantzun dioten pertsonen altuerak.

a) Kalkula itzazu pertsona horien altueraren batezbestekoa eta desbideratze estandarra.

b) % 95eko konfiantza-mailaz, estima ezazu EAEko biztanleen batez besteko altuera.

c) % 95eko konfiantza-mailaz, estima ezazu EAEko biztanleen altueren desbideratze estandarra. Baldintzaren bat bete behar al du altuera aldagaiak?

4. 8. EUSTATek argitaraturiko kutsadurari buruzko datuen arabera, 1995. urtean, Bilbon dauden sentsoreek neurturiko SO_2 sufre-dioxidoaren kantitateak (mg/Nm^3) honako hauek izan ziren: 15 – 12 – 24 – 28 – 20 – 24 – 31 – 24 – 27 – 14 – 19. Sufre-dioxidoaren kantitatea banaketa normalari darraiola suposatuz, era ezazu 1995ean Bilbon zegoen SO_2 sufre-dioxidoaren batez besteko kantitaterako % 95eko konfiantza-tartea. (Iturria: EUSTAT. *EUE* 97/98).

4. 9. 1998. urteko otsailean zehar, Sondikako aireportuan neurturiko batez besteko gehieneko tenperatura 16,9º izan zen (Iturria: EUSTAT. *EUE* 97/98). 2000. urteko otsaileko hamar egun aukeratu ditugu zoriz, eta honako hauek izan dira egun horietan Sondikako aireportuan neurturiko gehieneko tenperaturak: 17 – 20 – 17 – 9 – 13 – 13 – 9 – 11 – 12 – 19. Datu horietan oinarrituta, eta gehieneko tenperaturaren banaketa normala dela suposatuz, % 90eko konfiantza-mailaz baiezta al dezakegu 2000. urteko otsaileko batez besteko gehieneko tenperatura 1998. urtekoa baino txikiagoa izan denik?

4. 10. EAEko mendien batez besteko altitudea 965 metrokoa da (Iturria: Euskal Mendizaleen Federazioa). Berrogeita bederatzi mendi aukeratu ditugu zoriz, eta honako datu hauek lortu ditugu:

$$\sum_{i=1}^{49} x_i = 50.085 \qquad \sum_{i=1}^{49} x_i^2 = 54.107.530 \ (x_i = i. \text{ mendiaren altitudea})$$

a) Datu horien arabera, % 99ko konfiantza-mailaz, zein balioren artean egongo da EAEko mendien batez besteko altitudea?

b) Onar al dezakegu aukeratu dugun lagina, % 99ko konfiantza-mailaz, EAEko mendi guztien adierazgarri denik? Zergatik?

4. 11. 1991. urtean, EAEko familien batez besteko kide kopurua 3,32 zen. EAEko 36 herri aukeratu ditugu zoriz eta, 1996an, hauexek izan ziren herri haietako familien batez besteko kide kopurua: 3,81 – 3,08 – 2,95 – 3,27 – 3,25 – 3,24 – 3,36 – 3,09 – 3,06 – 3,34 – 2,91 – 2,71 – 3,76 – 3,27 – 3,27 – 3,33 – 3,69 – 3,08 – 3,76 – 3,07 – 3,20 – 3,33 – 3,24 – 3,45 – 3,22 – 3,28 – 2,99 – 3,23 – 3,63 – 3,27 – 2,86 – 3,02 – 3,18 – 3,04 – 3,13 – 3,40. (Iturria: EUSTAT. *EUE* 97/98).

a) Era ezazu 1996ko familien batez besteko kide kopururako % 95eko konfiantza-tartea.

b) % 95eko konfiantza-mailaz, onar al dezakegu EAEko familien batez besteko kide kopurua gutxitu egin denik 1991tik 1996ra?

4. 12. EUSTATek neurturiko etxebizitzen erosotasun-indizea banaketa normalari darraio. EAEko hogeita bost etxebizitza aukeratu ditugu zoriz, eta haien erosotasun-indizeak honako hauek izan dira: 67 – 58,8 – 52,3 – 63,9 – 71,4 – 63,9 – 60,5 – 62 – 57 – 65,3 – 58,8 – 62,3 – 59,8 – 63,1 – 63,7 – 59,8 – 65,6 – 56,9 – 66,8 – 62,7 – 69,3 – 63,3 – 61,4 – 62,4 – 71,7 (Iturria: EUSTAT. *EUE* 97/98).

a) Informazio hori erabiliz, estima itzazu μ, σ^2 eta σ.

b) Lor itzazu μ eta σ-rako % 90eko konfiantza-tarteak.

c) Etxebizitzen erosotasun-indize egokia gutxienez 70 dela jotzen dugu. Onar al dezakegu EAEko etxebizitzen erosotasun-indizea egokia denik? Azal ezazu erantzuna, lortutako konfiantza-tartean oinarriturik.

4. 13. EUSTATek emandako datuen arabera, pago-azalera 100 Ha baino gutxiago eta 2000 Ha baino gehiago duten herriak kenduta, 1996. urtean pagoek osaturiko EAEko baso-azalera 400 Ha-ko desbideratze estandarra duen banaketa normalari darraio. Hamar herri aukeratu ditugu zoriz, eta hauexek izan dira horietan dauden pago-azalerak: 548 – 1.403 – 1.318 – 203 – 694 – 298 – 162 – 712 – 262 – 274 (Iturria: EUSTAT. *EUE* 97/98).

Datu horietan oinarrituta, kalkula ezazu, 100 Ha baino gutxiago eta 2.000 Ha baino gehiago dituzten herriak kenduta, EAEko batez besteko pago-azalerako % 95eko konfiantza-tartea. EUSTATek emandako datuak erabiliz, badakigu 1996. urtean EAEko batez besteko pago-azalera 535 Ha zela. Zoriz aukeraturiko hamar herri horiek, % 95eko konfiantza-mailaz, EAE osoaren adierazgarri al dira? Zergatik?

4. 14. EAEn zegoen eukalipto-azalera aztertzen ari gara. EUSTATek emandako datuen arabera, 1996. urtean, zuhaitz espezie guztiak kontuan izanda, eukaliptoko azalera-proportzioa % 1 baino handiagoa zen EAEko 69 herritan. Herri horiek osaturiko populazio horretan, eukaliptoko azalera-proportzioa banaketa normalari darraiola egiaztatu dugu. Hirurogeita bederatzi herri horien artean hamabost herri aukeratu ditugu zoriz, eta hauexek dira horietan dauden eukaliptoko azalera-proportzioak: 14,65 – 16,15 – 60,5 – 3,3 – 39,21 – 39,01 – 39,7 – 31,85 – 5,07 – 20,74 – 2,99 – 2,15 – 68,95 – 21,57 – 55,68 (Iturria: EUSTAT. *EUE* 97/98).

a) Datu horietan oinarrituta, era itzazu EAEn dagoen batez besteko eukaliptoko azalera-proportziorako eta proportzio horren bariantzarako % 90eko konfiantza-tarteak.

b) Eukaliptoko batez besteko azalera-proportzioa % 50 baino handiagoa bada, beste espezieak landatzeko kaltegarria dela jotzen da. Aurreko atalean oinarrituta, azal ezazu ea arrisku horren aurrean gauden ala ez.

4. 15. Lehenengo gaiko 13. ariketan, egunero erretzen den puru kopurua aztertu dugu.

a) Han lortutako emaitzen arabera, populaziotzat hartuz egunero pururen bat erretzen dutenek osaturikoa, zein dira μ-ren eta σ^2-ren estimazio alboragabeak?

b) Egunero pururen bat erretzen dutenen artean, berrogeita bost pertsona aukeratu ditugu zoriz, eta honako informazio hau lortu dugu haiek egunero erretzen duten puru kopuruari buruz:

$$\sum_{i=1}^{45} x_i = 121 \qquad \sum_{i=1}^{45} x_i^2 = 815$$

(x_i = i. pertsonak egunero erretzen duen puru kopurua)

Informazio horren arabera, % 95eko konfiantza-mailaz, zein balioren artean egongo da μ?

c) Posible al da σ^2-rako konfiantza-tartea eratzea? Zergatik?

4. 16. Lehenengo gaiko 14. ariketako datuetan oinarrituta, 1996. urtean 1.800 tona antxoa baino gehiago lehorreratu zituzten bost itsas portuak kontuan harturik, honako datu hauek lortu ditugu: $\bar{x}$ = 2.705,5 tona eta s = 1.172 tona. Edozein itsas portutan lehorreraturiko antxoa kantitatearen banaketa normala dela suposatzen dugu.

a) % 95eko konfiantza-mailaz, estima ezazu EAEn lehorreraturiko batez besteko antxoa kantitatea.

b) % 90eko konfiantza-mailaz, estima ezazu EAEn lehorreraturiko antxoa kantitatearen desbideratze estandarra.

c) Antxoaren lehorreratzeari dagokionez, Europar Batasunak EAEri ezartzen dion kupoa itsas portuko 4.000 tonako batezbestekoa dela jo dezakegu. % 95eko konfiantza-mailaz, onar al dezakegu kupo hori ez dela gainditukо?

4. 17. Lehenengo gaiko 15. ariketan, 1996. urtean EAEko behatoki publiometrikoetan behatu diren euria izandako egun kopuruak agertzen dira.

a) Datu horietan oinarrituta, era ezazu EAEko edozein tokitan euria izandako batez besteko egun kopururako % 95eko konfiantza-tartea.

b) Arabako edozein tokitan euria izandako batez besteko egun kopururako % 95eko konfiantza-tartea eratu nahi badugu, zer baldintza bete behar da? Zergatik? Beharrezkoa den baldintza kontuan hartuz, era ezazu konfiantza-tartea.

c) Beharrezkoak diren baldintzak kontuan hartuz, egizu gauza bera Bizkaiko eta Gipuzkoako herrialdeetarako.

d) % 95eko konfiantza-mailaz, onar al daiteke Arabako, Bizkaiko eta Gipuzkoako herrialdeetan euria izandako batez besteko egun kopuruen artean erlazioa dagoela? Zein?

4. 18. Zuhaitzaren kalibrea lurretik 3 cm-ra neurtutako bere diametroa da, eta banaketa normalari darraio. Baso pribatu bateko 3 eta 4 metro bitarteko hamasei zuhaitzek osaturiko lagina aukeratu, eta haien kalibrea neurtu da. Hauexek dira lortutako datuak (cm-tan): 2,3 – 1,9 – 1,7 – 2,1 – 1,5 –1,8 –1,8 – 1,1 – 2,1 – 1,5 – 2,0 – 1,6 – 1,3 – 1,6 – 1,5 – 1,3 (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

Kalkula ezazu baso horretan landaturiko zuhaitzen batez besteko kalibrerako % 95eko konfiantza-tartea. Zuhaitzen batez besteko tamaina enborraren erresistentziarekiko proportzionala dela ziurtatzeko, tamaina horretako zuhaitzen batez besteko kalibreak 2 cm-koa izan beharko luke. Hemen landatuko zuhaitzek betetzen al dute baldintza hori?

4. 19. Landare baten ohiko zelulak mintz plasmatikoa deritzon zelula-mintzaren bidez mugaturiko zitoplasma kantitate handia du. Mintz horren batez besteko lodiera espezieen arabera aldatzen da. Hogei espeziek osaturiko zorizko laginean, zelula-mintzaren lodiera adierazten duen X aldagaiaren honako balio hauek lortu dira (angstrom-etan): 80 – 90 – 85 – 82 – 75 – 58 – 70 – 84 – 87 – 81 – 87 – 61 – 73 – 84 – 85 – 70 – 78 – 95 – 77 – 52 (Iturria: J. S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994).

a) Lor itzazu μ-ren eta σ^2-ren puntu-estimazioak.

b) Demagun X normalki banatuta dagoela. Lor itzazu μ-rako, σ^2-rako eta σ-rako % 95eko konfiantza-tarteak.

4. 20. Hamabi urte baino gutxiago dituzten umeengan loditasunari buruzko ikerlana egin da. Oso lodi dauden ehun umek osaturiko lagina aukeratu da, eta loditasuna agertu zen adina aztertu. Laginaren batezbestekoa eta desbideratze estandarra 4 eta 1.5 urte dira, hurrenez hurren. Loditasunaren hasierako adina banaketa normalari darraiola jotzen da. (Iturria: B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, 1994).

a) Kalkula ezazu umeen loditasunaren hasierako batez besteko adinerako % 95eko konfiantza-tartea.

b) Zehaztu ezazu umeen loditasunaren hasierako adinaren bariantzarako % 95eko konfiantza-tartea.

c) Zehaztu ezazu umeen loditasunaren hasierako adinaren desbideratze estandarrerako % 95eko konfiantza-tartea.

5 HIPOTESI-KONTRASTEA. LAGIN BAKAR BATEN AZTERKETA

5.1 SARRERA

Laugarren gaian banaketa batzuen parametroen puntu- eta tarte-estimaziorako metodoak ezarri ditugu; baina, sarritan, ikertzailea, aurresperientzia edo bere intuizioa erabilita, esperientzia-plangintza diseinatzerakoan zenbait hipotesi formulatzen ausartzen da. Hipotesi horiek, gehienean, populazioaren parametro baterako zenbakizko balioak finkatzean dautza.

Adibidez, burdin faltaren ondorioz pairatzen den anemia osasun-arazo larria da munduan, batez ere, haur txiroen artean. Demagun, egoera normalean, 9 eta 11 urte bitarteko haurren eguneroko batez besteko burdin-kontsumoa 14,44 mg dela. Haur txiroen batez besteko burdin-kontsumoa hori baino baxuagoa dela susmatzen da. Hau da, honako bi baieztapen hauen artean egin nahi dugu hautaketa:

1. Haur txiroen eguneroko batez besteko burdin-kontsumoa ohikoa da (14,44).
2. Haur txiroen eguneroko batez besteko burdin-kontsumoa ohikoa baino baxuagoa da.

Hipotesi-kontrastearen ereduan, mota horretako hautaketa egiteko bi hipotesiak planteatzen dira: bata, orain arte egiatzat hartu dena, eta bestea, ikertzaileak proposatzen duena. Lehenengoari hipotesi nulua (H_0) deritzo, eta bigarrenari hipotesi alternatiboa (H_1), eta elkarren osagarriak izan behar dute. Hipotesi nuluak ordura arte gertatzen zenarekiko edota egiatzat hartzen dugunarekiko «aldaketarik ez» adierazten du. Praktikan, horrek berdintasuna H_0-ren parte izatea behartzen du. Aurreko adibidean, hipotesi nulua litzateke haur txiroen eguneroko batez besteko burdin-kontsumoa 14,44 edo gehiago dela, eta hipotesi alternatiboa, 14,44 baino txikiagoa dela. Bi hipotesien artean hautatzeko, lagin bat aukeratzen dugu populaziotik. Gure adibidean, eskualde txiro bateko 9 eta 11 urte bitarteko haur taldea aukeratu beharko dugu. Ondoren, egiatzat hartutako hipotesiaren menpe, lagineko datuak lortzeko probabilitatea kalkulatuko dugu eta, balio horren arabera, laginak hipotesi nulua indartu edo baztertu egiten duen egiaztatuko dugu.

Hipotesi-kontrasteak, populazioaren adierazgarri den laginean baturiko informazioan oinarrituz, hasierako hipotesia errefusatzeko edo ez errefusatzeko (beti termino probabilistikoetan) metodo zientifiko egokia eskaintzen digu. Erabakiak hartzerakoan, ideia subjektiboaz fidatu beharrean, probabilitate-teorian oinarrituriko metodo objektiboa eskaintzean datza hipotesi-kontrastearen garrantzia.

Hipotesi-kontrastearen metodoaren formulazio orokorra eman baino lehen, era intuitiboan azalduko dugu.

5.2 LEHENDABIZIKO ADIBIDEA: HIPOTESI-KONTRASTEAREN KONTZEPTU INTUITIBOAK

Demagun X zorizko aldagaia σ ezaguneko populazio normalari darraiola, hau da, X: $N(\mu,\sigma)$. Jo dezagun H_0: $\mu = \mu_0$ hipotesi «nulua», H_1: $\mu \neq \mu_0$ hipotesi «alternatibo»arekin kontrastatu nahi dugula. Horretarako, n tamainako lagina (X_1, X_2, …X_n) aterako dugu populaziotik, eta haren $\overline{X}$ batezbestekoa kalkulatuko dugu.

Era intuitiboan, horrela arrazoitu dezakegu: H_0 hipotesia egiazkoa izango balitz, $\mu = \mu_0$ beteko litzateke, eta, beraz, $\overline{X}$: $N(\mu_0,)$ litzateke. Hori dela eta, esperoko dugu «gehienean» $\overline{X}$ -ren balioak μ_0-rekiko «hurbil» (gehiegi zein gutxiegi) egongo direla eta «gutxitan» aldenduko direla μ_0-tik.

Orduan, honako jokabide-arau induktibo hau har genezake: esperimentalki, n tamainako lagina aukeraturik, haren $\overline{x}$ -k μ_0-rekiko balio «urruna» hartzen badu, $\mu = \mu_0$ hipotesia errefusatuko da; lorturiko $\overline{x}$ -ren balioa, ordea, μ_0-tik «hurbil» badago, ezin izango da $\mu = \mu_0$ hipotesia errefusatu. Arrazoibide horren oinarria «hurbil» eta «urrun» definitzean datza.

$\overline{x}$ μ_0-rekiko «hurbil» dagoela esango dugu, $\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \in (a,b)$ betetzen baldin bada. (a,b) tarteari onarpen-eskualde esango diogu eta, oro har, S_0-ren bidez adieraziko da.

Aldiz, $\overline{x}$ μ_0-rekiko «urrun» dagoela esango dugu, $\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \notin (a,b)$ denean; alegia, $\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \in \overline{S}_0$ betetzen baldin bada. Oro har, $\overline{S}_0 = S_1$-i eskualde kritikoa esango diogu.

5.3 PLANTEAMENDU OROKORRA

Hipotesi-kontrastea gauzatzeko lehenengo pausoa kontrastatu nahi ditugun hipotesiak formulatzea da. Alde batetik, egiaztatu nahi dugun hipotesia **hipotesi nulua** izango da eta $\boldsymbol{H_0}$-ren bidez adieraziko dugu. Sarreran aipatu dugunez, hipotesi nuluak orain arte gertatzen zenarekiko edota egiatzat hartzen dugunarekiko «aldaketarik ez» adierazten du, eta, beraz, berdintasunak H_0-ren parte izan behar du. Bestetik, **hipotesi alternatiboa** H_0-ren hipotesi osagarria izango da eta $\boldsymbol{H_1}$-en bidez adieraziko dugu.

Hipotesiak formulatu ondoren, beren arteko bat hautatzeko lagin bat aukeratu behar dugu eta, beren balioen arabera, hipotesi egokia hautatzeko metodo bat finkatu behar dugu. Horretarako, laginaren balioen funtzio erreala den eta H_0 egia izanik probabilitate-banaketa ezaguna duen estatistiko bat erabiliko dugu (aurreko atalean, $\overline{X}$). Estatistiko

horri **test estatistikoa** edo **pibot estatistikoa** deritzo eta, kontrastatu nahi dugun populazioaren parametroaren arabera, test estatistiko ezberdinak izango ditugu.

Test estatistikoaren balioen arabera, erabakia hartuko dugu honako era honetara jokatuz: test estatistikoaren balioa H_0 egia deneko baldintzaren menpe itxarotako baliotik aldentzen bada, H_0 hipotesia errefusatuko dugu eta H_1 hipotesia aukeratuko dugu; aitzitik, estatistikoaren balioa aipaturiko baldintzaren menpe itxarotako baliotik hurbil badago, ezin izango dugu H_0 hipotesia errefusatu.

H_0 errefusatzen den test estatistikoaren balioen heinak **eskualde kritikoa** osatzen du eta $\boldsymbol{S_1}$ adierazten da. H_0 errefusatzen ez den test estatistikoaren balioen heinak **onarpen-eskualdea** osatzen du eta $\boldsymbol{S_0}$ adierazten da. Bi eskualde horiek osagarriak izango dira.

Kontuan hartu behar dugu hipotesi-kontrasteetan ez dela hipotesia egiaztatzen, baizik eta onartu ala errefusatu egiten dela. Horrexegatik, aurreko jokabide-arauan oinarrituta, 5.1 taulan agertzen diren lau egoerak gerta daitezke.

		Egia	
		$\boldsymbol{H_0}$	$\boldsymbol{H_1}$
Erabakia	H_0 ez da errefusatzen	H_0 egiazkoa da eta ez da errefusatzen.	H_1 egiazkoa da, baina H_0 ez da errefusatzen.
	H_0 errefusatzen da	H_0 egiazkoa da, baina errefusatu egiten da.	H_1 egiazkoa da eta H_0 errefusatzen da.

5.1 taula. Hipotesi-kontrastearen jokabide-arauan aurki ditzakegun egoerak.

Era horretan, bi errore mota ditugu: H_0 egia denean, hura errefusatzea, edo H_0 egia ez denean, hura ez errefusatzea. **I motako errorea**, hipotesi nulua egiazkoa denean, hura errefusatzean datza. **II motako errorea**, aldiz, hipotesi nulua egia ez denean, hura ez errefusatzean datza.

Bestalde, kontuan hartu behar dugu, edozein erabaki hartuta ere, beti akatsa izateko aukera dagoela. Hau da, H_0 errefusatzen dugun bakoitzean, I motako errorea izan dezakegu, eta H_0 errefusatzen ez dugun bakoitzean, II motako errorea izan dezakegu. Beraz, errore mota horiek izateko probabilitateez hitz egin beharko dugu.

I motako errorearen probabilitateari hipotesi-kontrastearen **adierazgarritasun-maila** deritzo, eta α adierazten da. II motako errorearen probabilitatea β adierazten da, parametroaren balio alternatiboaren menpean dago, eta ez du izen berezirik. Hori baino interesgarriagoa eta erabilgarriagoa da bere osagarria, hots, $(1 - \beta) = 1 - P$(II motako

errorea). Probabilitate horrek, hipotesi nulua faltsua denean, bera errefusatzeko probabilitatea ematen digu, eta hipotesi-kontrastearen **ahalmen** deritzo.

Hipotesi-kontrastearen helburu orokorra da I eta II motako erroreak izateko probabilitateak diren α eta β ahalik eta txikienak izatea. Helmuga horrek konponbide bat eskatzen du, zeren eta α txikitzeak hipotesi nulua gutxiagotan errefusatzea, eta β txikitzeak, ordea, hipotesi nulua gutxiagotan ez errefusatzea ondorioztatzen baitute. Ekintza horiek kontrakoak dira; alegia, α handitzen denean β txikitu egin beharko da, eta α txikitzen denean, ordea, β handitu beharko da. Estrategia orokorra, α maila finkatu ondoren (esate baterako, 0,05 , 0,01…), β minimizatzen duen edo baliokidea den ahalmena maximizatzen duen eskualde kritikoa aurkitzean datza; hau da, hipotesi-kontraste ahaltsuena eratu behar dugu.

Estrategia horrek (α finkatuaren arabera S_1 eratzeak) arazoak sortzen ditu. Esate baterako, hipotesi-kontrastearen emaitzak menpekotasun handia sortzen du arbitrarioa den α balioarekiko, eta posible da H_0 errefusatzea $\alpha = 0{,}05$ denean, baina ez errefusatzea $\alpha = 0{,}04$ denean.

Arazo hori ekiditeko, hipotesi-kontrastearen metodologia beste era honetara azal dezakegu. Lehenengo, H_0 egia dela suposatuz, laginean behaturiko test estatistikoaren balioa bezain arraroa edo arraroagoa den emaitza zoriz lortzeko probabilitatea kalkulatuko dugu. Probabilitate horri ***p*-balioa** edo **testaren balio kritikoa** deritzo. Hodges-ek eta Lehman-ek «hipotesi nuluaren aldeko pertsona batengan, esperimentuak eragingo lukeen ustekabe-mailaren neurketa» gisa deskribatu dute *p*-balioa (Iturria: *Basic Concepts of Probability and Statistics*. Holden Day, San Francisco, 1970).

Probabilitate hori oso txikia bada, laginean lorturiko balioak oso arraroak direla pentsatuko dugu, eta, beraz, H_0 hipotesia errefusatzera joko dugu. Probabilitatea «txiki»tzat hartuko dugu, aurrefinkaturiko α baino txikiagoa bada. Hau da, $p \leq \alpha$ bada, H_0 errefusatuko dugu, eta $p > \alpha$ bada, ezin izango dugu H_0 errefusatu. Jokabide-arau horrek laginak duen adierazgarritasuna nabarmentzen du, eta ikertzailearen eskuetan uzten du ondorioak ateratzea.

5.4 HIPOTESI-KONTRASTEA EGITEKO URRATSAK

R.A. Fisher, J. Neyman eta E.S. Pearson matematikariek 1920 eta 1933 urteen artean egindako lanen emaitza da aurreko atalean azaldu dugun hipotesi-kontrastearen metodologia. Laburbilduz, honako hauek dira hipotesi-kontrastea gauzatzeko urratsak:

1. Parametroari buruzko hipotesiak formulatu. Bi hipotesi-kontraste mota daude:

   i) **Aldebikoa**: H_0: $\theta = \theta_0$
   H_1: $\theta \neq \theta_0$

ii) **Aldebakarrekoa**: a) **Ezkerretikoa**: H_0: $\theta \geq \theta_0$

H_1: $\theta < \theta_0$

b) **Eskuinetikoa**: H_0: $\theta \leq \theta_0$

H_1: $\theta > \theta_0$

2. Parametroaren arabera, test estatistikoa aukeratu eta haren balioa kalkulatu.

Hortik aurrera bi bide daude (*p*-balioan edo eskualde kritikoan oinarriturik), eta bietatik edozein aukera daiteke:

a) *p*-balioaren bidez hipotesi-kontrastea egiteko urratsak:

1. *p*-balioa kalkulatu. Eskuinerantz isatsa duen kontrastean, *p*-balioa da test estatistikoaren baliotik eskuinerantz doan dentsitate-kurbaren peko azalera; ezkerrerantz isatsa duen kontrastean, ordea, test estatistikoaren baliotik ezkerrerantz doan dentsitate-kurbaren peko azalera da *p*-balioa. Aldebiko (edo isats biko) kontrasteetan, *p*-balioa aldebakarreko kontrastearen *p*-balioaren bikoitza da.

2. α adierazgarritasun-maila finkatu. H_0 egia denean, hura errefusatzeko probabilitatea ikertzaileak aurrefinkaturik dago, eta, praktikan, 0,05 edo 0,01 izaten da.

3. *p*-balioa α adierazgarritasun-mailarekin konparatu. Orduan:

   - $p \leq \alpha \Rightarrow H_0$ errefusatu
   - $p > \alpha \Rightarrow H_0$ ez errefusatu (H_0 onartu)

   H_0 egia izango balitz, lortutako *p*-balioa txikiegia dela uste dugunean, hau da, dagoen ezberdintasuna zoriz agertzeko probabilitatea oso txikia denean, H_0 errefusatu egiten da eta adierazgarritasuna dagoela esaten dugu.

   α balioa aurrefinkaturik ez dagoenean, honako jokabide-arau honi jarraitzen zaio:

   - Baldin eta 0,01 < *p*-balioa ≤ 0,05 bada, emaitzak **adierazgarriak** direla esaten da.
   - Baldin eta 0,001 < *p*-balioa ≤ 0,01 bada, emaitzak **oso adierazgarriak** direla esaten da.
   - Baldin eta *p*-balioa ≤ 0,001 bada, emaitzak **guztiz adierazgarriak** direla esaten da.

- Baldin eta p-balioa > 0,05 bada, emaitzak ez dira adierazgarriak; hau da, ezberdintasuna zoriz ager daiteke eta ez dago H_0 errefusatzeko nabaritasunik.

Batzuetan, 0,05 < p-balioa < 0,10 denean, pentsa dezakegu nolabaiteko joera dagoela adierazgarritasun estatistikorantz.

b) Eskualde kritikoaren bidez hipotesi-kontrastea egiteko urratsak:

1. α adierazgarritasun-maila finkatu. Lehen aipatu dugunez, ikertzaileak aurrefinkaturik dago eta, praktikan, 0,05 edo 0,01 izaten da.
2. S_1 eskualde kritikoa kalkulatu. Eskualde kritikoa test estatistikoaren lagin-banaketan eta adierazgarritasun-mailan oinarrituta dago. Hipotesi-kontrastea aldebikoa edo aldebakarrekoa den, eskualde kritikoa tarte finitua edo infinitua izango da.
3. Test estatistikoaren balioa eskualde kritikoan dagoen ala ez egiaztatu. Orduan:

   - Test estatistikoaren balioa S_1 eskualde kritikoan baldin badago $\Rightarrow$ H_0 errefusatu.

   - Test estatistikoaren balioa S_1 eskualde kritikoan ez badago $\Rightarrow$ H_0 ez errefusatu (H_0 onartu).

Aurreko atalean aipatu dugunez, lehenengo metodoaren abantaila nagusia, adierazgarritasun-maila aurrefinkatu beharrean, p-balioaren bidez emaitzen adierazgarritasuna neurtzean datza. Zenbat eta p-balio txikiagoa izan, orduan eta handiagoa izango dugu hipotesi nulua egokia ez deneko nabaritasuna. Azken metodoak, ordea, adierazgarritasun-maila aurrefinkatzea behartzen du, beharrezkoa baita eskualde kritikoa kalkulatzeko. Dena dela, bi metodoen abantailen eta desabantailen inguruko eztabaida handia dago oraindik, gai honen amaieran dagoen eranskinean ikus dezakezuenez.

Hipotesi-kontrasteen planteamendu orokorra azaldu ondoren, populazioaren parametro bati (hain zuzen, μ batezbestekoari) buruzko hipotesi-kontrasteak garatuko ditugu hurrengo ataletan, populazioaren baldintzen arabera. Lehenengo kasua bariantza ezaguna duen populazioari dagokio.

5.5 BARIANTZA EZAGUNEKO POPULAZIOAREN BATEZBESTEKORAKO HIPOTESI-KONTRASTEA

Demagun populazioaren σ^2 bariantza ezaguna dela. Populazioa normala izan daiteke, hau da, X: $N(\mu,\sigma)$; edo bestela, laginaren tamaina aski handia hartuko dugu, hau da, $n \geq 30$ izango da, Limitearen Teorema Zentrala betetzen dela ziurtatzeko (hots, $X \approx$

$N(\mu, \frac{\sigma}{\sqrt{n}})$ izango da). Populazioaren batezbestekoari buruzko hipotesi-kontrastea egiteko erabiltzen den test estatistikoa $Z_p = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ da, eta beronen banaketa $N(0,1)$ da (X normala denean banaketa zehatza da eta, normala ez denean, $n \geq 30$ izanik, asintotikoa).

5.5.1 Aldebakarreko kontrasteak

5.5.1.1 Ezkerretikoa

Honako hau da hipotesi-kontrastea: H_0: $\mu \geq \mu_0$

H_1: $\mu < \mu_0$

H_0 egiazkoa dela suposatuz, $z_p = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$ test estatistikoaren balioa eta $p = P(Z < z_p)$ p-balioa kalkulatzen dira.

Eskualde kritikoa honako era honetara kalkulatuko genuke:

H_0 errefusatu $\Leftrightarrow p \leq \alpha \Leftrightarrow P(Z < z_p) \leq P(Z > z_\alpha) = P(Z < -z_\alpha) \Leftrightarrow z_p \leq -z_\alpha \Leftrightarrow$

$\Leftrightarrow \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \leq -z_\alpha$

Beraz, $Z_p = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ test estatistikoaren eskualde kritikoa $S_1 = (-\infty, -z_\alpha]$ da.

5.1 adibidea

Laugarren gaiko 4. 7 adibidean adierazten denez, EAEko biztanleek paseo batean ematen duten denboraren desbideratze estandarra 45 minutukoa dela suposatzen dugu. Gai horren 4.3 taulan agertzen diren datu guztiak lagin bakartzat harturik, onar al dezakegu EAEko biztanleek paseo batean ematen duten batez besteko denbora 75 minutukoa baino laburragoa denik, % 5eko adierazgarritasun-mailaz?

Honako hipotesi-kontraste hau gauzatuko dugu: $H_0 : \mu \geq 75$

$H_1 : \mu < 75$

Aipatutako 4.3 taulako datuetan oinarrituz, n = 30 eta $\bar{x}$ = 71,83 dira. Beraz, hauxe izango da test estatistikoaren balioa:

$$z_p = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{71{,}83 - 75}{45/\sqrt{30}} = -0{,}39$$

Orduan, $p = P(Z < z_p) = P(Z < -0{,}39) = P(Z > 0{,}39) = 0{,}3483 > 0{,}05 \Rightarrow$ % 5eko adierazgarritasun-mailaz ezin da H_0 errefusatu; hau da, ezin dugu onartu EAEko biztanleek paseo batean ematen duten batez besteko denbora 75 minutukoa baino laburragoa denik.

5.5.1.2 Eskuinetikoa

Hipotesi-kontrastea honako hau da: H_0: $\mu \leq \mu_0$

H_1: $\mu > \mu_0$

H_0 egiazkoa dela suposatuz, $z_p = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ test estatistikoaren balioa eta $p = P(Z > z_p)$ p-balioa kalkulatzen dira.

Eskualde kritikoa honako era honetara kalkulatuko genuke:

$$H_0 \text{ errefusatu} \Leftrightarrow p \leq \alpha \Leftrightarrow P(Z > z_p) \leq P(Z > z_\alpha) \Leftrightarrow z_p \geq z_\alpha \Leftrightarrow \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$$

Beraz, $Z_p = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ test estatistikoaren eskualde kritikoa $S_1 = [z_\alpha, +\infty)$ da.

5.2 adibidea

Aurreko adibideko datuen arabera, onar al dezakegu % 5eko adierazgarritasun-mailaz EAEko biztanleek paseo batean ematen duten batez besteko denbora 45 minutukoa baino luzeagoa denik?
Hauxe da kasu horretan gauzatuko dugun hipotesi-kontrastea:

H_0 : $\mu \leq 45$

H_1 : $\mu > 45$

Test estatistikoaren balioa $z_p = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{71{,}83 - 45}{45/\sqrt{30}}$ = 3,27 izango da.

Honela kalkulatzen da hipotesi-kontraste horren p-balioa:

p = P(Z > z_p) = P(Z > 3,27) = 0,000483 < 0.05 ⇒ % 5eko adierazgarritasun-mailaz H_0 errefusatzen dugu; hau da, EAEko biztanleek paseo batean ematen duten batez besteko denbora 45 minutu baino luzeagoa dela onartzen dugu, % 5eko adierazgarritasun-mailaz.

5.5.2 Aldebiko kontrastea

Kontrastatzeko hipotesiak honako hauek dira: H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

H_0 egiazkoa dela suposatuz, $z_p = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ test estatistikoaren balioa eta $p = 2 \cdot P(Z > |z_p|)$ p-balioa kalkulatzen dira.

Eskualde kritikoa honako era honetara kalkulatuko genuke:

H_0 errefusatu $\Leftrightarrow p \leq \alpha \Leftrightarrow P(Z > |z_p|) \leq P(Z > |z_{\alpha/2}|) \Leftrightarrow z_p \geq z_{\alpha/2}$ edo $z_p \leq -z_{\alpha/2} \Leftrightarrow$

$$\Leftrightarrow \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq z_{\alpha/2} \text{ edo } \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq -z_{\alpha/2}$$

Beraz, $Z_p = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ test estatistikoaren eskualde kritikoa $S_1 = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, +\infty)$ da.

5.3 adibidea

Adibide berarekin jarraituz, onar al dezakegu % 5eko adierazgarritasun-mailaz EAEko biztanleek paseo batean ematen duten batez besteko denbora 75 minutukoa denik?

Honako hipotesi-kontraste hau gauzatuko dugu: $H_0 : \mu = 75$
$H_1 : \mu \neq 75$

Test estatistikoaren balioa 5.1 adibidean kalkulatutakoa da, hots, $z_p = -0{,}39$.

Kasu horretan, $p = 2 \cdot P(Z > |z_p|) = 2 \cdot P(Z > 0{,}39) = 2 \cdot 0{,}3483 = 0{,}6966 > 0{,}05$ ⇒ % 5eko adierazgarritasun-mailaz ezin da H_0 errefusatu; hau da, EAEko biztanleek paseo batean ematen duten batezbesteko denbora 75 minutukoa

dela onartzen dugu, % 5eko adierazgarritasun-mailaz.

5.6 TESTAREN AHALMENA

Hirugarren atalean emandako definizioari eutsiz, hipotesi nulua faltsua denean, hura errefusatzeko probabilitatea da kontrastearen ahalmena. Alegia:

$1 - \beta = 1 - P$(II motako errorea) $= 1 - P(H_0$ ez errefusatu$|\ H_0$ faltsua$) = P(H_0$ errefusatu$|$ H_0 faltsua)

Bariantza ezaguneko populazioaren batezbestekoari buruzko hipotesi-kontrastea gauzatu denean, ikus dezagun nola kalkulatzen den egindako kontrastearen ahalmena. Horretarako, hipotesi alternatiboaren balio posible bat (μ_1) aurrefinkatu behar dugu, kontrastearen ahalmena ez baita α bezalako balio zehatza, baizik eta hipotesi alternatiboaren barruko balioen funtzioa.

5.6.1 Ezkerretiko aldebakarreko kontrastea

Kontrastatzeko hipotesi nulua H_0: $\mu \geq \mu_0$ da, H_1: $\mu = \mu_1 < \mu_0$ alternatiboaren aurka. Laugarren atalean aztertu dugunez, p-balioa α adierazgarritasun-maila baino txikiagoa denean, H_0 errefusatuko dugu. Hau da, $p = P(Z < z_p) \leq \alpha$ denean, edo eskualde kritikoaren arabera, $Z_p = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \in (-\infty, -z_\alpha]$ denean.

Ahalmena, beraz, honako hau izango da: $1 - \beta = P(Z_p \leq -z_\alpha \mid \mu = \mu_1) =$

$$P\left(\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \leq -z_\alpha \mid \mu = \mu_1 \right).$$

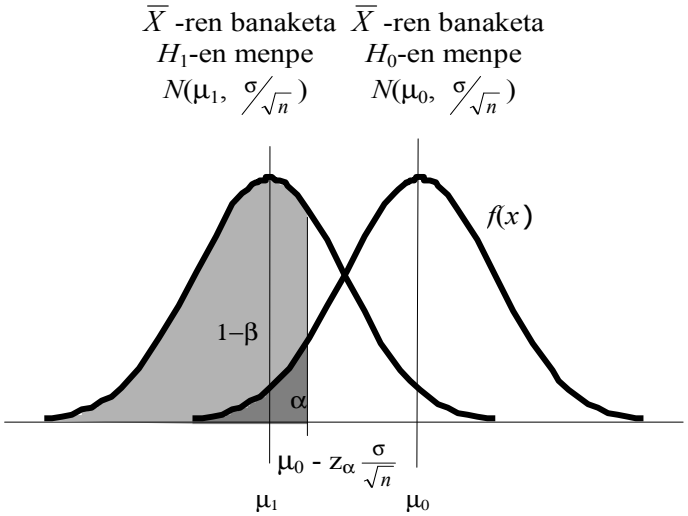
H_0 hipotesiaren baldintzapean, badakigu $\dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = Z$: $N(0,1)$ banaketari darraiola. Baina H_1 baldintzapean hori ez da egia izango. H_1 hipotesia egiazkoa denean, $\overline{X}$: $N(\mu_1,\sigma)$ da, hots, $\dfrac{\overline{X} - \mu_1}{\sigma / \sqrt{n}} = Z$: $N(0,1)$.

Ondorioz:

$$1-\beta = P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le -z_\alpha \,\middle|\, \mu=\mu_1\right) = P\left(\overline{X} \le \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \,\middle|\, \mu=\mu_1\right) =$$

$$P\left(\frac{\overline{X}-\mu_1}{\sigma/\sqrt{n}} \le -z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right) = P\left(Z \le -z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$$

Orduan, kontrastearen ahalmena μ_1 balioaren menpekoa da. Grafikoki, 5.1 irudian agertzen den bezala adieraz daiteke.



5.1 irudia. Bariantza ezaguneko banaketaren batezbestekorako, ezkerretiko aldebakarreko kontrastearen ($\mu_1 < \mu_0$) ahalmenaren adierazpide grafikoa.

5.4 adibidea

Kalkula ezazu 5.1 adibideko kontrastearen ahalmena, hipotesi alternatiboaren barruan $\mu_1 = 60$ minutu aukeratzen dugunean.

Kontraste mota horretan, eskualde kritikoa $S_1 = (-\infty, -z_{0.05}] = (-\infty, -1.645]$ da, hau da, H_0 errefusatzen dugu $Z_p \le -1.645$ denean.

Beraz, $1 - \beta = P(H_0$ errefusatzeko $\mid H_0$ faltsua bada$) = P(Z_p \le -1{,}645 \mid \mu = 60)=$

$$= P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le -1{,}645 \,\middle|\, \mu = 60\right) = P\left(\frac{\overline{X}-75}{45/\sqrt{30}} \le -1{,}645 \,\middle|\, \mu = 60\right) =$$

$$= P\left(\overline{X} \le 75 - 1{,}645\frac{45}{\sqrt{30}} \,\middle|\, \mu = 60\right) =$$

$$= P\left(\overline{X} \le 61{,}49 \,\middle|\, \mu = 60\right) = P\left(\frac{\overline{X}-60}{45/\sqrt{30}} \le \frac{61{,}49-60}{45/\sqrt{30}} \,\middle|\, \mu = 60\right) =$$

$$= P(Z \le 0{,}18) = 1 - \mathrm{P}(Z > 0{,}18) = 1 - 0{,}4286 = 0{,}5714$$

5.6.2 Eskuinetiko aldebakarreko kontrastea

Kontrastatzeko hipotesi nulua H_0: $\mu \le \mu_0$ da, H_1: $\mu = \mu_1 > \mu_0$ alternatiboaren aurka. H_0 errefusatuko dugu, p-balioa α adierazgarritasun-maila baino txikiagoa denean. Hau da, $p = P(Z > z_p) \le \alpha$ denean. Test estatistikoaren eskualde kritikoaren arabera, $Z_p = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \in [z_\alpha, +\infty)$ baldin bada, H_0 errefusatu egingo dugu.

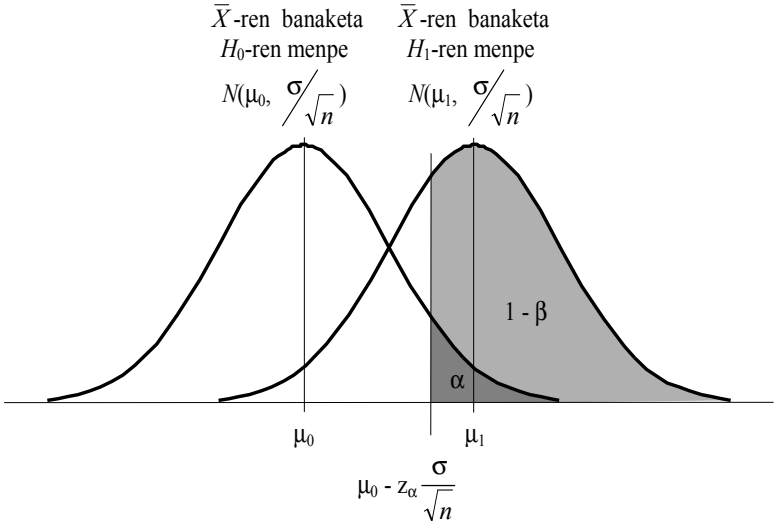
Ahalmena, beraz, honako hau izango da:

$$1 - \beta = P(Z_p \ge z_\alpha \mid \mu = \mu_1) = P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \ge z_\alpha \,\middle|\, \mu = \mu_1\right) =$$

$$= P\left(\overline{X} \ge \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \,\middle|\, \mu = \mu_1\right) =$$

$$= P\left(\frac{\overline{X}-\mu_1}{\sigma/\sqrt{n}} \ge z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_1\right) = P\left(Z \ge z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$$

Kasu horretan, 5.2 irudian agertzen da kontrastearen ahalmenaren adierazpide grafikoa.



5.2 irudia. Bariantza ezaguneko banaketaren batezbestekorako, eskuinetiko aldebateko kontrastearen ($\mu_1 > \mu_0$) ahalmenaren adierazpide grafikoa.

5.5 adibidea

Gai honen 5.2 adibidean, hipotesi alternatiboaren barruan μ_1= 70 minutu aukeratzen badugu, zenbatekoa da kontrastearen ahalmena?

Kasu honetan, eskualde kritikoa $S_1 = [z_{0.05}, +\infty) = [1{,}645, +\infty)$ da.

Beraz, $1 - \beta = P(H_0$ errefusatzeko $\mid H_0$ faltsua bada$) = P(Z_p \geq 1{,}645 \mid \mu = 70) =$

$$P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq 1{,}645 \mid \mu = 70\right) = P\left(\frac{\overline{X} - 45}{45/\sqrt{30}} \geq 1{,}645 \mid \mu = 70\right) =$$

$$= P\left(\overline{X} \geq 45 + 1{,}645\,\frac{45}{\sqrt{30}} \mid \mu = 70\right) =$$

$$P\left(\overline{X} \geq 58{,}52 \mid \mu = 70\right) = P\left(\frac{\overline{X} - 70}{45/\sqrt{30}} \geq \frac{58.52 - 70}{45/\sqrt{30}} \mid \mu = 70\right) = P(Z \geq -1.4)$$

$$= 1 - P(Z > 1{,}4) = 1 - 0{,}0808 = 0{,}9192$$

<u>Oharra</u>: aldebakarreko bi kontrasteak batera hartuta, bientzako da baliagarria honako adierazpen hau:

$$1-\beta = P\left(Z \le -z_\alpha + \frac{|\mu_0-\mu_1|}{\sigma/\sqrt{n}}\right)$$

5.6.3 Aldebiko kontrastea

Kontrastatzeko hipotesi nulua H_0: $\mu = \mu_0$ da, H_1: $\mu = \mu_1$ alternatiboaren aurka.

Kasu horretan, honako hau da ahalmenaren kalkulurako adierazpena:

$$1-\beta = P\left(Z \le -z_{\alpha/2} + \frac{(\mu_0-\mu_1)}{\sigma/\sqrt{n}}\right) + P\left(Z \le -z_{\alpha/2} + \frac{(\mu_1-\mu_0)}{\sigma/\sqrt{n}}\right)$$

5.6 adibidea

Kalkula ezazu 5.3 adibidean egindako kontrastearen ahalmena, hipotesi alternatiboaren balioa 75 minututik 10 minutu (edozein norantzatan) aldentzen dela jakinik.
Aldebiko kontrastearen eskualde kritikoa $S_1 = (-\infty, -z_{0.025}] \cup [z_{0.025}, +\infty) = (-\infty, -1{,}96] \cup [1{,}96\ , +\infty)$ da.

Beraz, $1-\beta = P(H_0$ errefusatzeko $|\ H_0$ faltsua bada$) =$

$P(Z_p \le -1{,}96 \mid \mu = 75-10) + P(Z_p \ge 1{,}96 \mid \mu = 75+10) =$

$$P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le -1{,}96 \,\middle|\, \mu = 65\right) + P\left(\frac{\overline{X}-75}{45/\sqrt{30}} \ge 1{,}96 \,\middle|\, \mu = 85\right) =$$

$$P\left(\frac{\overline{X}-75}{45/\sqrt{30}} \le -1{,}96 \,\middle|\, \mu = 65\right) + P\left(\frac{\overline{X}-75}{45/\sqrt{30}} \ge 1{,}96 \,\middle|\, \mu = 85\right) =$$

$$P\left(\overline{X} \le 75 - 1{,}96\frac{45}{\sqrt{30}} \,\middle|\, \mu = 65\right) + P\left(\overline{X} \ge 75 + 1{,}96\frac{45}{\sqrt{30}} \,\middle|\, \mu = 85\right) =$$

$$= P\left(\overline{X} \le 58{,}9 \,\middle|\, \mu = 65\right) + P\left(\overline{X} \le 91{,}1 \,\middle|\, \mu = 85\right) =$$

$$P\left(\frac{\overline{X}-65}{45/\sqrt{30}} \le \frac{58{,}9-65}{45/\sqrt{30}} \,\middle|\, \mu = 65\right) + P\left(\frac{\overline{X}-78}{45/\sqrt{30}} \ge \frac{91{,}1-85}{45/\sqrt{30}} \,\middle|\, \mu = 85\right) =$$

$$= P(Z \le -0{,}74) + P(Z \ge 0{,}74) = 2\cdot P(Z \ge 0{,}74) = 2\cdot 0{,}2296 = 0{,}4592$$

Oharra. Kontrastearen ahalmenean eragina duten faktoreak dira adierazgarritasun-maila, batez besteko nuluaren eta alternatiboaren arteko diferentzia, eta laginaren tamaina:

1. Adierazgarritasun-maila txikitzen denean, z_α handitu egiten da, eta, orduan, ahalmena txikitu egiten da.
2. Batez besteko alternatiboa nulutik urruntzen baldin bada, alegia, $|\mu_0 - \mu_1|$ handitzen baldin bada, ahalmena handitu egingo da.
3. Laginaren tamaina handitzen denean, ahalmena ere handitu egiten da.

5.7 LAGIN-TAMAINAREN KALKULUA

Orain arte, lagin-tamaina finkoarekin jokatu dugu. Baina, esperimentua diseinatzerakoan, ikertzailearen eskuetan dago laginaren tamaina egokia aukeratzea.

Laginaren tamaina egokia aukeratzerakoan, honako lau faktore hauek izan behar ditugu kontuan: α adierazgarritasun-maila; μ_0 eta μ_1 balioen arteko diferentziaren tamaina; laginketa gauzatu den populazioaren bariantza, eta kontrastearen ahalmena.

Demagun H_0 faltsua dela, hots, $\mu \neq \mu_0$ dela. Logikoki, μ_0 eta μ_1 balioen aldea handia bada errazago nabarituko da, txikia bada baino. Diferentzia handiak detektatzeko ez dugu lagin-tamaina handirik behar; diferentzia txikiak detektatzeko, ordea, lagin-tamaina handiagoa beharko dugu.

Laginaren tamaina aukeratzerakoan, populazioaren aldakortasun-mailak ere eragina du. Populazioak egonkorrak direnean, errazagoa da ondorio zehatzak ateratzea. Zenbat eta populazioaren bariantza handiagoa izan, orduan eta laginaren tamaina handiagoa beharko dugu ondorio zehatzak ateratzeko.

Beste aldetik, α txikitzen denean β handitzen da, eta, beraz, ahalmena txikitzen da. Ahalmena hobetzeko, laginaren tamaina egokitu beharko dugu, aukeraturiko α adierazgarritasun-mailaren arabera.

Azken finean, arazoa honako era honetara plantea daiteke: kontrastearen adierazgarritasun-maila α eta itxarotako batez besteko alternatiboa μ_1 badira, zein da erabili behar dugun laginaren tamaina, $1 - \beta$ probabilitatearekin ezberdintasun adierazgarria aurkitzeko? Galdera horri erantzuteko, hainbat kontraste mota aztertuko ditugu.

5.7.1 Aldebakarreko kontrasteak

Aurreko atalean ikusi dugunez, honako adierazpen hau betetzen da aldebakarreko kontrasteen ahalmenerako: $1 - \beta = P\left(Z \leq -z_\alpha + \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}\right)$.

Alegia, $\beta = P\left(Z > -z_\alpha + \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}\right)$.

Beraz, $z_\beta = -z_\alpha + \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}$ da, eta n isolatuz, honako hau lortuko dugu:

$$n = \frac{\sigma^2 (z_\alpha + z_\beta)^2}{(\mu_0 - \mu_1)^2}.$$

5.7 adibidea

EAEko biztanleek paseo batean ematen duten batezbesteko denbora 45 minutukoa ala ordubetekoa den kontrastatu nahi dugu, % 5eko adierazgarritasun-mailaz eta % 90eko ahalmenaz. Zer tamainatako lagina aukeratu behar dugu?

$\alpha = 0{,}05 \Rightarrow z_\alpha = 1{,}645$

$1 - \beta = 0{,}90 \Rightarrow \beta = 0{,}1 \Rightarrow z_\beta = 1{,}28$

Beraz, aldebakarreko kontrastea denez,

$$n = \frac{\sigma^2 (z_\alpha + z_\beta)^2}{(\mu_0 - \mu_1)^2} = \frac{45^2 (1{,}645 + 1{,}28)^2}{(45 - 60)^2} = 77 \text{ da.}$$

Hots, zoriz aukeraturiko EAEko 77 biztanlek paseo batean ematen duten denbora aztertu beharko dugu, aipaturiko kontrastea emandako baldintzetan gauzatzeko.

5.7.2 Aldebiko kontrastea

Kasu honetan, $n = \frac{\sigma^2 (z_{\alpha/2} + z_\beta)^2}{(\mu_0 - \mu_1)^2}$ dela froga daiteke.

5.8 adibidea

Demagun EAEko biztanleek paseo batean ematen duten batez besteko

denbora 75 minututik 5 minutu (edozein norantzatan) aldentzen den kontrastatu nahi dugula, % 1eko adierazgarritasun-mailaz eta % 95eko ahalmenaz. Zer tamainatako lagina aukeratu behar dugu?

$\alpha = 0{,}01 \Rightarrow z_{\alpha/2} = 2{,}575$

$1 - \beta = 0{,}95 \Rightarrow \beta = 0{,}05 \Rightarrow z_\beta = 1{,}645$

Aldebiko kontrastearen kasuan,

$$n = \frac{\sigma^2 \left(z_{\alpha/2} + z_\beta\right)^2}{\left(\mu_0 - \mu_1\right)^2} = \frac{45^2 \left(2{,}575 + 1{,}645\right)^2}{5^2} = 160{,}28 \approx 161 \text{ da.}$$

Hots, zoriz aukeraturiko EAEko 161 biztanlek paseo batean ematen duten denbora aztertu beharko dugu aipaturiko kontrastea emandako baldintzetan gauzatzeko.

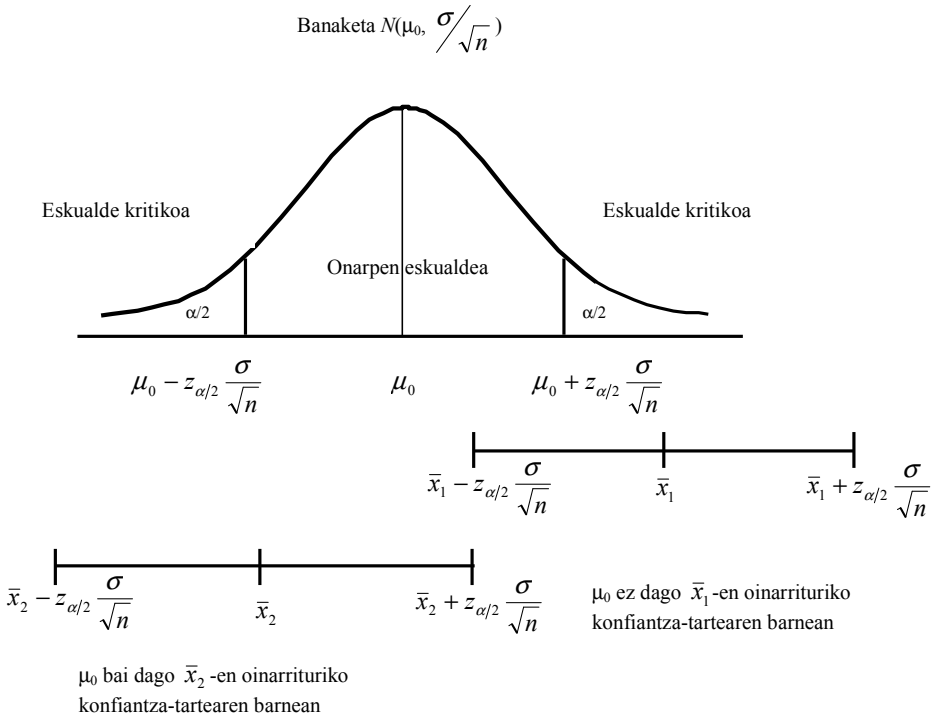
Oharra: edozein kontraste mota izanda ere, hasieran aipatutako faktoreek (adierazgarritasun-maila, ahalmena eta batezbesteko nuluaren eta alternatiboaren arteko distantzia) honako eragin hau dute laginaren tamainan:

1. Adierazgarritasun-maila txikitzen denean, z_α handitu egiten da, eta, orduan, laginaren tamaina ere handitu egiten da.
2. Ahalmena handitzen denean, β txikitu eta z_β handitu egiten da. Beraz, laginaren tamaina ere handitu egiten da.
3. Batez besteko alternatiboa nulutik urruntzen baldin bada, alegia, $|\mu_0 - \mu_1|$ handitzen baldin bada, laginaren tamaina txikitu egiten da.

5.8 KONFIANTZA-TARTEAREN ETA HIPOTESI-KONTRASTEAREN ARTEKO ERLAZIOA

Konfiantza-tarteen teknika eta hipotesi-kontrastearen teknika bi era dira populaziotik ateratako laginak emandako datuetan oinarrituriko populazioari buruzko ondorioak ateratzeko. Populazioaren parametroa estimatzeko, bi erak dira aplikagarriak, eta elkarren osagarri dira.

Demagun H_0: $\mu = \mu_0$ hipotesia H_1: $\mu = \mu_1$ aukeraren aurka dugula. α adierazgarritasun-maila erabiliz, H_0 errefusatuko dugu baldin eta soilik baldin μ-rako %100 $\times$ $(1 - \alpha)$ konfiantza-mailako tarteak ez badauka μ_0 bere barnean; eta ez dugu H_0 errefusatuko, baldin eta soilik baldin konfiantza-tarte horrek μ_0 bere barnean badauka. Jokabide hori 5.3 irudian erakusten da grafikoki.



5.3 irudia. Aldebiko hipotesi-kontrastearen eta konfiantza-tartearen arteko erlazioa

5.9 adibidea

Laugarren gaiko 4. 7 adibidean ikusi dugunez, gai horren 4.3 taulan agertzen diren datu guztiak lagin bakartzat harturik, EAEko biztanleek paseo batean ematen duten batez besteko denborarako % 95eko konfiantza-tartea (55,73 , 87,93) da.

H_0 : μ = 75 hipotesi nulua H_1 : $\mu \neq 75$ hipotesi alternatiboaren aurka kontrastatu nahi badugu, honako arrazoibide hau egingo dugu: $\mu_0 = 75$ balioa konfiantza-tarte horren barruan dagoenez, H_0 : μ = 75 hipotesi nulua onar dezakegu % 5eko adierazgarritasun-mailaz.

Gogora dezagun hipotesi-kontraste hori baliozkoa dela populazioa normala denean, edo, normala ez izanik, laginaren tamaina $n \geq 30$ denean, baina beti populazioaren bariantza ezaguna izanik. Ondoren, bariantza ezagutzen ez dugunerako eratuko dugu populazioaren batezbestekorako hipotesi-kontrastea.

5.9 BARIANTZA EZEZAGUNEKO POPULAZIO NORMALAREN BATEZBESTEKORAKO HIPOTESI-KONTRASTEA

Populazioaren bariantza ezezaguna baldin bada, bi kasu bereiziko ditugu: $n \geq 30$ edo $n < 30$ denean.

5.9.1 Lagin-tamaina handia: $n \geq 30$

Kasu honetan, batezbestekorako hipotesi-kontrastea gauzatzeko erabiltzen den test estatistikoa $Z_p = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}}$ da, eta bere banaketa $N(0,1)$ da gutxi gorabehera. Beraz, 5.4 atalean azaldutakoaren arabera egiten da hipotesi-kontrastea, kontuan izanik σ-ren ordez s (laginaren kuasibariantzaren erro karratua) erabili behar dugula.

5.9.2 Lagin-tamaina txikia: $n < 30$

Kasu honetan, nahitaez, populazioak normala izan behar du, hots, X: $N(\mu,\sigma)$. Populazioaren batezbestekorako hipotesi-kontrastea gauzatzeko erabiltzen den test estatistikoa $T_p = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}}$ da, eta beronen banaketa t_{n-1} da, hots, $(n - 1)$ askatasun graduko Student-en t banaketa.

5.9.2.1 Aldebakarreko kontrasteak

5.9.2.1.1Ezkerretikoa:

Honako hau da hipotesi-kontrastea: H_0: $\mu \geq \mu_0$

H_1: $\mu < \mu_0$

H_0 egiazkoa dela suposatuz, test estatistikoaren balioa kalkulatzen da: $t_p = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$

Kasu horretan, honako hau da p-balioa: $p = P(t_{n-1} < t_p)$

5.10 adibidea

Laugarren gaiko 4.3 taulan agertzen den Gipuzkoako lagina erabiliz, gauza ezazu, % 5eko adierazgarritasun-mailaz, Gipuzkoako biztanleek paseo batean

ematen duten batez besteko denborari buruzko honako hipotesi-kontraste hau:

$H_0 : \mu \geq 80$
$H_1 : \mu < 80$

Desbideratze estandarra ezezaguna eta lagin-tamaina txikia direnez, hauxe izango da test estatistikoaren balioa:

$$t_p = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{73{,}5 - 80}{30{,}37/\sqrt{10}} = -0{,}68$$

Orduan, $p = P(t_{n-1} < t_p) = P(t_9 < -0{,}68) = P(t_9 > 0{,}68) > P(t_9 > 0{,}883) = 0{,}2 > 0{,}05 \Rightarrow$ % 5eko adierazgarritasun-mailaz ezin da H_0 errefusatu; hau da, % 5eko adierazgarritasun-mailaz onar dezakegu Gipuzkoako biztanleek paseo batean ematen duten batezbesteko denbora 80 minutukoa edo hori baino luzeagoa dela.

5.9.2.1.2 Eskuinetikoa

Hipotesi-kontrastea honako hau da: H_0: $\mu \leq \mu_0$

H_1: $\mu > \mu_0$

H_0 egiazkoa dela suposatuz, test estatistikoaren balioa kalkulatuko dugu: $t_p = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$.

Kasu honetan, p-balioa $p = P(t_{n-1} > t_p)$ da.

5.11 adibidea

Aurreko adibideko datuak erabiliz, % 1eko adierazgarritasun-mailaz onar al dezakegu Gipuzkoako biztanleek paseo batean ematen duten batezbesteko denbora ordubete baino gehiago denik?

Hauexek dira kontrastatu behar ditugun hipotesiak: $H_0 : \mu \leq 60$
$H_1 : \mu > 60$

Kasu horretan, test estatistikoaren balioa $t_p = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{73{,}5 - 60}{30{,}37/\sqrt{10}} = 1{,}41$ da.

Orduan, $p = P(t_{n-1} > t_p) = P(t_9 > 1{,}41) > P(t_9 > 1{,}833) = 0{,}05 > 0{,}01 \Rightarrow$ % 1eko adierazgarritasun-mailaz, ezin dugu H_0 errefusatu; hau da, Gipuzkoako biztanleek paseo batean ematen duten batez besteko denbora ordubete baino gehiago izatea errefusatzen dugu, % 1eko adierazgarritasun-mailaz.

5.9.2.2 Aldebiko kontrastea

Kontrastatzeko hipotesiak honako hauek dira: H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$

H_0 egiazkoa dela suposatuz, test estatistikoaren balioa kalkulatuko dugu: $t_p = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$.

Kasu horretan, honako hau da p-balioa: $p = 2 \cdot P(t_{n-1} > | t_p |)$

5.12 adibidea

Adibide berbereko datuak erabiliz, % 5eko adierazgarritasun-mailaz, onar al dezakegu Gipuzkoako biztanleek paseo batean ematen duten batez besteko denbora 75 minutukoa denik?

Honako hipotesi-kontraste hau gauzatuko dugu: $H_0 : \mu = 75$
$H_1 : \mu \neq 75$

Test estatistikoaren balioa, $t_p = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{73{,}5 - 75}{30{,}37/\sqrt{10}} = -0{,}16$ da.

Kasu horretan, $p = 2 \cdot P(t_{n-1} > | t_p |) = 2 \cdot P(t_9 > 0{,}16) > 2 \cdot P(t_9 > 0{,}261) = 2 \cdot 0{,}4 = 0{,}8 > 0{,}05 \Rightarrow$ % 5eko adierazgarritasun-mailaz ezin da H_0 errefusatu; hau da, Gipuzkoako biztanleek paseo batean ematen duten batez besteko denbora 75 minutukoa dela onartuko dugu, % 5eko adierazgarritasun-mailaz.

Oharra. Laginaren tamaina handia baldin bada, $n \geq 30$ hain zuzen ere, honako hau da test estatistikoaren banaketa normala izatearen arrazoia: n handitzean, t aldagaia $N(0,1)$ normal tipifikatura hurbiltzen da. Kontura zaitez, Student-en t banaketaren taulan $n = \infty$ agertzen den errenkadan, aldagai normal tipifikatuari dagozkion balioak agertzen direla.

Orain arte, batezbestekoari buruzko hipotesiak kontrastatu ditugu; baina batezbestekoa ez da kontrasta dezakegun populazioaren parametro bakarra. Ikus dezagun nola eratzen diren populazioaren bariantzari buruzko hipotesi-kontrasteak.

5.10 BANAKETA NORMALAREN BARIANTZARAKO HIPOTESI-KONTRASTEA

Hauexek dira, kasu honetan, kontrastatu nahi ditugun hipotesiak:

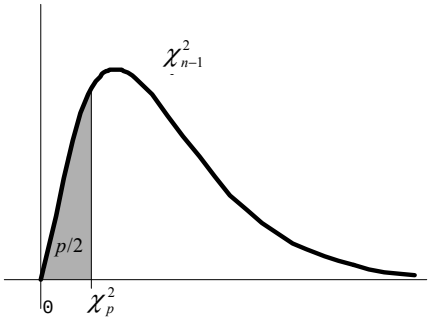
H_0: $\sigma^2 = \sigma_0^2$
H_1: $\sigma^2 \neq \sigma_0^2$

Kontraste hori gauzatzeko, honako test estatistiko hau erabiliko dugu: $\frac{(n-1)\cdot S^2}{\sigma_0^2}$, zeina, H_0 egiazkoa izatearen baldintzapean, χ^2_{n-1} $(n-1)$ askatasun-graduko khi karratuaren banaketari baitarraio. Lagin batean beronek hartzen duen balioa $\chi^2_p = \frac{(n-1)\cdot s^2}{\sigma_0^2}$ adieraziko dugu.

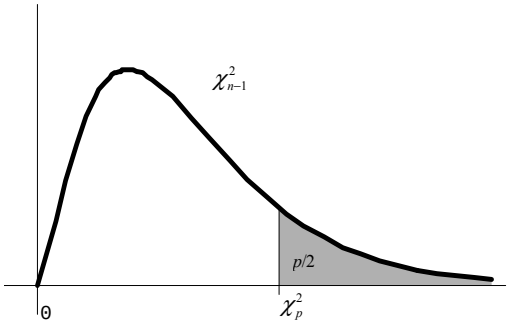
Test estatistikoari dagokion *p*-balioa kalkulatzeko, khi karratu banaketa simetrikoa ez denez, bi kasu bereizi behar ditugu, eta *p*-balioa honako era honetara kalkulatuko dugu:

Baldin eta $s^2 \leq \sigma_0^2$ bada, $p = 2\cdot P(\chi^2_{n-1} < \chi^2_p)$

Baldin eta $s^2 > \sigma_0^2$ bada, $p = 2\cdot P(\chi^2_{n-1} > \chi^2_p)$



$s^2 \leq \sigma_0^2$ bada, *p*-balioa = 2· Azalera marraztua



$s^2 > \sigma_0^2$ bada, *p*-balioa = 2· Azalera marraztua

5.4 irudia. Banaketa normalaren bariantzarako, aldebiko kontrastearen p-balioaren adierazpide grafikoa

Aurrefinkaturiko α adierazgarritasun-maila baino p-balio txikiagoa ateratzen denean, H_0 hipotesi nulua errefusatu egingo dugu, eta $\sigma^2 \neq \sigma_0^2$ onar daitekeela esango dugu.

5.13 adibidea

Laugarren gaiko 4. 12 adibideko datuak erabiliz, onar al dezakegu EAEko biztanleak txirrinduz ibilitako denboraren desbideratze estandarra ordubetekoa denik, % 5eko adierazgarritasun-mailaz?

Hauexek dira kontrastatuko ditugun hipotesiak:
$H_0 : \sigma = 60 \Leftrightarrow \sigma^2 = 60^2$
$H_1 : \sigma \neq 60 \Leftrightarrow \sigma^2 \neq 60^2$

Honako hau izango da test estatistikoaren balioa:

$$\chi_p^2 = \frac{(n-1)\cdot s^2}{\sigma_0^2} = \frac{19\cdot 5019{,}74}{60^2} = 26{,}49$$

Kasu honetan, $p = 2\cdot P(\chi_{n-1}^2 > \chi_p^2) = 2\cdot P(\chi_{n-1}^2 > 26{,}49) > 2\cdot P(\chi_{n-1}^2 > 27{,}204) = 2\cdot 0{,}1 = 0{,}2 > 0{,}05 \Rightarrow$ % 5eko adierazgarritasun-mailaz ezin da H_0 errefusatu; hau da, EAEko biztanleak txirrinduz ibilitako denboraren desbideratze estandarra ordubetekoa dela onartzen dugu, % 5eko adierazgarritasun-mailaz.

<u>Oharra</u>: aldebakarreko kontrasteak ere eraiki ditzakegu, nahiz eta hain erabiliak ez izan:

- Ezkerretiko aldebakarreko kontrastea: H_0: $\sigma^2 \geq \sigma_0^2$

  H_1: $\sigma^2 < \sigma_0^2$

  Test estatistikoa $\chi_p^2 = \dfrac{(n-1)\cdot s^2}{\sigma_0^2}$ da, eta dagokion p-balioa: $p = P(\chi_{n-1}^2 < \chi_p^2)$.

- Eskuinetiko aldebakarreko kontrastea: H_0: $\sigma^2 \leq \sigma_0^2$

  H_1: $\sigma^2 > \sigma_0^2$

  Test estatistikoa $\chi_p^2 = \dfrac{(n-1)\cdot s^2}{\sigma_0^2}$ da, eta dagokion p-balioa: $p = P(\chi_{n-1}^2 > \chi_p^2)$.

5.11 ARIKETAK

5.1. Pago-azalera aztertzean, hipotesi-kontrastearen bidez erantzun nahi dugu 4.13 ariketan egiten den azken galdera (ea hamar herri horiek EAE osoaren adierazgarriak diren).

a) Zein dira hipotesi nulua eta aternatiboa?

b) Zenbatekoa da kontrastearen *p*-balioa?

c) Horren arabera, onar al daiteke zoriz aukeraturiko hamar herri horiek EAE osoaren adierazgarri direnik, % 5eko adierazgarritasun-mailaz?

5.2. 2000. urteko otsaileko batez besteko gehieneko tenperatura 1998. urtekoa baino txikiagoa denetz baieztatu nahi dugu, 4.9 ariketan emandako datuetan oinarrituz.

a) Eraiki itzazu baieztapen hori kontrastatzeko beharrezkoak diren hipotesi nulua eta aternatiboa.

b) Zenbatekoa da kontrastearen *p*-balioa?

c) Horren arabera, onar al dezakegu 2000. urteko otsaileko batez besteko gehieneko tenperatura 1998. urtekoa baino txikiagoa dela, % 5eko adierazgarritasun-mailaz? Eta adierazgarritasun-maila 0.01 izango balitz?

5.3. EAEn, 1996. urtean lehorreraturiko antxoa kantitatea aztertzean, 4.16 ariketako datuetan oinarrituko gara. Era itzazu honako galdera hauek erantzuteko beharrezkoak diren hipotesi-kontrasteak:

a) Europar Batasunak EAEri ezarritako lehorreratzeko antxoa-kupoa gainditzen denetz.

b) EAEn lehorreraturiko antxoa kantitatearen desbideratze estandarra 1.000 tona denetz.

c) Aurreko ataletako zein hipotesi errefusatuko ditugu, % 5eko adierazgarritasun-mailaz, 4.16 ariketan emandako datuen arabera?

5.4. EAEko etxebizitzen erosotasun-indizearen azterketan, 4.12 ariketaren datuak erabiliz, onar al daiteke, % 1eko adierazgarritasun-mailaz, EAEko etxebizitzen erosotasun-indizea egokia denik? Azal ezazu erantzuna *p*-balioa erabiliz.

5.5. Onar al dezakegu EAEn dagoen eukaliptoko batez besteko azalera-proportzioa, 4.14 ariketan emandako datuen arabera, beste espezieak landatzeko kaltegarria denik, % 5eko adierazgarritasun-mailaz? Azal ezazu erantzuna *p*-balioa erabiliz.

5.6. EAEko familien batez besteko kide kopurua 1991tik 1996ra gutxitu egin dela egiaztatu nahi dugu hipotesi-kontrastearen bidez, 4.11 ariketan emandako datuetan oinarrituz.

a) Eraiki itzazu egiaztapen hori kontrastatzeko beharrezkoak diren hipotesi nulua eta aternatiboa.

b) Zenbatekoa da kontrastearen *p*-balioa?

c) Adierazgarritasun-maila 0,05 bada, onar al dezakegu EAEko familien batez besteko kide kopurua gutxitu egin denik 1991tik 1996ra? Eta % 1eko adierazgarritasun-mailaz?

5.7. 1.15 ariketan azaldu dira EAEn dauden behatoki plubiometrikoetan 1996. urtean bildutako datuak. Datu horiek erabiliz eta % 1eko adierazgarritasun-mailaz, kontrasta itzazu honako baieztapen hauek:

a) Euria izandako batez besteko egun kopurua 150 baino txikiagoa da.

b) Ekaitza izandako batez besteko egun kopurua hamar baino handiagoa da.

c) Graduak zero azpitik izandako batez besteko egun kopurua hogei da.

5.8. Adierazgarritasun-maila % 5 izanik, errefusa al dezakegu zuhaitzen batez besteko kalibrea, 4.18 ariketan emandako datuen arabera, 2 cm-koa denik? Azal ezazu erantzuna *p*-balioa erabiliz.

5.9. Ullibarri urtegian 1997. urtean zehar bildutako ur-bolumena (Hm^3-tan) dugu aztergai, eta 4.5 ariketan emandako datuak erabiliko ditugu.

a) Datu horien arabera, errefusa al dezakegu bildutako batez besteko ur-bolumena 100 Hm^3 baino txikiagoa izan denik, % 5eko adierazgarritasun-mailaz?

b) Eta ur-bolumenaren desbideratze estandarra 15 Hm^3 denik?

Azal itzazu erantzunak *p*-balioak erabiliz.

5.10. 4.8 ariketan azaldu dira kutsadurari buruzko zenbait datu. Hipotesi-kontrasteen bidez eta datu horiek erabiliz, erantzun itzazu honako galdera hauek:

a) Bilbon 1995ean zegoen SO_2 sufre-dioxidoaren batez besteko kantitatea 20 mg/Nm^3 baino handiagoa al zen?

b) Bilbon 1995ean zegoen SO_2 sufre-dioxidoaren kantitatearen desbideratze estandarra 5 mg/Nm^3 baino txikiagoa al zen?

Azal itzazu erantzunak *p*-balioak erabiliz.

5.11. HIESaren birusaren presentzia hautemateko proba egiterakoan, honako test hau egin zuten:

H_0 : Birusik ez dago.

H_1 : Birusa dago.

Azal ezazu zer gertatu den I motako errorea eginez gero. Zer gertatu da II motako errorea eginez gero?

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994)

5.12. EUSTATek argitaraturiko kutsadurari buruzko datuen arabera, 1995ean Bilbon zegoen SO_2 sufre-dioxidoaren batez besteko kantitatea 20 mg/Nm^3 zela onartzen da. Ingurumena babesteko hartutako neurriak direla kausa, 2000. urtean batez besteko kantitate hori gutxitu izana espero da.

a) Eraiki itzazu baieztapen hori defendatzeko hipotesi nulu eta alternatibo egokiak.

b) Azal itzazu praktikoki I motako errorea eta II motako errorea jaulkitzearen ondorioak.

5.13. Emakume gazte osasuntsuen eguneroko batez besteko elikagai-kontsumoa 2.300 kcal-koa dela iragarri da. Erregistraturiko desbideratze estandarra 237 kcal-koa da. Azterketa hori lagin txiki batean oinarritzen da, beraz, ikerketa errepikatzea interesatzen zaigu. Zer tamainatako lagina aukeratu beharko da, erregistraturiko batezbestekotik 100 kcal (edozein norantzatan) aldentzen den batezbestekoa detektatzeko, $\alpha = 0.1$ eta ahalmena 0.9 izanik? (Iradokizuna: aldebiko kontrastea da). (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994)

5.14. Egunero pururen bat erretzen duten EAEko pertsonek erretako batez besteko puru kopurua aztertu nahi dugu. Horretarako, 4.15 ariketaren b) atalean azaldutako datuak erabiliko ditugu.

a) Onar al dezakegu egunero batez beste bi puru baino gehiago erretzen dituztenik, % 1eko adierazgarritasun-mailaz? Azal ezazu erantzuna, *p*-balioa erabiliz.

b) Noiz jaulkiko dugu I motako errorea?

c) Eta II motakoa?

d) Lehenengo ataleko kontrastea gauzatu nahi dugu, eta $\mu = 3$ hipotesi alternatiborako II motako errorea jaulkitzeko probabilitatea 0.01 da. Zenbat pertsona aukeratu beharko ditugu, egunero pururen bat erretzen dutenen artean?

5.15. Pertsonen altueraren azterketarako, 4.7 ariketan agertu den taula erabiliko dugu.

a) Onar al dezakegu EAEko biztanleen batezbesteko altuera 165 cm baino handiagoa denik, % 1eko adierazgarritasun-mailaz? Azal ezazu erantzuna *p*-balioa erabiliz.

b) Hipotesi alternatiboan aukeratzen dugun balioa 166 cm bada, zenbatekoa da kontrastearen ahalmena?

c) Hipotesi alternatiboa μ = 166 cm izanik, kontrastearen ahalmena 0,95 izan dadin, zenbatekoa da aukeratu behar dugun laginaren tamaina?

5.16. Sondikako aireportuan, 1998. urtean zehar, hilero neurturiko euri kantitateak (mm^3-tan) 65,7 – 43,3 - 43,3 – 177,4 – 90,5 – 38,2 – 39,1 – 62,4 – 109,8 – 311,3 – 173,6 – 57,7 izan ziren. (Iturria: EUSTAT. *EUE 97/98*).

Jausitako euri kantitatea 100 mm^3-ko desbideratze estandarra duen banaketa normalari darraiola suposatzen da.

a) Adierazgarritasun-maila 0,05 izanik, onartuko al dugu edozein urtetan jausitako batez besteko euri kantitatea 100 mm^3 baino txikiagoa denik?

b) Kontrastearen ahalmena % 95 izateko, hipotesi alternatiboan zein balio hartu beharko genuke jausitako batez besteko euri kantitatetzat?

c) Hipotesi alternatiboan μ = 90 mm^3 hartzean, zenbat neurketa egin beharko genituzke kontrastearen adierazgarritasun-maila 0,05 eta ahalmena 0,95 izan daitezen?

5.17. EAEko pertsonen osasun-egoera orokorra adierazten duen bizi-kalitate indizearen banaketa $N(66,18)$ dela adierazi dugu 4.2 ariketan. Bertan azaltzen denez, 74,9 izan zen zoriz aukeraturiko hamar pertsonaren batez besteko bizi-kalitate indizea.

a) Bizi-kalitate indizeari dagokionez, hamar pertsona horiek EAEko biztanleen adierazgarri al dira? Zenbateko adierazgarritasun-mailaz?

b) Egindako hipotesia indize horren batezbestekoa 75 den hipotesiarekin kontrastatzen badugu, % 5eko adierazgarritasun-mailaz, zer probabilitate dago II motako errorea jaulkitzeko?

c) Aurreko kontrastea gauzatu nahi dugu, II motako errorea jaulkitzeko probabilitatea 0.05 izanik. Zenbat pertsona aukeratu beharko ditugu zoriz?

5.18. Zelula-mintzaren lodiera aztertzeko, 4.19 ariketan emandako datuak erabiliko ditugu, lodieraren banaketa normala dela suposatuz.

a) Onartuko al dugu zelula-mintzaren batez besteko lodiera 75 mm baino handiagoa denik, % 5eko adierazgarritasun-mailaz?

b) Adieraz ezazu egindako kontrastearen ahalmena.

c) Hipotesi alternatiboan 85 mm balioa aukeratzen badugu, zenbatekoa da kontrastearen ahalmena?

d) Lehenengo ataleko kontrastea gauzatu nahi dugu, hipotesi alternatiboaren balioa μ = 80 mm izanik. Ahalmena 0,99 izateko, zenbat espezie aztertu beharko ditugu?

5.19. EAEko mendien batez besteko altitudea 965 metrokoa dela adierazi dugu 4.10 ariketan. Bertan emandako datuen arabera, onar al dezakegu baieztapen hori, % 5eko adierazgarritasun-mailaz?

5.20. Orain arte, adineko pazienteek hamazazpi egun egon behar zuten klinikan, batez beste, etxeratu baino lehen. Programa berri batek egonaldi hori murriztea espero dute. Honako hauek dira zoriz aukeraturiko hamasei pazientek osaturiko laginean lortutako klinikako egun kopuruak:

3	5	12	7	22	6	2	18
9	8	20	15	3	36	38	43

Frogatzen al dute datu horiek ikerketako hipotesia, 0,05eko adierazgarritasun-mailaz? Jo dezagun egonaldia banaketa normalari darraiola.

5.12 ERANSKINA

(*Iturria*: J.A.G. Sterne, G. Davey Smith. *Sifting the evidence—what's wrong with significance tests?*. BMJ, 322 lib., 2001eko urtarrilak 27)

5.12.1 *p*-balioaren eta adierazgarritasun-testen historia-laburpena

Duela 60 urte baino gehiago inferentzia estatistikoaren sortzaileen artean sortu zen eztabaidatik dator hipotesi-testen egungo praktikan dagoen nahasketa.

R.A. Fisher-ek sortu zuen adierazgarritasun-testen ideia. Demagun sendagai berriak bihotzekoa izan ondorengo biziraupen-maila emendatzen duen egiaztatu nahi dugula. Sendagai berriarekin trataturiko gaixo talde bat aztertu dugu, eta plazeboarekin trataturiko beste talde batekin konparatu dugu. Sendagai berriarekin trataturikoen artean hilkortasun-maila erdia dela behatu dugu, plazeboarekin trataturikoen aldean. Egindako hipotesia berresten al du horrek, edo kasualitatea izan daiteke? Galdera horri erantzuna emateko *p*–balioa kalkulatuko dugu; hau da, sendagaiak eraginik ez duela suposatuz, bizirik dirauen gaixo kopurua bikoitza baino handiagoa izateko probabilitatea.

Fisher-ek hipotesi nuluaren aurrean (gure adibidean, sendagai berriak eraginik ez izatea) datuek azaltzen duten neurritzat hartu du *p*–balioa. Haren ustez, $p < 0{,}05$ (% 5eko adierazgarritasun-maila) estandar gisa hartu behar da kontrastaturiko hipotesia egokia ez dela nabarmentzeko, baina ez da arau absolutua. Horrela zioen: «*p*-ren balioa 0,1 eta 0,9 balioen artean baldin badago, ez dugu hipotesi nuluarekiko susmorik izan behar. *p* probabilitatea 0,02 baino txikiagoa bada, argi eta garbi adierazten digu hipotesia ez dela egokia errealitatea azaltzeko. Horrelako erabakia hartzeko 0,05 maila hartzen badugu, ez gara askotan tronpatuko». Azken finean, Fisher-ek azpimarratzen duenez, ikertzailearen eskuetan dago *p*–balioaren interpretazioa. Adibidez, *p*–balioa gutxi gorabehera 0,05 bada, pentsa genezake hipotesi nulua egokia den ala ez, baina dudarik gabe beste esperimentu bat egitera eramango gaitu.

Aurreko metodoari guztiz lotuta dagoen interpretazio subjektiboa errefusatzeko asmoz, Neyman-ek eta Pearson-ek «adierazgarritasun-testen» ordez «hipotesi-testak» izendatzea proposatu zuten. Neyman-ek eta Pearson-ek, esperimentuaren emaitzak interpretatzean bi motako erroreak jaulki ditzakegula argudiatzen dute. Fisher-en metodoak, I motako errorearen probabilitatea —hau da, hipotesi nulua (sendagaiak ez du eraginik) egiazkoa denean, hura errefusatzeko probabilitatea— soilik hartzen du kontuan. Neyman eta Pearson II motako errorearen probabilitateaz ere kezkati ziren, hots, hipotesi nulua (beraz, sendagai berriak huts egin du) faltsua denean, hura onartzeko probabilitateaz. Bi errore horien probabilitateen mugak finkatzen baditugu, mugatua izan daiteke esperimentu kopuru handia gauzatzean jaulkitako akats kopurua.

Neyman-ek eta Pearson-ek ziotenez, «probabilitate-teorian oinarrituriko testek soilik eman dezakete hipotesiaren egiazkotasunaren ala faltsutasunaren nabaritasuna. Baina testen helburua beste ikuspuntu batetik azter dezakegu. Hipotesi bakoitza egiazkoa ala faltsua den jakitearen zain egon gabe, gure jokaera bideratzeko arauak bila ditzakegu, kontuan hartuz gure asmoa dela, esperientzian zehar, akats asko egingo ez dugun ziurtasuna izatea».

Horrela, Neyman-en eta Pearson-en metodoan, gure esperimentuaren emaitzak interpretatzeko erabaki-araua finkatzen dugu; hau da, hipotesi nulua noiz onartuko eta noiz errefusatuko dugun. Fisher-en ikuspuntu subjektiboagoaren aurrean, Neyman-ek eta Pearson-ek ez dute *p*–balioa erabiltzen, esperimentu indibidual batean, hipotesi nuluaren egokitasuna neurtzeko.

Neyman-en eta Pearson-en metodoa erabili ahal izateko, hipotesi alternatiboa zehaztu behar dugu. Beste era batera esanda, ez da nahikoa sendagai berriak eragina duela esatea, baizik eta zenbateko eragina duen ere esan behar dugu (esate baterako, sendagai berriak % 60an gutxitzen duela hilkortasuna). Ikertzaileak, hipotesi alternatiboa eta I eta II motako erroreen probabilitateak zehaztuz, erabaki-araua alda dezake; baina, aldi berean, aurrera egin dezake esperimentuan. Zoritxarrez, ikertzaileek helburu horiek betetzeko oztopoak dituzte eta, beren ikerketak gauzatu edo azterketak zehaztu baino lehen, oso gutxitan hartzen dute kontuan, hipotesi alternatiboaren barruan, tratamenduaren eraginaren balio zehatzik. Guztiz onartuta dago, aldiz, Neyman-en eta Pearson-en metodoaren alderik errazena; hau da, $p < 0{,}05$ bada (I motako errorearen probabilitatea = 0,05), hipotesi nulua errefusa dezakegula. Horren ondorioz, Neyman-en eta Pearson-en metodoak Fisher-en metodoaren antzekoa dela ematen du, nahiz eta filosofia ezberdinak izan.

5.12.2 *p*-balioen gaizki-ulertzeak

Gaizki-ulertzerik zabalduena honetan datza: *p*-balioa hipotesi nulua egiazkoa izateko probabilitatetzat hartzeak eta, beraz, emaitza adierazgarriak adieraziko luke hipotesi nulua egiazkoa izatea gertagaitzagoa dela. Ideia hori faltsua dela frogatuko dugu.

Demagun izatez faltsuak diren hipotesi nuluen proportzioa % 10 dela; hau da, kontrastatzen ditugun hipotesien % 90 ez da zuzena. Beste aldetik, kontrastearen ahalmena % 50 dela suposatuko dugu.

Orain, jo dezagun 1.000 saio egiten ditugula hipotesia kontrastatzeko eta, $p < 0{,}05$ bada, hipotesi nulua errefusatzen dugula. Faltsuak diren hipotesi nuluen proportzioa % 10 denez, 100 saiotan hipotesi nulua izatez faltsua izango da. II motako errorearen probabilitatea % 50 denez, 100 saio horietako 50ean errefusatuko dugu hipotesi nulua. Hipotesi nulua egiazkoa den 900 saioetan, adierazgarritasun-maila % 5 izanik, 45 saiotan errefusatuko dugu hipotesi nulua.

Esperimentuaren ondorioa	Egiazkoak diren hipotesi nuluak	Faltsuak diren hipotesi nuluak	Totala
Onartutako hipotesi nuluak	855	50	905
Errefusatutako hipotesi nuluak	45	50	95
Totala	900	100	1000

5.2 taula. Izandako usteen arabera, hipotesi nulua onartutako eta errefusatutako aldi kopuruak.

Beraz, 5.2 taulan ikusten dugunez, 95 saioak adierazgarriak dira eta horietako 45etan (% 47an) hipotesi nulua egiazkoa da, hau da, «alarma faltsuak» dira (H_0 errefusatu dugu eta ez genuela horrela egin behar).

5.12.3 Zer adierazten du adierazgarritasunak?

Praktikan, Medikuntzan eta Biologian erabilitako estatistikan, emaitzak adierazgarri ala ez-adierazgarri sailkatu izan dira nagusiki, eta II motako errorearen probabilitatea ez da ia-ia kontuan hartu. Horrela izatetik, honako bi ondorio arrunt eta potentzialki larri hauek atera daitezke: bata, lagin txikietan behaturiko diferentziak, klinikoki garrantzitsuak izan daitezkeenak, ez-adierazgarritzat hartuak izatea, eta, beraz, baztertuak izatea; bestea, laginaren tamaina handia dela-eta lortutako emaitza adierazgarriak tratamenduaren benetako eraginaren ondoriotzat hartuak izatea.

Beste aldetik, laginaren tamaina edo neurrien zehaztasuna emendatuz, hipotesi-kontrasteen ahalmena emenda dezakegu. Eta 5.3 taulan ikus dezakegunez, hautatutako *p*-balio bakoitzerako, ahalmena emendatzean, nabarmen gutxituko da positibo faltsuak diren emaitza «adierazgarrien» proportzioa. Halaber, hipotesi faltsuen proportzioaz oso ezkorrak ez bagara behintzat, 5.3 taulak azaltzen diguna da 0.001 baino txikiagoak diren *p*-balioek hipotesi nulua egokia ez deneko nabaritasun handia adierazten digutela.

Hipotesi-testen ahalmena	$p = 0,05$	$p = 0,01$	$p = 0,001$
% 80 (hipotesi nulu faltsuak)			
20	5,9	1,2	0,10
50	2,4	0,5	0,05
80	1,5	0,3	0,03
% 50 (hipotesi nulu faltsuak)			
20	20,0	4,8	0,50
50	9,1	2,0	0,20
80	5,9	1,2	0,10
% 10 (hipotesi nulu faltsuak)			
20	69,2	31,0	4,30
50	47,4(*)	15,3	1,80
80	36,0	10,1	1,10
% 1 (hipotesi nulu faltsuak)			
20	96,1	83,2	33,10
50	90,8	66,4	16,50
80	86,1	55,3	11,00

(*) 5.2 taularen suposizioari dagokio.

5.3 taula. Positibo faltsuak diren emaitza «adierazgarrien» proportzioa, hiru adierazgarritasun-mailaren arabera.

5.12.4 Zer egin?

Hiru era daude adierazgarritasun-testen praktikan dagoen nahasketa-maila gutxitzeko. Lehenengoz, taulak adierazten digunez, kontuan hartzea $p < 0,05$ izatea ez dela hipotesi nulua errefusatzeko erabateko irizpidea, erabaki hori ahalmenaren menpe ere bai baitago. Bigarrenez, nabaria da kontrastaturiko hipotesien arteko hipotesi adierazgarrien proportzioa emandatzeak nahasketa-maila gutxi dezakeela ere. Zoritxarrez, oso zaila da hori aurrera eramatea. Hipotesiak formulatzea ez da nahikoa. Alde batetik, eraginik ez duten ehun zorizko proba baldin baditugu, bakoitzean hipotesi-kontraste bakar bat gauzatuko dugu. Eta bestetik, ez da posible izango balizko hipotesi guztiak kontrastatzea. Hirugarrenez, akatsik larriena honako hau litzateke: paradigma estatistikoak aldatu ordez, saioen kalitatea hobetzeko laginaren tamaina eta neurrien zehaztasuna emendatzea.

Soluziobide erraza edo bakarra ez dagoen bitartean, hipotesi-testen emaitzek gu nahasteko arriskua murrizteko biderik egokiena honetan datza: azterketa estatistikoen aurkezpenetan, *p*-balioen ordez, edo horiez gainera, konfiantza-tarteak sartzea. Baina tarte horiek ezin ditugu erabili adierazgarritasun-maila aztertzeko (5.8 atalean adierazi dugun erara) bide laguntzaile huts moduan, baizik eta tarteen balioen heinek izan ditzaketen ondorioak aztertzeko era gisa. Hala ere, % 95eko konfiantza-tarteek, inplizituki, % 5eko adierazgarritasun-muga erabiltzen dutenez, nahasketa handiago sor dezakete. Horrexegatik, konfiantza-mailatzat % 90 hartzea gomendatzen da, % 95 hartu ordez.

Adibide gisa, gai honen 5.2 adibidea berregingo dugu gomendaturiko erara.

5.14 adibidea

EAEko biztanleek paseo batean ematen duten batezbesteko denbora 45 minutukoa baino luzeagoa den kontrastatu nahi dugu.

Hauxe da gauzatu nahi dugun hipotesi-kontrastea: $H_0 : \mu \leq 45$
$H_1 : \mu > 45$

Horretarako, 4. gaiko 4. 7 adibidean adierazten denez, EAEko biztanleek paseo batean ematen duten denboraren desbideratze estandarra 45 minutukoa dela suposatuko dugu, eta lagin bakartzat hartuko ditugu gai horren 4.3 taulan agertzen diren datu guztiak.

Horrela eginez, test estatistikoaren balioa $z_p = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{71{,}83 - 45}{45/\sqrt{30}} = 3{,}27$ izango da.

Honela kalkulatzen da hipotesi-kontraste horren *p*-balioa:

$p = P(Z > z_p) = P(Z > 3{,}27) = 0{,}000483 < 0{,}05$

Beraz, % 5eko adierazgarritasun-mailaz (eta baita % 1eko adierazgarritasun-mailaz ere), H_0 errefusatuko dugu; hau da, EAEko biztanleek paseo batean ematen duten batezbesteko denbora 45 minutu baino luzeagoa dela onartuko dugu, % 5eko (edo % 1eko) adierazgarritasun-mailaz.

Konfiantza-tarteak ere kalkulatuko ditugu. Honako hau izango da % 90eko konfiantza-tartea:

$$\left(\bar{x} - 1{,}645 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}645 \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(71{,}83 - 1{,}645 \cdot \frac{45}{\sqrt{30}}, 71{,}83 + 1{,}645 \cdot \frac{45}{\sqrt{30}}\right) =$$

$(71{,}83 - 13{,}52\ ,\ 71{,}83 + 13{,}52) = (58{,}31\ ,\ 85{,}35)$

Konfiantza-tarte horren heina aztertzen badugu, argi dago hein osoa 45

minutu baino nabarmen handiagoa dela. Beraz, konfiantza-tarteak berresten digu EAEko biztanleek paseo batean ematen duten batezbesteko denbora 45 minutu baino luzeagoa dela.

6 BI POPULAZIORAKO INFERENTZIA

Aurreko bi gaietan lagin bakar baten azterketa estatistikoaz aritu gara, populazioaren parametroari buruzko inferentziak lortzeko asmoz, bai estimazioaren bidez bai hipotesi-kontrastearen bidez.

Gai honetan, bi populazioren batezbestekoak, alde batetik, eta bariantzak, bestetik, konparatu nahi ditugu. Horretarako, populazioetatik ateratako lagin banatik abiatuko gara, inferentzia estatistikoak gauzatzeko. Populazioen batezbestekoak zein bariantzak konparatzeko, estimazio eta hipotesi-kontrastearen metodoak erabiliko ditugu.

Bi egoera ezberdinetan konparatuko ditugu populazioen parametroak: lehenengoz, populazioak askeak direnean; bigarrenez, binakako datuak ditugunean.

6.1 BI LAGIN ASKEREN BATEZBESTEKOEN ARTEKO DIFERENTZIAREN BANAKETA

Demagun bi populazio aske ditugula, X_1 eta X_2, non batezbestekoak μ_1 eta μ_2 eta bariantzak σ_1^2 eta σ_2^2 baitira, hurrenez hurren.

Bi populazio askeren batezbestekoak konparatzen hasiko gara. Horretarako, bi batezbestekoen arteko $\mu_1 - \mu_2$ diferentzia erabiliko dugu.

Laugarren gaian ikusi dugunez, populazioaren batezbestekoaren estimatzailea laginaren batezbestekoa da. Bi populazio askeren batezbestekoen arteko $\mu_1 - \mu_2$ diferentziaren estimatzatilea kalkulatzeko, honako era honetara jokatuko dugu:

Populazio bakoitzetik n_1 eta n_2 tamainako laginak aterako ditugu, eta beren batezbestekoak $\overline{X}_1$ eta $\overline{X}_2$ adieraziko ditugu, hurrenez hurren. $\overline{X}_1$ lagin-batezbestekoa μ_1-en estimatzailea izango da, eta $\overline{X}_2$, μ_2-rena. Beraz, $\mu_1 - \mu_2$ diferentziaren estimatzatilea $\overline{X}_1 - \overline{X}_2$ izango da.

Inferentziak gauzatu ahal izateko, estimatzaile horren banaketa aztertzen hasiko gara.

6.1.1 Bariantza ezaguneko populazioak

Laugarren gaian, populazioa normala denean, edo $n \geq 30$ denean, laginaren batezbestekoaren banaketa $\overline{X} \approx N(\mu, \sigma / \sqrt{n})$ dela ikusi dugu.

Demagun orain <u>bi populazio aske</u> ditugula, X_1 eta X_2, non batezbestekoak μ_1 eta μ_2 eta bariantzak σ_1^2 eta σ_2^2 baitira, hurrenez hurren. Populazio bakoitzetik n_1 eta n_2 tamainako

laginak ateratzen baditugu, hurrenez hurren, honako hauek izango dira $\overline{X}_1$ eta $\overline{X}_2$ batezbestekoen lagin-banaketak:

$$\overline{X}_1 \approx N(\mu_1, \sigma_1 / \sqrt{n_1}) \text{ eta } \overline{X}_2 \approx N(\mu_2, \sigma_2 / \sqrt{n_2})$$

Banaketa horiek baliozkoak izateko, populazioek normalak izan behar dute, edo, bestela, laginen tamainak 30 edo handiagoak.

$\overline{X}_1 - \overline{X}_2$ zorizko aldagaia kontuan hartzen badugu, itxaropenaren eta bariantzaren propietateak erabiliz, honako emaitza hauek lor ditzakegu:

$$E(\overline{X}_1 - \overline{X}_2) = E[\overline{X}_1 + (-1)\cdot \overline{X}_2] = E(\overline{X}_1) + (-1)\, E(\overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$$

$$\mathrm{Var}(\overline{X}_1 - \overline{X}_2) = \mathrm{Var}[\overline{X}_1 + (-1)\,\overline{X}_2] = \mathrm{Var}(\overline{X}_1) + (-1)^2\cdot\mathrm{Var}(\overline{X}_2) = \mathrm{Var}(\overline{X}_1) + \mathrm{Var}(\overline{X}_2)$$
$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Beraz, populazioak normalak baldin badira, edo bestela, laginen tamainek $n_1 \geq 30$ eta $n_2 \geq 30$ betetzen baldin badute, honako hau izango da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\overline{X}_1 - \overline{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \text{ edo } \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0,1)$$

6.1.2 Bariantza ezezaguneko populazio normalak, $n_1 \geq 30$ eta $n_2 \geq 30$

Kasu honetan, bariantzak laginen kuasibariantzen bidez estimatzen ditugu, eta honako hau da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\overline{X}_1 - \overline{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) \text{ edo } \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0,1)$$

6.1.3 Bariantza ezezaguneko populazio normalak, eta $n_1 < 30$ edo $n_2 < 30$

Bi laginetatik bat, edo biak, 30 baino tamaina txikiagokoa denean, bariantza ezezagunak berdinak edo ezberdinak izan daitezke. Bi kasuetan, batezbestekoen arteko diferentziaren banaketa ez da berbera, eta bi kasuak bereizi beharko ditugu.

6.1.3.1 Bariantza ezezagun baina berdineko populazio normalak

Demagun bi populazio normal eta aske ditugula, X_1 eta X_2, non batezbestekoak μ_1 eta μ_2 baitira, hurrenez hurren, eta bariantza berdina baina ezezaguna σ^2 baita. Populazio bakoitzetik n_1 eta n_2 tamainako laginak aterako ditugu, hurrenez hurren; laginen kuasibariantzak bariantzaren estimazioak izango dira, hau da: $\hat{\sigma}_1^2 = S_1^2$ eta $\hat{\sigma}_2^2 = S_2^2$.

Estimazioak gauzatzean, zenbat eta handiagoa izan laginaren tamaina, orduan eta estimazio hobea izango dugu. Bi kuasibariantzek parametro ezezagun bera estimatzen dutenez, bi laginak batera har ditzakegu, eta bariantza komunaren estimaziorik hoberena aurreko estimazio bien batezbesteko haztatua izango da; alegia:

$$\hat{\sigma}^2 = S_p^2 = \frac{(n_1 - 1)\cdot S_1^2 + (n_2 - 1)\cdot S_2^2}{n_1 + n_2 - 2}$$

Kasu horretan, honako era honetara estimatuko da $\overline{X}_1 - \overline{X}_2$ zorizko aldagaiaren bariantza:

$$\mathrm{Var}(\overline{X}_1 - \overline{X}_2) = \mathrm{Var}[\overline{X}_1 + (-1)\overline{X}_2] = \mathrm{Var}(\overline{X}_1) + (-1)^2 \cdot \mathrm{Var}(\overline{X}_2) =$$

$$= \mathrm{Var}(\overline{X}_1) + \mathrm{Var}(\overline{X}_2) \approx \frac{S_p^2}{n_1} + \frac{S_p^2}{n_2} = S_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Itxaropena, berriz, aurreko kasuan bezala, $E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$ izango da. Orduan, honako hau da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = t_{n_1 + n_2 - 2}$$

6.1.3.2 Bariantza ezezagun eta ezberdineko populazio normalak

Demagun bi populazio normal eta aske ditugula, X_1 eta X_2, non batezbestekoak μ_1 eta μ_2 baitira, hurrenez hurren. Bariantza ezezagunak ditugunean, eta, gainera, berdintasuna

onartu ezin denean, laginen kuasibariantza bakoitza bere bariantzaren estimazioa izango da. Alegia, $\hat{\sigma}_1^2 = S_1^2$ eta $\hat{\sigma}_2^2 = S_2^2$.

Beraz, $\mathrm{Var}(\overline{X}_1 - \overline{X}_2) = \mathrm{Var}(\overline{X}_1) + \mathrm{Var}(\overline{X}_2) \approx \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}$.

Itxaropena, berriz, aurreko kasuetan bezala, $E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$ izango da.

Kasu honetan, honako hau da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = t_g \qquad \text{non} \qquad g = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$$

g zenbaki osoa ez denean, zenbaki osorik hurbilena erabiliko da.

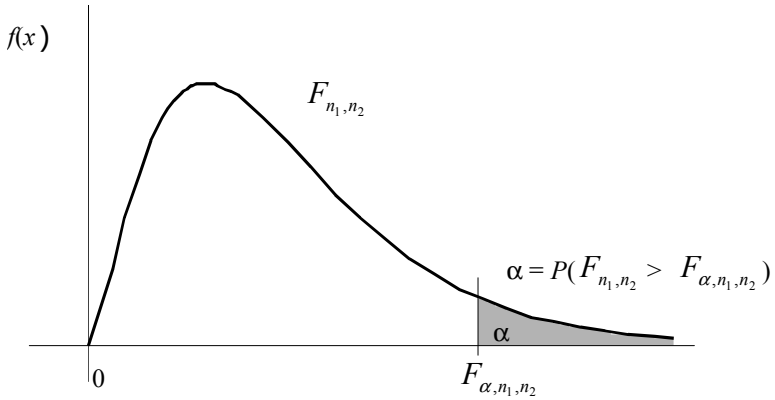
Baina, nola dakigu bariantza ezezagunak berdinak ala ezberdinak diren? Bi bariantza konparatzeko, beren arteko zatidura erabiliko dugu, hau da, $\dfrac{\sigma_1^2}{\sigma_2^2}$. Bariantzak berdinak badira, zatidura 1 izango da.

Zatidura horren estimatzailea kuasibariantzen arteko zatidura izango da, eta bere lagin-banaketa ezagutzeko, banaketa berri bat behar dugu, Fisher-Snedecor-en banaketa, hain zuzen.

6.2 FISHER-SNEDECOR-EN F BANAKETA

Izan bitez $\chi^2_{n_1}$ eta $\chi^2_{n_2}$ n_1 eta n_2 askatasun-graduko khi karratuaren banaketei darraizkien aldagai askeak, hurrenez hurren. Honako adierazpen honen bidez definitzen da **n_1 eta n_2 askatasun-graduetako Fisher-Snedecor-en *F*** zorizko aldagaia:

$$F_{n_1,n_2} = \frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$$



6.1 irudia. $F_{\alpha;n_1,n_2}$ balioaren adierazpen grafikoa.

Propietateak:

1. [0, +∞) tarteko balioak hartzen ditu.
2. Ez da simetrikoa.
3. n_1 eta n_2-ren balioen pean dago.
4. $P(F_{n_1,n_2} > A) = P(F_{n_2,n_1} < \frac{1}{A})$

<u>Froga</u>:

$$P(F_{n_1,n_2} > A) = P\left(\frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2} > A\right) = P\left(\frac{\chi^2_{n_2}/n_2}{\chi^2_{n_1}/n_1} < \frac{1}{A}\right) = P(F_{n_2,n_1} < \frac{1}{A})$$

n_1 eta n_2 askatasun-graduetako F banaketaren dentsitate-funtzioaren eta abzisa-ardatzaren bitartean eskuinerantz aurrefinkaturiko azalera uzten duen balioari **n_1 eta n_2 askatasun-graduetako *F* banaketaren balio kritiko** deritzo (banaketa-funtzioaren balioaren osagarria da). Aurrefinkaturiko azalera hori α baldin bada, balio kritikoa $F_{\alpha;n_1,n_2}$ adierazten da, hots, $P(F_{n_1,n_2} > F_{\alpha;n_1,n_2}) = \alpha$. Erlazio hori grafikoki 6.1 irudian erakusten da.

6.1 adibidea

Izan bedi $F_{10,15}$ 10 eta 15 askatasun-graduetako F banaketari darraion zorizko aldagaia. Kalkula itzazu:

a) $P(F_{10,15} \leq a) = 0{,}95$

b) $P(F_{10,15} \geq b) = 0{,}025$

c) $P(F_{10,15} < d) = 0{,}01$

d) $P(F_{10,15} > e) = 0{,}975$

Emaitza:

a) $P(F_{10,15} \leq a) = 0{,}95 \Rightarrow P(F_{10,15} > a) = 1 - 0{,}95 = 0{,}05 \Rightarrow a = 2{,}54$

b) $P(F_{10,15} \geq b) = 0{,}025 \Rightarrow b = 3{,}06$

c) $P(F_{10,15} < d) = 0{,}01 \Rightarrow P(F_{10,15} > d) = 1 - 0{,}01 = 0{,}99 = P(F_{15,10} < \frac{1}{d}) \Rightarrow$

$P(F_{15,10} > \frac{1}{d}) = 0{,}01 \Rightarrow \Rightarrow \frac{1}{d} = 4{,}56 \Rightarrow d = \frac{1}{4.56} = 0{,}219$

d) $P(F_{10,15} > e) = 0{,}975 = P(F_{15,10} < \frac{1}{e}) \Rightarrow P(F_{15,10} > \frac{1}{e}) = 0{,}025 \Rightarrow$

$\Rightarrow \frac{1}{e} = 3{,}52 \Rightarrow e = 0{,}284$

6.2.1 Kuasibariantzen arteko zatiduraren banaketa

σ_1^2 eta σ_2^2, hurrenez hurren, bariantzako bi populazio normal eta aske emanik, n_1 eta n_2 tamainako laginak aterako ditugu, beren kuasibariantzak kalkulaturik. Orduan, $\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$ adierazpena $(n_1 - 1)$ eta $(n_2 - 1)$ askatasun-graduetako F banaketari darraio.

Froga:

$\frac{(n_1 - 1)S_1^2}{\sigma_1^2} = \chi^2_{n_1 - 1}$ eta $\frac{(n_2 - 1)S_2^2}{\sigma_2^2} = \chi^2_{n_2 - 1}$ direla ezagutzen dugu.

Orduan: $\frac{S_1^2}{\sigma_1^2} = \frac{\chi^2_{n_1 - 1}}{(n_1 - 1)}$ eta $\frac{S_2^2}{\sigma_2^2} = \frac{\chi^2_{n_2 - 1}}{(n_2 - 1)}$

Beraz, $\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\chi^2_{n_1 - 1} / (n_1 - 1)}{\chi^2_{n_2 - 1} / (n_2 - 1)} = F_{(n_1 - 1),(n_2 - 1)}$

6.3 BANAKETA NORMALARI DARRAIZKION BI POPULAZIO ASKEREN BARIANTZEN KONPARAKETA

Bi populazio normal eta askeren bariantzak, σ_1^2 eta σ_2^2, konparatzeko, bariantzen arteko berdintasunerako aldebiko kontrastea erabiliko dugu. Alegia, honako hauek dira kontrastatzeko hipotesiak:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

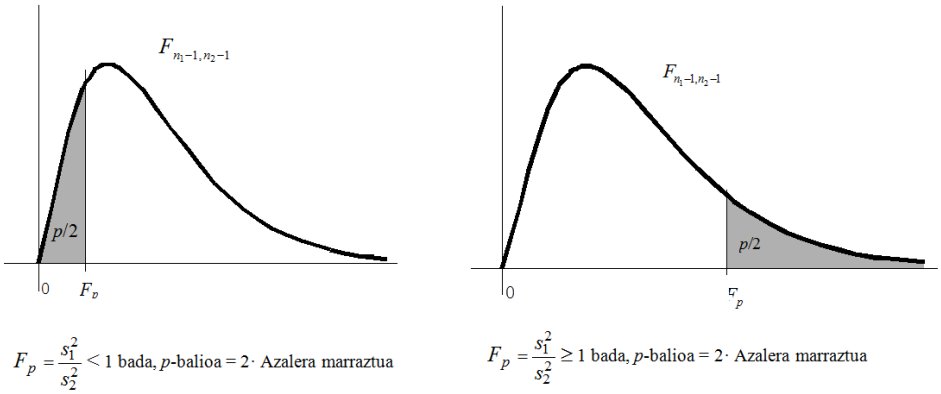
Horretarako, $\dfrac{S_1^2}{S_2^2}$ estatistikoa erabiliko da. Aurreko atalean ikusi dugunez, $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = F_{(n_1-1),(n_2-1)}$ da, eta, beraz, H_0 egiazkoa izatearen baldintzapean, $\sigma_1^2 = \sigma_2^2$ -ren ondorioz, $\dfrac{S_1^2}{S_2^2} : F_{(n_1-1),(n_2-1)}$.

Laginak aukeratu ondoren, estatistikoaren balioa kalkulatzen da: $F_p = \dfrac{s_1^2}{s_2^2}$

Kasu honetan, estatistikoari dagokion p-balioa honako era honetara kalkulatuko da:

Baldin eta $s_1^2 \geq s_2^2$ bada, $p = 2 \cdot P(F_{(n_1-1),(n_2-1)} > F_p)$

Baldin eta $s_1^2 < s_2^2$ bada, $p = 2 \cdot P(F_{(n_1-1),(n_2-1)} < F_p)$



6.2 irudia. Bariantzen arteko berdintasunerako hipotesi-kontrastearen p-balioaren adierazpen grafikoa.

6.2 adibidea

1997ko Osasun Inkestako datuak erabiliz, azken bi asteetan paseoan ibili diren pertsonen artean, 25 gizon eta 30 emakume aukeratu ditugu zoriz. Gizonek paseo batean emandako batez besteko denbora 75 minutu izan da, eta desbideratze estandarra 53 minutu. Paseoan ibili diren emakumeen kasuan, paseo batean emandako batez besteko denbora 77 minutu eta desbideratze estandarra 38 minutu izan dira. Paseoan ibilitako denboraren banaketa (gizonezkoen zein emakumezkoen kasuan) normala dela suposatuz, onar al dezakegu gizonek eta emakumeek paseatzen emandako denboren sakabanatzeak berdinak direnik, % 1eko adierazgarritasun-mailaz?

Emaitza:

Kontrastatu nahi dugun hipotesi nulua H_0: $\sigma_1^2 = \sigma_2^2$ da, eta alternatiboa H_1: $\sigma_1^2 \neq \sigma_2^2$

Estatistikoaren balioa $F_p = \dfrac{s_1^2}{s_2^2} = \dfrac{\frac{25}{24}53^2}{\frac{30}{29}38^2} = \dfrac{25 \cdot 29 \cdot 53^2}{30 \cdot 24 \cdot 38^2} = 1{,}96$ da, eta

H_0 egiazkoa denean, bere banaketa $F_{24,\,29}$ da.

Taulan begiratzen badugu: $0{,}01 < P(F_{24,\,29} > 1{,}96) < 0{,}05$

Beraz, $p = 2 \cdot P(F_{24,\,29} > 1{,}96)$ denez, $0{,}02 < p < 0{,}1 \Rightarrow p > 0{,}02 > 0{,}01 \Rightarrow$ Ezin dugu H_0 errefusatu; hau da, % 1eko adierazgarritasun-mailaz, sakabanatzeak berdinak direla onar dezakegu.

Oharra**:** aldebakarreko kontrasteak ere eraiki ditzakegu, nahiz eta ez izan oso erabiliak:

- **Ezkerretiko aldebakarreko kontrastea**:

  H_0: $\sigma_1^2 \geq \sigma_2^2$

  H_1: $\sigma_1^2 < \sigma_2^2$

  Estatistikoa $F_p = \dfrac{s_1^2}{s_2^2}$ da, eta dagokion p-balioa $p = P(F_{(n_1-1),(n_2-1)} < F_p)$

- **Eskuinetiko aldebakarreko kontrastea:**

  H_0: $\sigma_1^2 \leq \sigma_2^2$

  H_1: $\sigma_1^2 > \sigma_2^2$

Estatistikoa $F_p = \dfrac{s_1^2}{s_2^2}$ da, eta dagokion p-balioa $p = P(F_{(n_1-1),(n_2-1)} > F_p)$

6.4 BI BANAKETA ASKEREN BATEZBESTEKOEN ARTEKO DIFERENTZIARAKO KONFIANTZA-TARTEA

Demagun bi populazio aske ditugula, X_1 eta X_2, non batezbestekoak μ_1 eta μ_2, bariantzak σ_1^2 eta σ_2^2, eta populazio bakoitzetik n_1 eta n_2 tamainako laginak ateratzen baititugu, hurrenez hurren.

6.4.1 Bariantza ezaguneko populazioak

σ_1^2 eta σ_2^2 bariantzak ezagunak badira, eta populazio normalak izanik, edo bestela, laginen tamainek $n_1 \geq 30$ eta $n_2 \geq 30$ betetzen baldin badute, 6.1.1 atalean ikusi dugunez, honako hau da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \approx N(0,1)$$

Beraz, hauxe da bariantza ezaguneko bi populazio askeren ($\mu_1 - \mu_2$) batezbestekoen arteko diferentziarako **%100 × (1 − α) mailako konfiantza-tartea** (KT):

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left(\overline{x}_1 - \overline{x}_2 - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \overline{x}_1 - \overline{x}_2 + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Oharra: tarte hori populazioak normalak direnean da baliozkoa, edo bestela, bi laginen tamainak 30 edo handiagoak direnean.

Badakigu, banaketa normalerako, honako hau betetzen dela: $1 - \alpha = P(-z_{\alpha/2} < Z < z_{\alpha/2})$.

Ekuazio horretan, Z-ren ordez $\overline{X}_1 - \overline{X}_2$ laginen batezbestekoen arteko diferentziaren lagin-banaketa estandarizatua, hau da, $\dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = Z$ jartzen badugu, hauxe lortuko dugu:

$$1 - \alpha = P\left(-z_{\alpha/2} < \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < z_{\alpha/2} \right) =$$

$$= P\left(-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right) < z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) =$$

$$= P\left(-\left(\overline{X}_1 - \overline{X}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < -\left(\mu_1 - \mu_2\right) < -\left(\overline{X}_1 - \overline{X}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) =$$

$$= P\left(\left(\overline{X}_1 - \overline{X}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Hau da, (L_1, L_2) tarte bat lortu dugu, non $P(L_1 < \mu_1 - \mu_2 < L_2) = 1 - \alpha$ baldintza betetzen baita. Orduan:

$$\mathrm{L}_1 = \left(\overline{X}_1 - \overline{X}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\mathrm{L}_2 = \left(\overline{X}_1 - \overline{X}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Zorizko laginak aukeratu ondoren, $\overline{x}_1$ eta $\overline{x}_2$ batezbestekoak kalkulatzen dira; aurreko adierazpeneko behe- eta goi-muturretan ordezkatuz, %100 × (1 − α) mailako ($\mu_1 - \mu_2$)-ren konfiantza-tartea lortuko dugu. Alegia:

$$I^{1-\alpha}_{\mu_1-\mu_2} = \left(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right)$$

6.3 adibidea

Inolako tratamendurik gabeko akurien (Indiako ontzien) pisu irabazia 10 g-ko desbideratze estandarra daukan banaketa normalari darraio. Akuriak bazkaz elikatu ondoren, berriz, irabazi horren banaketak 15 g-ko desbideratze estandarra dauka. Tratamendurik gabeko hamasei akuriren batez besteko pisu-irabazia 105,3 g-koa izan da; bazkaz elikaturiko zortzi akurirena, berriz, 175 g-koa. Eraiki ezazu tratatu gabeko eta bazkaz elikaturiko akurien batez besteko pisu-irabazien arteko diferentziarako % 95eko konfiantza-tartea. Zer ondorio atera dezakegu?

Emaitza:

$$I^{0.95}_{\mu_1-\mu_2} = \left(105{,}3 - 175 - 1{,}96\sqrt{\frac{10^2}{16}+\frac{15^2}{8}}\ ,\ 105{,}3 - 175 + 1{,}96\sqrt{\frac{10^2}{16}+\frac{15^2}{8}} \right) =$$

$$= \left(-69{,}7 - 11{,}49\ , -69{,}7 + 11{,}49\right) = (-81{,}19\ , -58{,}21)$$

Tartearen bi muturrak negatiboak direnez, % 95eko konfiantza-mailaz $\mu_1 < \mu_2$ dela ondoriozta dezakegu. Alegia, bazkak akurien batez besteko pisu-irabazia emendatzen duela.

6.4.2 Bariantza ezezaguneko populazio normalak, $n_1 \geq 30$ eta $n_2 \geq 30$

Kasu honetan, bariantza ezezagun bakoitza dagokion laginaren kuasibariantzaz ordezkatzea besterik ez dugu egin behar eta, 6.1.2 atalean ikusi dugunez, hauxe da $\bar{X}_1 - \bar{X}_2$ laginen batezbestekoen arteko diferentziaren lagin-banaketa:

$$\frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \approx N(0,1)$$

Beraz, zorizko laginak aukeratu ondoren, $\bar{x}_1$ eta $\bar{x}_2$ batezbestekoak eta s_1^2 eta s_2^2 kuasibariantzak kalkulatuko ditugu, eta hauxe da lortuko den bariantza ezezaguneko bi populazio normal eta askeren ($\mu_1 - \mu_2$) batezbestekoen arteko diferentziarako **%100 × (1 − α) mailako konfiantza-tartea (KT)**:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

6.4 adibidea

1997ko Osasun Inkestaren datuetan oinarriturik, zoriz aukeraturiko berrogei emakumek eta berrogeita hamar gizonek mendira joaten emandako denbora aztertu dugu. Emakumeen kasuan, batezbestekoa 160 minutu eta desbideratze estandarra 100 minutu izan dira; gizonen kasuan, aldiz, batezbestekoa 205 minutu eta desbideratze estandarra 93 minutu. Badakigu mendira joaten emandako denboraren banaketa normala dela. Datu horietan oinarrituta, % 90eko konfiantza-mailaz, ba al dago diferentziarik gizonek eta emakumeek mendira joaten emandako batezbesteko denboren artean?

Emaitza:

Populazioen bariantzak ezezagunak direnez, laginen kuasibariantzak erabiliko ditugu haien ordez. Hauxe izango da batezbestekoen diferentziarako konfiantza-tartea:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

$1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = z_{0.05} = 1{,}645$

Beraz,

$$I_{\mu_1-\mu_2}^{0.90} = \left(205 - 160 - 1{,}645 \cdot \sqrt{\frac{93^2}{50-1} + \frac{100^2}{40-1}}, 205 - 160 + 1{,}645 \cdot \sqrt{\frac{93^2}{50-1} + \frac{100^2}{40-1}} \right)$$

$$= (45 - 34{,}23 \ , \ 45 + 34{,}23) = (11{,}77 \ , \ 79{,}23)$$

Hau da, tartearen bi mugak positiboak direnez, % 90eko konfiantza-mailaz ondorioztatzen dugu gizonezkoek mendira joaten emandako batezbesteko denbora luzeagoa dela emakumezkoena baino.

6.4.3 Bariantza ezezaguneko populazio normalak eta $n_1 < 30$ edo $n_2 < 30$

Bi laginen batezbestekoen diferentziaren lagin-banaketa aztertzean (6.1.3 atalean) ikusi dugunez, laginen tamainak txikiak direnean, bi kasu bereizi behar ditugu: bariantzak

berdinak izatea ala ezberdinak izatea. Aldez aurretik, beraz, bariantzen konparaketa egin beharko dugu, 6.3 atalean azaldutako hipotesi-kontrastea gauzatuz.

6.4.3.1 Bariantza ezezagun baina berdineko populazio normalak

Demagun X_1: $N(\mu_1, \sigma)$ eta X_2: $N(\mu_2, \sigma)$ bi populazio normal eta aske direla, eta σ ezezaguna dela. Populazio bakoitzetik n_1 eta n_2 tamainako laginak ateratzen baditugu, $n_1 < 30$ edo $n_2 < 30$ izanik, 6.1.3.1 atalean ikusi dugunez, honako hau da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = t_{n_1 + n_2 - 2}$$

Beraz, bariantza ezezagun baina berdineko <u>bi populazio normal eta askeren</u> $(\mu_1 - \mu_2)$ batezbestekoen arteko diferentziarako, hauxe da **%100 × (1 − α) mailako konfiantza-tartea (KT)**:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left(\bar{x}_1 - \bar{x}_2 - t_{\alpha/2;(n_1+n_2-2)} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{\alpha/2;(n_1+n_2-2)} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

Izan ere, honako hau beteko da Student-en t banaketarako:

$$1 - \alpha = P\left(-t_{\alpha/2;\, n_1 + n_2 - 2} < t_{n_1 + n_2 - 2} < t_{\alpha/2;\, n_1 + n_2 - 2}\right).$$

Adierazpen horretan, t banaketaren ordez $\dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = t_{n_1+n_2-2}$ jarriz, honako hau lortuko dugu:

$$1 - \alpha = P\left(-t_{\alpha/2;n_1+n_2-2} < \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\alpha/2;n_1+n_2-2}\right) =$$

$$= P\left(-t_{\alpha/2;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right) < \right.$$

$$\left. < t_{\alpha/2;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right)$$

$$= P\left(-\left(\overline{X}_1 - \overline{X}_2\right) - t_{\alpha/2;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < -\left(\mu_1 - \mu_2\right) < \right.$$

$$\left. < -\left(\overline{X}_1 - \overline{X}_2\right) + t_{\alpha/2;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right)$$

$$= P\left(\left(\overline{X}_1 - \overline{X}_2\right) - t_{\alpha/2;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < \right.$$

$$\left. < \left(\overline{X}_1 - \overline{X}_2\right) + t_{\alpha/2;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right)$$

Alegia, (L_1, L_2) tartea lortu dugu, non $P(L_1 < \mu_1 - \mu_2 < L_2) = 1 - \alpha$ betetzen baita:

$$L_1 = \left(\overline{X}_1 - \overline{X}_2\right) - t_{\alpha/2;n_1+n_2-2} \cdot S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$L_2 = \left(\overline{X}_1 - \overline{X}_2\right) + t_{\alpha/2;n_1+n_2-2} \cdot S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Zorizko laginak aukeratu eta $\overline{x}_1$ eta $\overline{x}_2$ batezbestekoak kalkulatu ondoren, aurreko adierazpeneko behe- eta goi-muturrak aukeratuz, $(\mu_1 - \mu_2)$-rako konfiantza-tartea lortuko dugu. Alegia,

$$I^{1-\alpha}_{\mu_1-\mu_2} = \left(\overline{x}_1 - \overline{x}_2 - t_{\alpha/2;(n_1+n_2-2)} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \overline{x}_1 - \overline{x}_2 + t_{\alpha/2;(n_1+n_2-2)} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right)$$

6.5 adibidea

Aurreko 6.2 adibidean azaldu dugunez, 25 gizon eta 30 emakume aukeratu ditugu zoriz 1997ko Osasun Inkestaren datuetatik. Gizonek paseo batean emandako batez besteko denbora 75 minutukoa izan da, eta desbideratze estandarra 53 minutukoa. Paseoan ibili diren emakumeen kasuan, paseo batean emandako batezbesteko denbora 77 minutukoa eta desbideratze estandarra 38 minutukoa izan dira.

Onar al dezakegu gizonek eta emakumeek paseatzen emandako batezbesteko denboren arteko diferentziarik dagoenik, % 99ko konfiantza-mailaz?

Emaitza:

$\bar{x}_1 = 75$ $\quad s_1 = \sqrt{25/24}\cdot 53$ $\quad \bar{x}_2 = 77$ $\quad s_2 = \sqrt{30/29}\cdot 38$

6.2 adibidean frogatu dugunez, % 1eko adierazgarritasun-mailaz, populazioen bariantzak berdinak direla onar dezakegu. Orduan, hauxe izango da batezbestekoen arteko kendurarako konfiantza-tartea:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left(\bar{x}_1 - \bar{x}_2 - t_{\alpha/2;(n_1+n_2-2)}\cdot s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 - t_{\alpha/2;(n_1+n_2-2)}\cdot s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

$$s_p = \sqrt{\frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}} = \sqrt{\frac{25\cdot 53^2 + 30\cdot 38^2}{25+30-2}} = 46{,}2856$$

$t_{0\cdot 005;53}$ balio kritikoa kalkulatzeko, n handia denez, Z-ren bidezko hurbilketa erabiliko dugu. Hau da, $t_{0\cdot 005;53} \approx z_{0,005} = 2{,}575$

Beraz,

$$I_{\mu_1-\mu_2}^{0.95} = \left(75 - 77 - 2{,}575\cdot 46{,}2856\sqrt{\frac{1}{25}+\frac{1}{30}}, 75 - 77 + 2{,}575\cdot 46{,}2856\sqrt{\frac{1}{25}+\frac{1}{30}}\right) =$$

$= (-2 - 32{,}28, -2 + 32{,}28) = (-34{,}28\ ,\ 30{,}28)$

Hau da, $0 \in (-34{,}28\ ,\ 30{,}28)$ denez, gizonek eta emakumeek paseatzen emandako batezbesteko denboren arteko diferentzia adierazgarririk ez dagoela ondorioztatuko dugu, % 99ko konfiantza-mailaz.

6.4.3.2 Bariantza ezezagun eta ezberdineko populazio normalak

Demagun X_1: $N(\mu_1, \sigma_1)$ eta X_2: $N(\mu_2, \sigma_2)$ bi populazio normal eta aske direla, eta σ_1^2 eta σ_2^2 bariantzak ezezagunak eta ezberdinak direla. Populazio bakoitzetik n_1 eta n_2 tamainako laginak ateratzen baditugu, hurrenez hurren, $n_1 < 30$ edo $n_2 < 30$ izanik, 6.1.3.2 atalean ikusi dugunez, honako hau da $\overline{X}_1 - \overline{X}_2$ batezbestekoen arteko diferentziaren lagin-banaketa:

$$\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}=t_g \qquad \text{non} \qquad g=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1}+\frac{\left(s_2^2/n_2\right)^2}{n_2}}$$

Beraz, hauxe da bariantza ezezagun eta ezberdineko <u>bi populazio normal eta askeren</u> ($\mu_1 - \mu_2$) batezbestekoen arteko diferentziarako **%100 × (1 − α) mailako konfiantza-tartea (KT)**:

$$I_{\mu_1-\mu_2}^{1-\alpha}=\left(\overline{X}_1-\overline{X}_2-t_{\alpha/2;g}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\overline{X}_1-\overline{X}_2+t_{\alpha/2;g}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)$$

Badakigu Student-en t banaketak honako hau betetzen duela:

$$1-\alpha=P\left(-t_{\alpha/2;g}<t_g<t_{\alpha/2;g}\right).$$

Adierazpen horretan, t_g banaketaren ordez batezbestekoen arteko diferentziaren lagin-banaketa jartzen badugu, honako hau daukagu:

$$1-\alpha=P\left(-t_{\alpha/2;g}<\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}<t_{\alpha/2;g}\right)=$$

$$=P\left(-t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)<-t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)=$$

$$=P\left(-\left(\overline{X}_1-\overline{X}_2\right)-t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<-\left(\mu_1-\mu_2\right)<-\left(\overline{X}_1-\overline{X}_2\right)+t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)=$$

$$=P\left(\left(\overline{X}_1-\overline{X}_2\right)-t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<\left(\overline{X}_1-\overline{X}_2\right)+t_{\alpha/2;g}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)$$

Alegia, (L_1, L_2) tartea lortu dugu, non $P(L_1 < \mu_1 - \mu_2 < L_2) = 1 - \alpha$ betetzen baita:

$$L_1 = \left(\bar{X}_1 - \bar{X}_2\right) - t_{\alpha/2;g} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$L_2 = \left(\bar{X}_1 - \bar{X}_2\right) + t_{\alpha/2;g} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Zorizko laginak aukeratu eta, $\bar{x}_1$ eta $\bar{x}_2$, batezbestekoak kalkulatu ondoren, aurreko adierazpeneko behe- eta goi-muturrak aukeratuz ($\mu_1 - \mu_2$)-rako konfiantza-tartea lortuko dugu. Alegia,

$$I^{1-\alpha}_{\mu_1 - \mu_2} = \left(\bar{x}_1 - \bar{x}_2 - t_{\alpha/2;g} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{\alpha/2;g} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

6.6 adibidea

Habiaren hazkuntzaren eta mantenimenduaren energia-beharrei buruzko ikerketa egin dute. *X*, txori bakoitzak orduko eta gramoko behar duen kilokaloria kopuru aldagai normala aztertu dute, eta honako datu hauek lortzen dituzte:

Helduak txitatzen	Helduak hazkuntza aurretik
$n_1 = 57$	$n_2 = 12$
$\bar{x}_1 = 0{,}0167$ Kcal/(g)(ordu)	$\bar{x}_2 = 0{,}014$ Kcal/(g)(ordu)
$s_1 = 0{,}0042$ Kcal/(g)(ordu)	$s_2 = 0{,}0024$ Kcal/(g)(ordu)

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

% 95eko konfiantza-tartea erabiliz, zein ondorio atera dezakegu txitatzen ari diren eta hazkuntzaren aurretik dauden txori helduen gramoko eta orduko batez besteko kilokalorien arteko diferentziari buruz?

Emaitza:

Lehendabizi, bariantzen berdintasunerako aldebiko kontrastea egingo dugu:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

H_0 egia dela suposatuz, estatistikoaren balioa $F_p = \dfrac{0.0042^2}{0.0024^2} = 3.0625$ da. Eta dagokion p-balioa: $p = 2 \cdot P(F_{56,11} > 3{,}0625)$.

F banaketaren taulan, ez dago $F_{56,11}$ banaketarik, eta $F_{60,11}$ banaketaren bidez hurbilduko dugu p-balioa:

$P(F_{56,11} > 3{,}0625) \approx P(F_{60,11} > 3{,}0625) < P(F_{60,11} > 3{,}00) = 0{,}025$.

Beraz, $p < 2 \cdot 0{,}025 = 0{,}05$. Eta, ondorioz, % 5eko adierazgarritasun-mailaz, bariantzen berdintasuna errefusatu egiten da.

Bariantzak ezezagunak eta ezberdinak badira, laginen tamainak txikiak izanik, honako hau da $\overline{X}_1 - \overline{X}_2$ lagin batezbestekoan arteko diferentziaren lagin-banaketa:

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = t_g$$

non

$$g = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}} = \frac{\left(\dfrac{0{,}0042^2}{57} + \dfrac{0{,}0024^2}{12}\right)^2}{\dfrac{(0{,}0042^2 / 57)^2}{56} + \dfrac{(0{,}0024^2 / 12)^2}{11}} = 27{,}49 \approx 27$$

$t_{0.025;27} = 2{,}052$ da.

Hauxe da % 95eko konfiantza-tartea:

$$I_{\mu_1 - \mu_2}^{0.95} = \left(0{,}0167 - 0{,}0144 - 2{,}052 \cdot \sqrt{\frac{0{,}0042^2}{57} + \frac{0{,}0024^2}{12}}, \right.$$

$$\left. 0{,}0167 - 0{,}0144 + 2{,}052 \cdot \sqrt{\frac{0{,}0042^2}{57} + \frac{0{,}0024^2}{12}} \right) =$$

$$= (0{,}0023 - 0{,}0018\,,\, 0{,}0023 + 0{,}0018) = (0{,}0005\,,\, 0{,}0041)$$

$0 \notin I^{0.95}_{\mu_1 - \mu_2}$ betetzen denez, % 95eko konfiantza-mailaz, txitatzen ari diren txori helduen gramoko eta orduko batez besteko kilokaloria kopurua txitatzen hasi gabe daudenena baino handiagoa dela ondoriozta dezakegu.

6.5 BI POPULAZIO ASKEREN BATEZBESTEKOEN BERDINTASUNERAKO KONTRASTEA

Bi populazioen batezbestekoak konparatzeko, konfiantza-tarteak erabiltzeaz gainera, hipotesi-kontrasteak ere erabil ditzakegu. Kontuan hartu behar dugu, populazioen baldintzen eta laginen tamainen arabera, estatistikoen lagin-banaketak ezberdinak izango direla eta, beraz, konfiantza-tarteekin egin dugun moduan, kasu ezberdinak bereizi beharko ditugula.

Demagun <u>bi populazio aske</u> ditugula, X_1 eta X_2, non batezbestekoak μ_1 eta μ_2 eta bariantzak σ_1^2 eta σ_2^2 baitira, eta populazio bakoitzetik n_1 eta n_2 tamainako laginak ateratzen ditugula, hurrenez hurren.

6.5.1 Bariantza ezaguneko populazioak

Populazioak normalak badira, edo bestela, lagin-tamainak Limitearen Teorema Zentrala betetzen dela ziurtatzeko aski handiak badira -hau da, $n_1 \geq 30$ eta $n_2 \geq 30$, hipotesi-kontrastea egiteko erabiltzen den estatistikoa $Z_p = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ da-,

6.1.1 atalean azaldutakoaren arabera, eta H_0 hipotesiaren pean, estatistiko horren banaketa $N(0, 1)$ izango da (X_1 eta X_2 normalak direnean, banaketa zehatza eta, normalak ez direnean, $n_1 \geq 30$ eta $n_2 \geq 30$ badira, asintotikoki normala).

Laginak aukeratu ondoren, estatistikoaren balio zehatza kalkulatuko dugu: $z_p = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Hipotesi-kontraste motaren arabera, *p*-balioa kalkulatzeko era aldatuko da, hipotesi alternatiboaren arabera kalkulatu behar baita; baina estatistiko berbera erabiliko da aldebakarreko zein aldebiko kontrasteetan.

6.5.1.1 Aldebakarreko kontrasteak

1. <u>Ezkerretikoa</u>

H_0: $\mu_1 \geq \mu_2$

H_1: $\mu_1 < \mu_2$

Kasu honetan, honako hau da *p*-balioa: $p = P(Z < z_p)$.

6.7 adibidea

6.3 adibideko datuetan oinarriturik, ondoriozta al dezakegu, % 5eko adierazgarritasun-mailaz, akuriek bazkaz elikatzeak batez besteko pisu-irabazia gehitzen duenik?

<u>Emaitza</u>:

Honako hipotesi-kontraste hau egiten dugu:

H_0: $\mu_1 \geq \mu_2$ (Bazkak ez du batez besteko pisu-irabazia handitzen).
H_1: $\mu_1 < \mu_2$ (Bazkak batez besteko pisu-irabazia handitzen du).

H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{105{,}3 - 175}{\sqrt{\frac{10^2}{16} + \frac{15^2}{8}}} = -11{,}89$$

p-balioa, $p = P(Z < -11{,}89) < 0{,}001$, guztiz adierazgarria da; orduan, H_0 hipotesia errefusatuko dugu, eta bazkak akurien batez besteko pisu-irabazia handitzen duela ondorioztatuko dugu.

2. <u>Eskuinetikoa</u>

H_0: $\mu_1 \leq \mu_2$

H_1: $\mu_1 > \mu_2$

Kasu honetan, honako hau da *p*-balioa: $p = P(Z > z_p)$.

6.8 adibidea

Mongolismoa pairatzen duten hamabi pertsonen serumeko azido urikoaren batez besteko kontzentrazioa neurtu dute, eta 4,5 mg/100 ml lortu dute. Hamabost pertsona normalek osaturiko taldeko azido urikoaren batez besteko kontzentrazioa 3,4 mg/100 ml-koa da. Edozein pertsonaren (bai pertsona

normalak bai mongolikoak) azido urikoaren kontzentrazioa 1 mg/100 ml-ko desbideratze estandarreko banaketa normalari darraio. Onar al dezakegu mongolismoak gaixotasun hori pairatzen duten pertsonen serumeko azido urikoaren batez besteko kontzentrazioaren gehikuntza eragiten duela, % 1eko adierazgarritasun-mailaz?

Emaitza:

Honako hipotesi hauek kontrastatu dira:

H_0: $\mu_1 \leq \mu_2$

H_1: $\mu_1 > \mu_2$ (mongolismoa pairatzeak azido urikoaren batez besteko kontzentrazioa handitzen du)

H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa: $z_p = \dfrac{4.5 - 3.4}{\sqrt{\dfrac{1^2}{12} + \dfrac{1^2}{15}}} = 2{,}84$.

p-balioa $p = P(Z > 2{,}84) = 0{,}00226 < 0{,}01$ da. Ondorioz, H_0 hipotesia errefusatu egiten dugu, eta mongolismoa pairatzeak serumeko azido urikoaren batez besteko kontzentrazioa handitu egiten duela ondorioztatzen dugu, % 1eko adierazgarritasun-mailaz.

6.5.1.2 Aldebiko kontrastea

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

H_0 egia dela suposatuz, honako hau da estatistikoaren balioa: $z_p = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Kasu horretan, honako hau da p-balioa: $p = 2 \cdot P(Z > |z_p|)$.

6.9 adibidea

Berrogei emakumek osaturiko taldearen batez besteko pisua 74 kg da; berrogeita hamar gizonek osaturikoarena, berriz, 78 kg. Emakumezkoen populazioaren pisua 8 kg-ko desbideratze estandarreko banaketa normalari darraio; gizonezkoen populazioarena, berriz, 7 kg-ko desbideratze

estandarrekoari. Onar al daiteke emakumeen eta gizonen batez besteko pisuen arteko ezberdintasun adierazgarririk dagoenik, % 5eko adierazgarritasun-mailaz?

Emaitza:

Kasu honetan, honako hauek dira kontrastatzen diren hipotesiak:

H_0: $\mu_1 = \mu_2$ (ezberdintasuna zoriaren eragina da)

H_1: $\mu_1 \neq \mu_2$ (ezberdintasuna adierazgarria da)

H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{74 - 78}{\sqrt{\frac{8^2}{40} + \frac{7^2}{50}}} = -2{,}49 \,.$$

p-balioa kalkulatzeko: $p = 2 \cdot P(Z > 2{,}49) = 2 \cdot 0{,}00639 = 0{,}01278 < 0{,}05$ da. Beraz, H_0 errefusatuko dugu, eta % 5eko adierazgarritasun-mailaz emakumeen eta gizonen batez besteko pisuen artean ezberdintasun adierazgarria dagoela onartuko dugu.

6.5.2 Bariantza ezezaguneko populazio normalak, $n_1 \geq 30$ eta $n_2 \geq 30$

Kasu honetan, bariantzak ezezagunak direnez, kuasibariantzen bidez hurbilduko ditugu, eta hipotesi-kontrastea egiteko erabiliko dugun estatistikoa $Z_p = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ izango da.

6.1.2 atalean esandakoaren arabera, H_0 hipotesiaren pean, estatistiko horren banaketa gutxi gorabehera $N(0, 1)$ izango da, baldin eta populazioak normalak badira.

Beraz, hipotesi-kontrastea 6.5.1 atalean azaldutako era berean egingo da, kontuan izanik σ_1^2 -en eta σ_2^2 -ren ordez s_1^2 eta s_2^2 lagin-kuasibariantzak erabili behar ditugula.

6.10 adibidea

Aurreko 6.4 adibideko datuetan oinarriturik, eta % 5eko adierazgarritasun-mailaz, onar al dezakegu gizonezkoek mendira joaten emandako batez besteko denbora emakumezkoena baino luzeagoa denik?

Emaitza:

Honako hipotesi hauek kontrastatzen dira:

H_0: $\mu_1 \leq \mu_2$ (gizonezkoen batezbestekoa ez da emakumezkoena baino luzeagoa).

H_1: $\mu_1 > \mu_2$ (gizonezkoen batezbestekoa emakumezkoena baino luzeagoa da).

H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{205 - 160}{\sqrt{\frac{93^2}{49} + \frac{100^2}{39}}} = 2{,}16$$

p-balioa = $P(Z > 2{,}16) = 0{,}0154 < 0{,}05$ da. Beraz, H_0 hipotesi nulua errefusatuko dugu, eta gizonezkoek mendira joaten emandako batez besteko denbora emakumezkoena baino luzeagoa dela onartuko dugu, % 5eko adierazgarritasun-mailaz.

6.5.3 Bariantza ezezaguneko populazio normalak eta $n_1 < 30$ edo $n_2 < 30$

Konfiantza-tarteak kalkulatzerakoan egin dugun bezala, bi kasu bereiziko ditugu, bariantza berdinak ala ezberdinak izatearen arabera. Aldez aurretik, beraz, bariantzen konparaketa egin beharko dugu.

6.5.3.1 Bariantza ezezagun baina berdineko populazio normalak

Kasu honetan, hipotesi-kontrastea egiteko erabiltzen den estatistikoa $T_p = \dfrac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ da eta, 6.1.3.1 atalean azaldutakoaren arabera, H_0 hipotesiaren pean, estatistiko horren banaketa $(n_1 + n_2 - 2)$ askatasun-graduko Student-en t izango da.

Kontraste mota ezberdinetan (aldebikoan zein aldebakarrekoan), estatistiko berbera erabiltzen da, eta bere balioa, laginak aukeratu ondoren, hauxe izango da: $t_p = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Kontraste mota bakoitzean, hipotesi alternatiboaren arabera kalkulatuko da *p*-balioa.

6.5.3.1.1Aldebakarreko kontrasteak

1. <u>Ezkerretikoa</u>

H_0: $\mu_1 \geq \mu_2$

H_1: $\mu_1 < \mu_2$

Kasu honetan, hauxe da *p*-balioa: $p = P(t_{n_1+n_2-2} < t_p)$.

6.11 adibidea

Arratoien bularreko anginaren azterketan, hemezortzi animalia zoriz banandu ziren bederatziko bi taldetan. Lehenengo taldeari plazeboa eman zioten, eta bigarrenari FL113 deritzon sendagai berria, esperimentalki probatzen ari direna. Onarturiko oxigeno kantitatea minutuko mililitrotan aztertu zen, eta honako informazio hau lortu zen:

Plazeboa	FL113
$n_1 = 9$	$n_2 = 9$
$\bar{x}_1 = 1509$ ml/min	$\bar{x}_2 = 1702$ ml/min
$s_1 = 169$ ml/min	$s_2 = 181$ ml/min

(<u>Iturria</u>: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Onarturiko oxigeno kantitateak banaketa normalari darraio. Onar al dezakegu, % 5eko adierazgarritasun-mailaz, FL113 hartzen duten arratoiek onarturiko batez besteko oxigeno kantitatea altuagoa denik plazeboa hartzen dutenek onarturikoa baino?

<u>Emaitza</u>:

Honako hipotesi hauek kontrastatu nahi ditugu:

H_0: $\mu_1 \geq \mu_2$ (ez da altuagoa).

H_1: $\mu_1 < \mu_2$ (bada altuagoa).

Lehendabizi, bariantzak berdinak diren hipotesia onar daitekeen ala ez aztertu beharko dugu. Horretarako, bariantzen berdinketarako kontrastea egiten da:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

H_0 egia dela suposatuz, kuasibariantzen arteko zatiduraren lagin-banaketa $F_{8,8}$ izango da, eta hauxe da estatistikoaren balioa: $F_p = \frac{s_1^2}{s_2^2} = \frac{169^2}{181^2} = 0{,}872$

p-balioa kalkulatzeko, F banaketa simetrikoa ez denez, kuasibariantzen arteko erlazioa hartu behar dugu kontuan. Kasu honetan, $s_1^2 < s_2^2$ denez, honela kalkulatuko dugu p-balioa:

$$p = 2\cdot P(F_{8,8} < 0{,}872) = 2\cdot P\left(F_{8,8} > \frac{1}{0{,}872}\right) = 2\cdot P(F_{8,8} > 1{,}147)) \approx 2\cdot P(F_{8,8} > 1)$$

F banaketaren propietateak aplikatuz, askatasun-graduen kopurua izendatzailean eta zenbakitzailean berdina dela kontuan izanik, honako hau dugu:

$P(F_{8,8} > 1) = P(F_{8,8} < 1) = 1 - P(F_{8,8} > 1) \Rightarrow 2\cdot P(F_{8,8} > 1) = 1 \Rightarrow P(F_{8,8} > 1) = 0{,}5$

Beraz, ezin dugu baztertu $p \approx 2\cdot 0{,}5 = 1$ eta bariantzen berdinketa.

Orduan, batezbestekoak konparatzeko estatistikoaren balioa kalkulatzeko, lehenengo, desbideratze estandarraren s_p estimazio komuna kalkulatu behar dugu:

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{8\cdot 169^2 + 8\cdot 181^2}{9+9-2}} = 175{,}1$$

Eta estatistikoaren balioa honako hau da: $t_{16} = \frac{1509-1702}{175.1\cdot\sqrt{\frac{1}{9}+\frac{1}{9}}} = -2.34$

Beraz, hauxe da p-balioa: $p = P(t_{16} < -2{,}34) = P(t_{16} > 2{,}34)$.

Baina, t banaketaren taulan $P(t_{16} > 2{,}12) = 0{,}025$ eta $P(t_{16} > 2{,}583) = 0{,}01$ ditugu.

Beraz, $0{,}01 < p < 0{,}025 < 0{,}05$ betetzen da, eta horrek H_0 hipotesi nulua errefusatzera garamatza. Ondorioz, nahiko arrazoi dugu pentsatzeko, % 5eko adierazgarritasun-mailaz, FL113 hartzen duten arratoiak onarturiko batez besteko oxigeno kantitatea plazeboa hartzen dutenek onarturikoa baino altuagoa dela.

2. <u>Eskuinetikoa</u>

H_0: $\mu_1 \leq \mu_2$

H_1: $\mu_1 > \mu_2$

Kasu honetan, hauxe da p-balioa: $p = P(t_{n_1+n_2-2} > t_p)$.

6.12 adibidea

Aurreko 6.11 adibideko esperimentuarekin jarraituz, arratoi bakoitzak gurpilean kontrolpean egindako ariketaren ondoren, errekuperazio-denbora neurtu da. FL113-k batez besteko errekuperazio-denbora murriztuko duela susmatu da. Errekuperazio-denbora banaketa normalari darraio, eta honako informazio hau lortu da:

Plazeboa	FL113
$n_1 = 9$	$n_2 = 9$
$\bar{x}_1 = 329$ segundo	$\bar{x}_2 = 238$ segundo
$s_1 = 45$ segundo	$s_2 = 43$ segundo

(<u>Iturria</u>: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Onar al dezakegu FL113-k arratoien batezbesteko errekuperazio denbora murrizten duela, % 1eko adierazgarritasun-mailaz?

<u>Emaitza</u>:

Honako hipotesi-kontraste hau egin behar da:

H_0: $\mu_1 \leq \mu_2$ (ez da murrizten).

H_1: $\mu_1 > \mu_2$ (murrizten da).

Lehendabizi, barientzen berdinatsunerako testa egin behar da:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

Hauxe da bariantzak konparatzeko erabili behar dugun estatistikoaren balioa: $F_p = \frac{s_1^2}{s_2^2} = \frac{45^2}{43^2} = 1.095$

Eta p-balioa era honetan kalkulatzen da: $p = 2 \cdot P(F_{8,8} > 1.095) \approx 2 \cdot P(F_{8,8} > 1) = 1$

Beraz, ezin dugu bariantzen berdintasuna errefusatu; eta batezbestekoak konparatzeko erabiliko dugun estatistikoa $T_p = \frac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ izango da.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{8 \cdot 45^2 + 8 \cdot 43^2}{9 + 9 - 2}} = 44{,}01$$

Beraz, hauxe da batezbestekoak konparatzeko erabiltzen den estatistikoaren balioa: $t_p = \frac{329 - 238}{44.01 \cdot \sqrt{\frac{1}{9} + \frac{1}{9}}} = 4.386$

Eta honela kalkulatzen da p-balioa: $p = P(t_{16} > 4{,}386) < P(t_{16} > 4{,}015) < 0{,}0005$ guztiz adierazgarria da. Orduan, H_0 errefusatzen dugu, eta % 1eko adierazgarritasun-mailaz FL113-k arratoien batez besteko errekuperazio-denbora laburtzen duela ondoriozta dezakegu.

6.5.3.1.2Aldebiko kontrastea

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Kasu honetan, hauxe da p-balioa: $p = 2\cdot P(t_{n_1+n_2-2} > |t_p|)$

6.13 adibidea

Aurreko 6.2 adibideko datuetan oinarriturik, ondoriozta al daiteke, % 1eko adierazgarritasun-mailaz, gizon eta emakumeek paseatzen emandako batez besteko denboren arteko diferentzia adierazgarririk dagoenik?

Emaitza:

Hauxe da egingo dugun hipotesi-kontrastea:

H_0: $\mu_1 = \mu_2$ (diferentzia zoriaren eragina da)

H_1: $\mu_1 \neq \mu_2$ (diferentzia adierazgarria da)

6.5 adibidean ikusi dugunez, % 1eko adierazgarritasun-mailaz, onartzen da bariantza ezezagunak berdinak direla.

Konfiantza-tartean eraikitzeko, 6.5 adibidean, s_p estimazio komuna ere kalkulatu dugu: $s_p = 46{,}2856$.

Beraz, hauxe izango da estatistikoaren balioa:

$$t_p = \frac{75-77}{46{,}2856\sqrt{\frac{1}{25}+\frac{1}{30}}} = -0{,}16$$

p-balioa kalkulatuko dugu: $p = 2\cdot P(t_{53} > 0{,}16)$ da.

Baina, n handia denean, Student-en t_n banaketa Z-ren bidez hurbildu behar dugu probabilitateak kalkulatzeko. Hau da, $P(t_{53} > 0{,}16) \approx P(Z > 0{,}16) = 0{,}4364$.

Beraz, $p \approx 2\cdot 0{,}4364 = 0{,}8728 > 0{,}01$. Ondorioz, % 1eko adierazgarritasun-mailaz, gizon eta emakumeek paseatzen emandako batez besteko denboren arteko diferentzia adierazgarririk ez dagoela ondorioztatzen dugu.

6.5.3.2 Bariantza ezezagun eta ezberdineko populazio normalak

Kasu honetan, hipotesi-kontrastea egiteko erabiltzen den estatistikoa $T_p = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$ da eta, 6.1.3.2 atalean azaldutakoaren arabera, H_0 hipotesiaren pean, estatistiko horren

banaketa g askatasun-graduko Student-en t banaketa da, non $g = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2}}$

baita.

Aurreko hipotesi-kontrasteetan ikusi dugun bezala, aldebiko zein aldebakarreko kontrasteetan erabiltzen den estatistikoa berbera da, eta hauxe izango da bere balioa, laginak aukeratu ondoren: $t_p = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Kontraste mota bakoitzean, hipotesi alternatiboaren arabera kalkulatuko dugu p-balioa.

6.5.3.2.1Aldebakarreko kontrasteak

1. <u>Ezkerretikoa</u>

H_0: $\mu_1 \geq \mu_2$

H_1: $\mu_1 < \mu_2$

Kasu honetan, hauxe da p-balioa: $p = P(t_g < t_p)$.

6.14 adibidea

Hegazti espezie ezberdinen hegada-abiadurari buruzko azterketa egin dute. Pelikano nabarra eta amerikar itsas mika konparatu nahi dituzte. Aurreko ikerketetan, hegaztien hegada-abiadura banaketa normalari darraiola egiaztatu dute. Ondokoa taulan agertzen diren datuek egiaztatzen al dute pelikano nabarraren batez besteko abiadura amerikar itsas mikarena baino txikiagoa dela, % 5eko adierazgarritasun-mailaz?

Pelikano nabarra	Amerikar itsas mika
$n_1 = 9$	$n_2 = 12$
$\bar{x}_1 = 26{,}05$ m/ordu	$\bar{x}_2 = 30{,}19$ m/ordu
$s_1 = 6{,}34$ m/ordu	$s_2 = 3{,}20$ m/ordu

(<u>Iturria</u>: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Emaitza:

Hauxe da kontrastatu nahi den hipotesia:

H_0: $\mu_1 \geq \mu_2$ (ez txikiagoa)
H_1: $\mu_1 < \mu_2$ (txikiagoa)

Baina, lehendabizi, bariantza ezezagunak berdinak diren ala ez egiaztu behar dugu, eta, horretarako, bariantzen berdintasunerako testa egingo dugu:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

H_0 egia dela suposatuz, honako hau da estatistikoaren behaturiko balioa:

$$F_p = \frac{s_1^2}{s_2^2} = \frac{6.34^2}{3.20^2} = 3.925$$

Eta p-balioa: $p = 2 \cdot P(F_{8,11} > 3.925)$.

F banaketaren taulan, $P(F_{8,11} > 3.66) = 0{,}025$ eta $P(F_{8,11} > 4.74) = 0{,}01$ ditugu. Beraz, $0{,}01 < P(F_{8,11} > 3.925) < 0{,}025 \Rightarrow 0{,}02 < p < 0{,}05$. Hau da, % 5eko adierazgarritasun-mailaz, bariantzen berdintasuna errefusatu egiten dugu.

Ondorioz, batezbestekoak konparatzeko erabili behar den estatistikoa $T_p = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ da, eta haren balioa $t_p = \dfrac{26{,}05 - 30{,}19}{\sqrt{\dfrac{6{,}34^2}{9} + \dfrac{3{,}20^2}{12}}} = -1{,}795$ da.

Honela kalkulatuko dugu askatasun-graduen kopurua:

$$g = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}} = \frac{\left(\dfrac{6.34^2}{9} + \dfrac{3.20^2}{12}\right)^2}{\dfrac{\left(6.34^2 / 9\right)^2}{8} + \dfrac{\left(3.20^2 / 12\right)^2}{11}} = 11.05 \approx 11$$

Eta hauxe izango da p-balioa: $p = P(t_{11} < -1{,}795) = P(t_{11} > 1{,}795) = 0{,}05$.

Beraz, % 5eko adierazgarritasun-mailaz, ez daukagu informazio nahikorik H_0 errefusatzeko, eta pelikano nabarraren batez besteko abiadura ez dela amerikar itsas mikarena baino txikiagoa ondorioztatu dugu.

2. <u>Eskuinetikoa</u>

H_0: $\mu_1 \leq \mu_2$

H_1: $\mu_1 > \mu_2$

Kasu honetan, hauxe da p-balioa: $p = P(t_g > t_p)$.

6.15 adibidea

6.6 adibideko datuetan oinarriturik, esan al dezakegu, % 5eko adierazgarritasun-mailaz, txitatzen ari diren txori helduen orduko eta gramoko batez besteko kilokaloria kopurua hazkuntza aurretik daudenena baino handiagoa dela?

<u>Emaitza</u>:

Honako test hau eraikitzen da:

H_0: $\mu_1 \leq \mu_2$ (ez da handiagoa)

H_1: $\mu_1 > \mu_2$ (handiagoa da)

6.6 adibidean, egiaztatu dugu % 5eko adieragarritasun-mailaz bariantzak ezberdinak direla. Beraz, batezbestekoak konparatzeko estatistiko egokia

$$T_p = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$ da, eta haren balioa $$t_p = \frac{0{,}0167 - 0{,}0144}{\sqrt{\frac{0{,}0042^2}{57} + \frac{0{,}0024^2}{12}}} = 2{,}59$$ da.

Askatasun-graduen kopurua ere kalkulaturik dago 6.6 adibidean: $g = 27$.

Orduan, hauxe da p-balioa: $p = P(t_{27} > 2{,}59)$

t banaketaren taulan, $P(t_{27} > 2{,}473) = 0{,}01$ eta $P(t_{27} > 2{,}771) = 0{,}005$ ditugu. Ondorioz, $0{,}005 < p < 0{,}01 < 0{,}05$, eta horrek H_0 errefusatzera garamatza; beraz, % 5eko adierazgarritasun-mailaz, txitatzen ari diren txori helduen orduko eta gramoko batez besteko kilokaloria kopurua hazkuntza aurretik daudenena baino handiagoa dela ondorioztatuko dugu. Hau da, konfiantza-tartearen bidez lortu dugun ondorio berbera.

6.5.3.2.2 Aldebiko kontrastea

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Kasu honetan, hauxe da p-balioa: $p = 2 \cdot P(t_g > |t_p|)$.

6.16 adibidea

Karbohidratoen metabolismoaren azterketan, 6 gradu zentigraduan dagoen uretan landaturiko ilarren sustraiaren hazkundea eta tenperatura berdineko fruktosaren soluzioan landaturikoena konparatu dituzte. Ilarren sustraiaren hazkundea banaketa normalari darraio. Honako informazio hau lortu dute:

Uretako landareak	Fruktosako landareak
$n_1 = 16$	$n_2 = 25$
$\bar{x}_1 = 9{,}48$ mm/120 ordu	$\bar{x}_2 = 9{,}46$ mm/120 ordu
$s_1 = 0{,}53$ mm/120 ordu	$s_2 = 0{,}25$ mm/120 ordu

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Ondoriozta al dezakegu, % 1eko adierazgarritasun-mailaz, uretan eta fruktosa soluzioan landaturiko ilarren sustraiaren batez besteko hazkundeen arteko diferentzia adierazgarririk dagoenik?

Emaitza:

Honako hipotesi hauek kontrastatu ditugu:

H_0: $\mu_1 = \mu_2$ (zorizko diferentzia da)

H_1: $\mu_1 \neq \mu_2$ (diferentzia adierazgarria da)

Lehendabizi, bi populazio normalen bariantzen arteko berdintasunerako testa egin behar da:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

H_0 egia dela suposatuz, honako hau da estatistikoaren behaturiko balioa:

$$F_p = \frac{s_1^2}{s_2^2} = \frac{0.53^2}{0.25^2} = 4.49$$

Eta hauxe da dagokion p-balioa: $p = 2 \cdot P(F_{15,24} > 4{,}49)$. F banaketaren taulan, $P(F_{15,24} > 3{,}25) = 0{,}005$ dugu. Orduan, $p = 2 \cdot P(F_{15,24} > 4{,}49) < 2 \cdot P(F_{15,24} > 3{,}25) = 0{,}01$ betetzen da. Hau da, % 1eko adierazgarritasun-mailaz, bariantzen berdintasuna errefusatu egingo dugu.

Beraz, batezbestekoak konparatzeko estatistiko egokia, H_0 hipotesiaren pean, $T_p = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ da, eta haren balioa $t_p = \dfrac{9{,}48 - 9{,}46}{\sqrt{\dfrac{0{,}53^2}{16} + \dfrac{0{,}25^2}{25}}} = 0{,}14$ da.

Honako era honetan kalkulatuko dugu askatasun-graduen kopurua:

$$g = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}} = \frac{\left(\dfrac{0.53^2}{16} + \dfrac{0.25^2}{25}\right)^2}{\dfrac{\left(0.53^2 / 16\right)^2}{15} + \dfrac{\left(0.25^2 / 25\right)^2}{24}} = 14.8 \approx 15$$

Eta hauxe izango da p-balioa: $p = 2 \cdot P(t_{15} > 0{,}14)$.

t banaketaren taulan $P(t_{15} > 0{,}258) = 0{,}40$ dugu.

Orduan, $p = 2 \cdot P(t_{15} > 0{,}14) > 2P(t_{15} > 0{,}258)) = 0{,}80$, eta, beraz, ezin da hipotesi nulua errefusatu. Hau da, uretan eta fruktosa soluzioan landaturiko ilarren sustraiaren batez besteko hazkundeen artean ez dagoela diferentzia adierazgarririk ondorioztatzen dugu.

Bi populazio askeen batezbestekoak konparatzeko *SPSS* pakete estatistikoa erabiltzen badugu, agindu bakar bat dago; bertan, aurrenik, bariantzak konparatzen dira, eta gero, (bariantzak berdinak izan ala ez) populazioen batezbestekoak konparatu eta emaitzak ematen dira. Egileak erabaki behar du emandako emaitzen artean zein diren egokiak.

SPSS pakete estatistikoaren erabilera azalduko dugu honako adibide honetan.

6.17 adibidea

Hamar urtean zehar, haritzek 975 metroko altitudean izandako hazkundea eta 675 metrokoan izandakoa konparatu nahi dituzte. Horretarako, azken hamar urteetako hazkunde-eraztunak estaltzen dituen nukleoaren luzera neurtuko dute

zentimetrotan. 975 metroko altitudeko hamabost haritz hartuko dituzte, eta beste hamabost haritz 675 metroko altitudean. Honako hauek dira *SPSS*ren bidez lorturiko datuak.

Estadísticos de grupo

	ALTURA	N	Media	Desviación típ.	Error típ. de la media
LONGITUD	975	15	2,567	1,3515	,3490
	675	15	1,860	,6334	,1635

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
LONGITUD	Se han asumido varianzas iguales	4,325	,047	1,834	28	,077	,707	,3854	-,0828	1,4961
	No se han asumido varianzas iguales			1,834	19,866	,082	,707	,3854	-,0976	1,5109

a) Onar al dezakegu, % 5eko adierazgarritasun-mailaz, haritzaren hazkundea neurtzen duen nukleoaren luzeraren bariantzak, aipaturiko bi altitudeetan, berdinak direnik? Adierazgarritasun-maila % 1ekoa bada, zer ondorio aterako genuke? Zergatik?

b) Onar al dezakegu, % 5eko adierazgarritasun-mailaz, haritzaren nukleoaren batez besteko luzera 975 metroko altitudean 675 metrokoan baino handiagoa denik?

<u>Emaitza</u>:

a) Bigarren taulan agertzen den *Levene-ren Proba*ko izenburupean dauden bi zutabeetan agertzen da bariantzak konparatzeko kontrastea. Bertan, honako hipotesi hauek kontrastatu dira:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Taulako bi zutabe horietan agertzen dira estatistikoa eta *p*-balioa. Estatistikoa kuasibariantzen arteko zatidura da, eta *F*-ren bidez azaldu da taulan: F = 4,325.

p-balioa *Sig*. laburduraren bidez azaltzen da, hau da, $p = 0{,}047 < 0{,}05$. Beraz, H_0 errefusatzen dugu, eta % 5eko adierazgarritasun-mailaz bariantzak ezberdinak direla ondorioztatzen dugu.

b) Orain, honako hauek dira batezbestekoak konparatzeko kontrastatu behar

ditugun hipotesiak:

H_0: $\mu_1 \leq \mu_2$

H_1: $\mu_1 > \mu_2$ (nukleoaren batezbesteko luzera handiagoa da 975 metroko altitudean)

Bigarren taulan agertzen dira batezbestekoak konparatzeko hipotesi-kontrastearen emaitzak, *Batezbestekoen berdintasunerako T proba* izenburupean dauden zutabeetan. Bertan agertzen dira, beste batzuekin batera, estatistikoaren balioa, askatasun-graduen kopurua, *p*-balioa eta batezbestekoen arteko diferentziarako konfiantza-tartea.

Taulan emaitza guztiak agertzen dira, bai bariantzak berdinak direnean bai ezberdinak direnean. Gure kasuan, azken lerroko emaitzak erabili behar ditugu, hau da, bariantza ezberdinei dagozkienak.

Beraz, estatistikoaren balioa t_p = 1,834 da, eta banaketaren askatasun-graduen kopurua g = 19,866 ≈ 20.

Sig.(*bilateral*) izeneko zutabean agertzen den balioa *p*-balioa da aldebiko kontrasterako. Planteatu dugun hipotesi-kontrastea aldebakarrekoa denez, dagokion *p*-balioa aurrekoaren erdia izango da, hau da, gure kontrastearen *p*-balioa = $\frac{0{,}082}{2}$ = 0,041 < 0,05.

Beraz, H_0 errefusatzen dugu eta, % 5eko adierazgarritasun-mailaz, haritzaren nukleoaren batez besteko luzera 975 metroko altitudean 675 metrokoan baino handiagoa dela ondorioztatzen dugu.

Taularen azken bi zutabeetan, populazioen batezbestekoen arteko diferentziarako % 95eko konfiantza-tartea agertzen da:

$$I^{0.95}_{\mu_1 - \mu_2} = (-0{,}0976\ ,\ 1{,}5109)$$

Konfiantza-tarte horren arabera, % 95eko konfiantza-mailaz, populazioen batezbestekoak berdinak direla onartuko genuke, $0 \in I^{0.95}_{\mu_1 - \mu_2}$ baita. Hau da, ondorio ezberdinak ateratzen ditugu, baina kontuan hartu behar dugu konfiantza-tartearen parekoa aldebiko kontrastea dela, ez aldebakarrekoa.

6.6 BATEZBESTEKOEN ARTEKO BERDINTASUNA: BINAKAKO DATUAK

Sarritan, laginak **ez dira askeak** izaten, baizik eta lagin bateko behaketa bakoitza beste lagin bateko behaketa batekin lotuta egoten da. Adibidez, problema askotan, faktore batek (gaixotasunak, sendagaia hartzeak, landareak ongarriztatzeak, eta abarrek) populazioaren elementuetan duen eragina aztertzeko, komenigarria da elementu bakoitzean bi neurri hartzea: lehena, faktorerik gabe, eta bigarrena, faktorea kontuan hartuta. Eragina aztertzeko, bi neurrien arteko diferentzia aztertu beharko dugu.

6.18 adibidea

Ariketa fisikoak serumeko kolesterol-mailan duen eragina neurtzeko ikerketa egin da, eta hamaika pertsonak parte hartu dute. Ariketa fisikoa egin aurretik, pertsona bakoitzaren kolesterol-maila neurtzeko odol-laginak hartu ziren. Ostean, pertsona horiek ibilaldi eta lasterketetan oinarrituriko eguneroko programa jarraitu zuten. Ariketa aldiaren bukaeran odol-lagin berriak hartu ziren, eta berriz neurtu zen pertsona bakoitzaren kolesterol-maila. Serumeko kolesterol-maila banaketa normalari darraiola suposa daiteke. Honako taula honetan agertzen dira behaturiko balioak:

Pertsona	Aurreko maila X (mg/dl)	Osteko maila Y (mg/dl)
1	182	198
2	232	210
3	191	194
4	200	220
5	148	138
6	249	220
7	276	219
8	213	161
9	241	210
10	480	313
11	262	226

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Horrela, pertsonen serumeko kolesterol-mailaren bi neurketa multzo ditugu. Datu multzoak (laginak) ez dira askeak, zeren eta, nahiz eta aldi ezberdinetan

neurtu, gizabanako berengan hartu baitira. Beraz, pertsona bakoitzaren bi datuak ditugu.

Oro har, populazioan, azterketa fisikoa egin aurretiko eta osteko batez besteko kolesterol-mailen arteko diferentzia estimatu nahi dugu. Batezbestekoak konparatzeko edozein teknikak kontuan hartu beharko du, kasu honetan, populazioak ez direla askeak, baizik eta binakako datuak direla.

Askeak ez diren bi populazioen batezbestekoak konparatzeko, bi populazioen arteko diferentzia adierazten duen populazio berria definituko dugu: $D = X - Y$. Aldi berean, binakako datuen arteko diferentziak kalkulatuko ditugu: $d_i = x_i - y_i$, $i = 1, 2, \ldots, n$ eta diferentzia horiek D populaziotik ateratako zorizko lagintzat hartuko ditugu.

X eta Y populazioak normalak badira, D populazioa banaketa normalari darraio, eta hauxe izango da haren batezbestekoa: $\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_Y$.

Era horretan, bi populazioen batezbestekoak konparatzeko hasierako arazoa populazio **bakar bateko** arazo bihurtu dugu; hain zuzen, D diferentzia populazioaren μ_D batezbestekorako inferentzia.

D diferentzia populazioaren bariantza ezezaguna izango da beti (nahiz eta X eta Y populazioen bariantzak ezagutu), eta, beraz, bariantza ezezaguneko populazio normalaren batezbestekorako inferentzia egiten duen kasua aukeratu beharko dugu.

6.6.1 Batezbestekoen arteko diferentziarako konfiantza-tartea

Demagun X eta Y populazio normalak direla, eta n tamainako (x_i, y_i) binakako lagina ateratzen dela. Alde batetik, $D = X - Y$ diferentzia populazio normala da, eta haren batezbestekoa $\mu_D = \mu_X - \mu_Y$ da; bestetik, binakako datuen arteko diferentziak ditugu: $d_i = x_i - y_i$, $i = 1, 2, \ldots, n$.

μ_D batezbestekoaren estimatzailea laginaren $\overline{D}$ batezbestekoa izango da.

D diferentzia populazioaren bariantza ezezaguna denez, 4. gaiko 4.6.2 atalean ikusi dugunez, hauxe izango da laginaren $\overline{D}$ batezbestekoaren banaketa: $\dfrac{\overline{D} - \mu_D}{S_d / \sqrt{n}} = t_{n-1}$

Beraz, laginak aukeratu ondoren, hauxe da μ_D diferentzia populazioaren batezbestekorako **%100 × (1 – α) mailako konfiantza-tartea (KT)**:

$$I_{\mu_D}^{1-\alpha} = \left(\bar{d} - t_{\alpha/2;(n-1)} \cdot \frac{s_d}{\sqrt{n}}, \bar{d} + t_{\alpha/2;(n-1)} \cdot \frac{s_d}{\sqrt{n}} \right)$$

Kontuan hartu behar da, n handia denean, $t_n \approx N(0,1)$ dela, eta beraz, $n \geq 30$ bada, $\frac{\bar{D} - \mu_D}{S_d / \sqrt{n}} \approx Z$ izango da, eta hauxe izango da %100 × (1 − α) mailako konfiantza-tartea (KT): $I_{\mu_D}^{1-\alpha} = \left(\bar{d} - z_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}, \bar{d} + z_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}} \right)$

6.19 adibidea

Aurreko 6.18 adibideko datuetan oinarriturik, kalkula ezazu azterketa fisikoa egin aurreko eta osteko batez besteko kolesterol-mailen arteko diferentziarako % 90eko konfiantza-tartea. Zer ondorio atera dezakegu?

Emaitza:

Konfiantza-tartea kalkulatzeko, diferentzien laginaren batezbestekoa eta kuasibariantza kalkulatu beharko ditugu:

Pertsona	Aurreko maila x (mg/dl)	Osteko maila y (mg/dl)	Diferentzia $d = x - y$	d^2
1	182	198	-16	256
2	232	210	22	484
3	191	194	-3	9
4	200	220	-20	400
5	148	138	10	100
6	249	220	29	841
7	276	219	57	3249
8	213	161	52	2704
9	241	210	31	961
10	480	313	167	27889
11	262	226	36	1296
			365	38189

$$\bar{d} = \frac{\sum_{i=1}^{11} d_i}{11} = \frac{365}{11} = 33{,}18$$

$$s_d^2 = \frac{11 \cdot \sum_{i=1}^{11} d_i^2 - \left[\sum_{i=1}^{11} d_i\right]^2}{11 \cdot 10} = \frac{11 \cdot 38189 - 365^2}{110} = 2607{,}76$$

$$s_d = \sqrt{2607{,}76} = 51{,}066$$

$t_{0.05;10} = 1{,}812$

Beraz, hauxe da μ_D-rako % 90eko konfiantza-tartea:

$$I_{\mu_D}^{0.90} = \left(\bar{d} - t_{0.05;10} \cdot \frac{s_d}{\sqrt{11}}, \bar{d} + t_{0.05;10} \cdot \frac{s_d}{\sqrt{11}}\right) =$$
$$= \left(33{,}18 - 1{,}812 \cdot \frac{51{,}066}{\sqrt{11}}, 33{,}18 + 1{,}812 \cdot \frac{51{,}066}{\sqrt{11}}\right) =$$
$$= (33{,}18 - 27{,}9\,, 33{,}18 - 27{,}9) = (5{,}28\,, 61{,}08)$$

Ondorioz, % 90eko konfiantza-mailaz, μ_D positiboa da, eta, beraz, batez besteko kolesterol-maila jaitsi egin dela baiezta dezakegu.

6.6.2 Batezbestekoen arteko berdintasunerako testa

Demagun X eta Y populazio normalak direla, eta n tamainako (x_i, y_i) binakako lagina ateratzen dela. Kasu honetan, hipotesi-kontrastea egiteko erabiltzen den estatistikoa $T_p = \dfrac{\overline{D} - \mu_D}{S_D / \sqrt{n}}$ da, eta bere banaketa (n – 1) askatasun-graduko Student-en t.

H_0 hipotesia egiazkoa denean, estatistikoaren balioa $t_p = \dfrac{\bar{d}}{s_d / \sqrt{n}}$ da.

Aurreko hipotesi-kontrasteetan ikusi dugun bezala, aldebiko zein aldebakarreko kontrasteetan erabiltzen den estatistikoa berbera da eta, kontraste mota bakoitzean, hipotesi alternatiboaren arabera kalkulatuko dugu *p*-balioa.

6.6.2.1 Aldebakarreko kontrastea

1. <u>Ezkerretikoa</u>

H_0: $\mu_D \geq 0$

H_1: $\mu_D < 0$

Kasu honetan, hauxe da *p*-balioa: $p = P(t_{n-1} < t_p)$

6.20 adibidea

Kolesterol-mailan duen eraginaz gainera, 6.18 adibideko hamaika pertsonengan ariketa fisikoak triglizerido-mailan duen eragina aztertu zuten. Honako neurri hauek lortu zituzten (100 mililitro odoleko triglizerido miligramotan), ariketa egin aurretik eta ostean:

Pertsona	Aurreko maila *x*	Osteko maila *y*
1	68	95
2	77	90
3	94	86
4	73	58
5	37	47
6	131	121
7	77	136
8	24	65
9	99	131
10	629	630
11	116	104

(<u>Iturria</u>: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Ondoriozta al dezakegu, % 5eko adierazgarritasun-mailaz, ariketa fisikoak batez besteko triglizerido-maila emendatzen duenik?

<u>Emaitza</u>:

Honako hipotesi hauek kontrastatu behar ditugu:

H_0: $\mu_D \geq 0$ (ez handitu)
H_1: $\mu_D < 0$ (handitu)

Lehenengo eta behin, lagineko diferentzien batezbestekoa eta kuasibariantza kalkulatuko ditugu:

Pertsona	Aurreko maila x	Osteko maila y	Diferentzia $d = x - y$	d^2
1	68	95	-27	729
2	77	90	-13	169
3	94	86	8	64
4	73	58	15	225
5	37	47	-10	100
6	131	121	10	100
7	77	136	-59	3481
8	24	65	-41	1681
9	99	131	-32	1024
10	629	630	-1	1
11	116	104	12	144
			-138	7718

$$\bar{d} = \frac{\sum_{i=1}^{11} d_i}{11} = \frac{-138}{11} = -12{,}55$$

$$s_d^2 = \frac{11 \cdot \sum_{i=1}^{11} d_i^2 - \left[\sum_{i=1}^{11} d_i\right]^2}{11 \cdot 10} = \frac{11 \cdot 7718 - (-138)^2}{110} = 598{,}67 \Rightarrow$$

$$s_d = \sqrt{598.67} = 24{,}47$$

Behaturiko estatistikoaren balioa $t_p = \dfrac{-12{,}55}{24{,}47/\sqrt{11}} = -1{,}7$ da.

Eta p-balioa hauxe da: $p = P(t_{10} < -1{,}7) = P(t_{10} > 1{,}7)$.

t banaketaren taulan, $P(t_{10} > 1{,}372) = 0{,}10$ eta $P(t_{10} > 1{,}812) = 0{,}05$ ditugu. Beraz, $0{,}05 < p < 0{,}10$. Ondorioz, ezin dugu H_0 errefusatu, eta ezin dugu esan % 5eko adierazgarritasun-mailaz azterketa fisikoak triglizerido-maila handitzen duenik.

2. <u>Eskuinetikoa</u>

H_0: $\mu_D \leq 0$

H_1: $\mu_D > 0$

Kasu honetan, hauxe da *p*-balioa: $p = P(t_{n-1} > t_p)$

6.21 adibidea

6.18 adibideko datuetan oinarriturik, ondoriozta al daiteke, % 5eko adierazgarritasun-mailaz, ariketa fisikoak batez besteko kolesterol-maila txikitzen duenik?

<u>Emaitza</u>:

Honako hipotesi hauek kontrastatzen dira:

H_0: $\mu_D \leq 0$ (ez du txikitzen)

H_1: $\mu_D > 0$ (txikitzen du)

6.19 adibidean $\bar{d} = 33{,}18$ eta $s_d = 51{,}066$ kalkulatu ditugu.

Orduan, estatistikoaren balioa $t_p = \dfrac{33{,}18}{51{,}066/\sqrt{11}} = 2{,}155$ da.

Eta hauxe da *p*-balioa: $p = P(t_{10} > 2{,}155)$.

t banaketaren taulan, $P(t_{10} > 1{,}812) = 0{,}05$ eta $P(t_{10} > 2{,}228) = 0{,}025$ ditugu. Orduan, $0{,}025 < p < 0{,}05$. Ondorioz, % 5eko adierazgarritasun-mailaz, H_0 hipotesia errefusatu egingo dugu, eta ariketa fisikoak batez besteko kolesterol-maila txikitu egiten duela ondoriozta dezakegu.

6.6.2.2 Aldebiko kontrastea

H_0: $\mu_D = 0$

H_1: $\mu_D \neq 0$

Kasu honetan, hauxe da *p*-balioa: $p = 2 \cdot P(t_{n-1} > |t_p|)$

6.22 adibidea

Antisorgailuak hartu aurretik eta ondoren, hamar emakumeren odolaren presioa neurtu da (mm Hg-tan), eta honako emaitza hauek lortu dira:

Emakumea	Antisorgailuak hartu aurreko odol-presioa: x	Antisorgailuak hartu osteko odol-presioa: y
1	115	128
2	112	115
3	107	106
4	119	128
5	115	122
6	138	145
7	126	132
8	105	109
9	104	102
10	115	117

Ondoriozta al dezakegu, % 1eko adierazgarritasun-mailaz, antisorgailuek odolaren presioan eragina dutenik?

Emaitza:

Hauexek dira kontrastatzen ditugun hipotesiak:

H_0: $\mu_D = 0$ (ez dute eraginik)

H_1: $\mu_D \neq 0$ (badute eragina)

Emakumea	Antisorgailuak hartu aurreko odol-presioa: x	Antisorgailuak hartu osteko odol-presioa: y	Diferentzia $d = x - y$	d^2
1	115	128	-13	169
2	112	115	-3	9
3	107	106	1	1
4	119	128	-9	81
5	115	122	-7	49
6	138	145	-7	49
7	126	132	-6	36
8	105	109	-4	16
9	104	102	2	4
10	115	117	-2	4
			-48	418

Lagineko diferentzien batezbestekoa eta kuasibariantza kalkulatu behar dira:

$$\bar{d} = \frac{\sum_{i=1}^{10} d_i}{10} = \frac{-48}{10} = -4{,}8$$

$$s_d^2 = \frac{10 \cdot \sum_{i=1}^{10} d_i^2 - \left[\sum_{i=1}^{10} d_i\right]^2}{10 \cdot 9} = \frac{10 \cdot 418 - (-48)^2}{90} = 20{,}844$$

$$s_d = \sqrt{20{,}844} = 4{,}566$$

Behaturiko estatistikoaren balioa $t_p = \frac{-4{,}8}{4{,}566/\sqrt{10}} = -3{,}32$ da. Eta hauxe da *p*-balioa: $p = 2 \cdot P(t_9 > 3{,}32)$.

t banaketaren taulan $P(t_9 > 3{,}25) = 0{,}005$ dugu. Orduan, $P(t_9 > 3{,}32) < P(t_9 > 3{,}25) = 0{,}005 \Rightarrow p < 0{,}01$.

Beraz, % 1eko adierazgarritasun-mailaz, H_0 hipotesia errefusatzen dugu eta antisorgailuek odolaren presioan eragina dutela ondoriozta dezakegu. Kontraste horrek ez du adierazten zein noranzkotan den eragin hori, alegia, odolaren presioa txikitu ala handitu egiten den antisorgailuak hartu ondoren. Galdera horri erantzuteko, aldebakarreko kontrasteak erabili beharko genituzke.

SPSS pakete estatistikoa erabil dezakegu binakako datuak aztertzeko. Adibide honen datuak sortuz, hauexek dira *SPSS*ren bidez lorturiko emaitzak:

Prueba de muestras relacionadas

		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típ.	Error típ. de la media	99% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Antisorgailuak hartu aurreko odol-presioa: x - Antisorgailuak hartu osteko odol-presioa: y	-4,800	4,566	1,444	-9,492	-,108	-3,325	9	,009

Taula horretan agertzen dira estatistikoa (*t* = -3,325), askatasun-graduen kopurua (*gl* = 9) eta aldebiko kontrasterako *p*-balioa (*Sig.*(*bilateral*) = 0,009).

Beraz, *p*-balioa = 0,009 < 0,01 denez, H_0 hipotesia errefusatzen dugu eta antisorgailuek odolaren presioan eragina dutela ondorioztatzen dugu, eskuz egindakoan ondorioztatzen dugun bezalaxe.

*SPSS*ren emaitzen taulan, % 99ko konfiantza-tartea agertzen da:

$$I_{\mu_D}^{0.99} .= (-9{,}492,\ -0{,}108)$$

Konfiantza-tarte horretan oinarrituz, % 99ko konfiantza-mailaz, $\mu_D < 0$ dela baiezta dezakegu, hau da, $\mu_{aurretik} - \mu_{ostean} < 0$ dela. Beste era batera esanda, antisorgailuek handitu egin dute odolaren batez besteko presioa.

Oharra. Konfiantza-tartea kalkulatzerakoan aipatu dugun moduan, hipotesi-kontrasteak egiterakoan, $n \geq 30$ denean, kontuan hartu beharko dugu estatistikoaren banaketa asintotikoa $N(0,\ 1)$ dela, hots, $\frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \approx Z$ dela. Estatistikoaren balioa $z_p = \frac{\overline{d}}{s_d / \sqrt{n}}$ izango da, eta p-balioa kalkulatzeko Z banaketa erabili beharko dugu.

6.7 ARIKETAK

6.1. *A* (normala) eta *B* (adrenalizatua) bi arratoi motatan egindako esperimentuan, honako odol-biskositatea aurkitu zuten:

Arratoia	Odol-biskositatea										
A	3,29	3,91	4,64	3,55	3,67	4,18	3,74	4,67	3,03	4,61	3,84
B	3,35	4,26	4,71	3,14	3,45	5,01	4,43	4,91	4,22	4,83	3,55

(Iturria: A. Urkaregi, *Bioestatistika*. UEU, 1991)

Odol-biskositatea banaketa normalari darraio.

a) Zehaztu itzazu bi populazioen batezbestekoen arteko diferentziarako % 95eko konfiantza-mugak. Interpreta itzazu emaitzak.

b) Baiezta al dezakegu, % 5eko adierazgarritasun-mailaz, adrenalinak batez besteko odol-biskositatea emendatzen duela?

6.2. Bi armiarma-espezieren elikatze-ohiturak aztertu dituzte. *Dinopis* eta *Menneus* izeneko espezie horiek Australiako ekialdean bizi dira. Banaketa normalari darraion espezie bakoitzaren harrapakinen tamainan interesaturik gaude. *Menneus* helduaren tamaina gutxi gorabehera *Dinopis* gaztearena da. Ezaguna da *Dinopis* helduaren eta gaztearen harrapakinen tamainen artean diferentzia dagoela, beren tamaina dela kausa. Ba al dago *Dinopis* helduaren eta *Menneus* helduaren harrapakinen batez besteko tamainaren arteko diferentzia adierazgarririk? Horrela baldin bada, zein da arrazoia? Galdera horiek erantzuteko, bi espezieen harrapakinen tamainari buruzko honako informazio hau lortu zuten (zentimetrotan):

Dinopis heldua		*Menneus* heldua	
12,9	11,9	10,2	5,3
10,2	7,1	6,9	7,5
7,4	9,9	10,9	10,3
7,0	14,4	11,0	9,2
10,5	11,3	10,1	8,8

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Erabil itzazu datu horiek populazioen bariantzak konparatzeko (kontuan hartu $P(F_{9;9} > 3{,}18) = 0{,}05$ dela).

b) Populazioen bariantzen ezberdintasunik ez bada aurkitzen, lor ezazu ($\mu_1 - \mu_2$)-rako % 90eko konfiantza-tartea.

c) Aurreko ataleko konfiantza-tartean oinarriturik, ba al dago bi espezieen harrapakinen batez besteko tamainaren arteko diferentziaren ziurtasunik? Arrazoitu ezazu. (Biologoek uste dute aurkitutako edozein ezberdintasun egoeraren eta sarearen tamainaren diferentziagatik dela)

d) Ondoriozta al daiteke, % 5eko adierazgarritasun-mailaz, *Menneus* helduaren harrapakinen batez besteko tamaina *Dinopis* helduarena baino txikiagoa denik?

6.3. Onarturik dago beroak eragindako tentsioari emandako erantzunean sexuaren araberako diferentziak daudela. Gurpil-tresnak erabiltzea behartzen zuen ariketa gogorreko programaren pean hamar gizon eta zortzi emakume jarri ziren. Ingurunea beroa eta ur kantitateak minimoak ziren. Aztertuko den aldagaia galdutako gorputzaren pisu-portzentajea da. Bi sexuek galdutako pisu-portzentajeen banaketak normalak dira; gizonezkoen kasuan desbideratze estandarra 0,3 da, eta emakumezkoen kasuan 0,5. Honako datu hauek lortu ziren:

Gizonak		Emakumeak	
2,9	3,7	3,0	3,8
3,5	3,8	2,5	4,1
3,9	4,0	3,7	3,6
3,8	3,6	3,3	4,0
3,6	3,7		

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Kalkula ezazu baldintza horietan ariketa fisikoa egiten duten gizon eta emakumeen gorputzeko batez besteko pisu-galeraren portzentajeen arteko kendurarako puntu-estimazioa.

b) Kalkula ezazu populazioen batezbestekoen arteko kenduraren % 95eko konfiantza-tartea. Konfiantza-tarte horretan oinarrituz, uste duzu inguruko beroak eragindako estresari emandako erantzunean, gizonen eta emakumeen arteko diferentziarik dagoenik? Azal ezazu erantzuna.

c) Onar al dezakegu, % 5eko adierazgarritasun-mailaz, batez beste gizonek emakumeek baino pisu gehiago galtzen dutenik?

6.4. Igerilari eta korrikalari olinpiarren atributu fisiko batzuk konparatzeko azterketa egin dute. Intereseko aldagaietariko bat gorputzeko koipe-edukia (kilogramotan) da. Hamabi korrikalari eta hamar igerilarik osaturiko laginetan, honako datu hauek lortu dituzte:

Korrikalariak		Igerilariak	
11,2	7,6	14,1	12,7
10,1	7,3	15,1	13,7
9,4	6,9	11,4	11,9
9,2	5,5	14,3	10,7
8,3	5,0	9,2	8,7
8,2	3,7		

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Lor ezazu igerilari eta korrikalari olinpiarren batez besteko koipe-eduki osoen arteko kendurarako puntu-estimazioa.

b) Normaltasuna onartuz, kalkula ezazu populazioen batezbestekoen arteko kenduraren % 99ko konfiantza-tartea. Konfiantza-tarte horretan oinarrituz, uste duzu igerilarien eta korrikalarien gorputzeko batez besteko koipe-eduki osoen arteko diferentziarik dagoenik? Azal ezazu erantzuna. (Kontuan hartu $P(F_{9;11} > 2{,}9) = 0{,}05$ dela).

c) Igerilariek korrikalariek baino batez besteko koipe-eduki handiagoa al daukate gorputzean? Arrazoitu ezazu erantzuna p-balioaren arabera.

6.5. Osagai ezberdinetan oinarriturik dauden bihotzeko transplanterako bi tratamenduren erabilera potentzialaren azterketa egin da. Osagaien helburua inmunokentzaileak izatea da: transplantea errefusatzeko gorputzaren joera naturala eragoztea. ACI arratoi arrak emaileak dira, eta Lewis Brown Norway arratoi arrak jasotzaileak. Bi arratoi espezie horien bateragarritasuna eskasa da. Aztertuko den aldagai normala, X, bizirik iraundako egun kopurua da. Honako emaitza estatistiko hauek lortu dira:

Sodio-salikalatoa soilik	Sodio-salikalatoa eta azatiopirina
$n_1 = 90$	$n_2 = 90$
$\bar{x}_1 = 16$ egun	$\bar{x}_2 = 15$ egun
$s_1 = 10{,}1$ egun	$s_2 = 10$ egun

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Kalkula ezazu bi tratamenduen batez besteko biziraupen-denboren arteko kendurarako % 90eko konfiantza-tartea. Interpreta ezazu konfiantza-tarte hori, bere ondorio praktikoen arabera.

b) Egiazta itzazu a) atalean lortutako emaitzak hipotesi-kontrastearen bidez.

6.6. Dietari eta dibertikuluari buruzko gaixotasunaren azterketan hogeita hiru barazkijale erabili ziren, eta haien zuntz dietetiko eduki osoa aztertu zen. Demagun normaltasuna onar daitekeela. Gaixotasuna pairatzen dutenentzat eta ez dutenentzat honako datu hauek lortu ziren:

Osasuntsuak	Gaixoak
$n_1 = 18$	$n_2 = 5$
$\bar{x}_1 = 42{,}7$ g	$\bar{x}_2 = 27{,}7$ g
$s_1 = 9{,}9$ g	$s_2 = 9{,}5$ g

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Kontrasta ezazu populazioen bariantzen berdintasuna (kontuan hartu $P(F_{17;4} > 5{,}83) = 0{,}05$ dela).

b) Nahikoa arrazoirik ba al dago zuntz dietetikoaren batez besteko eduki osoa gaixotasuna pairatzen ez dutenen artean gaixoen artean baino altuagoa dela pentsatzeko? Azal ezazu erantzuna, kontrastearen p-balioan oinarriturik.

c) Ebatz ezazu b) atala, % 95eko konfiantza-tartea erabiliz.

6.7. Igerilari eta korrikalari olinpiarren atributu fisikoak konparatzeko erabiltzen den beste aldagai bat da, erlaxaturik daudenean, besoaren goialdeko zirkunferentzia (zentimetrotan). Kirolari mota bietan, zirkunferentziaren banaketa 2 cm-ko desbideratze estandarra duen banaketa normalari darraio. Aurretik azterturiko laginetan, honako datu hauek lortu ziren:

Igerilariak	Korrikalariak
$n_1 = 10$	$n_2 = 12$
$\bar{x}_1 = 27{,}3$ cm	$\bar{x}_2 = 23{,}5$ cm

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Nahikoa arrazoirik ba al dago igerilarien besoaren goialdeko batez besteko zirkunferentzia korrikalariena baino handiagoa dela pentsatzeko? Azal ezazu erantzuna kontrastearen p-balioan oinarrituz.

b) Konpara itzazu igerilarien eta korrikalarien batez besteko zirkunferentziak % 95eko konfiantza-tartea erabiliz. Zer ondorio atera genezake?

6.8. Mutil gazteak erretzen hasten diren adina neska gazteena baino goiztiarragoa dela uste da. Honako datu hauek egiaztatzen al dute uste hori?

Mutilak	Neskak
$n_1 = 33$	$n_2 = 44$
Erretzen hasten diren batez besteko adina = 11.3 urte	Erretzen hasten diren batez besteko adina = 12.6 urte
$s_1^2 = 4$	$s_2^2 = 3{,}5$

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

6.9. Ikerketa bat egin da gibeleko zirrosia pairatzen duten gaixoek dituzten zelula batzuk hiltzeko monozitoen ahalmena aztertzeko. Zelula horiek kaltegarriak dira, zeren eta gaixoa hainbat zoldura mota harrapatzeko arriskuan uzten baitute. 16 zirrosidun gaixo eta 9 kontrolen odol-laginak hartu dira, eta monozitoak hildako zelula mota horretako portzentajea neurtu da (onar ezazu normaltasuna). Honako datu hauek lortu dira:

Kontrolak	Gaixoak
$n_1 = 9$	$n_2 = 16$
$\bar{x}_1 = \%\ 44{,}22$	$\bar{x}_2 = \%\ 28{,}22$
$s_1 = \%\ 6{,}17$	$s_2 = \%\ 4{,}11$

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Konpara itzazu populazioen bariantzak. Azken kontraste horretan oinarriturik, aukera ezazu populazioen batezbestekoak konparatzeko t estatistiko egokia. Nahikoa arrazoirik ba al dago kontroletan monozitoek hildako batez besteko zelula-portzentajea gaixoengan hildakoena baino handiagoa dela pentsatzeko? Arrazoitu ezazu erantzuna konfiantza-tartearen eta hipotesi-kontrastearen bidez.

6.10. Bi etxebizitza motaren uretako berun-kontzentrazioa konparatzeko ikerketa egin da. Etxebizitza mota batera heltzen diren ur-tutuak % 50 berunez eta % 50 eztainuz egindakoak dira. Beste etxebizitzakoak, ordea, ez. Datuak bi etxeetako sukaldeetan eta komunetan hartutako ur-laginetakoak dira.

Lehenengo tokiko uraren batez besteko berun-kontzentrazioak bigarren tokikoa gainditzen duela uste da. Berun-kontzentrazioa banaketa normalari darraiola suposatuz, egiazta ezazu aurreko baieztapena, hipotesi-kontrastearen eta konfiantza-tartearen bidez.

Etxebizitza 1 (% 50 beruna eta % 50 eztainua)	Etxebizitza 2
$n_1 = 25$	$n_2 = 25$
$\bar{x}_1 = 390$ ppb	$\bar{x}_2 = 217.5$ ppb
$s_1 = 10$ ppb	$s_2 = 5$ ppb

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

6.11. Ikerketa bat egin da itsas txakur austral gazteen larruazaleko plasmaren sodio-edukia beren esnearen sodio-edukiarekin konparatzeko. Sodio-edukia banaketa normalari darraio. Honako hauek dira zoriz aukeraturiko behaketak:

Alea	Esnea	Plasma
1	93	147
2	104	157
3	95	142
4	81,5	141
5	95	142
6	95	147
7	76,5	148
8	90,5	144
9	79,5	144
10	87,0	146

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Kalkula ezazu bi gorputz-fluidoetako sodio-edukien batez besteko kendurarako % 95eko konfiantza-tartea. Ezberdintasunik frogatzen al da? Horrela baldin bada, noranzkoa da ezberdintasuna?

b) Egiazta ezazu a) atalean erantzundakoa, hipotesi-kontrastearen bidez.

6.12. Hamar partzela-bikote daude. Bikote bakoitzeko batek (1. partzelak) ongarri fosforikoa dauka, eta besteak (2. partzelak) ez; gaineratiko baldintzak berberak dira. Laborearen etekina adierazten du honako taula honek:

1. partzela	6,6	6,5	5,6	6,6	6,1	6	5,8	6,1	6,3	6,1
2. partzela	6	5,4	5,8	5,4	5,8	5,4	5,7	5,7	6	5,3

(Iturria: A. Urkaregi, *Bioestatistika*. UEU, 1991)

Etekinaren banaketa normala dela eta ongarriak etekinaren desbideratze estandarrean eraginik ez daukala onartzen da.

a) Kalkula itzazu ongarririk gabeko partzelen etekinaren bariantzarako % 99ko konfiantza-tartea, eta batezbestekoen arteko kendurarako maila bereko tartea. Emaitza horietatik, ondoriozta al daiteke ongarri fosforikoaren eraginkortasuna? ($P(\chi^2_9 > 23{,}59) = 0{,}005$)

b) Azter ezazu ongarri fosforikoaren eraginkortasuna, α = % 5eko adierazgarritasun-mailako hipotesi-kontrastea erabiliz. Erabilitako metodoa egokia al da eraginkortasuna aztertzeko? Zergatik? Azter ezazu berriro ongarri fosforikoaren eraginkortasuna, metodo egokiena erabiliz.

6.13. Hautazko kirurgian, ebakuntza aurreko zehar-berdinketari buruzko ikerketa egin da. Azterturiko ebakuntza hautazko sabel-histerektomia da. *X* aldagaia, berehala prest dauden zeharka kontrastaturiko odol unitate kopurua, aztertuko da. Helburua da 1990. urtean prest zeuden batez besteko unitate kopurua gaur egunekoarekin konparatzea. Unitate kopurua banaketa normalari darraio.

1990	Gaur egun
$n_1 = 25$	$n_2 = 25$
$\bar{x}_1 = 2{,}73$	$\bar{x}_2 = 1{,}27$
$s_1 = 0{,}65$	$s_2 = 1{,}10$

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Populazioen bariantzak konparatu ondoren, kalkula ezazu ($\mu_1 - \mu_2$)-rako % 95eko konfiantza-tartea. Ba al dago 1990etik gaur egunera prest dauden batez besteko unitate kopuruaren beherapena izan denaren nabaritasunik? Arrazoitu ezazu erantzuna.

b) Azter ezazu 1990. urtetik gaur egun arte prest egon diren batez besteko unitate kopuruaren beherapena, hipotesi-kontrastea erabiliz.

6.14. Diabetikoei beren odoleko glukosa-edukia menperatzen laguntzeko etxeko neurgailuaren eragina zehazteko ikerketa egin da. Zoriz aukeraturiko hogeita hamasei diabetikok hartu dute parte ikerketan. Paziente bakoitzaren odoleko glukosa-edukiaren bi neurketa egin dira: bat neurgailua erabiltzen irakatsi baino lehen, eta beste bat neurgailua hainbat astean erabili ondoren. Diferentzien

laginean 2,78 mmol/litroko batezbestekoa eta 6,05 mmol/litroko desbideratze estandarra agertu dira (kenduraren ordena «aurrekoa» ken «ostekoa» da). Nahikoa seinalerik ba al daukagu neurgailua pazienteen odoleko glukosa-edukia murrizten laguntzeko eraginkorra dela esateko? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

6.15. Nahiko bizkorrak diren ariketa arruntez osaturiko programa komenigarria izango litzatekeela uste da aurretik bihotzekoren bat eduki izan duten gaixoentzat. Baieztapen hori egiaztatzeko, hamaika gaixok parte hartu zuten programan. Programa bera hasi baino lehen, pertsona bakoitzaren lan egiteko ahalmena neurtu zen, ibiltzeko zintaren gainean minutuko 160 taupadako tasara heltzeko beharrezkoa zen denbora neurtuz. Hogeita bost asteko kontrolpeko ariketa egin ondoren, berriz hartu ziren ibiltzeko zintaren gaineko neurriak, eta honako datu hauek lortu ziren:

Pertsona	Aurretik	Ostean
1	7,6	14,7
2	9,9	14,1
3	8,6	11,8
4	9,5	16,1
5	8,4	14,7
6	9,2	14,1
7	6,4	13,2
8	9,9	14,9
9	8,7	12,2
10	10,3	13,4
11	8,3	14,0

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Mantentzen al dute datu horiek ikertzaileen argudioa?

6.16. 1950. urtean, mundu osoko lurreko eta itsasoko behaketa meteorologikorako 1.000 estaziotan lorturiko tenperaturen datuek 14 ºC-ko batez besteko tenperatura eman ziguten. 1988. urtean, estazio beretako batez besteko tenperatura 14,5 ºC zen. 1950eko eta 1988ko irakurketak estaziorako binakatuz, irakurketen diferentziaren s_d = 0,2 ºC-koa dela estimatzen da. Tenperaturaren banaketa normala dela suposatuz, datu horiek sostengatzen al dute 1988ko batez besteko tenperatura 1950ekoa baino altuagoa den hipotesia? (Iturria: J.S.

Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

7 POPULAZIO BINOMIALETARAKO INFERENTZIA

Aurreko gaietan aztertu ditugun inferentzia-metodoak populazio baten batezbestekoari eta bariantzari buruzkoak izan dira, edo, bi populazioren kasuan, batezbestekoak eta bariantzak konparatzeko metodoak eman ditugu. Beti populazioak normalak izan dira, edo bestela, lagin-tamaina handiak erabili ditugu, banaketa normalaren hurbilketa erabili ahal izateko.

Gai honetan aztertuko dugun populazioa ez da normala izango, binomiala baino. Hau da, aztertuko dugun populazioa dikotomikoa izango da, bi klasez osaturikoa. Populazio horietan bi aukera soilik ditugu: zerbait gertatzea ala ez gertatzea. Esate baterako, sexuaren arabera azterturiko populazioa binomiala da, bi aukera besterik ez baitaude: emakumea ala gizona, emea ala arra. Gauza bera gertatzen da gaixotasun bat aztertzen dugunean ere; bi aukera izango ditugu: gaixoa ala osasuntsua.

Bi aukera horien arteko bat izango da gure interesekoa (*arrakasta* esango diogu), eta hori gertatzeko probabilitateari p esango diogu. Populazio binomialetik n tamainako zorizko lagina aukeratu eta bertan agertzen den arrakasta kopurua aztertzen badugu, arrakasta kopuruaren banaketa Bin(n, p) izango da. Gai honetan, p parametro horri buruzko inferentzia aztertuko dugu.

7.1 BANAKETA BINOMIALAREN p PARAMETROAREN ESTIMAZIOA

7.1.1 Puntu-estimazioa

Demagun p parametroko populazio binomiala dugula eta n tamainako lagina zoriz aukeratu dugula, bertan X arrakasta kopurua zenbatuz.

$\hat{P} = \frac{X}{n}$ lagin-proportzioa hartuko dugu p parametroaren estimatzailetzat.

7.1.2 Lagin-proportzioaren banaketa

Demagun $n \cdot \hat{p} > 5$ eta $n \cdot \hat{q} > 5$ baldintzak betetzen direla; orduan, $\hat{P} = \frac{X}{n}$ lagin-proportzioaren banaketa, p batezbestekoa eta $\sqrt{\frac{pq}{n}}$ desbideratze estandarreko banaketa normala da, gutxi gorabehera.

Hain zuzen, 3. gaian ikusi dugunez, $n \cdot \hat{p} > 5$ eta $n \cdot \hat{q} > 5$ betetzen direnean, Moivre-ren teorema aplikatuz,

X: $Bin(n,p) \approx N(np, \sqrt{npq})$.

$\hat{P} = \frac{X}{n}$ hartuz, honako hau daukagu:

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p$$

$$Var(\hat{P}) = Var\left(\frac{X}{n}\right) = \frac{Var(X)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

Beraz, $n \cdot \hat{p} > 5$ eta $n \cdot \hat{q} > 5$ betetzen direnean, $\hat{P} \approx N(p, \sqrt{\frac{pq}{n}})$.

7.1.3 Banaketa binomialaren *p* parametrorako konfiantza-tartea

Demagun X p parametroko populazio binomialetik n tamainako lagina aukeratzen dugula, $n \cdot \hat{p} > 5$ eta $n \cdot \hat{q} > 5$ baldintzak beterik. Kasu horretan, $\hat{P}$-ren banaketa ezagutzen dugu, eta horrek p-rako konfiantza-tartea kalkulatzea baimentzen digu.

Hain zuzen, $\hat{P} \approx N(p, \sqrt{\frac{pq}{n}})$ banaketa tipifikatzen badugu, $\frac{\hat{P} - p}{\sqrt{pq/n}} \approx Z$ izango da.

Badakigu banaketa normalerako $1 - \alpha = P(-z_{\alpha/2} < Z < z_{\alpha/2})$ betetzen dela. Ekuazio horretan, Z-ren ordez lagin-proportzioaren banaketa tipifikatua jartzen badugu, honako hau lortuko dugu:

$$1 - \alpha = P\left(-z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{p \cdot q / n}} < z_{\alpha/2}\right) = P\left(-\hat{P} - z_{\alpha/2}\sqrt{\frac{pq}{n}} < -p < -\hat{P} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) =$$

$$= P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$$

Beraz, (L_1, L_2) tartea lortu dugu, non $P(L_1 < p < L_2) = 1 - \alpha$ betetzen baita:

$$L_1 = \hat{P} - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

$$L_2 = \hat{P} + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

$\frac{pq}{n}$ kantitatea ezezaguna denez, $\frac{\hat{p}\hat{q}}{n}$ -ren bidez estimatzen da, hau da, lagina aukeratu ondoren lortu dugun $\hat{p}$ lagin-proportzioa erabiliz; eta, beraz, honako hau da *p*-rako **%100 × (1 – α) mailako konfiantza-tartea**:

$$I_p^{1-\alpha} = \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\,,\, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

7.1 adibidea

50 urtetik gorako umedun emakumeen bularreko minbizi-tasa estimatu nahi dugu. Datu estatistikoetatik jakin dugu baldintza horiek betetzen dituzten 10.000 emakumeen artetik 400ek minbizia pairatzen dutela.

Emaitza:

Datu horiek erabilirik, 50 urtetik gorako umedun emakumeen bularreko minbizi-tasaren estimazioa $\hat{p} = \frac{400}{10000} = 0{,}04$ izango da.

Hauxe da 50 urtetik gorako umedun emakumeen bularreko minbizi-tasarako % 95eko konfiantza-tartea:

$$I_p^{0.95} = \left(0{,}04 - 1{,}96 \cdot \sqrt{\frac{0{,}04 \cdot 0{,}96}{10000}}\,,\, 0{,}04 + 1{,}96 \cdot \sqrt{\frac{0{,}04 \cdot 0{,}96}{10000}} \right) =$$

$$= (0{,}04 - 0{,}004,\, 0{,}04 + 0{,}004) = (0{,}036,\, 0{,}044)$$

Demagun 50 urtetik gorako emakumeen (umedun zein umegabeen) bularreko minbizi-tasa orokorra % 2koa dela. 0,02 balioa ez dago % 95eko konfiantza-tartearen barnean, eta, tarte horren behe-muturra baino txikiagoa denez, ondoriozta dezakegu, % 95eko konfiantza-mailaz, 50 urtetik gorako umedun emakumeen bularreko minbizi-tasa adin bereko emakumeena (umedun zein umegabeena) baino handiagoa dela.

7.1.4 Lagin-tamainaren kalkulua, banaketa binomialaren *p* parametrorako konfiantza-tartearen luzeran oinarriturik

Demagun, p proportzioa estimatu nahi dugula, %100 × (1 – α) mailako konfiantza-tartearen luzera L izatean baldintzaturik. Zenbatekoa izan beharko du laginaren tamainak, baldintza hori bete dadin?

$L = 2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$ bete behar denez, n isolatuz lortuko dugu beharrezkoa den lagin-tamaina: $n = 4 z_{\alpha/2}^2 \frac{\hat{p}\hat{q}}{L^2}$.

7.2 adibidea

Akne deritzon gaixotasunaren aurkako azken aurrerakuntza cis-13-retinoico azidoa izeneko sendagaia da. Orain dela gutxi egindako ikerketan, sendagai hori erabili zen akne larria zeukaten hamalau gaixorengan. Horietako hamairu gaixoen lesioak erabat sendatu ziren. Osteko azterketan, akne pairatzen duten eta sendagai horren bidez sendatzen diren gaixoen proportzioa estimatu nahi dugu, % 2ko errorea eta % 90eko konfiantza-maila erabiliz. Zenbat paziente tratatu behar izango ditugu? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994)

Emaitza:

Konfiantza-tartearen luzera $L = 2 \cdot 0{,}02 = 0{,}04$ izango da.

$1 - \alpha = 0{,}90$ konfiantza-mailarako, $z_{\alpha/2} = z_{0.05} = 1{,}645$.

Hamalau gaixoek osaturiko laginean lortutako p-ren estimazioa kalkulatuko dugu: $\hat{p} = \frac{13}{14} = 0{,}93$

Beraz, hauxe da beharrezkoa den laginaren tamaina:

$$n = 4 \cdot z_{\alpha/2}^2 \cdot \frac{\hat{p} \cdot \hat{q}}{L^2} = 4 \cdot 1{,}645^2 \cdot \frac{0{,}93 \cdot 0{,}07}{0{,}04^2} = 440{,}4$$

Hau da, 441 paziente tratatu beharko ditugu.

7.2 BANAKETA BINOMIALAREN p PARAMETRORAKO HIPOTESI-KONTRASTEA

Demagun X p parametroko populazio binomialetik n tamainako lagina aukeratzen dugula. $n \cdot \hat{p} > 5$ eta $n \cdot \hat{q} > 5$ baldintzak betetzen badira, badakigu hauxe dela $\hat{P}$-ren banaketa: $\hat{P} \approx N(p, \sqrt{\frac{pq}{n}})$ edo $\frac{\hat{P} - p}{\sqrt{pq/n}} \approx Z$

Atal honetan aztertuko ditugun hipotesi-kontrasteetan, p parametroari aurrefinkaturiko p_0 balioa esleituko diogu, eta erabiliko dugun estatistikoa $Z_p = \frac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}}$ izango da.

Estatistiko horren banaketa, H_0 egia dela suposatuz, asintotikoki $N(0,1)$ izango da.

Lagina aukeratu ondoren, eta H_0 hipotesiaren pean, hauxe izango da estatistikoaren balioa: $z_p = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$

Hipotesi-kontraste motaren arabera, p-balioa kalkulatzeko era aldatu egingo da, hipotesi alternatiboaren arabera kalkulatu behar baita; baina estatistiko berbera erabiliko da aldebakarreko zein aldebiko kontrasteetan.

7.2.1 Aldebakarreko kontrastea

7.2.1.1 Ezkerretikoa

H_0: $p \geq p_0$
H_1: $p < p_0$

Kasu honetan, hauxe da p-balioa: $p = P(Z < z_p)$.

7.3 adibidea

Burmuineko gaixotasun birikoak eragindako heriotza-tasa % 70ekoa da. *Vidarabina* deritzon sendagai berriaren ikerketa egin da. Berrogeita hamar paziente *vidarabinaz* tratatu ondoren, horietariko hogeita lau hil egin ziren. Ondoriozta dezakegu, % 5eko adierazgarritasun-mailaz, gaixotasun hori tratatzeko eraginkorra denik *vidarabina*? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Emaitza:

Honako hipotesi hauek kontrastatu nahi ditugu:

H_0: $p \geq 0{,}70$ (*vidarabina* ez da eraginkorra)
H_1: $p < 0{,}70$ (*vidarabina* eraginkorra da)

Laginean hildakoen proportzioa $\hat{p} = \frac{24}{50} = 0{,}48$ da.

Beraz, H_0 egiazkoa dela suposatuz, behaturiko estatistikoaren balioa $z_p = \frac{0{,}48 - 0{,}70}{\sqrt{0{,}70 \cdot 0{,}30 / 50}} = -3{,}4$ da.

Hauxe da p-balioa: $p = P(Z < -3{,}4) = P(Z > 3{,}4) = 0{,}000337 < 0{,}05$.

Ondorioz, hipotesi nulua errefusatzen da eta, % 5eko adierazgarritasun-mailaz, baiezta daiteke *vidarabina*k gaixotasun horren heriotza-tasa txikitzen duela, eta, beraz, eraginkorra dela.

7.2.1.2 Eskuinetikoa

H_0: $p \leq p_0$
H_1: $p > p_0$

Kasu honetan, hauxe da p-balioa: $p = P(Z > z_p)$.

7.4 adibidea

«Kanibalismoaren bidez transferituriko irakaskuntza» deritzona nahiko eztabaida sortu duen ikasteari buruzko teoria da. Teoria horren balioa estatistikoki frogatzeko asmoz egindako ikerketan, harrapari taldearen heziketa egin zen, *shock* elektrikoa saihets zezaten. Harrapari horiek zatitu ondoren, hezi gabeko ehun harraparik osaturiko taldeari janaritzat eskaini zitzaien, eta horietako berrogeita hamazazpik *shock* elektrikoa saihestu zuten. % 5eko adierazgarritasun-mailaz, ikerketa horrek egiaztatzen al du «kanibalismoaren bidez transferituriko irakaskuntza» deritzon teoriaren baliozkotasuna? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Emaitza:

Irakaskuntza transferentziarik ez bada egon, heziketarik gabeko harrapariek *shock*a saihesteko probabilitatea 1/2 dela onartuko dugu. Ikerketa horren

helburua teoria baieztatzea denez, hipotesi alternatiboa $p > 1/2$ izango da.

Kontrastatu nahi ditugun hipotesiak, beraz, honako hauek dira:

H_0: $p \leq 1/2$

H_1: $p > 1/2$ (Kanibalismoak *shock*a saihesteko probabilitatea handitzen du)

H_0 egia dela suposatuz, behaturiko estatistikoaren balioa $z_p = \frac{0{,}57 - 0{,}5}{\sqrt{0{,}5 \cdot 0{,}5 / 100}} = 1{,}4$ da.

Hauxe izango da p-balioa: $p = P(Z > 1{,}4) = 0{,}0808 > 0{,}05$, beraz, ezin dugu hipotesi nulua errefusatu. Ondorioz, % 5eko adierazgarritasun-mailaz, ikerketa horrek ez du egiaztatzen «kanibalismoaren bidez transferituriko irakaskuntza» deritzon teoriaren baliozkotasuna.

7.2.2 Aldebiko kontrastea

H_0: $p = p_0$

H_1: $p \neq p_0$

Kasu honetan, hauxe da p-balioa: $p = 2 \cdot P(Z > |z_p|)$.

7.5 adibidea

7.1 adibideko datuak erabilirik, esan al daiteke 50 urtetik gorako umedun emakumeen bularreko minbizi-tasa % 5ean ezberdina denik, % 5eko adierazgarritasun-mailaz?

Emaitza:

Hauxe da egin beharreko hipotesi-kontrastea:

H_0: $p = 0{,}05$

H_1: $p \neq 0{,}05$

H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0{,}04 - 0{,}05}{\sqrt{0{,}05 \cdot 0{,}095 / 10000}} = -4{,}588$$

p-balioa horrela kalkulatzen da: $p = 2\cdot P(Z > 4{,}588) \approx 0 < 0{,}05$. Beraz, hipotesi nulua errefusatzen da eta, % 5eko adierazgarritasun-mailaz, esan daiteke 50 urtetik gorako umedun emakumeen bularreko minbizi-tasa ez dela % 5 .

7.2.3 Kontrastearen ahalmena

Hipotesi-kontrasteak azaldu ditugunean, kontrastearen ahalmena definitu dugu era honetan: faltsua den hipotesia errefusatzeko probabilitatea da, hau da, $1 - \beta = P(H_0$ errefusatu$|\ H_0$ faltsua).

7.2.3.1 Ezkerretiko aldebakarreko kontrastea

H_0: $p \geq p_0$ hipotesi nulua H_1: $p = p_1 < p_0$ alternatiboaren aurka kontrastatu nahi dugu.

H_0 hipotesia errefusatuko dugu, p-balioa aurrefinkaturiko α adierazgarritasun-mailaren berdina edo txikiagoa denean, hau da, $p = P(Z < z_p) \leq \alpha$ denean. Beste era batera esanda, $Z_p = \dfrac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}}$ estatistikoaren balioa α adierazgarritasun-mailan oinarrituz kalkula daitekeen a konstantearen berdina edo hura baino txikiagoa denean.

Hain zuzen, $\alpha = P(H_0$ errefusatu$|\ H_0$ egia$) = P\left(\dfrac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \leq a \,\middle|\, p = p_0\right) = P(Z \leq a) = P(Z \geq -a)$.

Beraz, $a = -z_\alpha$ da eta eskualde kritikoa $S_1 = (-\infty, -z_\alpha]$ izango da.

Kontrastearen ahalmena faltsua den hipotesia errefusatzeko probabilitatea denez, kasu horretan:

$1 - \beta = P(H_0$ errefusatu$|\ H_0$ faltsua$) = P(Z \leq -z_\alpha \mid p = p_1) =$

$$P\left(\frac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \leq -z_\alpha \,\middle|\, p = p_1\right)$$

H_0 egiazkoa denean, $\dfrac{\hat{P}-p_0}{\sqrt{p_0q_0/n}} = Z$: $N(0,1)$ dela betetzen da. Baina H_1 egiazkoa denean, aurrekoa baizik honako hau betetzen da: $\dfrac{\hat{P}-p_1}{\sqrt{p_1q_1/n}} = Z : N(0,1)$.

Ondorioz:

$$1-\beta = P\left(\frac{\hat{P}-p_0}{\sqrt{p_0q_0/n}} \le -z_\alpha \mid p=p_1\right) = P\left(\hat{P} \le p_0 - z_\alpha \cdot \sqrt{\frac{p_0q_0}{n}} \mid p=p_1\right) =$$

$$= P\left(\frac{\hat{P}-p_1}{\sqrt{p_1q_1/n}} \le -z_\alpha \cdot \sqrt{\frac{p_0q_0}{p_1q_1}} + \frac{(p_0-p_1)\sqrt{n}}{\sqrt{p_1q_1}} \mid p=p_1\right) =$$

$$= P\left(Z \le -z_\alpha \sqrt{\frac{p_0q_0}{p_1q_1}} + \frac{(p_0-p_1)\sqrt{n}}{\sqrt{p_1q_1}}\right)$$

7.6 adibidea

7.3 adibideko datuetan oinarriturik, kalkula ezazu egindako kontrastearen ahalmena, % 1eko adierazgarritasun-maila eta % 50eko heriotza-tasa alternatiboa erabilirik.

<u>Emaitza</u>:

Honako hau da hipotesi-kontrastea:

H_0: $p \ge 0{,}70$ ($p_0 = 0{,}70$)
H_1: $p < 0{,}70$ ($p_1 = 0{,}50$)

H_0 hipotesia errefusatuko dugu, $Z_p = \dfrac{\hat{P}-0{,}70}{\sqrt{0{,}70 \cdot 0{,}30/50}}$ estatistikoaren behaturiko balioa a konstantea baino txikiago edo berdina denean. Alegia, $S_1 = (-\infty, a]$.

a balioa kalkulatzeko, $\alpha = 0{,}01$ adierazgarritasun-mailan oinarrituko gara:

$$0{,}01 = P(H_0 \text{ errefusatu} \mid H_0 \text{ egia}) = P\left(\frac{\hat{P}-0{,}70}{\sqrt{0{,}70 \cdot 0{,}30/50}} \le a \mid p=0{,}70\right) =$$

$=P(Z \leq a) = P(Z > -a)$.

Beraz, $a = -z_{0.01} = -2{,}3267$.

Ondorioz, hauxe izango da kontrastearen ahalmena:

$$1 - \beta = P(\mathrm{H}_0 \text{ errefusatu} \mid \mathrm{H}_0 \text{ gezurra}) =$$

$$P\left(\frac{\hat{P} - 0{,}70}{\sqrt{0{,}70 \cdot 0{,}30/50}} \leq -2{,}3267 \,\middle|\, p = 0{,}5\right) =$$

$$= P\left(\hat{P} \leq 0{,}7 - 2{,}3267 \cdot \sqrt{\frac{0{,}7 \cdot 0{,}3}{50}} \,\middle|\, p = 0{,}5\right) =$$

$$= P\left(\frac{\hat{P} - 0{,}5}{\sqrt{0{,}5 \cdot 0{,}5 / 50}} \leq -2{,}3267 \cdot \sqrt{\frac{0{,}7 \cdot 0{,}3}{0{,}5 \cdot 0{,}5}} + \frac{(0{,}7 - 0{,}5)\sqrt{50}}{\sqrt{0{,}5 \cdot 0{,}5}} \,\middle|\, p = 0{,}5\right) =$$

$$= P(Z \leq 0{,}7) = 1 - 0{,}2420 = 0{,}7580$$

7.2.3.2 Eskuinetiko aldebakarreko kontrastea

H_0: $p \leq p_0$ hipotesi nulua H_1: $p = p_1 > p_0$ alternatiboaren aurka kontrastatu nahi dugu.

H_0 hipotesia errefusatuko dugu aurrefinkaturiko α adierazgarritasun-maila baino p-balio berdina edo txikiagoa ateratzen denean, hau da, $p = P(Z > z_p) \leq \alpha$ denean. Beste era batera esanda, $Z_p = \dfrac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}}$ estatistikoaren balioa α adierazgarritasun-mailan oinarrituz kalkula daitekeen b konstantea baino handiago edo berdina denean.

Honako era honetara kalkulatuko dugu b konstantea:

$$\alpha = P(\mathrm{H}_0 \text{ errefusatu} \mid \mathrm{H}_0 \text{ egia}) = P\left(\frac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \geq b \,\middle|\, p = p_0\right) = P(Z \geq b).$$

Beraz, $b = z_\alpha$ da, eta eskualde kritikoa $\mathrm{S}_1 = [z_\alpha, +\infty)$ izango da.

Ondorioz, ahalmena era honetara kalkulatuko dugu:

$$1 - \beta = P(H_0 \text{ errefusatu} \mid H_0 \text{ faltsua}) = P\left(\frac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \geq z_\alpha \,\middle|\, p = p_1\right) =$$

$$= P\left(\hat{P} \geq p_0 + z_\alpha \sqrt{\frac{p_0 q_0}{n}} \,\middle|\, p = p_1 \right) =$$

$$= P\left(\frac{\hat{P} - p_1}{\sqrt{p_1 q_1 / n}} \geq z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{(p_0 - p_1)\sqrt{n}}{\sqrt{p_1 q_1}} \,\middle|\, p = p_1 \right) =$$

$$= P\left(Z \geq z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{(p_0 - p_1)\sqrt{n}}{\sqrt{p_1 q_1}} \right) = P\left(Z \leq -z_\alpha \cdot \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{(p_1 - p_0)\sqrt{n}}{\sqrt{p_1 q_1}} \right)$$

7.7 adibidea

7.4 adibideko datuetan oinarriturik, kalkula ezazu kontrastearen ahalmena, % 5eko adierazgarritasun-maila eta 0.6 proportzio alternatiboa erabilirik.

Emaitza:

Honako hau da hipotesi-kontrastea:

H_0: $p \leq 0{,}5$ ($p = 0{,}5$)
H_1: $p > 0{,}5$ ($p = 0{,}6$)

$Z_p = \dfrac{\hat{P} - 0.5}{\sqrt{0{,}5 \cdot 0{,}5/100}}$ estatistikoaren balioa b konstantea baino handiago edo berdina denean errefusatuko dugu H_0 hipotesia. Alegia, $S_1 = [b, +\infty)$.

$\alpha = 0{,}05$ adierazgarritasun-mailan oinarrituz kalkulatuko da b balioa:

$$0{,}05 = P(H_0 \text{ errefusatu} \mid H_0 \text{ egia}) = P\left(\frac{\hat{P} - 0{,}5}{\sqrt{0{,}5 \cdot 0{,}5 / 100}} \geq b \,\middle|\, p = 0{,}5 \right) = P(Z \geq b).$$

Beraz, $b = z_{0.05} = 1{,}645$.

Honako hau da kontrastearen ahalmena:

$$1 - \beta = P(H_0 \text{ errefusatu} \mid H_0 \text{ gezurra}) = P\left(\frac{\hat{P} - 0{,}5}{\sqrt{0{,}5 \cdot 0{,}5 / 100}} \geq 1{,}645 \,\middle|\, p = 0{,}6 \right)$$

$$= P\left(\hat{P} \geq 0{,}5 + 1{,}645 \cdot \sqrt{\frac{0{,}5 \cdot 0{,}5}{100}} \mid p = 0{,}6 \right) =$$

$$= P\left(\frac{\hat{P} - 0{,}6}{\sqrt{0{,}6 \cdot 0{,}4 / 100}} \geq 1{,}645 \cdot \sqrt{\frac{0{,}5 \cdot 0{,}5}{0{,}6 \cdot 0{,}4}} + \frac{(0{,}5 - 0{,}7)\sqrt{100}}{\sqrt{0{,}6 \cdot 0{,}4}} \mid p = 0{,}6 \right) =$$

$$= P(Z \geq -2{,}4) = P(Z \leq 2{,}4) = 1 - 0{,}0082 = 0{,}9918$$

Oharra: honako era honetara laburbil daiteke aldebakarreko bi kontrasteen ahalmenaren kalkulua:

$$1 - \beta = P\left(Z \leq -z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{|\, p_0 - p_1 \,| \sqrt{n}}{\sqrt{p_1 q_1}} \right)$$

Formula hori aztertuz, kontrastearen ahalmenean eragina duten faktoreak zein diren eta nola eragiten duten egiazta dezakegu:

1. Adierazgarritasun-maila: α txikitzerakoan z_α handitu egiten da, eta, beraz, ahalmena txikitu egiten da.
2. Proportzio alternatiboa: p_1 balioa H_0 hipotesi nuluaren p_0 baliotik urrun baldin badago, alegia $|p_0 - p_1|$ handitzen bada, ahalmena ere handitu egingo da.
3. Laginaren tamaina: n handitzerakoan, ahalmena ere handitu egingo da.

7.2.4 Lagin-tamainaren kalkulua

Ikusi dugu laginaren tamaina handitzerakoan ahalmena handitu egiten dela, baina ikerketa bat gauzatu nahi denean, laginaren tamaina egokia zenbatekoa den zehaztu behar dugu. Horretarako, honako faktore hauek hartu behar ditugu kontuan:

1. α adierazgarritasun-maila.
2. p_0 eta p_1 balioen arteko diferentzia.
3. Kontrastearen ahalmena.

Arazoa honako era honetara planteatuko da: kontrastearen adierazgarritasun-maila α eta proportzio alternatiboa p_1 izanik, zenbatekoa izango da erabili behar dugun laginaren tamaina, diferentzia adierazgarria $1 - \beta$ probabilitateaz aurkitzeko?

7.2.4.1 Aldebakarreko aukerak

Kontrastearen ahalmena, $1-\beta = P\left(Z \le -z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_1 q_1}}\right)$ dela frogatu dugu.

Alegia, $\beta = P\left(Z > -z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_1 q_1}}\right)$.

Beraz, $z_\beta = -z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_1 q_1}}$, eta n isolatuz honako hau lortuko dugu:

$$n = \frac{\left(z_\beta + z_\alpha \sqrt{\frac{p_0 q_0}{p_1 q_1}}\right)^2 p_1 q_1}{(p_0 - p_1)^2} = \frac{\left(z_\beta \sqrt{p_1 q_1} + z_\alpha \sqrt{p_0 q_0}\right)^2}{(p_0 - p_1)^2}$$

7.8 adibidea

Demagun 7.3 adibidean *vidarabina*k heriotza-tasa % 60ra txikitzen duen egiaztatu nahi dugula, % 1eko adierazgarritasun-maila eta % 95eko ahalmena erabiliz. Zenbat gaixori eman beharko diegu sendagai hori?

Emaitza:

Honako hau da hipotesi kontrastea:

H_0: $p = 0{,}70$ (*vidarabina* ez da eraginkorra)
H_1: $p = 0{,}60$ (*vidarabina* eraginkorra da)

$\alpha = 0{,}01$ eta $1 - \beta = 0{,}95$ direnez, datu hauek lortuko ditugu banaketa normalaren taulan: $z_\alpha = z_{0.01} = 2{,}3267$ eta $z_\beta = z_{0.05} = 1{,}645$.

Beraz:

$$n = \frac{\left(z_\beta \sqrt{p_1 q_1} + z_\alpha \sqrt{p_0 q_0}\right)^2}{(p_0 - p_1)^2} =$$
$$= \frac{\left(1{,}645 \cdot \sqrt{0{,}6 \cdot 0{,}4} + 2{,}3267 \cdot \sqrt{0{,}7 \cdot 0{,}3}\right)^2}{(0{,}7 - 0{,}6)^2} = 350{,}48$$

Hau da, *vidarabina* 351 gaixori eman beharko genieke.

7.3 BI PROPORTZIOREN KONPARAKETA

Biologia eta Medikuntza arloetan, maiz agertzen da bi proportzio konparatu beharra. Bi populazio aske daude eta bietan ezaugarri berbera aztertuko dugu. Populazioko ale bakoitzak ezaugarri hori izan lezake ala ez; hau da, populazio binomialak dira. Demagun p_1 aztertzen den ezaugarria daukaten lehen populazioko ale proportzioa dela. Era berean, p_2 ezaugarria daukaten bigarren populazioko ale proportzioa izango da. Arazoa p_1 eta p_2 konparatzean datza.

7.3.1 Bi proportzioen diferentziaren puntu-estimazioa

Izan bitez X_1 eta X_2 bi populazio binomial aske, eta beren parametroak p_1 eta p_2, hurrenez hurren. Honako hau da $(p_1 - p_2)$ proportzioen diferentziaren estimatzaile alboragabea:

$$\left(\widehat{P_1 - P_2}\right) = \hat{P}_1 - \hat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

Orduan, laginetan agertutako arrakasta-proportzioen arteko diferentzia izango da $(p_1 - p_2)$-ren puntu-estimazioa; alegia:

$$\left(\widehat{p_1 - p_2}\right) = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$$

n_1 eta n_2 populazio bakoitzetik ateratako laginen tamainak dira, eta x_1 eta x_2 lagin horietan agerturiko intereseko ezaugarridun ale kopurua (arrakasta kopurua), hurrenez hurren.

7.9 adibidea

Giltzurrunetako gaixotasunerako erabiltzen den sendagaiaren azterketan, hemeretzi ospitaletatik hirurogeita hamabi gaixo aukeratu zituzten. Horietatik hogeita hamalauk sendagaia hartu zuten, eta seik bakarrik izan zuten

giltzurrun-krisia. Baina plazeboa hartu zuten hogeita hemezortzi gaixoen artean, hogeik izan zuten giltzurrun-krisiren bat. (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Sendagaia hartu edo ez, hauxe izango da giltzurrun-krisiren bat pairatzen duten gaixo-proportzioen arteko kenduraren estimazioa:

$$\widehat{(p_1 - p_2)} = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2} = \frac{6}{34} - \frac{20}{38} = -0{,}3498$$

7.3.2 Lagin-proportzioen arteko diferentziaren banaketa

Izan bitez X_1 eta X_2 bi populazio binomial aske, eta beren parametroak p_1 eta p_2, hurrenez hurren. Horietatik n_1 eta n_2 tamainako laginak aukeratu ditugu, $n_1\hat{p}_1 > 5$, $n_1\hat{q}_1 > 5$, $n_2\hat{p}_2 > 5$ eta $n_2\hat{q}_2 > 5$ baldintzak beterik.

Baldintza horien menpe, populazio binomial baten kasuan ikusi dugunez, hauxe beteko da bi laginetarako:

$$n_1\hat{p}_1 > 5 \text{ eta } n_1\hat{q}_1 > 5 \;\Rightarrow\; \hat{P}_1 \approx N\left(p_1, \sqrt{\frac{p_1 q_1}{n_1}}\right)$$

$$n_2\hat{p}_2 > 5 \text{ eta } n_2\hat{q}_2 > 5 \;\Rightarrow\; \hat{P}_2 \approx N\left(p_2, \sqrt{\frac{p_2 q_2}{n_2}}\right)$$

Bi estimatzaileak gutxi gorabehera normalak direnez, $\hat{P}_1 - \hat{P}_{21}$ bien arteko diferentziaren banaketa ere normala izango da gutxi gorabehera.

Itxaropenaren propietateak erabiliz: $E(\hat{P}_1 - \hat{P}_2) = E(\hat{P}_1) - E(\hat{P}_2) = p_1 - p_2$. Hau da, estimatzailea alboragabea da.

Populazioak askeak direnez, $\hat{P}_1$ eta $\hat{P}_2$ ere askeak izango dira, eta beraz, honako hau beteko da bariantzarako:

$\mathrm{Var}(\hat{P}_1 - \hat{P}_2) = \mathrm{Var}[\hat{P}_1 + (-1)\ \hat{P}_2] = \mathrm{Var}(\hat{P}_1) + (-1)^2 \cdot \mathrm{Var}(\hat{P}_2) = \mathrm{Var}(\hat{P}_1) + \mathrm{Var}(\hat{P}_2) =$

$$= \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Orduan, honako hau da $(\hat{P}_1 - \hat{P}_2)$ lagin-proportzioen arteko diferentziaren banaketa, $n_1\hat{p}_1 > 5$, $n_1\hat{p}_1 > 5$, $n_1\hat{q}_1 > 5$, $n_2\hat{p}_2 > 5$ eta $n_2\hat{q}_2 > 5$ baldintzen menpe:

$$\hat{P}_1 - \hat{P}_2 \approx N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right) \text{ edo } \frac{\left(\hat{P}_1 - \hat{P}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \approx Z : N(0,1)$$

7.3.3 Bi populazio binomial askeren proportzioen arteko diferentziarako konfiantza-tartea

Izan bitez X_1 eta X_2 bi populazio binomial aske, eta beren parametroak p_1 eta p_2, hurrenez hurren. Horietatik n_1 eta n_2 tamainako laginak aukeratuko ditugu, honako baldintza hauek beterik: $n_1\hat{p}_1 > 5$, $n_1\hat{q}_1 > 5$, $n_2\hat{p}_2 > 5$ eta $n_2\hat{q}_2 > 5$. Kasu horretan, $(\hat{P}_1 - \hat{P}_2)$-ren banaketa ezagutzen dugunez, $(p_1 - p_2)$-rako konfiantza-tartea kalkula dezakegu banaketa horretan oinarrituz.

Hain zuzen, $\dfrac{\left(\hat{P}_1 - \hat{P}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \approx Z : N(0,1)$ dela ikusi dugu, eta badakigu Z:$N(0,1)$ banaketak honako hau betetzen duela: $1 - \alpha = P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right)$

Ekuazio horretan, Z-ren ordez $(\hat{P}_1 - \hat{P}_2)$-ren banaketa tipifikatua jartzen badugu, honako hau lortuko dugu:

$$= P\left(-z_{\alpha/2} < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} < z_{\alpha/2}\right) =$$

$$= P\left(-z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < (\hat{p}_1 - \hat{p}_2) - (p_1 - p_2) < z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right) =$$

$$= P\left(-(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < -(p_1 - p_2) < -(\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} \right) =$$

$$= P\left(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} \right)$$

Hau da, (L_1, L_2) tartea lortu dugu, non $P(L_1 < p_1 - p_2 < L_2) = 1 - \alpha$ betetzen baita:

$$\mathrm{L}_1 = \left(\hat{P}_1 - \hat{P}_2\right) - z_{\alpha/2} \cdot \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

$$\mathrm{L}_2 = \left(\hat{P}_1 - \hat{P}_2\right) + z_{\alpha/2} \cdot \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

$\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$ kantitatea ezezaguna denez, $\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}$ estimazioaren bidez hurbiltzen da, eta, beraz, laginak aukeratu ondoren, hauxe da $(p_1 - p_2)$-rako **%100 × (1 – α) mailako konfiantza-tartea**:

$$I_{p_1-p_2}^{1-\alpha} = \left(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right)$$

7.10 adibidea

Aurreko adibidean oinarriturik, kalkula ezazu sendagaia hartuta edo hartu gabe, giltzurrun-krisia pairatzeko probabilitateen arteko diferentziarako % 95eko konfiantza-tartea.

Emaitza:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{6}{34} = 0{,}1765 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{20}{38} = 0{,}5263$$

$1 - \alpha = 0{,}95 \Rightarrow z_{\alpha/2} = z_{0.025} = 1{,}96.$

Beraz,

$$I_{p_1-p_2}^{0.95} = \left(0{,}1765 - 0{,}5263 - 1{,}96 \cdot \sqrt{\frac{0{,}1765 \cdot 0{,}8235}{34} + \frac{0{,}5263 \cdot 0{,}4737}{38}}, \right.$$

$$\left. 0{,}1765 - 0{,}5263 + 1{,}96 \cdot \sqrt{\frac{0{,}1765 \cdot 0{,}8235}{34} + \frac{0{,}5263 \cdot 0{,}4737}{38}} \right) =$$

$$= \left(-0{,}3498 - 0{,}2040, -0{,}3498 + 0{,}2040 \right) = (-0{,}5538\ , -0{,}1458)$$

Tartearen bi mugak negatiboak direnez, % 95eko konfiantza-mailaz, $(p_1 - p_2) < 0$ dela ondoriozta dezakegu; alegia, giltzurrun-krisia izateko probabilitatea txikiagoa dela sendagaia hartuz gero, hartu gabe baino. Beste era batera esanda, sendagaia eraginkorra dela ondorioztatu dugu, zeren eta giltzurruneko krisia izateko probabilitatea txikitu egin baitu.

7.3.4 Bi populazio binomial askeren parametroen arteko berdintasunerako kontrastea

Izan bitez X_1 eta X_2 bi populazio binomial aske, eta beren parametroak p_1 eta p_2, hurrenez hurren. Horietatik n_1 eta n_2 tamainetako laginak aukeratu ditugu, honako baldintza hauek beterik: $n_1\hat{p}_1 > 5,\ n_1\hat{q}_1 > 5,\ n_2\hat{p}_2 > 5$ eta $n_2\hat{q}_2 > 5$. Ikusi dugunez, kasu horretan, $\dfrac{\left(\hat{P}_1 - \hat{P}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}} \approx Z : N(0,1)$ da.

p_1 eta p_2 parametroak hipotesi-kontrasteen bidez konparatzeko, hipotesi nulutzat berdintasuna hartuko dugu; hau da, $p_1 - p_2 = 0$ dela suposatuko dugu. Era horretan, hipotesiak kontrastatzeko erabiliko dugun estatistikoa $Z_p = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}}$ izango da eta, proportzioen arteko diferentziaren lagin-banaketa kontuan hartuz, H_0 hipotesiaren pean, estatistiko horren banaketa Z: $N(0,1)$ izango da.

Lagina aukeratu ondoren, $\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}$ kantitatea ezezaguna denez, $\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}$ estimazioaren bidez hurbilduko gara. Horrela, eta beti H_0 hipotesiaren pean, hauxe izango da estatistikoaren balioa: $z_p = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}}$

Aurrekoetan bezala, hipotesi-kontrastearen mota bakoitzerako p-balioa hipotesi alternatiboaren arabera kalkulatuko dugu; baina estatistiko berbera erabiliko da aldebakarreko zein aldebiko kontrasteetan.

7.3.4.1 Aldebakarreko kontrasteak

7.3.4.1.1Ezkerretikoa

$$H_0: p_1 - p_2 \geq 0 \iff H_0: p_1 \geq p_2$$
$$H_1: p_1 - p_2 < 0 \iff H_1: p_1 < p_2$$

Kasu honetan, hauxe da p-balioa: $p = P(Z < z_p)$.

7.11 adibidea

Barraskiloen arerio nagusia birigarro kantaria da. Txori horiek kolonietatik barraskiloak aukeratu eta hurbileko haitzetara eramaten dituzte. Bertan, barraskiloak ireki ondoren, txoriek zati bigunak jan eta maskorrak uzten dituzte. Hautespen naturaleko azterketan, haitzetan zegoen marrarik gabeko maskor proportzioa eta hurbilen zegoen koloniako marrarik gabeko barraskilo-proportzioa konparatu ziren. Kolonia hori hondo uniformeko lupetza zen. Lupetzaren hondoarekin bat egiteko gaitasuna dela eta, marrarik gabeko barraskiloak marradunak baino hobeto babesturik egongo zirela harraparietatik pentsatu zen. Horren ondorioa, haitzetako marrarik gabeko maskor-proportzioa lupetzetako marrarik gabeko barraskilo-proportzioa baino txikiagoa izatea litzateke.

Haitzetan apurtuta zeuden 863 maskorretatik 377, eta lupetzetan aurkituriko 560 barraskiloetatik 296 marrarik gabekoak ziren. % 5eko adierazgarritasun-mailaz, datu horiek egiaztatzen al dute egindako hipotesia? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

Emaitza:

Demagun p_1 dela haitzetako marrarik gabeko maskor-proportzioa, eta p_2 lupetzetako marrarik gabeko barraskilo-proportzioa.

Honako hauek dira kontrastatu nahi diren hipotesiak:

$H_0: p_1 \geq p_2$
$H_1: p_1 < p_2$ (egindako hipotesia)

Hauxe daukagu lehenengo laginean: $\hat{p}_1 = \frac{x_1}{n_1} = \frac{377}{863} = 0{,}437$.

Bigarrenean, aldiz: $\hat{p}_2 = \frac{x_2}{n_2} = \frac{296}{560} = 0{,}529$.

Beraz, H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{0{,}437 - 0{,}529}{\sqrt{\frac{0{,}437 \cdot 0{,}563}{863} + \frac{0{,}529 \cdot 0{,}471}{560}}} = -3{,}4 \cdot$$

Eta p-balioa horrela kalkulatuko dugu: $p = P(Z < z_p) = P(Z < -3{,}4) = P(Z > 3{,}4) = 0{,}000337 < 0{,}05$.

Ondorioz, hipotesi nulua errefusatuko dugu, eta datu horiek, % 5eko adierazgarritasun-mailaz, haitzetako marrarik gabeko maskor-proportzioa lupetzetako marrarik gabeko barraskilo-proportzioa baino txikiagoa den hipotesia egiaztatuko dute.

7.3.4.1.2 Eskuinetikoa

H_0: $p_1 - p_2 \leq 0 \Leftrightarrow H_0$: $p_1 \leq p_2$

H_1: $p_1 - p_2 > 0 \Leftrightarrow H_1$: $p_1 > p_2$

Kasu honeta, hauxe da p-balioa: $p = P(Z > z_p)$.

7.12 adibidea

C bitamina gomendatzen dutenek baieztatzen dute minbiziari aurre egiteko probabilitatea hobetzen duela. Teoria hori frogatzeko asmoz, minbizia pairatzen duten 150 paziente zoriz aukeratu, eta tamaina bereko bi taldetan banatu dira. Lehenengo taldeak 10 g C bitamina eta bigarren taldeak plazeboa hartu zuten egunero. C bitamina hartu zutenen artetik berrogeita zazpik lau astean onera egin zuten; plazeboa hartu zutenen artetik, berriz, berrogeita hiruk soilik. Datu horiek egiaztatzen al dute C bitaminak minbiziari aurre egiteko laguntza ematen duen teoria? Erabil ezazu $\alpha = 0{,}05$ (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

Emaitza:

Demagun C bitamina hartu ondoren onera egiten duen paziente-proportzioa p_1 dela, eta C bitaminarik gabe suspertzen direnena p_2.

Hauek dira kontrastatu nahi ditugun hipotesiak:

H_0: $p_1 \leq p_2$
H_1: $p_1 > p_2$ (C bitaminak minbiziari aurre egiteko laguntza ematen du)

Lehenengo taldean, C bitamina hartu dutenen artean, $\hat{p}_1 = \frac{x_1}{n_1} = \frac{47}{75} = 0{,}627$ da, eta bigarrenean, C bitaminarik gabekoen artean, $\hat{p}_2 = \frac{x_2}{n_2} = \frac{43}{75} = 0{,}573$.

Beraz, H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{0{,}627 - 0{,}573}{\sqrt{\frac{0{,}627 \cdot 0{,}373}{75} + \frac{0{,}573 \cdot 0{,}427}{75}}} = 0{,}676 .$$

Eta hauxe izango da p-balioa: $p = P(Z > z_p) = P(Z > 0{,}676) = 0{,}25 > 0{,}05$.

Ondorioz, ezin dugu hipotesi nulua errefusatu, eta, beraz, % 5eko adierazgarritasun-mailaz, ezin dugu esan C bitaminak minbiziari aurre egiteko laguntza ematen duenik.

7.3.4.2 Aldebiko kontrastea

H_0: $p_1 - p_2 = 0 \Leftrightarrow H_0$: $p_1 = p_2$
H_1: $p_1 - p_2 \neq 0 \Leftrightarrow H_1$: $p_1 \neq p_2$

Kasu honetan, hauxe da p-balioa: $p = 2 \cdot P(Z > |z_p|)$.

7.13 adibidea

Hautespen naturaleko adibideen artean, izpildun sitsarena daukagu. 1845. urtera arte, ezagutzen ziren sits-espezie guztiak kolore argikoak ziren, baina urte hartan sits beltz bat harrapatu zen Manchesterren. Inguruko industrializazioaren ondorioz, zuhaitzen enborrak, haitzak eta lurra belztu egin ziren kedarrarekin. Espezie mutante beltz hori oso azkar zabaldu zen. Hasieran, kolore beltzak sitsa bere harraparietatik, batez ere txorietatik,

babesten zuelako zela pentsatu zen. Baina, inguruko entomologoek baieztatu zuten ez zutela inoiz ikusi txori bat inolako koloreko izpildun sitsik jaten, eta ideia hori baztertu egin zuten. Teoria hori baieztatzeko asmoz, kolore bakoitzeko (beltza eta argia) ehun sits hartu zituzten eta, markatu ondoren, aske utzi zituzten berriro. Gauez, argidun tranparen bidez, sits beltzen % 40 eta sits argien % 19 berreskuratu ziren. Ba al dago sits beltzen eta argien biziraupen-probabilitateen arteko diferentzia adierazgarririk? Erabil ezazu α = 0,05 (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

Emaitza:

Demagun p_1 eta p_2 bizirik irauten duten sits beltzen eta sits argien proportzioak direla, hurrenez hurren.

Hauek dira kontrastatu nahi diren hipotesiak:

H_0: $p_1 - p_2 = 0$ (ez dago diferentzia adierazgarririk)

H_1: $p_1 - p_2 \neq 0$ (diferentzia adierazgarria dago)

Sits beltzentzat $\hat{p}_1 = 0{,}40$ da eta sits argientzat $\hat{p}_2 = 0{,}19$.

Beraz, H_0 egia dela suposatuz, honako hau da behaturiko estatistikoaren balioa:

$$z_p = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{0{,}40 - 0{,}19}{\sqrt{\frac{0{,}4 \cdot 0{,}6}{100} + \frac{0{,}19 \cdot 0{,}81}{100}}} = 3{,}346 \cdot$$

Eta horrela kalkulatuko dugu p-balioa: $p = 2 \cdot P(Z > |z_p|) = 2 \cdot P(Z > 3{,}346) = 2 \cdot 0{,}000483 = 0{,}000966 < 0{,}05$.

Ondorioz, H_0 errefusatu egingo da, eta datu horiek egiaztatuko dute, % 5eko adierazgarritasun-mailaz, sits beltzen eta argien biziraupen-probabilitateen artean diferentzia adierazgarria dagoela.

7.4 ARIKETAK

7.1. Nazio-mailako ikerketa bat egin da hamasei urtetik beherako gazte erretzaileen *p* proportzioa estimatzeko. Elkarrizketatutako 1.000 pertsonetatik 200 erretzaileak ziren (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Aurki ezazu *p*-ren puntu-estimazioa.

b) Kalkula ezazu hamasei urtetik beherako gazte erretzaileen proportziorako % 99ko konfiantza-tartea. Proportzio hori 0,23 dela dioen artikulua irakurtzeak ustekabean harrapatuko zintuzkeen? Azal ezazu erantzuna.

7.2. Globulu zurien zenbaketa egiteko, odol tanta beirazko euskarri gainean uniformeki zabaldu, *Wright* tindagaiaz tindatu, eta mikroskopioan aztertu da. Zenbaturiko 200 globulu zurietatik 125 neutrofiloak ziren, hau da, hezur-muinean sortu eta odoleko zoldura-eragileak deuseztatzen dituzten globulu zurien motakoak (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Lor ezazu pertsonaren globulu zurien artean aurkituriko *p* neutrofilo-proportzioaren puntu estimazioa.

b) Lor ezazu *p*-ren % 90eko konfiantza-tartea.

c) Osasun normaleko gizabanakoen globulu zurien arteko neutrofilo-portzentajea ehuneko 60tik 70erakoa da. b) atalean kalkulaturiko tartean oinarrituz, ba al dago neutrofilo-desorekaren nabaritasunik pertsona horiengan?

7.3. 1987. urtean, haztegiko bi mila milioi landarerekin hiru milioi akre baso baino gehiago berritu ziren. Hurrengo urtaroan izandako izugarrizko lehortea zela eta, horietariko landare asko hil egin ziren. 1.000 landarek osaturiko lagina hartu zen, eta horietatik 300 hilda zeuden. Kalkula ezazu hildako haztegiko landare-proportziorako % 90eko konfiantza-tartea. Erabil ezazu informazio hori populazioan hildako haztegiko landare kopuru osoa estimatzeko (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

7.4. Bizidun batzuek harraparien kontra duten defentsa-sistema nazkagarriak egitean datza; hau da, zapore edo usain desatsegina hartu, edo likido zakarra jaurti. Ezaugarri hori duenetariko bat makila-intsektua da. Makila-intsektuak likidoa

jaurti zien berrogeita hamar eskinoso urdinen artetik berrogeita bi gutxienez bi astean egon ziren intsektuengandik urrun (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Kalkula ezazu makila-intsektuarekin kontaktuan egon ziren eskinoso urdinen artetik hurrengo bi asteetan kontaktuan jarraitu zutenen proportziorako % 94ko konfiantza-tartea.

b) Zenbatekoa izan beharko du laginaren tamainak, bi astean makila-intsektutik alde egin zuten eskinoso urdinen proportzioa % 3ko errore maximoaz eta % 94ko konfiantza-mailaz estimatzeko?

7.5. 1987. urteko lehorte larriak eragina izan zuen bai haztegiko landareen heriotza-tasan eta bai jadanik helduak ziren zuhaitzen hazkunde-tasan. Eragina izandako inguruetako zuhaitz gehienen 1987. urteko hazkunde-uztaia beste zuhaitz batzuen erdikoa baino txikiagoa dela susmatzen da. 250 zuhaitzek osaturiko laginean, ezaugarri hori zeukaten 150 aurkitu ziren. Oinarritu al daiteke emandako ideia datu horietan? Azal ezazu erantzuna (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

7.6. Estresak heriotza ekar dezaketen bihotz-arazoak sor ditzakeela dio orain dela gutxi egindako ikerketak. Besterik gabe heriotza sortzeko nahiko larria ez den eraso fisikoa jasan ondoren hil ziren hamazazpi pertsona aztertu ondoren, esandakoa egiaztatu zen. Horietatik hamaika pertsona endekapen miofibrilarra deritzon bihotzeko zelula-endekapenaren ondorioz hil ziren (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Aurki ezazu erasoa jasan ondoren endekapen miofibrilarraren ondorioz hildakoen p proportzioaren puntu-estimazioa.

b) Aurki ezazu erasoa jasan ondoren endekapen miofibrilarraren ondorioz hildakoen proportziorako % 95eko konfiantza-tartea. Tarte horretan oinarrituz, p-ren balioaren ideia ona izateko, nahiko handia al da n laginaren tamaina?

c) Zenbateko lagin-tamaina beharko genuke, 0.02-ko distantziara eta % 95eko konfiantza-mailaz, erasoa jasan ondoren endekapen miofibrilarraren ondorioz hildakoen proportzioa estimatzeko?

7.7. Espezie baten ar-portzentajea hogei deneko hipotesi nulua formulatu dugu hogeita hamar deneko hipotesi alternatiboaren aurka. Ehun tamainako lagina aukeratu dugu; eta laginean hogeita bost ar baino gehiago baldin badaude, hipotesia errefusatzea erabaki dugu (Iturria: A. Urkaregi, *Bioestatistika*. UEU, 1991).

a) Aurki itzazu I eta II motako erroreen arriskuak.

b) Hogeita hamar eta berrogei tamainako bi lagin aukeratu dira, eta sei eta hamar ar aurkitu dira, hurrenez hurren. % 1eko adierazgarritasun-mailaz, onar al dezakegu portzentajeen arteko kendura adierazgarria ez denik?

7.8. Kakalardoentzako tranpa japoniar bat probatzen ari dira. Fabrikatzaileak baieztatu du tranpatik 9 metrora hurbiltzen diren kakalardoen % 90 erakarri eta hil egiten dituela. Baieztapen hori egiaztatzeko esperimentua egin dute (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Eraiki itzazu hipotesi nulu eta alternatibo egokiak.

b) Tranpatik hurbil 900 kakalardo uzten dira esperimentua egiteko. H_0 egiazkoa baldin bada, gehienez zenbat kakalardo erakarriko ditu tranpak? 900en artetik 825 erakarri ziren, eta hil egin ziren. Baieztapena mantentzeko nahiko nabaria al da emaitza hori? Azal ezazu erantzuna, kontrastearen *p*-balioan oinarrituz.

c) Zein errore mota jaulki dezakegu? Azal itzazu errore hori jaulkitzearen ondorioak.

7.9. Ikerketa bat egin da VIH positibo (HIESa garatzeko arriskuan daudenak) direnen artean dagoen zain barneko drogazale-portzentajea VIH negatibo direnen arteko zain barneko drogazale-portzentajearekin konparatzeko (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Zenbatekoa izan beharko du laginaren tamainak, portzentajeen arteko kendura estimatzeko % 2ko distantzian eta % 95eko konfiantza-mailan? Horretarako, erabil ezazu $p_1 = p_2 = 1/2$ eta $n_1 = n_2$.

b) Uste al duzu p_1 eta p_2-rako aurrestimazioak ezagutzeak abantaila praktikorik eskaintzen digunik?

7.10. Bularreko mina duten umeen artetik % 85 baino gehiagok, hala eta guztiz ere, ekokardiograma normala izango dutela pentsatzen da. Horrelako 139 umek osaturiko laginean, 123 ekokardiograma normalak izan ziren (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Eraiki ezazu ikerketa-hipotesia.

b) Lor ezazu bularreko mina duten umeen artetik ekokardiograma normala dutenen proportzioaren puntu-estimazioa. Estimazio horretan oinarrituz, kontrastea egin ondoren uste al duzu ikerketako hipotesia eutsiko denik?

c) Kontrasta ezazu H_0: $p \leq 0{,}85$ hipotesia H_1: $p > 0{,}85$ alternatiboaren aurka. $\alpha = 0{,}10$ adierazgarritasun-mailaz, errefusa al daiteke H_0?

d) Azal ezazu b) eta c) atalen artean egon daitekeen kontraesana.

7.11. *Dioxiciclina* izeneko «bidaiariaren sabelekoa»ri aurre egiteko, antibiotikoaren probak egiten ari dira. Kenyara joan ziren 98 boluntarioren artetik, erdiak sendagai hori hartu zuen, eta gaineratikoek plazeboa. *Dioxiciclina* hartu zutenen artetik hamazazpik ez zuten arazorik izan; beste taldekoen artetik, ordea, hamaikak baino ez (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Aurki ezazu *dioxiciclina* erabiltzen dutenen eta ez dutenen babes-tasen arteko kenduraren puntu-estimazioa.

b) Kalkula ezazu *dioxiciclina* erabiltzen dutenen eta ez dutenen babes-tasen arteko kendurarako % 90eko konfiantza-tartea. Ba al dago arrazoirik *Dioxiciclina*k «bidaiariaren sabelekoa»ren aurkako babesa ezartzen laguntzen duen teoriari eusteko? Azal ezazu erantzuna.

7.12. *Anturana* sendagaia hezueria (gota) deritzon gaixotasunaren aurka erabili izan da, 1959. urtetik orain arte. Gaur egun, sendagai horren eraginkortasuna probatzen ari dira bihotzekoren bat jasan duten pazienteengan ere, bigarren bihotzekoaren ondorioz ustekabeko heriotzari aurre egiteko. Ikerketan, 733 pazientek *anturana* hartu zuten, eta 742k plazeboa. Zortzi hilabete pasa ondoren, bigarren bihotzekoaren ondorioz berrogeita bi heriotza zenbatu ziren: hogeita bederatzi plazeboaren taldean eta hamahiru *anturana*renean (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Erabil itzazu datu horiek, *anturana* hartu dutenen eta ez dutenen ustekabeko heriotza-portzentajeen arteko kendura estimatzeko.

b) Kalkula ezazu *anturana* hartu dutenen eta ez dutenen ustekabeko heriotza-portzentajeen arteko kendurarako % 95eko konfiantza-tartea.

c) Datu berdinekin % 90eko konfiantza-tartea kalkulatzen bada, zein izango da luzeagoa? Zergatik? Egiazta ezazu erantzuna.

d) Ba al dago arrazoirik bigarren bihotzekoaren ondoriozko heriotza-tasa *anturana* hartzen dutenen artean ez dutenen artean baino txikiagoa dela dioen teoriari eusteko? Azal ezazu erantzuna.

e) Erabil itzazu aurreko datuak ikerketa pilotu gisa, *anturana* hartu dutenen eta ez dutenen ustekabeko heriotza-portzentajeen arteko kendura, % 2ko distantzian eta % 95eko konfiantza-mailaz, estimatzeko beharrezkoa den laginaren tamaina kalkulatzeko. Eta % 90eko konfiantza-mailaz?

7.13. Ikerketa bat egin da kakalardo beltzen proportzioa toki batzuetatik beste batzuetara aldatu egiten delako argudioa mantentzeko seinaleak lortu nahian. *Providence*, Rhode Island, inguruan aurkituriko 500 kakalardoetatik 95 beltzak ziren. *Aqueduct*, New York, inguruan aurkituriko 112tik, berriz, 17 ziren beltzak (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

a) Eraiki ezazu aldebiko hipotesi-kontraste egokia.

b) Lor ezazu bi herrialdeetako kakalardo beltzen proportzioen arteko kenduraren puntu-estimazioa. Estimazio horretan oinarrituz, uste al duzu bi proportzioen arteko ezberdintasun adierazgarririk dagoenik?

c) Kontrasta itzazu a) atalean eraikitako hipotesiak. Zenbatekoa da kontrastearen *p*-balioa? Gogora ezazu aldebiko kontrastea dela.

7.14. 1970. urtean egindako ikerketan, odol-probak egin ziren odoleko zenbait infekzio pairatzen zituzten 759 pazienteengan. Horietatik 46 kasutan, gutxienez bi organismo ezberdin isolatu ziren odol-lagin berdinean. 1975. urtean ikerketa berdina egin zen 838 pazienteengan, eta horietatik 109 odol-laginek bi organismo ezberdin edo gehiago eduki zituzten. Bi lagin horiek erabiliz, ziurta al dezakegu horrelako kasu-proportzioa bost urteotan gutxienez ehuneko sei puntu gehitu dela? Azal ezazu erantzuna hipotesi-kontraste egokia eraikiz eta *p*-balioa kalkulatuz. Egiazta ezazu erantzuna konfiantza-tartea erabiliz (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

7.15. Belarretan futbolean jolasten duten jokalarien belauneko lesioak aztertzeko egindako ikerketan, bi oinetako mota konparatu zituzten. Txirikordaturiko futbol-zapatak erabili zituzten 266 jokalariren artean, hamalau belauneko lesio izan ziren. Zazpi puntuko ohiko futbol-zapatak erabili zituzten 2055 futbolariren artean, 162 lesio agertu ziren. Egiaztatzen al du emaitza horrek, $\alpha =$ 0,10 adierazgarritasun-mailaz, ohiko futbol-zapatak erabiliz belauneko lesioa pairatzeko probabilitatea handiagoa dela txirikordaturiko futbol-zapatak erabiliz baino? Ondorio bera ateratzen al da $\alpha = 0{,}05$ adierazgarritasun-mailarako? Eta $\alpha = 0{,}01$erako? (Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994).

7.16. Farmazia-laborategiak sendagai berri bat iragartzen duenean, barnean izaten du txostena. Txosten horretan, sendagaiaren bigarren mailako ondorioak hurbileneko lehiakideenarekin konparatzen dira. Aztertuko dugun sendagaia garraztasunari aurre egiteko da. Bigarren mailako ondorioak izan zituzten pertsonen honako portzentaje hauek agertzen ziren txostenean:

Bigarren mailako ondorioak	Bertokoa (*n* = 465), %	Lehiakidea (*n* = 195), %
Buruko mina	2,4	2,6
Sabelekoa	1,9	5,0

(Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud.* Interamericana-McGraw-Hill, 1994)

a) Egin ezazu buruko mindun pazienteen portzentajeen berdintasunerako testa. (Iradokizuna: aldebiko kontrastea da).

b) Egin ezazu sabelekoa pairatzen duten pazienteen portzentajeen arteko kendurarako testa.

8 KHI KARRATU BANAKETAREN APLIKAZIOAK

8.1 SARRERA

Orain arte landutako gaietan, inferentzia estatistikorako bi metodo nagusiak, estimazioa eta hipotesi-kontrasteak, azaldu ditugu; kasu gehienetan, aldagai jarraituen datuetan oinarrituta (4, 5 eta 6 gaiak). Horietan azaldu dugu nola egin populazio bakar bateko parametroari buruzko inferentziak, eta baita, bi populazio aske eta menpeko genituenean, bi parametro konparatzeko metodoak ere. Aurreko gaian, populazio binomial bakar baten parametroari buruzko zein bi populazio binomial askeren bi parametroak konparatzeko inferentzia-metodoak azaldu ditugu.

Gai honetan, hiru hipotesi-kontraste berezi azalduko ditugu. Bereziak dira hipotesiak, ez direlako parametro bati edo biri buruzkoak, baizik eta populazioaren banaketari edo bi aldagairen erlazioari buruzkoak. Alde batetik, bi aldagai kualitatiboren arteko independentzia eta laginen homogeneotasun-proba azalduko ditugu kontingentzia-taulak erabiliz; eta bestetik, doikuntza-egokitasuna azalduko dugu, non populazio baten banaketari buruz egindako hipotesia lagineko datuen bidez kontrastatuko baita.

Lehenengo kasuan, independentzia-probaren bidez, mota honetako galderak erantzuten saiatuko gara: ba al dago menpeko erlaziorik…

- Hipertentsioaren eta erretzeko ohituraren artean?
- Loditasunaren eta depresioaren artean?
- Zain barneko drogazaletasunaren eta HIESaren artean?
- Loreen kolorearen eta usainaren artean?

Bigarren kasuan, laginen homogeneotasun-proba egingo dugu, hau da, ezaugarri bakar bat k lagin ezberdinetan aztertuko dugu; eta hauxe da gure buruari egingo diogun galdera: homogeneoak al dira lagin horiek, hau da, onar al dezakegu populazio beretik aterata daudenik? Adibidez, Hego Euskal Herriko meteorologia-estazioetan neurturiko euria egin den egun-proportzioa dugu lau urtaroetan (udaberrian, udan, udazkenean eta neguan) eta jakin nahi dugu proportzio hori urtaroz urtaro aldatzen den ala ez. Ez bada aldatzen, laginak homogeneoak izango dira; kontrako kasuan, ez.

Hirugarren kasuan, doikuntza-egokitasuneko probak aztertuko ditugu. Horietan, populazio bakar bat daukagu, eta haren banaketari buruzko hipotesia egingo dugu. Hipotesi horren arabera maiztasun teorikoak kalkulatuko ditugu, eta jakin nahi dugu gure laginean behatu ditugun maiztasunekin komunztatzen duten ala ez.

Hiru kasu horietan, kontrasteak egiteko erabiltzen ditugun estatistikoak Pearson-en khi karratu banaketari (χ^2) jarraitzen diote. Orain arte, populazio normalaren bariantzarako inferentzia egiterakoan χ^2 banaketa aipatu da; beste aplikazio asko dituen banaketa bat da, tartean, gai honetan azalduko ditugunak.

8.2 INDEPENDENTZIA-PROBA

Independentzia-proba egiteko, bi aldagai kualitatiboak edo faktoreak kontingentzia-taula batean kokatu behar ditugu. Demagun A faktoreak r modalitate dituela, eta B faktoreak s modalitate. Ditugun n elementuak bi faktore horien arabera sailkatuko ditugu, eta r errenkada eta s zutabe dituen taula eratuko dugu; eta $r \times s$ gelaxka bakoitzean bi modalitateak aldi berean aurkezten dituen maiztasun absolutua zenbatuko dugu. Horrela eraikitzen dugu **$r \times s$ kontingentzia-taula**:

		***B* faktorea**					
		1	2	3	…	s	
***A* faktorea**	1	o_{11}	o_{12}	o_{13}	…	o_{1s}	$f_{1\cdot}$
	2	o_{21}	o_{22}	o_{23}	…	o_{2s}	$f_{2\cdot}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	o_{r1}	o_{r2}	o_{r3}	…	o_{rs}	$f_{r\cdot}$
		$z_{\cdot 1}$	$z_{\cdot 2}$	$z_{\cdot 3}$	…	$z_{\cdot s}$	n

8.1 taula. $r \times s$ kontingentzia-taularen ohiko adierazpena

non,

- o_{ij} : ij. gelaxkan behaturiko maiztasuna, hau da, j. zutabeari eta i. errenkadari dagokion maiztasuna.
- $z_{\cdot 1}$, $z_{\cdot 2}$, $z_{\cdot 3}$,…, $z_{\cdot s}$: **zutabeen bazter-totalak** deritze, eta 1go, 2., 3.,… s. zutabeen maiztasun totala adierazten dute, hurrenez hurren. Hau da:

$$z_{\cdot j} = o_{1j} + o_{2j} + o_{3j} + \ldots + o_{rj} \qquad \text{non } j = 1,2\ldots s$$

- $f_{1\cdot}, f_{2\cdot}, f_{3\cdot},\ldots, f_{r\cdot}$: **errenkaden bazter-totalak** deritze, eta 1go, 2., 3.,…, r. errenkaden maiztasun totala adierazten dute, hurrenez hurren. Hau da:

$$f_{i\cdot} = o_{i1} + o_{i2} + o_{i3} + \ldots + o_{is} \qquad \text{non } i = 1,2\ldots r$$

- n: behaturiko laginaren tamaina.

8.1 adibidea

Demagun psikologoak eskola-umeen pisua eta beren ikasketetako arrakasta goiztiarrari buruzko menpekotasuna aztertu nahi dituela. 500 eskola-umek osaturiko laginean, 8.2 taulan adierazten diren datuak lortu ditugu:

	Pisua			
Arrakasta goiztiarra	Gutxi	Ohikoa	Asko	
Bai	73	263	80	416
Ez	9	37	38	84
	82	300	118	500

8.2 taula. 500 eskola-umeen datuen adierazpena, pisuaren eta arrakasta goiztiarraren arabera.

8.2.1 Planteamendu orokorra

Demagun A eta B faktoreei buruzko datuak $r \times s$ kontingentzia-taulan bilduta daudela, eta beren artean erlazioa dagoen ala ez kontrastatu nahi dugula. **H_0 hipotesi nulua** izango da bi faktoreen arteko «menpekotasunik ez egotea». **H_1 hipotesi alternatiboa**, ordea, bien arteko menpekotasuna egotea izango da.

Beraz, **independentzia-proba** gauzatzen dugunean, honako era honetara adieraziko ditugu hipotesiak:

H_0: *A* eta *B* faktoreak askeak dira.

H_1: *A* eta *B* faktoreak menpekoak dira.

H_0 egiazkoa bada, hots, bi faktoreak askeak badira, $P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$ da, eta, beraz, honako era honetara kalkula daitezke gelaxka bakoitzerako itxarotako maiztasunak:

$$e_{ij} = n \cdot P(A_i \cap B_j) = n \cdot P(A_i) \cdot P(B_j)$$

$P(A_i)$ eta $P(B_j)$ ezagutzen ez ditugunez, laginaren datuetatik estimatu beharko ditugu maiztasun erlatiboen bidez, hau da:

$$P(A_i) \approx \frac{f_{i\cdot}}{n} \qquad \text{eta} \qquad P(B_j) \approx \frac{z_{\cdot j}}{n}$$

Beraz,

$$e_{ij} = n \cdot \frac{f_{i\cdot}}{n} \cdot \frac{z_{\cdot j}}{n} = \frac{f_{i.} \cdot z_{.j}}{n}$$

8.3 taulan, $r \times s$ kontingentzia-taularen adierazpen teorikoa erakusten da, eta, gelaxka bakoitzean, hari dagokion maiztasun teorikoa agertzen da.

		A faktorea					
		1	2	3	…	s	
B faktorea	1	e_{11}	e_{12}	e_{13}	…	e_{1s}	$f_{1\cdot}$
	2	e_{21}	e_{22}	e_{23}	…	e_{2s}	$f_{2\cdot}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	e_{r1}	e_{r2}	e_{r3}	…	e_{rs}	$f_{r\cdot}$
		$z_{\cdot 1}$	$z_{\cdot 2}$	$z_{\cdot 3}$	…	$z_{\cdot s}$	

8.3 taula. *rxs* kontingentzia-taularen adierazpen teorikoa, gelaxka bakoitzari dagokion itxarotako maiztasuna adierazirik.

Normalean, maiztasun teoriko horiek ez dira beste taula batean jartzen, baizik eta behaturiko kontingentzia-taulan, parentesien artean, bakoitzari dagokion gelaxkan.

Garrantzizkoa da azpimarratzea e_{ij} maiztasun teoriko guztiek bost baino handiagoak izan behar dutela, eta, beraz, $e_{ij} < 5$ denean, ondoz ondoko klaseak bildu beharko ditugula $e_{ij} > 5 \ \forall \ i, j$ lortu arte.

Honako hau da hipotesi-kontrastea egiteko erabiliko dugun estatistikoa:

$$\chi^2 = \sum_{j=1}^{s}\sum_{i=1}^{r}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{j=1}^{s}\sum_{i=1}^{r}\frac{O_{ij}^2}{E_{ij}} - n$$

eta bere banaketa $(r - 1)\cdot(s - 1)$ askatasun-graduko khi karratua da.

Estatistiko horren balioa kalkulatuko dugu kontingentzia-taulako datuak erabiliz:

$$\chi_p^2 = \sum_{j=1}^{s}\sum_{i=1}^{r}\frac{(o_{ij} - e_{ij})^2}{e_{ij}} = \sum_{j=1}^{s}\sum_{i=1}^{r}\frac{o_{ij}^2}{e_{ij}} - n$$

Eta hauxe da balio horri dagokion p-balioa: $p = P(\chi^2_{(r-1)\cdot(s-1)} > \chi^2_p)$.

Azkenik, α adierazgarritasun-maila zehaztuko dugu, eta bestelako hipotesi-kontrasteetan egin dugun moduan hartuko ditugu erabakiak:

- $p \leq \alpha \Rightarrow \alpha$ adierazgarritasun-mailaz, H_0 hipotesia errefusatuko dugu.
- $p > \alpha \Rightarrow \alpha$ adierazgarritasun-mailaz, ez dago H_0 errefusatzeko nabaritasunik.

Hau da, behaturiko estatistikoaren balioa χ^2_p «oso handia» bada, zilegi dirudi behaturiko maiztasunak eta itxarotakoak (faktoreak askeak direla joz) oso ezberdinak direla pentsatzea; beraz, faktoreak menpekoak direla ondorioztatuko dugu.

8.2 adibidea

Aurreko adibidean emandako kontingentzia-taulako datuetan oinarrituz, ba al dago menpekotasunik eskola-umeen pisuaren eta haien ikasketetako arrakasta goiztiarraren artean?

Emaitza:

8.4 taulan agertzen diren datuak: o_{ij}, $z_{.j}$, $f_{i.}$ eta n dira. Honako hau da hipotesi-kontrastea:

H_0: pisua eta eskolako arrakasta goiztiarra askeak dira.

H_1: pisua eta eskolako arrakasta goiztiarra menpekoak dira.

Itxarotako maiztasunak kalkulatuko ditugu, faktoreak askeak direla suposatuz:

$$e_{11} = \frac{416 \times 82}{500} = 68{,}22 \qquad e_{12} = \frac{416 \times 300}{500} = 249{,}6 \qquad e_{13} = \frac{416 \times 118}{500} = 98{,}18$$

$$e_{21} = \frac{84 \times 82}{500} = 13{,}78 \qquad e_{22} = \frac{84 \times 300}{500} = 504 \qquad e_{23} = \frac{84 \times 118}{500} = 19{,}82$$

Taula berean jarriko ditugu behaturiko o_{ij} eta itxarotako e_{ij} maiztasunak (azken horiek parentesi artean):

	Pisua			
Arrakasta goiztiarra	Gutxi	Ohikoa	Asko	
Bai	73 (68,22)	263 (249,6)	80 (98,18)	416
Ez	9 (13,78)	37 (50,4)	38 (19,82)	84
	82	300	118	500

8.4 taula. 500 eskola-umeen behaturiko eta itxarotako maiztasunen adierazpena.

Itxarotako maiztasun guztiak 5 baino handiagoak direnez, azaldutako teknika erabil dezakegu. Hauxe da estatistikoaren balioa:

$$\chi_p^2 = \sum_{j=1}^{3}\sum_{i=1}^{3}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \sum_{j=1}^{3}\sum_{i=1}^{3}\frac{o_{ij}^2}{e_{ij}} - n =$$

$$= \left(\frac{73^2}{68{,}22} + \frac{263^2}{249{,}6} + \frac{80^2}{98{,}18} + \frac{9^2}{13{,}78} + \frac{37^2}{50{,}4} + \frac{38^2}{19{,}82}\right) - 500 = 26{,}32$$

Kasu honetan $r = 2$ eta $s = 3$ da, eta askatasun-graduen kopurua $(r - 1)\cdot(s - 1) = (2 - 1)\cdot(3 - 1) = 2$ izango da.

Beraz, hauxe da kontrastearen p-balioa: $p = P(\chi_2^2 > 26{,}32)$

Khi karratu banaketaren taulan $P(\chi_2^2 > 9{,}21) = 0{,}01$ lortzen dugunez, $p < 0{,}01$ beteko da. Beraz, 0,01eko adierazgarritasun-mailaz, H_0 errefusatu egingo dugu, eta pisua eta eskolako arrakasta goiztiarra menpekoak direla ondorioztatuko dugu.

8.2.2 Yates-en zuzenketa

Batzuetan, aztertzen ditugun bi faktoreak bi klasetan edo bi mailatan soilik sailka daitezke. Datu gurutzatuak sailkatzerakoan, bi errenkadako eta bi zutabeko kontingentzia-taula lortuko da. Oro har, «2×2 taula» izenez ezagutzen da mota horretako taula, eta, kasu horretan, askatasun-graduen kopurua 1 da.

Arazo mota horretan, kontuan hartu behar dugu banaketa diskretua (zenbaki osoa den maila bakoitzaren maiztasuna) χ^2 banaketa jarraituaren bidez hurbiltzen dugula. Beraz, zuzendu beharrekoa da jaulkitzen dugun errore txikia.

Askatasun-graduen kopurua $(r - 1)\cdot(s - 1) > 1$ den kasuan, zuzenketa ez da beharrezkoa. Baina askatasun-graduen kopurua 1 denean, $\chi^2 = \sum_{j=1}^{s}\sum_{i=1}^{r}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$ estatistikoaren banaketa χ_1^2 delakoari darraiola suposatzerakoan, errore handiegia jaulkitzen dugu. Hori dela eta, Yates-ek beste estatistiko bat proposatu zuen, eta estatistiko hori hurbilketa hobea da χ_1^2 banaketarako. Estatistiko horren izena **Yates-en zuzenketa** da, eta hauxe da bere formula:

$$\chi_Y^2 = \sum_{j=1}^{2}\sum_{i=1}^{2}\frac{\left(|O_{ij}-E_{ij}|-\frac{1}{2}\right)^2}{E_{ij}}$$

Estatistiko horren banaketa askatasun-gradu bateko khi karratua da, eta honako hau izango da bere balioa:

$$\chi_Y^2 = \sum_{j=1}^{2}\sum_{i=1}^{2} \frac{\left(|o_{ij} - e_{ij}| - \frac{1}{2}\right)^2}{e_{ij}}$$

Yates-en zuzenketa erabiltzea beharrezkoa da $|o_{11}o_{22} - o_{12}o_{21}| > n/2$ betetzen denean.

Oharra**.** Aurreko formulak garatuz, honako formula hauen bidez kalkula daitezke 2×2 kontingentzia-taularako khi karratu estatistiko arrunta eta Yates-en estatistikoaren adierazpenak:

$$\chi_p^2 = \frac{n \cdot (o_{11}o_{22} - o_{12}o_{21})^2}{z_{\cdot 1} \cdot z_{\cdot 2} \cdot f_{1\cdot} \cdot f_{2\cdot}} \qquad |o_{11}o_{22} - o_{12}o_{21}| \le n/2 \text{ denean}$$

$$\chi_Y^2 = \frac{n \cdot (|o_{11}o_{22} - o_{12}o_{21}| - n/2)^2}{z_{\cdot 1} \cdot z_{\cdot 2} \cdot f_{1\cdot} \cdot f_{2\cdot}} \qquad |o_{11}o_{22} - o_{12}o_{21}| > n/2 \text{ denean}$$

Bi formula horiek 2×2 tauletako kalkuluak errazten dituzte.

8.2.3 Itxarotako maiztasun txikiak

Itxarotako maiztasunen bat 5 baino txikiagoa baldin bada, ezin dugu aurreko metodoa era zuzenean erabili, eta ondoz ondoko klaseak bildu beharko ditugu. Adibide baten bidez aztertuko dugu prozesu hori.

8.3 adibidea

Urin gastrikoaren azidotasun-maila ezberdinen (aklorhidria, hipoklorhidria, normala eta hiperklorhidria) eta gaixotasun motaren (ultzera gastrikoa eta minbizia) arteko menpekotasunik dagoen ala ez aztertu nahi dugu. Lorturiko emaitzak 8.5 taulan adierazten dira. Kontrasta ezazu, azidotasun gastrikoaren eta gaixotasun motaren arteko independentzia, % 1eko adierazgarritasun-mailaz.

	Azidotasun-maila			
	Aklorhidria	Hipoklorhidria	Normala	Hiperklorhidria
Ultzera gastrikoa	3	7	35	9
Minbizia	22	2	6	0

8.5 taula. Azidotasun-mailaren eta gaixotasun motaren arabera sailkaturiko datuen taula.

Emaitza:

Honako hauek dira kontrastatu behar ditugun hipotesiak:

H_0: urin gastrikoaren azidotasun-maila askea da gaixotasunarekiko.

H_1: urin gastrikoaren azidotasun-maila gaixotasunarekiko menpekoa da.

Itxarotako maiztasunak kalkulatuko ditugu:

$$e_{11} = \frac{54 \cdot 25}{84} = 16{,}07 \qquad e_{12} = \frac{54 \cdot 9}{84} = 5{,}78$$

$$e_{13} = \frac{54 \cdot 41}{84} = 26{,}36 \qquad e_{14} = \frac{54 \cdot 9}{84} = 5{,}78$$

$$e_{21} = \frac{30 \cdot 25}{84} = 8{,}93 \qquad e_{22} = \frac{30 \cdot 9}{84} = 3{,}21$$

$$e_{23} = \frac{30 \cdot 41}{84} = 14{,}64 \qquad e_{24} = \frac{30 \cdot 9}{84} = 3{,}21$$

Eta datu guztiak 8.6 taulan adieraziko ditugu.

	Azidotasun-maila				
	Aklorhidria	Hipoklorhidria	Normala	Hiperklorhidria	Totala
Ultzera gastrikoa	3 (16,07)	7 (5,78)	35 (26,36)	9 (5,78)	54
Minbizia	22 (8,93)	2 (3,21)	6 (14,24)	0 (3,21)	30
Totala	25	9	41	9	84

8.6 taula. Azidotasun-mailaren eta gaixotasun motaren arabera sailkaturiko gaixoen itxarotako eta behaturiko maiztasunen taula.

e_{22} eta e_{24} itxarotako maiztasunak 5 baino txikiagoak direnez, taula berria eratu behar dugu. Taula berri hori eratzeko, lehenengo zutabea bigarrenarekin eta laugarren zutabea hirugarrenarekin bilduko ditugu, eta 8.7 taulan erakutsiko dugu datuen antolamendu berria.

	Azidotasun-maila		
	Aklorhidria edo hipoklorhidria	Normala edo hiperklorhidria	Totala
Ultzera gastrikoa	10 (21,85)	44 (32,14)	54
Minbizia	24 (12,14)	6 (17,85)	30
Totala	34	50	84

8.7 taula. Gelaxkak bildu ondoren, azidotasun-mailaren eta gaixotasun motaren arteko independentzia aztertzeko taula.

2×2 taula lortu dugunez, Yates-en zuzenketa erabiltzeko beharra aztertuko dugu: $|o_{11}o_{22} - o_{12}o_{21}| = |10 \cdot 6 - 44 \cdot 24| = 996 > n/2 = 42$

Beraz, nahitaezkoa da Yates-en zuzenketa erabiltzea, eta hauexek dira behaturiko estatistikoaren balioa eta *p*-balioa:

$$\chi_Y^2 = \frac{n \cdot (|o_{11}o_{22} - o_{12}o_{21}| - n/2)^2}{z_{.1} \cdot z_{.2} \cdot f_{1\cdot} \cdot f_{2\cdot}} = \frac{84 \cdot (996 - 42)^2}{34 \cdot 50 \cdot 54 \cdot 30} = 27{,}75$$

$p = P(\chi_1^2 \geq 27{,}75)$

Khi karratu banaketaren taulan $P(\chi_1^2 \geq 6{,}635) = 0{,}01$ aurkitu dugunez, $p < 0{,}01$ beteko da. Beraz, % 1eko adierazgarritasun-mailaz, H_0 errefusatu egingo dugu, eta urin gastrikoaren azidotasun-maila eta gaixotasun mota menpekoak direla ondorioztatuko dugu.

8.3 HOMOGENEOTASUN-PROBA

Batzuetan, ikertzaileak aurrefinkatzen ditu *r*×*s* kontingentzia-taularen errenkaden total marjinalak. Kasu horietan, lagintzat hartuko ditugu aurrefinkaturiko errenkadak, eta lagin horietan zutabeetan agertzen den faktorea aztertuko dugu. Galdera honi erantzun behar diogu: lagin horiek populazio beretik aterata al daude? Galdera horri erantzuteko egin beharreko hipotesi-kontrasteari **homogenotasun-proba** deritzo, eta honako hauek dira hipotesiak:

H_0: laginak populazio beretik ateratzen dira.

H_1: laginak populazio desberdinetatik ateratzen dira.

8.4 adibidea

Ikerketa bat egin dute duodenoko ultzera eta odol-taldearen menpekotasuna aztertzeko. 1.301 gaixok eta 6.313 pertsona osasuntsuk (kontrolak) osaturiko laginak aukeratu dituzte, eta haien odol-taldea zehaztu dute. Gaixoen artean, 698 O taldekoak dira, 472 A taldekoak, 102 B taldekoak, eta gainerakoak AB taldekoak. Kontrolen artean, kopuruak 2.892, 2.625, 570 eta 226 dira, hurrenez hurren. Datu horiek honako era honetara antolatu dituzte 2×4 kontingentzia-taulan (bi errenkada eta lau zutabe):

	Odol-taldea				
	O	*A*	*B*	*AB*	Totala
Gaixoak	$o_{11} = 698$	$o_{12} = 472$	$o_{13} = 102$	$o_{41} = 29$	$f_{1.} = 1.301$ (aurrefinkaturik)
Kontrolak	$o_{21} = 2.892$	$o_{22} = 2.625$	$o_{23} = 570$	$o_{41} = 226$	$f_{2.} = 6.313$ (aurrefinkaturik)
Totala	$z_{.1} = 3.590$	$z_{.2} = 3.097$	$z_{.3} = 672$	$z_{.4} = 255$	$n = 7.614$

8.8 taula. Gaixoen eta kontrolen sailkapena, odol-taldearen arabera.

8.3.1 Planteamendu orokorra

Demagun aurrefinkatuta daudela $r \times s$ kontingentzia-taularen errenkaden total marjinalak. Kasu horretan, errenkada bakoitza lagin moduan hartuko dugu, eta lagin bakoitzean *B* faktorea aztertuko dugu. Hau da, $f_{1.}, f_{2.}, \ldots, f_{r.}$ tamainako *r* lagin ditugu, eta ea homogeneoak diren aztertu nahi dugu. Aipatu dugunez, hauexek dira homogeneotasun-probarako hipotesiak:

H_0: laginak populazio beretik aterata daude.

H_1: laginak populazio desberdinetatik aterata daude.

Homogeneotasun-proban, e_{ij} maiztasun teorikoa izango da *i*. laginean *B* faktorearen *j* modalitaterako itxarotako maiztasuna. H_0 egiazkoa bada, *B* faktorearen modalitateen proportzioak estimatzeko, lagin guztiak batera har ditzakegu, eta era horretan estimaziorik hoberenak lortuko ditugu, lagin-tamaina handiagoa baita:

$$\hat{P}(B_j) \approx \frac{z_{.j}}{n} \qquad j = 1,\ldots,r$$

Ondoren, honako era honetara kalkula ditzakegu e_{ij} maiztasun teorikoak:

$$e_{ij} = f_{i.} \cdot \; e_{ij} = f_{i.} \cdot \hat{P}(B_j) \approx f_{i.} \cdot \frac{z_{.j}}{n}$$

Hau da, homogeneotasun-probarako kalkulu matematikoak independentzia-probarako ikusitakoen berdinak dira. Eta, beraz, independentzia-probarako azaldutako estatistikoa, *p*-balioa, $e_{ij} > 5 \;\; \forall \; i, j$ izatearen baldintza edota Yates-en zuzenketa berdin aplikatzen dira homogeneotasun-probetan.

8.5 adibidea

Aurreko adibideko datuetan oinarriturik, odol-taldeak duodenoko ultzera pairatzean eraginik daukan ala ez aztertuko dugu. Beste era batera, haien odol-

taldearen arabera, gaixoak eta kontrolak populazio beretik ateratakoak diren ala ez aztertu nahi dugu.

Emaitza:

Bi lagin ezberdin ditugu, gaixoak eta kontrolak, 1.301 eta 6.313 tamainakoak, hurrenez hurren; hauek izango dira kontrastatu behar ditugun hipotesiak:

H_0: laginak populazio beretik aterata daude; edo, odol-taldearen arabera, ez dago gaixoen eta kontrolen arteko ezberdintasunik.
H_1: laginak populazio ezberdinetik aterata daude; edo, odol-taldearen arabera, gaixoak eta kontrolak ezberdinak dira.

Itxarotako maiztasunak kalkulatuko ditugu:

$$e_{11} = \frac{3590 \times 1301}{7614} = 613{,}42 \qquad e_{12} = \frac{3097 \times 1301}{7614} = 529{,}18$$

$$e_{13} = \frac{672 \times 1301}{7614} = 114{,}82 \qquad e_{14} = \frac{255 \times 1301}{7614} = 43{,}57$$

$$e_{21} = \frac{3590 \times 6313}{7614} = 2976{,}58 \qquad e_{22} = \frac{3097 \times 6313}{7614} = 2567{,}82$$

$$e_{23} = \frac{672 \times 6313}{7614} = 557{,}18 \qquad e_{24} = \frac{255 \times 6313}{7614} = 211{,}43$$

Taula berean jarriko ditugu behaturiko o_{ij} eta itxarotako e_{ij} maiztasunak (azken horiek parentesi artean):

	Odol-taldea				
	O	***A***	***B***	***AB***	**Totala**
Gaixoak	698 (613,42)	472 (529,18)	102 (114,82)	29 (43,57)	1.301 (aurrefinkatuta)
Kontrolak	2.892 (2976,58)	2.625 (2567,82)	570 (557,18)	226 (211,43)	6.313 (aurrefinkatuta)
Totala	3.590	3.097	672	255	7.614

8.9 taula. Gaixoen eta kontrolen odol-taldeetarako behaturiko eta itxarotako maiztasunak

Itxarotako maiztasun guztiak 5 baino handiagoak direnez, estatistikoaren balioa

era zuzenean kalkula dezakegu:

$$\chi_p^2 = \sum_{j=1}^{4}\sum_{i=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \sum_{j=1}^{4}\sum_{i=1}^{2}\frac{o_{ij}^2}{e_{ij}} - n =$$

$$= \left(\frac{698^2}{613{,}42} + \frac{472^2}{529{,}18} + \frac{102^2}{114{,}82} + \frac{29^2}{43{,}57} + \frac{2892^2}{2976{,}58} + \frac{2625^2}{2567{,}82} + \frac{570^2}{557{,}18} + \frac{226^2}{211{,}43}\right) - 7614 = 29{,}12$$

Kasu honetan, $r = 2$ eta $s = 4$ dira, eta askatasun-gradu kopurua $(r - 1)\cdot(s - 1) = (2 - 1)\cdot(4 - 1) = 3$ izango da.

Beraz, kontrastearen p-balioa $p = P(\chi_3^2 \geq 29{,}12)$ da.

Khi karratu banaketaren taulan $P(\chi_3^2 \geq 11{,}345) = 0{,}01$ dugunez, $p < 0{,}01$ beteko da. Beraz, % 1eko adierazgarritasun-mailaz, H_0 errefusatu egingo dugu, eta ondorioztatuko dugu gaixoen eta kontroletako odol-taldearen sailkapena ezberdina dela, edo bi laginak ez daudela odol-taldearen araberako populazio beretik aterata.

8.6 adibidea

108 pertsonetatik, tratamendu profilaktikoaren menpean egon eta gero, 20 gaixotu egin ziren; eta kontrasterako plazeboak hartu zituzten 237 pertsonetatik 74 gaixotu ziren. Adierazgarria al da sendagai errealaren (tratamendu profilaktikoaren) eta itxurazkoaren (kontrasterako plazeboen) arteko aktibitate ezberdintasuna? (har ezazu $\alpha = 0{,}05$)

Emaitza:

Sendagaiaren eta plazeboaren arteko aktibitate ezberdintasuna adierazgarria izatea, talde bietan gaixotzeko probabilitate bera izatea, edo bi laginak populazio beretik ateratakoak izatea da. Beraz, homogeneotasun-proba gauzatuko dugu, honako hipotesi-kontraste honen bidez:

H_0: laginak populazio beretik aterata daude; edo tratamenduak ez du eraginik.

H_1: laginak populazio ezberdinetatik aterata daude; edo tratamenduak eragina du.

Emandako datuak honako kontingentzia-taula honetan azaltzen dira:

	Gaixoak	Osasuntsuak	
Tratamendua	20	88	108
Tratamendurik ez	74	163	237
	94	251	345

8.10 taula. Tratamenduari eta gaixotzeari buruzko datuak.

2×2 taula daukagunez, Yates-en zuzenketa erabiltzeko beharra aztertuko dugu:

$$|o_{11}o_{22} - o_{12}o_{21}| = |20{,}163 - 74{,}88| = 3252 > n/2 = 1/172{,}5$$

Beraz, nahitaezkoa da Yates-en zuzenketa erabiltzea. Hauek izango dira estatistikoaren balio eta *p*-balioa:

$$\chi_Y^2 = \frac{n \cdot (|o_{11}o_{22} - o_{12}o_{21}| - n/2)^2}{z_{\cdot 1} \cdot z_{\cdot 2} \cdot f_{1\cdot} \cdot f_{2\cdot}} = \frac{345 \cdot (3252 - 172{,}5)^2}{94 \cdot 251 \cdot 108 \cdot 237} = 5{,}417$$

eta $p = P(\chi_1^2 \geq 5{,}417)$

Khi karratu banaketaren taulan $P(\chi_1^2 \geq 5{,}412) = 0{,}02$ dugu, eta, beraz, $p < 0{,}02 < 0{,}05$ beteko da. Ondorioz, % 5eko adierazgarritasun-mailaz, H_0 errefusatu egingo dugu, eta tratamenduaren eta plazeboaren eraginkortasunen arteko ezberdintasuna adierazgarria dela onartuko dugu.

8.3.2 Bi proportzioren konparazioa

2×2 taula daukagunean, bai independentzia- eta bai homogeneotasun-proba, bi proportzio konparatzean datza. Esate baterako, 8.6 adibidea begiratuz, baliokideak dira homogeneotasun-proba egitea eta tratamendupean egondakoen eta tratamendurik gabekoen gaixotuen proportzioen berdintasunerako hipotesi-kontrastea egitea. 7. gaian, bi proportzio konparatzeko beste metodo bat aipatu dugunez, aurreko adibidearen bidez egiaztatuko dugu bi metodo horien baliokidetasuna.

8.7 adibidea

Aurreko adibidearen datuetan oinarrituz, konpara itzazu tratamendupekoen eta tratamendurik gabekoen gaixo-proportzioak.

Emaitza:

Independentzia-probaren bidez, bi laginak (tratamendupekoak eta tratamendurik gabekoak) ez direla homogeneoak ondorioztatu dugu; hau da, proportzioak ezberdinak direla.

Orain, 7. gaian azterturiko proportzioak konparatzeko, hipotesi-kontrastea gauzatuko dugu.

p_1 eta p_2 izango dira tratamendupekoen eta tratamendurik gabekoen gaixo-proportzioak, hurrenez hurren. Planteaturiko arazoa bi proportzioen arteko aldebiko hipotesi-kontraste gisa azter daiteke:

H_0: $p_1 = p_2$

H_1: $p_1 \neq p_2$

Tratamendua izan dutenen artean, $\hat{p}_1 = \frac{20}{108} = 0{,}185$ eta, tratamendurik gabekoen artean, $\hat{p}_2 = \frac{74}{237} = 0{,}312$ dira.

H_0 egia dela suposatuz, honako hau da estatistikoaren balioa:

$$z_p = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}} = \frac{0{,}185 - 0{,}312}{\sqrt{\frac{0{,}185 \cdot 0{,}815}{108} + \frac{0{,}312 \cdot 0{,}688}{237}}} = -2{,}65 .$$

$p = 2 \cdot P(Z > 2{,}65) = 2 \cdot 0{,}00402 = 0{,}00804 < 0{,}05$. Beraz, % 5eko adierazgarritasun-mailaz, H_0 errefusatuko dugu. Hau da, tratamendupekoen eta tratamendurik gabekoen gaixo-proportzioak ezberdinak dira. Beraz, homogeneotasun-probaren bidez lorturiko emaitza berbera da.

8.3.3 Fisher-en proba zehatza

Orain arte gai honetan ikusi ditugun Pearson-en khi karratu banaketaren erabileretan, beharrezkoa izan da laginak handiak izatea. Zenbateko handiak? Galdera horren erantzun zuzenik ez egon arren, oro har, n handia suposatuko dugu, itxarotako maiztasun bat ere ez denean 5 baino txikiagoa. Baldintza hori ez bada betetzen, aurretik ere aipatu dugu klaseak bildu egin behar direla, betetzea lortu arte. Batzuetan konponbide hori ezinezkoa da, esate baterako 2×2 taula dugunean. Lagin txiki horietan, **Fisher-en proba zehatza** erabiltzen da independentzia- eta homogeneotasun-proba gauzatzeko.

Fisher-en proba zehatza banaketa hipergeometrikoan oinarritzen da; ez dugu hemen garatuko, duen zailtasunagatik. Dena den, egiten dugun hipotesi-kontrastea planteamendu orokorrean egindako berbera da; aldatzen diren gauza bakarrak estatistikoa, bere banaketa, eta *p*-balioa dira. Fisher-en proba behaturikoa baino taula «arraroagoa» lortzearen probabilitatea kalkulatzean datza. Horretarako, behaturiko maiztasun txikiena baino txikiagoak dituzten taula guztiak eraiki, eta bakoitza ateratzeko probabilitatea kalkulatuko da; probabilitate horien guztien batura izango da hain arraroa den taula lortzeko probabilitatea. Kalkulaturiko probabilitate hori *p*-balioa da. Gaur egun, ordenagailuen erabilerak *p*-balioaren kalkulua errazten duenez, honako adibide honetan Fisher-en proba zehatza erabili da, kalkuluak ordenagailuaren bidez eginik.

8.8 adibidea

Hilabetean zehar, 50 eta 54 urte bitartean hildako gizonezkoei buruzko azterketa gauzatu da. Gaixotasun kardiobaskularraren (GKB) ondorioz hildakoen eta beste gaixotasun baten (BGB) ondorioz hildakoen pertsona kopuru bera biltzen saiatu ziren ikertzaileak. GKBren ondorioz hildako 35 pertsonetatik 5ek eta BGBren ondorioz hildako 25etik 2k soilik zuten gatz gutxiko dieta. Datu horiek 8.10 taulan antolatu dira, 2×2 kontingentzia-taularen erara.

	Dieta mota		
heriotza-kausa	Gatz askokoa	Gatz gutxikoa	Totala
BGB	2	23	25
GKB	5	30	35
	7	33	60

8.11 taula. Heriotza-kausari eta dieta motari buruzko datuak

Azter ezazu dieta motaren eta heriotzaren kausaren arteko menpekotasuna, $\alpha = 0{,}05$ erabiliz.

Emaitza. Kasu honetan, gaixo bakoitzaren ondorioz hildakoen kopurua aurrefinkaturik dago, beraz, homogeneotasun-proba gauzatuko da. Honako hau da hipotesi-kontrastea:

H_0: laginak populazio beretik atera dira; ez dago dieta motaren eta heriotzaren kausaren arteko menpekotasunik.

H_1: laginak populazio ezberdinetatik atera dira; menpekotasuna dago dieta motaren eta heriotzaren kausaren artean.

$e_{12} = \frac{7 \times 25}{60} = 2{,}92 < 5$, beraz ezin dugu khi karratu banaketa erabili.

Ordenagailuaren bidez, H_0 egiazkotzat harturik, behaturikoa bezain taula arraroa edo arraroagoa lortzeko probabilitatea 0,75 da. 0,75 > 0,05 denez, H_0 ezin dugu errefusatu eta, beraz, % 5eko adierazgarritasun-mailaz, ezin da esan dieta motaren eta heriotzaren kausaren arteko menpekotasunik dagoenik.

8.4 DOIKUNTZA-EGOKITASUNERAKO PROBA

Populazioak banaketa-funtzio zehatz bat duela dakigunean edo gertaera posible guztiek probabilitate ezagunak dauzkaten hipotesia egiten dugunean, egindako hipotesia arrazoigarria den egiaztatu beharko dugu, populaziotik ateratako laginaren bidez. Banaketa teorikoa eta behaturiko banaketa konparatzen dituen probari doikuntza-egokitasuna deritzo.

Demagun Z_1, Z_2,…, Z_k klase baztertzaile eta exhaustiboetan sailkaturiko populazioa dugula, eta klase horien probabilitateak $p_1 + p_2 + ... + p_k = 1$ betetzen duten p_1, p_2,…, p_k diren hipotesia kontrastatu nahi dugula. Alegia,

H_0: k klaseen probabilitateak p_1, p_2,…, p_k dira.

H_1: k klaseen probabilitateak EZ dira p_1, p_2,…, p_k.

n elementu askek osaturiko lagina aukeratuko dugu zoriz, eta honako hauek izango dira klase bakoitzerako itxarotako maiztasunak edo maiztasun teorikoak:

$$e_1 = n \cdot p_1 \ ; \ e_2 = n \cdot p_2 \ ; \ldots; \ e_k = n \cdot p_k$$

Honako baldintza hau beteko dute maiztasun teoriko horiek:

$$e_1 + e_2 + ... + e_k = n \cdot p_1 + n \cdot p_2 + ... + n \cdot p_k = n(p_1 + p_2 + ... + p_k) = n \cdot 1 = n$$

Laginean behaturiko maiztasunak $o_1, o_2, ..., o_k$ -ren bidez adieraziko ditugu, $o_1 + o_2 + ... + o_k = n$ izanik.

Praktikan, hemen azalduko dugun teknika aplikatu ahal izateko, klase guztietako itxarotako maiztasunek 5 baino handiagoak izan behar dute ($e_i > 5 \ \forall \ i$). Baldintza hori betetzen ez denean, ondoz ondoko klaseak batuko dira, edozein i-tarako $e_i > 5$ lortu arte.

Orduan, hipotesi-kontrastea egiteko erabiliko dugun estatistikoa $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ izango da.

Estatistiko horren banaketa khi karratua da, eta askatasun-graduak p_1, p_2,…, p_k probabilitateak guztiz zehazturik egotearen ala ez egotearen menpean daude. Hain zuzen, behaturiko maiztasun askeen kopurutzat jotzen da askatasun-graduen kopurua.

- p_1, p_2,…, p_k probabilitateak zeharo espezifikaturik daudenean: kasu honetan, $k - 1$ maiztasuna emanik, $o_1 + o_2 + ... + o_k = n$ denez, beste maiztasuna zehazturik dago; beraz, khi karratuaren askatasun-graduak $k - 1$ dira.
- p_1, p_2,…, p_k probabilitateak zeharo espezifikatu gabe daudenean: itxarotako maiztasunak edo maiztasun teorikoak lortzeko l parametro behar baldin badira, parametro bakoitzaren estimazioak o_i-ren arteko erlazio berri bat ($o_1 + o_2 + ... + o_k = n$ erlazioaz gainera) ezarriko du, eta, beraz, estimaturiko parametro bakoitzeko unitate batez gutxituko dira khi karratu banaketaren askatasun-graduak. Kasu honetan, askatasun-graduen kopurua $k - l - 1$ da, non l estimaturiko parametro kopurua baita.

Oro har, estatistiko horren banaketaren askatasun-graduen kopurua $k - l - 1$ da, non k batu eta geroko klase kopurua eta l estimaturiko parametro kopurua baitira.

Era honetara kalkulatuko da behaturiko estatistikoaren balioa, hau da, lagineko datuetarako proba-estatistikoak hartzen duen balioa:

$$\chi_p^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \sum_{i=1}^{n} \frac{(o_i - np_i)^2}{np_i}$$

χ_p^2 «oso handia» bada, ezberdintasuna adierazgarria dela adierazten du; beraz, banaketa teorikoa eta behaturikoa batera datozen hipotesia errefusatu behar dugu. H_0 hipotesia onargarria izan dadin, χ_p^2-ren balioak «txikia» izan beharko du.

Balio horri dagokion p-balioa $p = P(\chi_{k-l-1}^2 > \chi_p^2)$ da.

Azkenik, α adierazgarritasun-maila zehaztuko dugu, eta honako era honetara arrazoituko dugu:

- $p \leq \alpha \Rightarrow \alpha$ adierazgarritasun-mailaz H_0 hipotesia errefusatuko dugu. Hau da, Z_1, $Z_2,\ldots, Z_k$ klaseen probabilitateak $p_1, p_2,\ldots, p_k$ diren hipotesia errefusatuko dugu.

- $p > \alpha \Rightarrow$ ezberdintasunak zoriz ager ditzake $\Rightarrow \alpha$ adierazgarritasun-mailaz ez dago H_0 errefusatzeko nabaritasunik. Hau da, Z_1, $Z_2,\ldots$, Z_k klaseen probabilitateak p_1, $p_2,\ldots, p_k$ diren hipotesia onartuko dugu.

Oharra: proba estatistikoaren kalkulua errazteko, honako formula baliokide hau erabil daiteke:

$$\chi_p^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{o_i^2 - 2o_ie_i + e_i^2}{e_i} = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - 2\sum_{i=1}^{k} o_i + \sum_{i=1}^{k} e_i = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - 2n + n =$$

$$= \sum_{i=1}^{k} \frac{o_i^2}{e_i} - n$$

Oharra. Komenigarria da azpimarratzea χ^2 testak ez direla behin betiko erabaki orokorrak. Aurrefinkaturiko arriskuz, zenbait hipotesi onartzeko ala errefusatzeko irizpide arrazoizkoak ditugu. Alegia, χ^2 testetik abiatuz, ezberdintasun adierazgarririk ez dagoela (edo doikuntza egokia dela) baiesten dugunean, ez dugu adierazten populazioak suposatu dugun banaketa betetzen duela matematikoki frogatu dugunik, baizik eta χ^2 testaren irizpidea erabiliz, ez dugula ezberdintasunik ikusi. Horrek ez du esan nahi beste irizpide baten bidez ezberdintasuna aurki dezakegunik; are gehiago, χ^2 test berberaz, baina beste lagin bat erabiliz, ezberdintasunak ager daitezke.

8.9 adibidea

Mendel-en legeak esperimentalki egiaztatu nahi ditugu. Horretarako, 500 landare gurutzatu ziren; teoriaren arabera, lore gorri, arrosa, hori eta zuriko landare kopuruak 8, 12, 10 eta 20 zenbakien proportzionalak izan beharko lirateke. 8.12 taulan adierazi dira lorturiko datuak.

Kolorea	Gorria	Arrosa	Horia	Zuria
Behaturiko maiztasunak	70	126	96	208

8.12 taula. 8.9 adibideko datuak

Emaitza: itxarotako maiztasunak 8–12–10–20 proportzionaltasunak ematen dizkigu; alegia,

e_G = lore gorrien kopurua = $500 \cdot \frac{8}{50} = 80$

e_A = lore arrosen kopurua = $500 \cdot \frac{12}{50} = 120$

e_H = lore horien kopurua = $500 \cdot \frac{10}{50} = 100$

e_Z = lore zurien kopurua = $500 \cdot \frac{20}{50} = 200$

Loreen kolorea	o_i	e_i	o_i^2/e_i
Gorriak	70	80	61,25
Arrosak	126	120	132,30
Horiak	96	100	92,16
Zuriak	208	200	216,32
Denetarikoak	500	500	502,03

8.13 taula. 8.9 adibideko behaturiko eta itxarotako maiztasunak.

Honako hau da estatistikoaren behaturiko balioa:

$$\chi_p^2 = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - n = 502{,}03 - 500 = 2{,}03$$

Askatasun-graduen kopurua $k - l - 1 = 4 - 0 - 1 = 3$ da (e_i maiztasun teorikoak kalkulatzeko ez da parametrorik estimatzen).

Eta p-balioa $p = P(\chi_3^2 > 2{,}03)$ da. Taulan $P(\chi_3^2 > 6{,}251) = 0{,}1$ agertzen da. Baina $p = P(\chi_3^2 > 2{,}03) > P(\chi_3^2 > 6{,}251) = 0{,}1$ denez, $p > 0{,}1$ lortuko dugu. Ondorioz, % 5eko adierazgarritasun-mailaz, H_0 ezin dugu errefusatu; beraz, Mendelen legea esperimentalki egiaztatu da, zeren lore gorri, arrosa, hori eta zuriko landare kopuruak 8, 12,10 eta 20 zenbakiei proportzionalak baitira.

8.10 adibidea

Kultibo tanta bat hedatzen denean, plaka laukidunaren erretikulu bakoitzeko bakterio kopurua Poisson-en banaketari darraion zorizko aldagaia den hipotesia

kontrastatu nahi dugu (α = 0,05 mailaz). Horretarako, kultiboaren disoluzio tanta bat 400 erretikulu txikiko plakan hedatu zen eta, mikroskopioaren laguntzaz, erretikulu bakoitzeko bakterio kopuruak zenbatu ziren. 8.15 taulan adierazi dira lorturiko o_i emaitzak.

Emaitza: itxarotako maiztasunak estimatu behar dugun λ parametroko Poisson-en banaketari jarraitzen zaizkio; o_i emaitzetatik abiatuz, lagin-batezbestekoaren bidez estimatuko dugu λ, hots:

$$\hat{\lambda} = \bar{x} = \frac{1}{400} \cdot \sum x_i \cdot o_i = 4{,}68$$

Hauek dira kontrastatu nahi ditugun hipotesiak:

H_0: plaka laukidunaren erretikulu bakoitzeko bakterio kopurua $\mathcal{P}(4.68)$ banaketari darraio.

H_1: plaka laukidunaren erretikulu bakoitzeko bakterio kopurua ez darraio $\mathcal{P}(4{,}68)$ banaketari.

8.14 taulan agertzen dira behaturiko maiztasunak, Poisson-en legearen arabera itxarotako maiztasunak, $e_i > 5$ lortzeko egin behar diren moldaketak, eta estatistikoaren kalkulua errazten duen azken zutabea.

$\forall\ i\ e_i > 5$ izateko, taulan azaltzen den moduan bildu ditugu klaseak, eta hamar klase lortu ditugu. Beraz, askatasun-graduen kopurua $k - l - 1 = 10 - 1 - 1 = 8$ da.

Itxarotako maiztasunen kalkulua honako Poisson-en probabilitate-legearen formula hau erabiliz egin dugu:

$$e_i = n \cdot P(X = i) = 400 \cdot \frac{e^{-4{,}68} \cdot 4{,}68^i}{i!}$$

non $i = 1, 2, 3, \ldots, 15$.

Esate baterako, $e_0 = n \cdot P(X = 0) = 400 \cdot \dfrac{e^{-4{,}68} \cdot 4{,}68^0}{0!} = 400 \cdot 9.27 \cdot 10^{-3} = 3{,}71$

Orduan, honako hau da behaturiko estatistikoaren balioa:

$\chi^2_p = \sum_{i=1}^{k} \dfrac{o_i^2}{e_i} - n = 404{,}39 - 400 = 4{,}39$ eta p-balioa $p = P(\chi^2_8 > 4{,}39)$ da.

Taulan $P(\chi^2_8 > 13{,}36) = 0{,}1$ agertzen da. Baina $p = P(\chi^2_8 > 4{,}39) > P(\chi^2_8 > 13{,}36) = 0{,}1$ denez, $p > 0{,}1$ lortuko dugu. Ondorioz, % 5eko adierazgarritasun-mailaz,

H_0 ezin dugu errefusatu; beraz, onartzen dugu erretikuluko bakterio kopurua λ = 4,68 parametroko Poisson-en banaketari darraion zorizko aldagaia dela.

	X bakterio kopurua daukan erretikulu kopurua				
***X*, bakterio kopurua**	**Behaturikoak (o_i)**		**Itxarotakoak (e_i)**		**o_i^2/e_i**
0	0	20	3,71	21,08	18,98
1	20		17,37		
2	43		40,65		45,49
3	53		63,41		44,30
4	86		74,19		99,69
5	70		69,44		70,56
6	54		54,16		53,84
7	37		36,21		37,81
8	18		21,18		15,30
9	10		11,02		9,07
10	5	9	5,16	8,66	9,35
11	2		2,19		
12	2		0,86		
13	0		0,31		
14	0		0,10		
15	0		0,03		
15 baino gehiago	0		0,01		
Baturak	400		400		404,39

8.14 taula. 8.10 adibideko datuen antolamendua.

8.5 ARIKETAK

8.1. Hepatitisaren aurkako txerto berriaren ikerketan 1.083 gizonezko boluntario erabili ziren. Horietatik 549 zoriz aukeratu eta sendagai berriarekin txertatu ziren, eta gainerako 534ak txertatu gabe gelditu ziren. Denbora pasatu ondoren, txertatuen artetik 11 eta txertatu gabekoen artetik 70 boluntario gaixotu zirela egiaztatu zen. Zehaztu ezazu txertaturik egotea eta hepatitisa pairatzea askeak diren ala ez.

8.2. Urte askoan zehar pentsatu da bai faktore genetikoak eta bai ingurunea ultzera peptikoaren kausa izan daitezkeela. Orain dela gutxi egindako ikerketan, odoleko I motako pepsinogeno-maila eta ultzera peptikoa edukitzearen menpekotasuna aztertu zen; ultzeradun 14 paziente 49 kontrolekin konparatu ziren. Horietariko bakoitzaren odoleko I pepsinogeno-maila neurtu, eta baxu edo altu gisa sailkatu zen. Honako taula honetan adierazi dira lorturiko datuak:

	Pepsinogeno-maila		
Ultzera	Altua	Baxua	
Bai	12		114
Ez		31	49

a) Osa ezazu taula.

b) Eraiki ezazu bi ezaugarrien arteko menpekotasuna kontrastatzeko hipotesi nulu egokia.

c) Kalkula ezazu gelaxka bakoitzerako itxarotako pertsona kopurua.

d) Egin ezazu hipotesi-kontrastea.

8.3. Genetika-ikertzaileak gaixotasunaren heredagarritasuna aztertzen du. Honako datu hauek bildu ditu 180 haurreko laginean:

	Gaixotasuna, haurrak pairatzen al du?	
Amak pairatzen al du?	Bai	Ez
Bai	10	70
Ez	8	92

0,01eko adierazgarritasun-mailaz, gaixotasuna amak pairatu badu, probableagoa al da haurrak pairatzea?

8.4. Bost eskualdetan jaiotzen diren haur deformatuak ez direla uniformeki banatzen susmatzen da; hots, zenbait industria kaltegarri direla kausa, haur deformatuen

portzentajea eskualde batzuetan beste batzuetan baino handiagoa dela susmatzen da. Susmoa egiaztatzeko ala baztertzeko, honako datu hauek hartu ziren kontuan:

Eskualdea	Haur deformatuen kopurua	Jaiotza-kopuru osoa
1	62	1.740
2	38	1.498
3	25	942
4	38	1.312
5	50	1.326
	213	6.818

Kontrasta ezazu eskualdearen eta haur deformatuen proportzioen arteko independentzia ($\alpha = 0{,}05$).

8.5. Adinaren eta tuberkulinari erreakzioa izatearen menpekotasuna aztertu nahi zen, eta, horretarako, ikerketa gauzatu zen. Adin ezberdineko 1.969 pertsona injektatu ziren, eta honako taula honetan adierazi da haien erreakzioa:

	Erreakzioa		
Adina	Positiboa	Negatiboa	Totala
17tik 19ra	285	983	1.268
20tik 24ra	172	401	573
25 edo gehiago	59	69	128
Totala	516	1.453	1.969

Datu horiek egiaztatzen al dute adinaren eragina tuberkulinari egindako erreakzioan? Zenbatekoa da *p*-balioa?

8.6. Edozein motatako alergia dela eta klinikara joandako 19 urtetik beherako pazienteen artetik 245 tamainako lagina aukeratu dute. Paziente bakoitza adinaren eta arrautzari alergia izatearen arabera sailkatu dute. Hiru urtetik gorako 133 umeen artean 30ek, eta 3 urtetik beherako 112 umeen artean 32k alergia izan zioten arrautzari.

a) Eraiki ezazu datu horien 2×2 kontingentzia-taula.

b) Kontrasta ezazu adinaren eta arrautzari alergia diotenen arteko ez-menpekotasunaren hipotesia. Independentzia- edo homogeneotasun-proba gauzatu al duzu?

8.7. Azalea basatien kolorea eta usain gozoaren arteko erlazioa aztertzeko ikerketa egin da. Loretan dauden 200 landare zoriz aukeratu, eta haien kolorearen eta

usain gozoaren presentziaren arabera sailkatu dira. Datuak honako taula honetan adierazi dira:

	Kolorea		
Usain gozoa	**Zuria**	**Arrosa**	**Laranja**
Bai	12	60	58
Ez	50	10	10

a) Ikertzaileak aurrefinkatu al du total marjinalen bat?

b) Zer gauzatu da, independentzia-proba ala homogeneotasun-proba?

c) Eraiki ezazu hipotesi nulu egokia.

d) Errefusa daiteke H_0? Zenbatekoa da testaren p-balioa? Zer ondorio atera dezakegu datu horietatik?

8.8. Zoriz aukeraturiko 200 haziri aplikaturiko hiru tratamendu kimiko probatzean, hozitze-probak gauzatuz, honako 3×2 kontigentzia-taulan adierazitako emaitzak lortu ziren.

Tratamendu kimikoa	Hozi zirenak	Hozi ez zirenak	Denetara
A	190	10	200
B	170	30	200
D	180	20	200
Denetara	540	60	600

Kontrasta ezazu tratamendu kimikoak hozitzean eragina daukan ala ez, % 5eko adierazgarritasun-mailaz.

8.9. Honako taula honetan adierazi dira eskualde bakoitzean aurkituriko amorrudun azerien maiztasunak. Kontrasta ezazu amorruaren eragina bi eskualdeetan berdina den hipotesia.

	Amorrua		
Eskualdea	Bai	Ez	Totala
I	14	29	43
II	12	38	50
Totala	26	67	93

8.10. Pearson-en χ^2 banaketa erabiliz, egiazta ezazu hemorragia anormalak dauzkaten emakumeen proportzioak berdinak diren hipotesia, haurdunaldian zehar droga hartu zutenentzat eta hartu ez zutenentzat.

	Droga hartu zutenak	Droga hartu ez zutenak	Denetara
Hemorragia normalak	14	6	20
Hemorragia ez-normalak	347	334	681
Denetara	361	340	701

8.11. Plaka laukidunaren gainean zortzi odol tanta zeuden eta, mikroskopioaren laguntzaz, tanta bakoitzeko globulu gorrien kopurua zenbatu zen:

1870 – 1994 – 1872 – 1880 – 1869 – 1820 – 1830 – 1836

Onar al dezakegu zortzi odol tantak (bolumen berekoak) Poisson-en populazio beretik datozela? Har ezazu $\alpha = 0{,}05$.

8.12. Kuien arteko gurutzamenduaren 64 ondorengoen artean, 34 gorri, 10 beltz eta 20 zuri lortu ziren. Genetikak emandako ereduaren arabera, zenbaki horiek 9:3:4 proportzioan agertu behar dute. $\alpha = 0{,}05$ izanik, azter ezazu datu horiek eta ereduak elkarrekin komunztatzen duten.

8.13. 250 gaixorengan jasotako azido urikoaren determinazioen banaketa adierazten digu honako taula honek:

Azido urikoaren determinazioak	Behaturiko maiztasunak
< 1	1
1 – 1,99	5
2 – 2,99	15
3 – 3,99	24
4 – 4,99	43
5 – 5,99	50
6 – 6,99	45
7 – 7,99	30
8 – 8,99	22
9 – 9,99	10
≥10	5

Froga ezazu, $\mu = 5{,}74$ eta $\sigma = 2{,}01$ parametroetako banaketa normalarekiko datu horien doikuntzaren egokitasuna ($\alpha = 0{,}01$).

8.14. Hirurogeita hamar gaixori bost proba egitean datzan esperimentua gauzatu da. Probak positibo emateari «arrakasta» deritzogu, eta honako hau da lorturiko arrakasta-banaketa:

Arrakasta kopurua	Gaixo kopurua
0	3
1	8
2	14
3	18
4	15
5	12

Onar al daiteke datu horiek banaketa binomiala betetzen duten hipotesia (α = 0,05)?

8.15. Honako taula honetan, ibai batetik ateratako 100 ur-laginetan aurkitutako organismo kopuruak azaldu dira:

Lagineko organismo kopurua	Behaturiko maiztasuna
0	15
1	30
2	25
3	20
4	5
5	4
6	1
7	0

Froga ezazu datu horiek Poisson-en banaketatik atera zireneko hipotesi nulua (α = 0,05).

9 BARIANTZA-ANALISIA

9.1 SARRERA

Seigarren gaian, bi populazio askeren batezbestekoak konparatzeko teknikak azaldu dira, *t* banaketa erabiliz. Orain, gai horretan garatutako metodo hura bi populazio baino gehiagora hedatuko dugu, *F* banaketa erabiliz bariantza-analisi deritzon teknikaren bitartez. Teknika estatistiko horren izena dagokio erantzunaren aldakortasun osoa jatorri ezberdin eta ezagunei loturik dauden osagarrietan zatitzeari. Bariantza-analisia oso analisi-metodo ahaltsua eta sendoa da. Erantzuna esango diogun aldagai jarraituaren batezbestekoa, faktore esango diogun aldagai diskretuaren modalitatean zehar ezberdina den aztertzea da bariantza-analisiaren helburua. Seigarren gaian aurkeztutako metodoa da horren kasu berezia, hain zuzen, faktorearen modalitate kopurua bi denean. Bariantza-analisia da datuak aztertzeko modu egokia izateko behar diren baldintzak erantzuna banaketa normalari darraion aldagaia izatea, eta erantzunaren bariantza-faktorearen modalitate ezberdinetan zehar berdina izatea. Hain zuzen, faktorea populazioaren azpitaldeak egiteko modura ulertzen bada, bariantza-analisiaren helburua da bariantza bereko populazio normalen batezbestekoen berdintasuna konparatzea. Kasurik sinpleena faktore bakar bateko bariantza-analisi deritzona da, eta horixe da, hain zuzen, gai honetako bigarren atalean garatuko duguna. Erantzunaren batezbestekoa faktore bat baino gehiagoren menpe aztertu nahi denean, bi faktoreko edo gehiagoko bariantza-analisiaren aurrean gaude, eta testu honetatik kanpo geratzen da helburu hori.

9.2 FAKTORE BAKARREKO BARIANTZA-ANALISIA

Atal honetan, faktore bakar bateko edo sailkapen bakuneko bariantza-analisi deritzona garatuko dugu. Lehendabizi, metodo horrekin aztertu beharko genukeen kasu praktiko bat aurkeztuko dugu adibide gisa.

9.1 adibidea

Aknearen tratamendurako, hiru terapia-metodoren eraginkortasuna konparatzeko esperimentua egin dute. Honako hauek dira erabilitako hiru metodoak:

1. Eguneko 250 mg tetraziklina erabiltzeaz gainera, polietilenozko eskuilaz eta xaboi urratzaile batekin egunean birritan garbitzea.
2. Tretinoina-krema ematea, eguzkia saihestea, urarekin eta xaboi emultsionatzaile batekin egunean birritan garbitzea, eta baita, egunean birritan, 250 mg tetraziklina erabiltzea ere.
3. Ura saihestea, lipidorik gabeko garbigarri batekin egunean birritan garbitzea, eta tretinoina-krema eta bentzoilo-preoxidoa ematea.

Hiru metodo horiek aknea daukaten pazienteengan konparatu dituzte, lesio

kopurua murrizteko beren eraginkortasunaren arabera. Esperimentuan 35 pazientek parte hartu dute, eta 10, 12 eta 13 tamainako hiru taldetan zoriz banatu dira, talde bakoitzari tratamendu bat esleituz. Hamasei aste pasatu ondoren, paziente bakoitzaren lesio kopuruaren hobekuntza portzentajea neurtu zuten.

Adibide honetan, $N = 35$ aknea pairatzen duten paziente eta $k = 3$ tratamendu daude konparatzeko. Pazienteak zoriz sailkatu dira $n_1 = 10$, $n_2 = 12$ eta $n_3 = 13$ tamainako taldeetan. Talde bakoitzak tratamendu bat jaso du, eta 16 aste pasatu ondoren neurtu den erantzuna da behaturiko lesio kopuruaren hobekuntza-portzentajea. I, II eta III tratamenduak hartu dituzten paziente guztien populazioetatik ateratako zorizko lagin asketzat hartzen dira hiru azpitaldeak, hurrenez hurren. Lorturiko datuetan oinarriturik, honako hau kontrastatu nahi dugu:

H_0: $\mu_1 = \mu_2 = \mu_3$ (hiru tratamenduen arteko diferentziarik ez dago batez besteko erantzunean)

H_1: $\mu_i \neq \mu_j$ i eta j batzuentzat (gutxienez tratamendu baten batez besteko erantzuna ezberdina da besteekiko)

Faktore bakarra esatearen arrazoia da esperimentu bakoitzean faktore edo ezaugarri bakar baten eragina aztertzea. Aipaturiko adibidean interesa daukan faktorea tratamendua da. Esperimentua faktorearen k mailatan oinarritzen da; aurreko adibidean $k = 3$ tratamenduan. Askotan erabiltzen den guztiz zorizkoa izatearen arrazoia da, baita ere, k laginak elkarrekiko askeak izatea. Kontuan hartutako datuak modu honetara adierazi dira 9.1 taulan.

Faktorearen maila				
1	2	3	...	k
X_{11}	X_{21}	X_{31}	…	X_{k1}
X_{12}	X_{22}	X_{32}	…	X_{k2}
X_{13}	X_{23}	X_{33}	…	X_{k3}
…	…	…	…	…
X_{1n_1}	X_{2n_2}	X_{3n_3}	…	X_{kn_k}

9.1 taula. Bariantza-analisirako datuen antolamendua.

Halaber, i. populaziotik ateratako laginaren tamaina n_i da, $N = \sum_{i=1}^{k} n_i$ erantzunaren kopuru osoa da, eta X_{ij} faktorearen i. maila j. alearen erantzuna da, non $i = 1, 2, ..., k$ eta j

=1, 2, ..., n_i baitira. Lagineko datuak kalkulatzen direnean, honako estatistiko hauek kalkulatuko dira:

$$T_{i\cdot} = \sum_{j=1}^{n_i} X_{ij} = i. \text{ mailako erantzunen batura osoa} \qquad (i = 1, 2, ..., k)$$

$$\overline{X}_{i\cdot} = \frac{T_{i\cdot}}{n_i} = i. \text{ mailako erantzunen batezbestekoa} \qquad (i = 1, 2, ..., k)$$

$$\mathrm{T}_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} = \sum_{i=1}^{k} T_{i\cdot} = \text{erantzunen batura osoa}$$

$$\overline{X}_{\cdot\cdot} = \frac{T_{\cdot\cdot}}{N} = \text{erantzunen batezbestekoa}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 = \text{erantzun bakoitzaren karratuen batura}$$

Idazkera horretan, puntuak batukariak eragina duen azpiindizea adierazten du.

Bariantza-analisia erabiliz kontrastatu nahi dugun hipotesia da populazioko batezbestekoen arteko berdintasuna. Eredua formalki idazteko, honako idazkera hau erabiliko dugu:

μ_i = *i.* mailan itxarotako erantzuna edo batezbesteko teorikoa (i = 1, 2, ..., k)
= *i.* populazioaren batezbestekoa (parametro ezezaguna)

μ = itxarotako erantzuna edo batezbesteko teorikoa, faktorearen maila kontuan hartu gabe
= *i* populazioak bakar baten bilduz lortutako batezbestekoa (parametro ezezaguna)

Beraz, faktorearen mailak erantzunean eraginik ez balu, μ_1, μ_2, ..., μ_k batezbestekoak berdinak eta μ-ren berdinak izango lirateke. Faktorearen mailak erantzunean eraginik balu, ordea, batezbestekoak ezberdinak izango lirateke, batzuk gutxienez. Ondorioz, *i.* mailako batezbestekoa eta batezbesteko osoaren arteko diferentzia $\mu_i - \mu$, faktorearen *i.* mailak erantzunean daukan eraginaren adierazlea da, baldin eta eragina badu. Eta nahiz eta *i.* populazioko ale bakoitzak tratamendu bera jasotzen duen, zorizkotasuna dela medio, erantzunak zerbait ezberdinak izango dira. Hau da, populazio bakoitzean zorizkotasunaren araberako aldakortasun naturala dago batezbestekoarekiko. Erantzun zehatz baterako, X_{ij}, aldakortasun horren adierazlea $X_{ij} - \mu_i$ da; diferentzia horri zorizko

errore deritzo. Aurreko guztia kontuan harturik, honako era honetara adieraziko da faktore bakar bateko bariantza-analisiaren eredua:

$$X_{ij} = \mu + (\mu_i - \mu) + (X_{ij} - \mu_i) \qquad \begin{array}{l} i = 1, 2, .., k \\ j = 1, 2, \ldots, n_i \end{array}$$

Eredu matematiko horren atzean honako ideia hau dago: erantzun bakoitza hiru osagaitan zatitu daiteke, modu honetara:

j. ale esperimentalaren erantzuna *i.* tratamenduari	=	Batezbesteko erantzun osoa	+	Batezbesteko osoarekiko desbideratzea *i.* tratamendua jasotzeagatik	+	*i.*batezbesteko arekiko zorizko desbideratzea
(X_{ij})		(μ)		$(\mu_i - \mu)$		$(X_{ij} - \mu_i)$

Aurreko gaietan ikusi dugun era berean, hipotesi-kontrastea gauzatzeko estatistiko bat aukeratu behar dugu. Estatistiko horren aukerak zentzuzkoa izan behar du; baina, hipotesi nulua egia balitz, alegia, *k* populazioetako batezbestekoak berdinak balira, bere probabilitate-banaketak ere ezaguna izan behar du. Estatistiko hori gauzatzea posible izateko, laginak atera diren populazioei buruzko hainbat baldintza ezarri behar dira; hain zuzen:

1. *k* populazio zehatzetan, $\mu_1, \mu_2, ..., \mu_k$ parametro ezezaguneko batezbestekoak dituen erantzunetik ateratako *k* laginak zorizko laginak dira.
2. Erantzunak banaketa normalari darraio *k* populazioetan.
3. Erantzunak bariantza berdina du *k* populazioetan; hain zuzen, σ^2.

Baldintza horiek seigarren gaian bi populazio askeen batezbestekoen konparazioa *t* banaketaren bidez egiteko ezarri diren baldintzen berdinak dira.

Orain arte, eredu teorikoa definitu dugu. Praktikan, batezbesteko teorikoak beren estimatzaileen bidez ordezkatzen dira, eta honako eredu hau lortuko dugu:

$$X_{ij} = \overline{X}_{..} + (\overline{X}_{i.} - \overline{X}_{..}) + (X_{ij} - \overline{X}_{i.}) \qquad \begin{array}{l} i = 1, 2, .., k \\ j = 1, 2, \ldots, n_i \end{array}$$

$(X_{ij} - \overline{X}_{i.})$ batugaiari hondar deritzo, eta ereduaren zorizkotasuna adierazten du.

Ereduaren adierazpen horretatik abiatuz, aldakortasunaren deskonposaketa deritzon adierazpena aterako da, hain zuzen:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{..})^2=\sum_{i=1}^{k}n_i(\overline{X}_{i\cdot}-\overline{X}_{..})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i\cdot})^2$$

Adierazpen horretan hiru batugai daude:

$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{..})^2$ = Batezbesteko osoarekiko behaketen desbideratze karratuen batura

= Datuen aldakortasun osoaren neurria
= Karratuen batura totala = *KBT*

$\sum_{i=1}^{k}n_i(\overline{X}_{i\cdot}-\overline{X}_{..})^2$ = Batezbesteko osoarekiko tratamendu edo maila bakoitzaren batezbestekoaren desbideratze karratuen batura haztatua
= Mailari edo tratamenduari dagokion aldakortasunaren neurria
= Tratamenduaren karratuen batura = *KB(Tr)*

$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i\cdot})^2$ = Tratamendu edo maila bakoitzaren batezbestekoarekiko behaketen desbideratze karratuen batura.
= Tratamendu berdinaren pean edo maila berean dauden aleen arteko zorizkotasunari dagokion aldakortasunaren neurria.
= Hondarraren edo erroreen karratuen batura = *EKB*

Idazkera hori erabiliz, modu honetan geratuko da aurreko berdintasuna:

$$KBT = KB(Tr) + EKB$$

Honako hipotesi-kontraste hau gauzatuko dugu:

H_0: $\mu_1 = \mu_2 = ... = \mu_k$
H_1: batezbestekoren bat ezberdina da

Hauxe da hipotesi-kontraste hori gauzatzeko erabiltzen den estatistikoa:

$$F_p = \frac{KB(Tr)/k-1}{EKB/N-k}$$

H_0 egia denean, estatistikoaren banaketa $k-1$ eta $N-k$ askatasun graduko F banaketa da, eta balio horri dagokion p-balioa $p = P(F_{k-1;\,N-k} > F_p)$ da.

Azkenik, α adierazgarritasun-maila zehaztuko dugu, eta honako era honetara arrazoituko dugu:

- $p \leq \alpha \Rightarrow p$-balioa zoriz agertzeko txikiegia dela uste dugu $\Rightarrow \alpha$ adierazgarritasun-mailaz H_0 hipotesia errefusatuko dugu.
- $p > \alpha \Rightarrow$ ezberdintasunak zoriz ager daitezke $\Rightarrow \alpha$ adierazgarritasun-mailaz ez dago H_0 errefusatzeko nabaritasunik.

Beraz, F_p-ren balio handien esanahia da faktoreak zorizko aldakortasunarekiko azaltzen duen aldakortasuna handia dela, eta, ondorioz, H_0 errefusatzera garamatza; aldiz, F_p-ren balio txikiak esan nahi du faktorearen bidez azaldutako aldakortasuna txikia dela eta, ondorioz, faktoreak eragin txikia (estatistikoki ez adierazgarria) duela erantzunean.

9.2 adibidea

Aurreko 9.1 adibidean azaldu den esperimentuan, 9.2 taulan agertzen diren datuak bildu dira. Gogora dezagun, hamasei asteko tratamenduaren ostean, aknearen lesio kopuruan pazienteak izan duen hobekuntza-portzentajea dela behaturiko erantzuna.

TRATAMENDUA					
I		II		III	
48,6	50,8	68,0	71,9	67,5	61,4
49,4	47,1	67,0	71,5	62,5	67,4
50,1	52,5	70,1	69,9	64,2	65,4
49,8	49,0	64,5	68,9	62,5	63,2
50,6	46,7	68,0	67,8	63,9	61,2
		68,3	68,9	64,8	60,5
				62,3	

9.2 taula. Pazienteak tratamenduarekin izandako hobekuntza-portzentajea.

Honako hauek dira behaturiko estatistikoen balioak:

$T_{1\cdot}$ = I tratamenduaren erantzunen batura = 48,6 + … + 46,7 = 494,6

$T_{2\cdot}$ = II tratamenduaren erantzunen batura = 68,0 + … + 68,9 = 824,8

$T_{3\cdot}$ = III tratamenduaren erantzunen batura = 67,5 + … + 60,5 = 826,8

$\overline{X}_{1\cdot} = \dfrac{T_{1\cdot}}{n_1}$ = I tratamenduaren batez besteko erantzuna = $\dfrac{494{,}6}{10} = 49{,}46$

$\overline{X}_{2\cdot} = \dfrac{T_{2\cdot}}{n_2}$ = II tratamenduaren batez besteko erantzuna = $\dfrac{824{,}8}{12} = 68{,}73$

$\overline{X}_{3\cdot} = \dfrac{T_{3\cdot}}{n_3}$ = III tratamenduaren batez besteko erantzuna = $\dfrac{826{,}8}{13} = 63{,}60$

$T_{\cdot\cdot}$ = Erantzunen batura osoa = 2146,2

$$\overline{X}_{..} = \frac{T_{..}}{N} = \text{batez besteko erantzun osoa} = \frac{2146,2}{35} = 61,32$$

Behaturiko batezbestekoek nolabaiteko ezberdintasunak erakusten dituzte batez besteko erantzunean, batez ere I tratamenduarenak beste biekiko (% 49,6 vs. % 68,73 eta % 63,60). Ikus dezagun behaturiko diferentzia horiek estatistikoki adierazgarriak diren. Bereziki, ea datu horiek erakusten duten batez besteko hobekuntza-portzentajea tratamenduaren menpekoa den (H_1); edo, alderantziz, batez besteko hobekuntza-portzentajea tratamenduarekiko askea den (H_0).

ANOVA deritzon 9.3 taulan azaldu dugu datu horiek erakusten duten aldakortasunaren deskonposaketa:

Iturria	Askatasun-graduak	KB	*F*	*P*
Tratamendua	2	2133,66	262,12	< 0,01
Errorea	32	130,30		
Totala	34	2263,96		

9.3 taula. Bariantza-analisiko taula nagusia.

Taula horrek, $p < 0,01$ denez, % 5eko adierazgarritasun-mailaz, batezbestekoen berdintasuna errefusatzera garamatza. Ondorioz, esan dezakegu aknearen lesio kopuruan pazienteak izan duen batez besteko hobekuntza-portzentajea tratamenduaren menpekoa dela, % 5eko adierazgarritasun-mailaz.

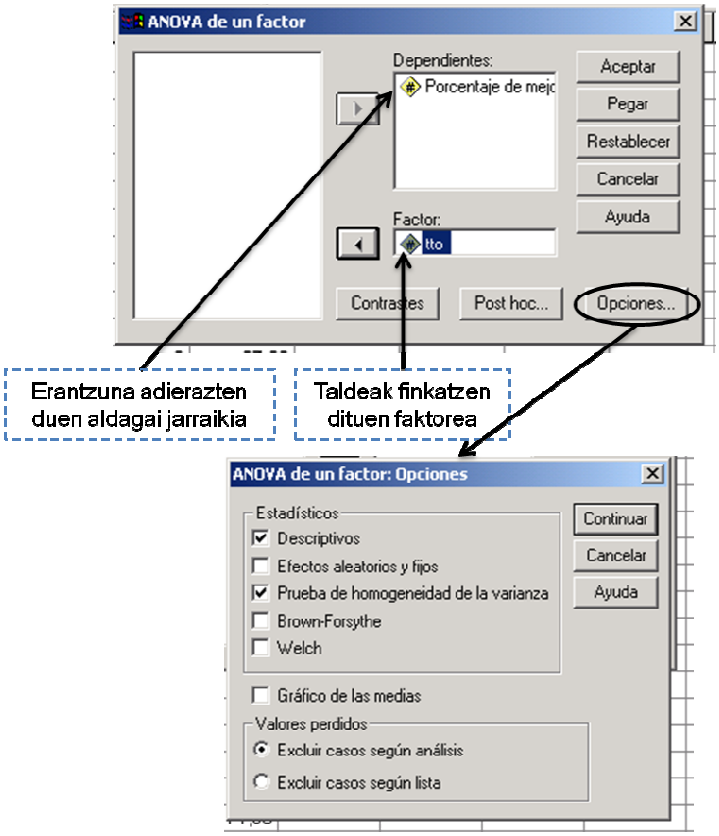
Merezi du bariantza-analisia erabiltzeko ezarri diren baldintzei buruzko hausnarketa txiki bat egitea. Erantzunaren banaketa normala eta bariantza berdinekoa suposatu da *k* populazioetan. Baldintza horiek egiaztatzeko metodoak daude. Normaltasuna kontrastatzeko, zortzigarren gaian garatu den doikuntza-egokitasuna deritzon teknika erabil daiteke. Bariantzen berdintasuna konparatzeko, seigarren gaian bi populaziotarako erabili den hipotesi-kontrastearen parekoa ere existitzen da, bi baino populazio gehiago dauden kasurako; apunte hauetan, ordea, ez dugu hipotesi-kontraste hori garatuko. Bariantza-analisia erabiltzerakoan kontuan izan behar den azken gauza da metodo estatistiko sendoa dela. Metodo estatistiko sendoa izateak esan nahi du baldintzak ez betetzearen ondorioak ez direla larriak. Batez ere, bariantzen berdintasuna edo erantzunaren normaltasuna ez betetzearen ondorioak ez dira larriak, baldin eta taldeen tamainak antzekoak badira. Beraz, komenigarria da taldeak pareko tamainakoak izatea; ez bakarrik metodoaren sendotasunagatik, baizik baita ondoren definituko diren

kontrasteak batezbestekoen arteko diferentziak daudenean non dauden jakiteko ere, tamaina bereko edo antzeko kasuetarako garatuak daudelako.

9.3 BARIANTZA-ANALISIA SPSS-REN BITARTEZ

Bariantza-analisiaren emaitzak lortzeko egin beharreko eragiketa kopurua nahiko handia da. Praktikan, emaitzak estatistikarako diseinaturiko *software*aren bidez lortzen dira; esate baterako, *SPSS* erabiliz, gure helburua izaten da ondorioak ateratzeko beharrezko emaitzak lokalizatzea eta interpretatzea.

Modu horretara, *SPSS*ren bitartez bariantza-analisia egiteko erabili behar den agindua da *>Analizar >Comparar medias >ANOVA de un factor*. Agertzen zaigun leihatilan adierazten dugu zein aldagai jarraitutan egin nahi den batezbestekoen diferentzia, eta zein den taldeak egiteko erabili behar den faktorea. Prozedura hori aurrera eramateko, nahitaezkoa da faktorea zenbakizko aldagai modura definituta egotea; bestela, aldagaiaren berkodetzea egin beharko da.



9.1 irudia. Bariantza-analisia *SPSS*rekin.

9.3 adibidea

Berriro 9.1 adibideko datuak erabiliz, ikus dezagun nola egingo genukeen analisia *SPSS* erabiliz, eta zein ondorio aterako genituzkeen. Analisia egiteko urratsak 9.1 irudian erakusten dira. Halaber, *Opciones* botoia sakatuz eskatzen den analisiari emaitza gehigarriak eska diezazkiokegu, 9.1 irudian erakusten den erara. Esate baterako, estatistiko deskribatzaileak eta bariantzen arteko berdintasunerako kontrastea eska daitezke. Leihatila horretan, batezbestekoen adierazpide grafikoa ere eska daiteke, eta faltako balioen azterketak zein izan behar duen agindu daiteke.

*SPSS*k eskaintzen dizkigun emaitzak honako taula honetan agertzen dira:

Descriptivos

Porcentaje de mejoría

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
1	10	49,4600	1,73218	,54776	48,2209	50,6991	46,70	52,50
2	12	68,7333	2,00061	,57753	67,4622	70,0045	64,50	71,90
3	13	63,6000	2,22224	,61634	62,2571	64,9429	60,50	67,50
Total	35	61,3200	8,16008	1,37931	58,5169	64,1231	46,70	71,90

9.4 taula. 9.3 adibideko datuen estatistika deskribatzailea.

Lehenengo taulan, 9.4 taulan, estatistiko deskribatzaile batzuk agertzen dira, tratamendu bakoitzerako eta denak batera. Ikusten denez, laginak 35 ale ditu, hiru tratamendutan banatuta; 10, 12 eta 13 gizaki bakoitzean, hurrenez, hurren. Tratamendu bakoitzerako, hobekuntza-portzentajearen batezbestekoa, % 95eko konfiantza-tartea, desbideratze estandarra, errore estandarra, baliorik baxuena eta altuena agertzen dira. Tratamendua kontuan hartu gabe, talde osorako balio horiek ere azken lerroan daude.

ANOVA

Porcentaje de mejoría

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	2133,665	2	1066,833	262,019	,000
Intra-grupos	130,291	32	4,072		
Total	2263,956	34			

9.5 taula. 9.3 adibideko datuen ANOVA taula.

Bigarrenik, ANOVA taula (9.5 taula) agertzen zaigu. Taula horretan, tratamenduen arteko batezbesteko hobekuntza-portzentajearen berdintasunerako hipotesi-kontrastearen emaitzak daude. Taula horren eta 9.2 adibidean eskuz egin denaren arteko berdintasuna ikus daiteke. Aurreko adibidean ikusi den era berean, $p < 0{,}0001$ denez, batezbestekoen berdintasuna errefusatuko dugu. Ondorioz, aknearen lesio kopuruan pazienteak izan duen batezbesteko hobekuntza-portzentajea tratamenduaren menpekoa dela esan dezakegu ($p < 0{,}0001$).

Prueba de homogeneidad de varianzas

Porcentaje de mejoría

Estadístico de Levene	gl1	gl2	Sig.
,491	2	32	,616

9.6 taula. 9.3 adibideko datuen bariantzen berdintasunerako hipotesi-kontrastea.

Azkenik, hirugarren taulan (9.6 taulan), bariantzen berdintasunaren hipotesi-kontrastearen emaitzak ikusten dira. Kasu honetan, $p = 0{,}616 > 0{,}05$ denez, bariantzen berdintasuna ez dugu errefusatzen eta, beraz, bariantza-analisia erabiltzea arrazoizkoa dela esan dezakegu.

9.4 KONPARAKETA ANITZAK

Demagun, behin populazioko *k* batezbestekoren arteko berdintasuna kontrastatzeko bariantza-analisia erabili eta gero, H_0 errefusatu egiten dela. Ondorioz, emaitzek erakusten dute *k* batezbestekoen artean ezberdintasun adierazgarriak daudela. Kasu honetan, datuen analisia hasi baino ez da egin, funtsezkoa baita ezberdintasun horiek zein populazioren artekoak diren zehaztasunez ezagutzeko ikertzen jarraitzea.

Behin *k* batezbestekoren arteko berdintasuna errefusatu denean, metodo ugari daude batezbesteko horien arteko diferentziak antzemateko. Hemen ez ditugu metodo horiek garatuko, ez eta guztiak deskribatuko ere; horietariko batzuk soilik aipatuko dira, irakurleak ideia orokorra izan dezan. Aipatzekoa da *SPSS*k horretariko hainbat erabiltzen dituela, eta horiek guztiak eskuliburuetan deskribaturik agertzen direla. Testu honetan, adibideen bidez eta *SPSS* erabiliz, metodo horietako batzuen erabilgarritasuna erakutsiko da.

Oro har, bi motatako kontraste anitzak daude. Heinen proben bitartez, elkarrekiko ezberdinak ez diren batezbestekoen talde homogeneoak izendatzen dira. Binakako konparaketa anitzek populazio bikote bakoitzeko batezbestekoen arteko diferentziak kontrastatzen dituzte. Bigarren metodo horien emaitza matrize erara adierazten den *p*-

balio mordoa da, eta marka baten bitartez α = 0,05eko adierazgarritasun-mailaz adierazgarriak diren diferentziak markatzen dira.

SPSS erabiliz eskuragarri ditugun heinen probak dira *Tukey b*, *S-N-K* (*Student-Newman-Keuls*), *Duncan*, *R-E-G-W* (*Ryan-Einot-Gabriel-Welsch*) eta *Waller-Duncan*. Konparaketa anitzak egiteko *SPSS*k eskaintzen dituen aukeretariko batzuk dira *Bonferroni*, *Tukey*ren diferentzia oneski adierazgarria, *Sidak*, *Gabriel*, *Hochberg*, *Dunnett* eta *Scheffé*. Azken horietariko batzuek (hain zuzen, *Tukey*, *Gabriel*, *Hochberg* eta *Scheffé* probek) bi gauzak egiten dituzte: heinen proba eta konparaketa anitzak. Azkenik, bariantzen berdintasuna errefusatzen denean, espreski erabiltzeko probak ere eskaintzen ditu *SPSS*k, hain zuzen *Tamhanen T2*, *Dunnett*en *T3* eta *C,* eta *Games-Howell*.

Praktikan, metodo bakoitzaren egokitasunean eta abantailetan oinarrituz, aldez aurretik finkatzea gomendatzen da ikertzaileak zein metodo erabili. Behin metodoa aukeratuta, beronek ematen dituen emaitzak solik interpretatuko dira. Emaitza adierazgarrien bilaketan murgilduz metodo bat eta bestea erabiltzea, baten batek emaitza adierazgarriak erakutsi arte, kontraste anizkoitzen filosofia beraren kontra doa. Hipotesi-kontraste ugari gauzatzen direnean, zoriaren fruitu soilik diren erlazio adierazgarriak aurkituko dira (% 5, hain zuzen). Kontraste anizkoitzen bidez, horixe ekiditen saiatuko gara. Baina, metodo bat baino gehiago erabiliz, arazo hori konpondu baino areagotu egiten da. Halaber, adibideren baten bidez erakutsiko dugu metodo bat baino gehiago aukeratzeak kontraesanera eraman gaitzakeen.

9.4.1 BONFERRONIren METODOA

Metodorik kontserbadoreena da; garatzen lehenengotarikoa izan zen eta oso logikoa dirudi. Oinarrian, bikote bakoitzerako kontrastea egiterakoan $\alpha = \alpha_T/m$ adierazgarritasun-maila aukeratzean datza, non m egin beharreko kontraste kopurua eta α_T osotasunean izan nahi den adierazgarritasun-maila baitira. Egin behar den kontraste kopurua, m, handia baldin bada, kontraste bakoitzerako eskatutako adierazgarritasun-maila oso txikia izango da, eta diferentziak antzematea ia-ia ezinezko bihurtuko da. Metodo hori oso ona da teorian, baina praktikan laguntza gutxikoa izaten da.

9.4.2 TUKEYren METODOA

*Tukey*k proposaturiko metodoaren aldaera eta zuzenketa dezente egin dira. Berau da taldeen arteko diferentziak antzematen metodorik ahaltsuenetarikoa. Hasieran, taldeen tamaina berdina denean soilik erabiltzeko garatu zen, eta horixe da bere eragozpenik handiena, nahiz eta, gaur egun, edozein kasutan erabil daitekeen orokortasuna ere garatu den.

9.4.3 SCHEFFÉren METODOA

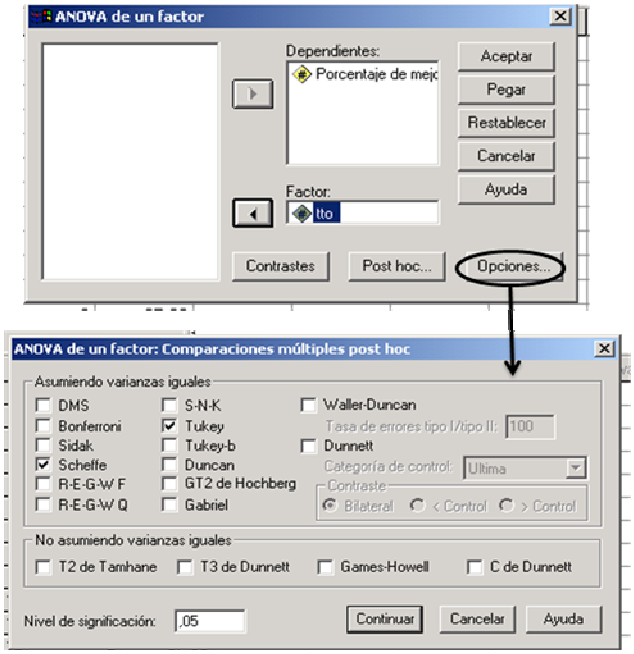
Metodo hori edozein kasutan erabil daiteke, taldeen tamainak berdinak izan zein ez, eta segur aski hedatuen dagoena eta gehien erabiltzen dena da. Kontuan hartu beharrekoa da metodo zeharo kontserbadorea dela, eta, beraz, taldeen tamainak berdin antzekoak direnean, *Tukey*ren metodoak baino *p*-balio handiagoak ematen dituela.

9.4.4 DUNCANen METODOA

Metodo hau zaharrena da eta, segur aski, zientzia-lanetan gehien aipatu izan dena. Hasieran, tamaina bereko taldeentzat diseinatu zen, baina *C.Y. Kramer*ek edozein kasutarako orokortu zuen urte batzuk geroago.

9.4 adibidea

Berriro 9.1 adibidean emandako datuetan oinarrituz, hamasei aste pasatu ondoren, batez besteko hobekuntza-portzentajea ezberdina dela hiru tratamenduentzat ondorioztatu dugu 9.3 adibidean. Behin hiru tratamenduen arteko berdintasuna batez besteko hobekuntza-portzentajean baztertuta, ezberdintasunak zein tratamenduren artean dauden jakin nahi dugu. Horri erantzuteko kontraste anitzak egin behar ditugu, eta hori bariantza-analisia egiterakoan agertzen zitzaigun leihatila nagusian *Post Hoc* botoia sakatuz lortuko dugu (9.2 irudia).



9.2 irudia. Bariantza-analisiko konparaketa anitzak *SPSS*rekin.

Leihatila berri horretan, konparaketa anitzak egiteko erabili nahi dugun metodoa aukera dezakegu. Ikusten dugunez, metodo ugari daude, eta zein aukeratu ikertzailearen menpe dago. Lehenengo irizpidea aukera egiterakoan bariantzen berdintasuna errefusatzearen menpe egongo da. Adibide honetan, bariantzen berdintasuna ez da errefusatu; beraz, lehenengo multzoan agertzen diren metodoetatik ikertzaileak nahi beste aukera ditzake. Hipotesi-kontraste horietarako nahi dugun adierazgarritasun-maila ere finkatzen da hemen; berez, % 5ean dago. Metodoak erakusteko helburuarekin, guk *Scheffé* eta *Tukey* metodoak aukeratu ditugu. Aurretik aipatu dugu metodo bakarra aukeratu beharko genukeela eta baten emaitzetan solik oinarritu beharko genituzkeela gure ondorioak; guk, ordea, bi aukeratu ditugu, ilustrazio modura. Baina, bi metodoek eskainitako emaitzen arteko kontraesanik existituko balitz, *Tukey*ren metodoaz baliatuko gara ondorioak ateratzeko. Hiru taldeetako tamainak berdin antzekoak izanik (10, 13 eta 12 dira), *Tukey*ren metodoa hartuko dugu onargarritzat. 9.7 taulan agertzen dira datu horiek sartu ondoren lortutako emaitzak.

Porcentaje de mejoría

	TTO	N	Subconjunto para alfa = .05		
			1	2	3
HSD de Tukey[a,b]	1	10	49,4600		
	3	13		63,6000	
	2	12			68,7333
	Sig.		1,000	1,000	1,000
Scheffé[a,b]	1	10	49,4600		
	3	13		63,6000	
	2	12			68,7333
	Sig.		1,000	1,000	1,000

Se muestran las medias para los grupos en los subconjuntos homogéneos.

a. Usa el tamaño muestral de la media armónica = 11,527.

b. Los tamaños de los grupos no son iguales. Se utilizará la media armónica de los tamaños de los grupos. Los niveles de error de tipo I no están garantizados.

9.7 taula. Konparaketa anitzak.

Bi metodoek, alegia, *Scheffé* eta *Tukey* metodoek hiru talde homogeneo sailkatzen dituzte. Hau da, hiru tratamenduen arteko diferentzia adierazgarriak antzematen dira batezbesteko hobekuntza-portzentajean, erabilitako metodoa edozein dela ere.

9.5 adibidea

Ospitale bateko Kirurgia Baskularreko Zerbitzuan ikerketa bat egin da, erretzeko ohiturak triglizerido-mailan duen eragina neurtzeko. Helburu horrekin, 105 pertsonaz osaturiko lagina hartu eta beren triglizerido-maila eta

erretzeko ohiturak batzen dira. Erretzeko ohitura neurtzeko, pertsona bakoitza erretzailea, erretzaile ohia edo ez erretzailea den biltzen da. Modu honetara eratuko dugu erretzearen menpe batez besteko triglizerido-mailan ezberdintasunik badagoen neurtzeko erabili behar dugun hipotesi-kontrastea:

H_0: $\mu_{\text{ERRETZAILEA}} = \mu_{\text{ERRETZAILE OHIA}} = \mu_{\text{EZ ERRETZAILEA}}$
H_1: batezbestekoak ezberdinak dira

Ikus ditzagun *SPSS*ren bidez lortutako emaitzak, hipotesi-kontrastea egiteko bariantza-analisia erabili ondoren.

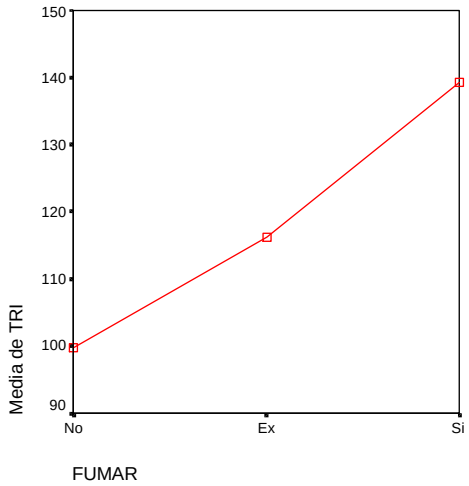
Descriptivos

TRI

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
No	29	99,72	50,398	9,359	80,55	118,89	33	255
Ex	20	116,25	40,214	8,992	97,43	135,07	49	208
Si	56	139,20	74,044	9,895	119,37	159,03	40	418
Total	105	123,92	64,645	6,309	111,41	136,43	33	418

9.8 taula. 9.5 adibideko estatistika deskribatzailea.

Lehendabizi daukagun 9.8 taula deskribatzaileak tamainari dagokionez hiru talde nahiko ez-orekatu erakusten dizkigu: 29 ez-erretzaile, 20 erretzaile ohi, eta 56 erretzaile. Batez besteko triglizerido-mailaren arabera ere, ezberdintasunek argiak dirudite: 99,7 ez-erretzaileak, 116,25 erretzaile ohiak eta 139,2 erretzaileak. Datu horien adierazpide grafikoari begiratuz, 9.3 irudian, batez besteko triglizerido-mailak goranzko joera daukala dirudi, erretzearen menpe.



9.3 irudia. Batez besteko triglizerido-maila, erretze-ohituraren arabera.

ANOVA

TRI

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	31223,008	2	15611,504	3,948	,022
Intra-grupos	403384,382	102	3954,749		
Total	434607,390	104			

9.9 taula. 9.5 adibideko ANOVA taula.

Bigarrenik, 9.9 taulan, ANOVA taulak batezbestekoen berdintasunerako hipotesi-kontrastearen emaitzak erakusten dizkigu. Ikusten denez, $p = 0{,}022 < 0{,}05$ eta, beraz, batezbestekoen arteko berdintasuna errefusatu egiten dugu. Ondorioz, batez besteko triglizerido-maila erretzeko ohiturarekiko askea dela errefusatu egiten dugu eta, aldiz, erretzeko ohiturak batez besteko triglizerido-mailan daukan eragina estatistikoki adierazgarria dela esango dugu, % 5eko adierazgarritasun-mailaz.

Prueba de homogeneidad de varianzas

TRI

Estadístico de Levene	gl1	gl2	Sig.
1,883	2	102	,157

9.10 taula. 9.5 adibideko datuen bariantzen homogeneotasun-proba.

9.10 taulan ikus dezakegu, baita ere, bariantzen arteko berdintasuna ez dela errefusatzen ($p = 0{,}157$), eta horrek bariantza-analisiaren egokitasuna bermatzen duela. Emaitza hori ikusita, *Scheffé* eta *Tukey* probak eskatu dizkiogu, ezberdintasunak zein talderen artean dauden kontrastatzeko. Berriro ere, bi metodoak ilustrazio modura erabiliko dira, baina biek eskainitako emaitzen arteko kontraesanik existituko balitz, *Scheffé*ren metodoaz baliatuko gara ondorioak ateratzeko, hiru talderen tamainak zeharo ezberdinak direlako (alegia, 29, 20 eta 56).

Binakako konparaketen emaitzak begiratuz 9.11 taulan, bi metodoek —*Scheffé* eta *Tukey* metodoek, alegia— erakusten dituzten emaitzek ondorio berdinetara garamatzate. Batez besteko triglizerido-mailaren ezberdintasunak adierazgarriak dira erretzaileen eta ez-erretzaileen artean (izartxo baten bidez taulan markatuak). Ordea, erretzaile ohien batez besteko triglizerido-mailak ez dauka ezberdintasun adierazgarririk ez ez-erretzaileenarekin ez erretzaileenarekin.

Comparaciones múltiples

Variable dependiente: TRI

	(I) FUMAR	(J) FUMAR	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95% Límite inferior	Límite superior
HSD de Tukey	No	Ex	-16,53	18,279	,639	-60,00	26,95
		Si	-39,47*	14,387	,019	-73,69	-5,25
	Ex	No	16,53	18,279	,639	-26,95	60,00
		Si	-22,95	16,382	,344	-61,91	16,02
	Si	No	39,47*	14,387	,019	5,25	73,69
		Ex	22,95	16,382	,344	-16,02	61,91
Scheffé	No	Ex	-16,53	18,279	,666	-61,93	28,88
		Si	-39,47*	14,387	,026	-75,21	-3,73
	Ex	No	16,53	18,279	,666	-28,88	61,93
		Si	-22,95	16,382	,378	-63,64	17,75
	Si	No	39,47*	14,387	,026	3,73	75,21
		Ex	22,95	16,382	,378	-17,75	63,64

*. La diferencia entre las medias es significativa al nivel .05.

9.11 taula. 9.5 adibideko datuen konparaketa anitzak.

Bi metodoen arteko ezberdintasun bakarra *p*-balioari dagokio. *Tukey*ren metodoaren bidez $p = 0{,}019$, eta *Scheffé*ren metodoaren bidez $p = 0{,}026$ dira. Detaile hori aurretik ere aipatu dugu; *Scheffé*ren metodoa kontserbadorea da eta, ondorioz, eskaintzen duen *p*-balioa bestea baino handiagoa da. Baina, taldeen tamainak hain ezberdinak izanik, kasu honetan *Scheffé*ren metodoa erabiliz lortutako emaitzek *Tukey*ren metodoarekin lortutakoek baino fidagarritasun handiagoa dute.

TRI

	FUMAR	N	Subconjunto para alfa = .05 1	2
HSD de Tukey[a,b]	No	29	99,72	
	Ex	20	116,25	116,25
	Si	56		139,20
	Sig.		,575	,346
Scheffé[a,b]	No	29	99,72	
	Ex	20	116,25	
	Si	56	139,20	
	Sig.		,060	

Se muestran las medias para los grupos en los subconjuntos homogéneos.

a. Usa el tamaño muestral de la media armónica = 29,314.

b. Los tamaños de los grupos no son iguales. Se utilizará la media armónica de los tamaños de los grupos. Los niveles de error de tipo I no están garantizados.

9.12 taula. 9.5 adibideko datuen heinen probak.

Heinen probaren emaitzak begiratzen baditugu, ordea (9.12 taula), ondorioak ezberdinak dira analisia egiteko aukeratzen dugun metodoaren arabera. *Tukey*ren metodoak bi talde homogeneo antzematen ditu: bata erretzaileek ordezkatzen dute, eta bestea ez-erretzaileak; erretzaile ohiak, berriz, batean zein bestean egon daitezke. Beraz, emaitza horietan oinarrituz bi modalitateko aldagaia eraiki nahi izango bagenu erretze-ohitura adierazteko, erretzaile ohiak edozein taldetan sartu ahal izango genituzke, erretzaileekin edo ez erretzaileekin. *Scheffé*ren metodoak talde homogeneo bakarra antzematen du; alegia, ez ditu erretze-ohituraren menpeko diferentzia adierazgarriak antzematen batez besteko triglizerido-mailan ($p = 0,06$).

Bi metodoen emaitzen arteko diferentzia beren oinarrian kokatzen da. Jadanik aipatu dugu p-balio ezberdinak lortzen dituztela ere, metodo ezberdinetan oinarriturik, eta baliteke, adibide horretan ikusten dugun moduan, adierazgarritasun-mailatik gertu dauden p-balioak izaterakoan bata txikiagoa eta bestea handiagoa izatea. Datuak aztertu aurretik finkatu dugun irizpidea jarraituz, taldeen tamainak hain ezberdinak izanik, *Scheffé*ren metodoak eskainitako emaitzak erabiliko ditugu ondorioak ateratzeko.

Edozein kasutan, *Scheffé*ren metodoari jarraituz kontraesana agertzen zaigu. Lehendabizi, bariantza-analisiak dio erretzeko ohiturak batez besteko triglizerido-mailan daukan eragina estatistikoki adierazgarria dela ($p = 0,022$). Gero, binakako hipotesi-kontrasteek diote diferentzia adierazgarriak erretzaileen eta ez-erretzaileen artean daudela ($p = 0,026$). Baina, heinen probak talde homogeneo bakarra eraikitzen du ($p = 0,06$). Horren arrazoia egindako probak ezberdinak izatean datza; hain zuzen, binakako ezberdintasunak lortzea eta talde homogeneoak egitea ez dira metodo baliagarriak. Bi taldeen arteko diferentzia adierazgarriak antzematea «errazagoa» da talde homogeneoak egitea baino. Kasu honetan, erretzaileen eta ez-erretzaileen arteko diferentziak ez dira oso adierazgarriak, $p = 0,026$ baita, eta eskakizun-maila apur bat gehitzen denean, p-balio hori $p = 0,06$ ez-adierazgarri izatera pasatzen da. Horrelako kasuetan, bereziki kontuan eduki behar da hasierako helburua. Hain zuzen, helburua talde homogeneoak egitea bada, talde bakarra utzi behar dugu; helburua binaka diferentziak antzematea bada, ordea, diferentzia adierazgarriak daude erretzaileen eta ez-erretzaileen artean. Laginaren tamaina da horrelako kontraesanen aurrean kontuan eduki behar den beste gauza bat; zenbat eta lagin tamaina handiagoa, orduan eta kontrastea ahaltsuagoa izango da eta, beraz, p-balioak txikiagoak izango dira. Diferentziak, laginean egoteaz gainera, populazioan ere egotekotan, lagin tamaina handiagoak diferentzia adierazgarriak erakutsiko lituzke.

9.5 ARIKETAK

9.1. Nahaste baten tenperatura lau termometro erabiliz neurtu da, eta honako datu hauek izan dira:

Termometroak			
1	2	3	4
63	64	58	61
63	64	59	61
62	63	59	62
65	64	68	60
66	65		63

a) Onar al daiteke lau termometroek neurtu duten batez besteko tenperatura berdina den hipotesia?

b) Bazter ezazu hirugarren termometroa eta kalkula ezazu ANOVA taula gainerako hirurentzat. Zer ondorio atera dezakegu?

9.2. Erretzen dituen hiru marka desberdineko zigarroen batez besteko nikotina-edukiaz oso kezkaturik dagoen erretzaile batek zoriz aukeraturiko zigarroen azterketa gauzatu du. Honako hauek izan ziren mg-tan adierazitako nikotina-edukiak:

Marka	Nikotina-edukiak (mg-tan)									
A	28	20	26	30	32	34	25	23		
B	29	31	28	30	24	26	25	25	28	28
C	32	30	28	30	25	27	30	26	24	22

Nikotina-edukia banaketa normalari darraiola eta hiru populazioen bariantza berdina dela suposatzen baldin bada, zer ondorio atera daiteke emaitza horietatik?

Iradokizuna: $KB(X) = 0{,}129$ eta $EKB = 292{,}3$.

9.3. Hiru analgesikoren artean desberdintasun adierazgarririk dagoen aztertu nahi dugu. Horretarako, buruko mina pairatzen duten pertsonengandik hiru lagin aukeratu ditugu, eta analgesiko bakoitzaren dosi egokia eman diegu, eragin lasaigarriaren iraupen-aldiak (ordutan) neurturik. Honako taula honetan azaltzen dira lorturiko emaitzak:

Analgesikoa		
A	*B*	*C*
2,8	3,6	2,8
3,4	4,1	1,9
3,1	3,9	2,7
4,2	4,3	3,1
2,9		

a) Eragin lasaigarriaren iraupen-aldiaren normaltasuna eta hiru populazioen bariantza berdina suposatuz, kontrasta ezazu, % 5eko adierazgarritasun-mailaz, minaren batez besteko inhibizio-aldien arteko berdintasunaren hipotesia.

Iradokizuna: $KB(X) = 3{,}646$ eta $EKB = 2{,}323$.

b) *SPSS*tik ateratako honako emaitza hauek erabiliz, zein ondorio atera dezakezu?

Comparaciones múltiples

Variable dependiente: ALDIA

Scheffé

(I) ANALGESI	(J) ANALGESI	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95% Límite inferior	Límite superior
1	2	-,6950	,323	,150	-1,6212	,2312
	3	,6550	,323	,179	-,2712	1,5812
2	1	,6950	,323	,150	-,2312	1,6212
	3	1,3500*	,341	,009	,3737	2,3263
3	1	-,6550	,323	,179	-1,5812	,2712
	2	-1,3500*	,341	,009	-2,3263	-,3737

*. La diferencia entre las medias es significativa al nivel .05.

9.4. Itsas artropodoen lau espezietarako aminoazido askeen edukia (µmol/g-tan, pisu sikua) zehaztu zen. Honako hauek dira datuak:

1. espeziea	2. espeziea	3. espeziea	4. espeziea
431,1	477,1	385,5	366,8
440,2	479,0	387,9	369,8
443,2	481,3	389,6	371,4
445,5	487,8	391,4	373,2
448,6	489,6	399,1	377,2
451,2	474,0	403,7	379,4
442,3	489,3	394,5	381,3

a) Aminoazido askeen edukiaren normaltasuna eta lau espezie-populazioen bariantza berdina suposatuz, kontrasta ezazu lau artropodo-espezietarako aminoazido askeen batez besteko edukia berdina den hipotesia. Har ezazu $\alpha = 0{,}05$.

Iradokizuna: *KB(X)* = 50.653,07 eta *EKB* = 914,10.

b) *SPSS*tik ateratako honako emaitza hauek erabiliz, zer ondorio atera dezakezu?

Comparaciones múltiples

Variable dependiente: EDUKINA
HSD de Tukey

(I) ESPEZIE	(J) ESPEZIE	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95% Límite inferior	Límite superior
1	2	-39,4286*	3,299	,000	-48,5288	-30,3283
	3	50,0571*	3,299	,000	40,9569	59,1574
	4	69,0000*	3,299	,000	59,8998	78,1002
2	1	39,4286*	3,299	,000	30,3283	48,5288
	3	89,4857*	3,299	,000	80,3855	98,5860
	4	108,4286*	3,299	,000	99,3283	117,5288
3	1	-50,0571*	3,299	,000	-59,1574	-40,9569
	2	-89,4857*	3,299	,000	-98,5860	-80,3855
	4	18,9429*	3,299	,000	9,8426	28,0431
4	1	-69,0000*	3,299	,000	-78,1002	-59,8998
	2	-108,4286*	3,299	,000	-117,5288	-99,3283
	3	-18,9429*	3,299	,000	-28,0431	-9,8426

*. La diferencia entre las medias es significativa al nivel .05.

9.5. Higadurak espora kopuruan daukan eragina larreen altitudearen menpekoa dela baieztatu dute hainbat ikertzailek. Hipotesi hori kontrastatzeko, itsasoaren mailatik altitude ezberdinera higaturiko lurretako laginak hartu dira, eta

haietako espora kopurua neurtu da: 200 metrotik behera, 200 eta 500 metro bitartean, eta 500 metrotik gora. Honako taula hauetan biltzen dira egindako analisi estatistikotik lortutako emaitzak:

100 gramo lurreko espora kopurua

	N	Media	Desviación típica	Error típico
500 metrotik gora	10	610,40	66,076	20,895
200 eta 500 metro bitartean	10	624,20	95,742	30,276
200 metrotik behera	10	664,90	120,410	38,077
Total	30	633,17	96,193	17,562

100 gramo lurreko espora kopurua

	Suma de cuadrados	Gl	Media cuadrática	F	Sig.
Inter-grupos	16057,267	2	8028,633	,859	,435
Intra-grupos	252280,900	27	9343,737		
Total	268338,167	29			

Datu horiek baieztatu al dute higadurak batez besteko espora kopuruan daukan eragina larreen altitudearen menpekoa dela, $\alpha = 0{,}05$ adierazgarritasun-mailarako? Idatz ezazu galdera zientifikoari erantzuteko egin beharreko hipotesi-kontrastea; adieraz itzazu estatistikoaren balioa eta *p*-balioa, eta interpreta ezazu emaitza erantzuna arrazoituz.

9.6. Herrialde bateko prezipitazioen jaitsiera kezkagarria izan da azken urtean. Neurririk hartu baino lehen, herrialdeko lau eskualdeetan jaitsiera homogeneoa izan den jakiteko ikerketa egitea erabaki da. Horretarako, bost meteorologia-estazio aukeratu dira zoriz eskualde bakoitzean, eta prezipitazioen jaitsiera-portzentajeak kalkulatu dira. Honako taula honetan agertzen da datu horien azterketa estatistikotik lortutako emaitzen zati bat.

	Karratuen batura	Askatasun graduak	Karratuen batezbestekoa	Estatistikoa	*p*
Eredua		3		3,03781	
Errorea	78,396				
Totala					

a) Osa ezazu aurreko ANOVA taula.

b) Zein da hondar-bariantzaren balioa?

c) Lau herrialdeetako batez besteko jaitsiera-portzentajeak berdinak al dira % 5eko adierazgarritasun-mailaz?

9.7. Demagun ikasketa-mailak bizi-kalitatean daukan eragina neurtzeko 193 tamainako lagina aztertzen dugula. Bizi-kalitatea 0 eta 100 bitarteko neurri jarraikia da, eta balio altuak bizi-kalitate hobea adierazten du. Ikasketa-maila hiru modalitateko aldagai diskretua da: 1 = ikasketarik ez edo lehen mailako ikasketak; 2 = bigarren mailako ikasketak; 3 = goi-mailako ikasketak. Datu horiekin bariantza-analisia egiten da.

a) Osa ezazu ANOVA taula honetan falta den informazioa:

	Askatasun-graduak	Karratuen batura	F estatistikoa	*p*-balioa
Eredua		3.975,73		
Errorea				
Totala		**83.961,99**		

b) Kalkula ezazu hondar-bariantza.

c) Egin ezazu bariantza-analisiaren hipotesi-kontrastea α = 0,05eko adierazgarritasun-mailarako. Zer ondorio aterako zenuke ikasketa-mailak bizi-kalitatean daukan eraginari buruz?

d) Scheffé-ren metodoa erabiliz egindako binakako konparazioek honako emaitza hauek eman dizkigute:

Konparazioak	Batezbestekoen arteko diferentzia	% 95eko baterako konfiantza-tarteak
3 – 2	5,832	**(-0,383 , 12,047)**
3 – 1	6,445	**(1,264 , 11,626)**
2 – 3	-5,832	**(-12,047 , 0,383)**
2 – 1	0,613	**(-4,084 , 5,311)**
1 – 3	-6,445	**(-11,626 , -1,264)**
1 - 2	**-0,613**	**(-5,311 , 4,084)**

Batez besteko bizi-kalitatean ikasketa-mailarekiko diferentziak egotekotan, zer ikasketa-mailen artean izango lirateke adierazgarriak diferentzia horiek? (α = 0,05).

e) Kasu horretan bariantza-analisia metodo onargarria izateko, zer hipotesi bete beharko lirateke?

9.8. Sufre-edukiari buruzko ikerketa egin da bost ikatz-meatokitan. Horretarako, sufre-edukiaren 7, 8, 9, 8 eta 10 tamainako laginak hartu dira, bost meatokietan. Honako datu hauek lortu dira, besteak beste:

KBT = 8,432 (Karratuen batura totala)

KB(F) = 3,935 (Meatokiari dagokion karratuen batura)

a) Sufre-edukiaren banaketa normala dela eta meatokietan zehar berdinki sakabanaturik dagoela suposatzen da. Onargarria al da batez besteko sufre-edukia berdina dela meatoki guztietan, % 5eko adierazgarritasun-mailaz? Arrazoitu ezazu erantzuna, egin beharreko hipotesi-kontrasteak eginik.
b) Bat eta lau meatokien arteko batez besteko sufre-edukien diferentziaren % 95eko konfiantza-tartea (0,195 , 1,365) dela jakinik, zein ondorio atera dezakezu?

9.9. Odoleko minbiziaren diagnosia izan ondoren, hasierako egoera fisikoaren arabera, biziraupen-denbora aztertu nahi dugu, 280 gaixok osaturiko laginean oinarrituta. Hasierako egoera fisikoa zenbait proba klinikoren bitartez zehaztu da, paziente bakoitza honako talde hauetakoren batean sailkatuz: mugarik gabe, muga arinak, muga ertainak eta muga larriak.

Odoleko minbiziaren diagnosia izan ondoren, biziraupen-denbora banaketa normalari darraiola onartzen dugu. Honako hauek dira talde bakoitzeko batez besteko biziraupen-denborak: mugarik gabe 1.028 egun (n_1 = 71), muga arinak 1.037 egun (n_2 = 80), muga ertainak 845 egun (n_3 = 76) eta muga larriak 788 egun (n_4 = 53).

a) Osa ezazu honako *SPSS* taula hau:

ANOVA

		Suma de cuadrados	gl	F	Sig.
Superbizipen denbora	Inter-grupos	3202961,5			,031
	Intra-grupos		276		
	Total	1,02E+08			

b) Emandako datuen arabera, zein azterketa estatistiko erabiliko zenuke batez besteko biziraupen-denboran hasierako egoera fisikoaren menpeko desberdintasunak detektatzeko? Egin ezazu planteaturiko hipotesi-kontrastea % 5eko adierazgarritasun-mailaz.

c) Aurreko kontrastea egin ondoren, zein beste proba estatistiko osagarri egingo zenuke?

9.10. Herrialde batean, berunik gabeko gasolinaren litroko prezioari buruzko ikerketa egin nahi dute merkataritza-guneetan (I), hiriguneetan (II) eta industrialdeetan (III) kokatuta dauden gasolindegietan. Ikerketa hori egiteko, lehen aipatutako eskualdeetatik zoriz gasolindegi batzuk aukeratu dira, eta berunik gabeko

gasolinaren litroko eta asteko batez besteko prezioa idatzi da, eurotan. Ikerketa estatistikoak eginez, normaltasuna suposatuz, honako emaitza hauek lortu dira:

PRECIO

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	2.383E-03	2	1.192E-03	6.067	.014
Intra-grupos	2.554E-03	13	1.964E-04		
Total	4.937E-03	15			

Variable dependiente: PRECIO

Tamhane

(I) ZONA	(J) ZONA	Diferencia de medias (I-J)	Error típico	Intervalo de confianza al 97.5%	
				Límite inferior	Límite superior
1.00	1.00				
	2.00	-3.0004E-02*	8.785E-03	-4.1720E-02	-1.82883E-02
	3.00	-1.6082E-02	8.206E-03	-6.3893E-02	3.17285E-02
2.00	1.00	3.00043E-02*	8.785E-03	1.82883E-02	4.17203E-02
	2.00				
	3.00	1.39222E-02	9.402E-03	-3.4920E-02	6.27648E-02
3.00	1.00	1.60821E-02	8.206E-03	-3.1729E-02	6.38927E-02
	2.00	-1.3922E-02	9.402E-03	-6.2765E-02	3.49205E-02
	3.00				

*. La diferencia entre las medias es significativa al nivel .025.

a) Ba al dago, % 5eko adierazgarritasun-mailaz, hiru eskualdetako gasolindegietan, berunik gabeko gasolinaren litroko batez besteko prezioen arteko desberdintasun adierazgarririk?

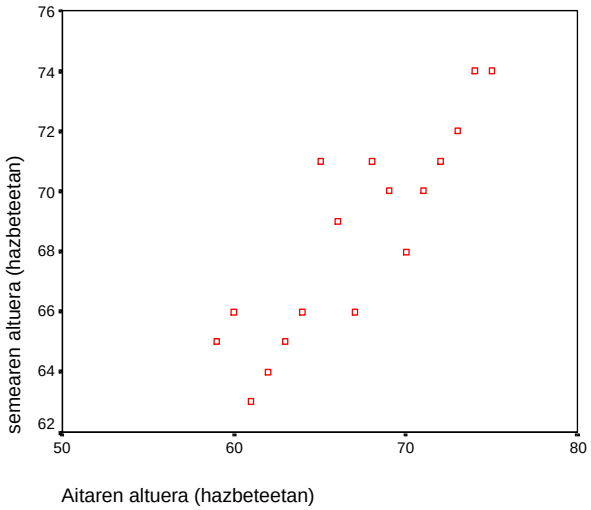
b) Zein da eskualderik merkeena?

10 ERREGRESIO LINEALA

10.1 SARRERA

Bi aldagai jarraituren arteko erlazio linealaren indarra neurtzean datza korrelazioa. Erregresioaren helburua, ordea, aldagai ezberdinek hartzen dituzten balioetatik abiatuta beste zorizko aldagai jarraitu baten itxarotako balioak aurresateko ekuazio bat lortzea da. Ekuazioaren bidez azaldu edo aurresan nahi den zorizko aldagaiari menpeko aldagai edo erantzun deritzo, eta hori azaltzeko erabiltzen diren aldagaiei askeak, azaltzaileak edo koaldagaiak deritze. Erregresio eta korrelazio kontzeptuek lotura estua daukate, bereziki erregresio lineal sinpleaz mintzatzen garenean, hau da, aldagai aske bakarra daukagunean.

Sir Francis Galton jaunak sartu zuen XIX. mendean erregresio lineal sinplea teknika estatistiko gisa. Semearen eta aitaren altueren arteko erlazioan interesaturik zegoen *Galton*. 10.1 irudiko puntu-hodeian ikusten den gisara, gizon altuek seme altuak izateko joera dute, semeen batez besteko altuera gurasoena baino txikiagoa den arren.



10.1 irudia. *Galton*en datuak, semearen eta aitaren altueren arteko erlazioa.

Honako hau da *Galton*ek erabilitako ekuazioa:

$$Semearen\ altuera = 28{,}63 + 0{,}595 \cdot Aitaren\ altuera$$

Gertakari horri **erregresio** deritzo *Galton*ek. Ordutik hona, izen hori izango da erantzun-aldagaia (adibide horretan, semearen altuera) koaldagai baten edo gehiagoren bidez (adibide horretan, aitaren altuera) aurresateko doitzen den eredurako erabiliko den izena.

*Galton*ek behaturiko erregresio deritzon gertakari horrek segidako ataletan eztabaidatuko diren ondorio garrantzizkoak dakartza lorturiko emaitzak azaltzerakoan.

Erregresio-ereduan zorizko aldagai baten interesatuak gaude gehienbat, Y (semearen altuera). Aldagai horren balioa beste aldagai batek edo batzuek eragin dutela pentsa dezakegu (aitaren altuerak). Gure intereseko Y aldagaiari menpeko aldagai edo erantzun-aldagai deritzo, eta Y aldagaian eragina daukaten beste aldagaiei askeak, azaltzaileak edo koaldagaiak deritze. Estimazioak edo predikzioak egiterakoan, aldagai askeak ez dira zorizkotzat hartzen. Balio ezberdinak hartzen dituzten ezaugarriak dira horiek, baina predikzioa egiterakoan balioa horiek ez dira zoriz hartutakoak. Esate baterako, *Galton*en datuetan oinarrituz berriro, semearen altuera neurri batean aitaren altueraren menpekoa izan daitekeenez, bi aldagai ari gara kontuan hartzen: X, aitaren altuera, eta Y, semearen altuera. Helburua ez da aitaren altueraz ezer aurresatea, baizik eta semearen altuera aurresatea; aitaren altuera aurretik zehaztasun osoz ezagutzen dela suposatzen da. Adibide horretan, semearen altuera erantzun-aldagai izango litzateke, eta aitaren altuera kontuan hartu dugun aldagai azaltzaile bakarra.

Has gaitezen erregresio lineala erabiliz deskriba daitekeen ikerketa aurkeztuz.

10.1 adibidea

Diabetikoei beren odoleko glukosa-maila neurtzen irakasteak abantaila ikaragarriak ditu. Ohiko prozedura baino errazagoa den teknika probatzeko ikerketa egin da. Teknika horrek oxidasa-entzima erabili du hagatxo batean. Hagatxoak bi kolore garatzen ditu aldi berean, eta kolore horiek glukosa-maila markatzen duen txartelarekin konparatu behar dira begirada batean. Teknika hori zehatza dela erakutsi ahal izanez gero, bere erabilera merkaturatuko litzateke. Helburu horrekin, paziente diabetikoaren odoleko glukosa-maila neurtzen da bi metodoekin, hagatxoarekin eta laborategian, ohiko litroko milimoletan.

Adibide honetan, hagatxoaren bidez neurtutako pazientearen odoleko glukosa-maila aldagai azaltzailea izango da; laborategian neurtutako pazientearen odoleko glukosa-maila, ordea, menpeko aldagaia.

Gai honetan erabiliko dugun idazkera ezarriko dugu orain. Y aldagaia X-ren balioen menpeko moduan erabiliko dugu. Beraz, finkaturiko x-rako aldagai baldintzatuaz mintzatzen ari garela onartuko dugu $Y|x$ (Y non $X = x$ baita). Zorizko aldagai baldintzatuaren itxaropena edo batez bestekoa $\mu_{Y|x}$ adieraziko dugu, eta logikoa dirudi $\mu_{Y|x}$ x-ren funtzioa dela pentsatzea. Funtzio horren adierazpen grafikoari, hain zuzen, X gaineko Y-ren erregresio zuzena deritzo.

Planteatzen den arazoa X aldagai azaltzailearen behaturiko balio batzuen menpe, x_1, x_2, ..., x_n, $\mu_{Y|x}$ funtzioaren itxura ikertzea da. Helburu nagusia da erregresio-ekuazio sinple baina arrazoizkoa aurkitzea.

10.2 DATUEN DESKRIBAPENA

Erregresio linealaren ereduan behaturiko datuak n neurketa pare dira, (y_i, x_i) non, $i = 1, 2, ..., n$ baita. Hots, y_i-k erantzun-aldagaiaren balioa eta x_i-k koaldagaiaren balioa adierazten dituzte; n laginaren tamaina da eta i azpindizeak lagineko ale bakoitza adierazten du.

Erregresio-ereduaren analisian, garrantzizkoa da X koaldagaiarekiko Y aldagaiaren puntu-hodeia eraikitzea. Ordenatu-ardatzean (bertikalean) Y erantzun aldagaiaren balioak eta abzisa-ardatzean (horizontalean) dagozkion X argibide-aldagaiaren balioak adierazten dituen grafikoa da hori. Irudikapen erraz horrek aldagai bien arteko erlazioaren ideia ematen du. *Software*aren aurrerapena kontuan izanik, nahitaezko bihurtzen da, inolako analisirik egiten ausartu baino lehen, grafiko arrunt hori egitea. Behin puntu-hodeia adierazten duen grafikoa egin eta gero, puntuan gutxi gorabehera zuzen bat erakusten dutela ikusten baldin bada, arrazoizkoa dela onartuko dugu puntu horiek doitzeko zuzen bat erabiltzea eta, beraz, metodo onargarritzat hartuko dugu erregresio lineala.

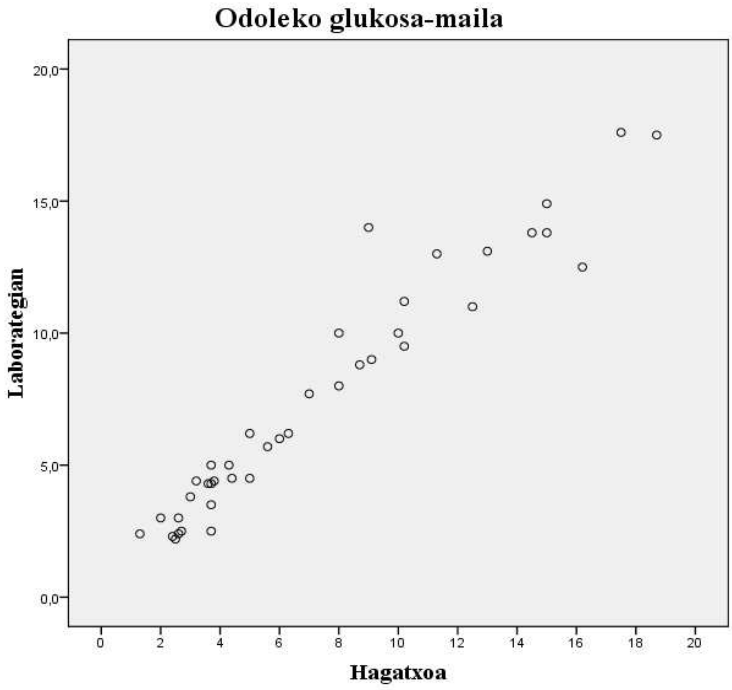
10.2 adibidea

Aurreko 10.1 adibidean deskribaturiko esperimentua egin ondoren, honako 10.1 taulako datuak lortu ziren. Gogora dezagun, 10.1 adibidean, pazientearen odoleko glukosa-maila neurtzen dela, hagatxoa erabiliz (X) eta laborategian (Y), litroko milimoletan.

x	*y*	*x*	*Y*	*X*	*Y*	*x*	*y*
1,3	2,4	3,2	4,4	7,0	7,7	15,0	14,9
2,0	3,0	3,6	4,3	8,0	8,0	15,0	13,8
2,4	2,3	3,7	4,3	8,0	10,0	17,5	17,6
2,6	3,0	3,7	5,0	10,0	10,0	18,7	17,5
2,5	2,2	3,8	4,4	10,2	9,5	6,0	6,0
2,6	2,4	4,4	4,5	10,2	11,2	8,7	8,8
2,7	2,5	4,3	5,0	12,5	11,0	5,6	5,7
3,0	3,8	5,0	4,5	11,3	13,0	9,1	9,0
3,7	2,5	5,0	6,2	13,0	13,1	16,2	12,5
3,7	3,5	6,3	6,2	14,5	13,8	9,0	14,0

10.1 taula. 10.1 adibideko datuak

Hurrengo grafikoak biltzen duen puntu-hodeia lortzen da taulako X-ren balioak ardatz horizontalean eta Y-ren balioak ardatz bertikalean adierazita. Nahiz eta puntu horiek zehatz-mehatz zuzen bat ez osatu, joera lineala erakusten du grafikoak. Joera hori bilatzen dugu; beraz, erregresio lineal sinplea onargarritzat hartzen dugu ikerketa honetan planteaturiko arazoa ebazteko.



10.2 irudia. 10.1 adibideko datuen adierazpide grafikoa.

10.3 KORRELAZIOA

Aurreko adibidean, *X* aldagaiaren balioen arabera *Y* beste aldagai baten balioak aurresaten saiatu gara, bata bestearen menpe deskribatuz. Askotan, interesa orokorragoa da, hain zuzen, bi aldagairen artean erlazio lineal bat dagoen jakitea, bata bestearen menpekotasuna landu barik. Halaber, *Y*-ren menpekotasun lineala *X*-rekiko gauzatzeko erregresio lineal eredua eraiki baino lehen, egokia da bien arteko baterako erlazio lineala deskribatzea ere. Helburu horretarako erabili beharreko parametroa da korrelazio linealaren koefizientea.

10.3.1 Korrelazio linealaren koefizientea

Bi aldagai jarraituren arteko erlazio lineala neurtzeko maizen erabiltzen den neurria da korrelazio linealaren *Pearson*en koefizientea. Parametro hori *X*-ren eta *Y*-ren arteko kobariantzaren bidez definitzen da, non kobariantza *X*-ren eta *Y*-ren baterako aldakuntzaren neurria baita. Kobariantzak daukan arazoetariko bat bornatua ez egotea da; beraz, neurri absolutuetan interpretatu beharko genuke. Arazo hori ekiditeko,

kobariantza «erlatibizatu» egiten da, eta bi aldagaien desbideratze estandarren biderkaduraz zatitzen da. Horretara, ***X*-ren eta *Y*-ren arteko *Pearson*en korrelazio linealaren koefizientea definitzen** da, eta ρ adierazten da:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}}$$

*Pearson*en korrelazio linealaren koefizientea [−1, 1] tartean dago beti. Balioa positiboa denean korrelazioa positiboa dela esaten dugu (*X* handitzerakoan *Y* ere handitu egiten da); balio negatiboek, ordea, korrelazio negatiboa adierazten dute (*X* handitzerakoan *Y* txikitu egiten da).

Definizio hori guztiz teorikoa dela konturatu behar dugu; beraz, ρ kalkulatu ahal izateko, *X*-ren eta *Y*-ren bariantzak eta bien arteko kobariantza kalkulatu behar dira; alegia, aldagaien probabilitate-banaketek ezagunak izan behar dute. Estatistikan, aurreko kantitateak estimatzean datza arazoa. Laugarren gaian azaldu dugu nola estimatu bariantza; era berean, gai honetan, *Pearson*en korrelazio linealaren ρ koefizientea estimatzen arituko gara.

Demagun behaturiko datuak *n* neurketa pare direla, (y_i, x_i) non, $i = 1, 2, ..., n$. Hots, y_i-k erantzun-aldagaiaren balioa eta x_i-k koaldagaiaren balioa adierazten dituzte. *Pearson*en korrelazio linealaren koefizientearen estimatzailea *r* adierazten da eta honako hau da:

$$r = \hat{\rho} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Oharra: *r* estatistikoak *X*-ren eta *Y*-ren arteko erlazio lineala neurtzen du. Alegia, *X*-ren eta *Y*-ren arteko erlazioa adierazgarria izan arren, lineala ez bada, *r*-k zero balioa har dezake.

10.3 adibidea

10.2 taulan agertzen diren datuetarako $r = 0$ da, nahiz eta *X*-ren eta *Y*-ren artean erlazio funtzionala dagoen, $y=x^2$ alegia.

x	-2	-1	0	1	2
y	4	1	0	1	4

10.2 taula. 10.3 adibideko datuak

*Pearson*en korrelazio linealaren koefizientea estimatzen dugunean, haren interpretazioak dagokion arloan oinarritu behar du. 10.3 irudian, interpretazio-laguntza gisa erabil daitekeen irizpidea eskaintzen da *r* interpretatzerakoan. Esate baterako, 0,5 korrelazio

positibo moderatua da; −0,9, ordea, korrelazio negatibo altua. Edozein kasutan, hemen aurkezten dugun irizpide hau zuhurtasunez erabili behar da, neurtzen ari garen ezaugarriaren menpekoa baita korrelazioaren indarra. Laborategiko esperimentu biologiko edo kimikoetan, batez ere oso giro kontrolatuetan egiten direnean, korrelazio-koefiziente altuak espero ditugu. Gizakiek parte hartzen duten ikerketetan, ordea, batez ere behaketan oinarritzen direnetan, dauden korrelazio-koefizienteak, baxuak izan arren, oso esanguratsuak izan daitezke.

10.3 irudia. Korrelazio linealaren interpretazio praktikoa. Iturria: J.S. Milton. *Estadística para Biología y Ciencias de la Salud*. Interamericana-McGraw-Hill, 1994.

X-ren eta *Y*-ren arteko korrelazioa altua izateko arrazoiak hainbat izan daitezke: kausazko erlazioa dagoela bi aldagaien artean; beste hirugarren aldagai batek aurreko bietan eragiten duela; edo baita, *X* eta *Y* erlazionatuta ez egon arren, lortutako lagina arraroa izateagatik erlazio faltsua erakusten duela.

Korrelazio linealaren koefizientea kalkulatzen duten agindu gehienek honako hipotesi-kontraste hau egiten dute, baita ρ-rako ere:

H_0: $\rho = 0$
H_1: $\rho \neq 0$

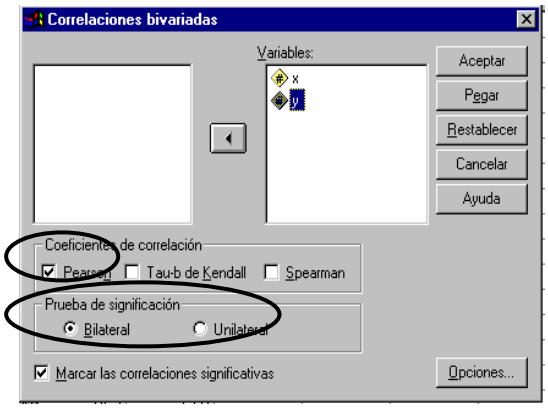
Hipotesi-kontraste horren emaitza arretaz interpretatu behar da. *X*-ren eta *Y*-ren arteko korrelazio lineala ezberdin zero den kontrastatzen du hain zuzen, baina inola ere ez du kontrastatzen korrelazioaren esanahi praktikoa. Laginaren tamaina handia denean, oso erraza da r = 0,05 izanik hipotesi-kontrastearen ondorioz korrelazioa adierazgarria izatea, nahiz eta korrelazio hori oso ahula izan.

Hauxe da *SPSS*ren bitartez korrelazio-analisia gauzatzeko erabili beharreko agindua: >*Analizar* >*Correlaciones* >*Bivariadas*. Agertzen zaigun leihatilan, erabili nahi ditugun aldagaiak soilik aukeratu behar ditugu. *Pearson* aukera hautatuz, *Pearson*en korrelazio linealaren koefizientea estimatuko dugu. *Prueba de significación* aukeratuz, korrelazio linealaren koefizientearen adierazgarritasunerako aipaturiko hipotesi-kontrastea egingo dugu.

10.4 adibidea

10.1 adibidean azaldutako ikerketarako 10.2 adibidean emandako datuak erabiliz, paziente diabetikoen odoleko glukosa-maila neurtzeko bi metodoen arteko korrelazio lineala kalkulatuko dugu, *SPSS* erabiliz. 10.4 irudiak korrelazio-analisia *SPSS*ren bitartez egiteko jarraitu beharreko urratsak

erakusten ditu.



10.4 irudia. Korrelazio linealaren kalkulua, *SPSS* erabiliz

*SPSS*k eskaintzen duen emaitza taula bakarra da, eta *X*-ren eta *Y*-ren arteko korrelazio-matrizea agertzen da (10.3 taula). Datu hauek agertzen dira matrizean: lehena estimaturiko *Pearson*en korrelazio linealaren koefizientea (r = 0,965) da; bigarrena H_0: $\rho = 0$ eta H_1: $\rho \neq 0$ hipotesi-kontrastearen p-balioa ($p < 0{,}05$) da; eta hirugarrena, laginaren tamaina (n = 40). Datu horien arabera, odoleko glukosa-maila neurtzeko bi metodoen arteko korrelazio lineala oso sendoa da.

Correlaciones

		X	Y
X	Correlación de Pearson	1	,965**
	Sig. (bilateral)	.	,000
	N	40	40
Y	Correlación de Pearson	,965**	1
	Sig. (bilateral)	,000	.
	N	40	40

**. La correlación es significativa al nivel 0,01

10.3 taula. 10.2 adibideko datuen korrelazio lineala, *SPSS*rekin

10.4 ERREGRESIO LINEAL SINPLEA

Erregresio lineal sinpleaz mintzo garenean, *Y* menpeko aldagaiari buruzko informazioa aurresateko *X* aldagai azaltzaile bakarra ari garela erabiltzen suposatzen dugu. Aldagai azaltzaileen kopurua bat baino handiagoa denean, ordea, erregresio lineal anizkoitza deritzogu.

10.4.1 Eredua

X gaineko Y-ren erregresio lineal sinplea deritzogun ereduak honako itxura hau du:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

(Y = Osagai sistematikoa + Zorizko osagaia)

Eredu horretan bi osagai daude. Bata osagai sistematikoa da, eta Y-ren eta X-ren arteko menpekotasun lineala definitzen du ($\beta_0 + \beta_1 X$). Bestea zorizko osagaia da, eta Y-ren balioak X-ren bidez estimatzerakoan jaulkitzen den errorea neurtzen du. Zorizko osagaiari **zorizko errore** edo **perturbazio** deritzo, eta ε bidez adierazten da.

Oro har, n alek osaturiko zorizko laginean, honako era honetara adierazten da erregresio linealaren eredua. Erantzun-aldagaiak, Y_1, Y_2, ..., Y_n, honako adierazpen hau dauka:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

non ε_i 0 itxaropena eta σ^2 bariantza duten zorizko aldagaiak normal eta askeak baitira. Hain zuzen, $E(Y_i) = \mu_i$ eta $Var(Y_i) = \sigma^2$.

Erregresio lineal sinplearen ereduko osagai sistematikoa E(Y)-ren eta X-ren arteko erlazioa lerro zuzenak emandakoa dela onartzen du. Alegia, Y menpeko aldagaiaren batez besteko balioen grafikoa lerro zuzena dela, aurresandako X aldagai azaltzailearen balioetarako.

$$\mu_{Y|X} = \beta_0 + \beta_1 X$$

Alegia, erregresio lineal sinplearen ereduak onartzen du X-ren eta Y-ren arteko erlazio teorikoa **erregresio zuzena** deritzon lerro zuzenaren bidez zehazturik dagoela. Eredu horretan hiru parametro daude: β_0,β_1 eta σ^2. Osagai sistematikoan agertzen diren β_0 eta β_1 parametroek adierazten dute lerro zuzenak y ardatza mozten duen puntua eta zuzenaren malda, hurrenez hurren. Zorizko osagaian agertzen den σ^2 parametroak, berriz, X koaldagaiaren edozein baliotarako Y-ren banaketaren sakabanatzea adierazten du.

Praktikan, erlazio funtzional zehatza ezezaguna da; alegia, ez da ezaguna β_0 eta β_1 parametroen balioak zein diren. Parametro horiek aukeraturiko zorizko laginean behaturiko X-ren eta Y-ren balioak erabiliz estimatzen dira.

10.4.2 Parametroen estimazioa

Erregresio zuzeneko parametroak, hain zuzen β_0 eta β_1, estimatzeko erabiltzen den metodoari **karratu txikienen metodo** deritzo. Metodo horren baitan dagoen arrazoibidea oso sinplea da. Puntu-hodeiaren inguruan marraztu daitezkeen hainbat lerro zuzenen artean, komenigarria da behaturiko datuak hoberen doitzen dituena aukeratzea. Doikuntza hobea izateko erabiltzen den irizpidea da behaturiko puntuen eta doituriko zuzenaren arteko distantzia karratuen batura txikiena izatea. Horrela, orotara, behaturiko puntu guztietatik ahalik eta hurbilen dagoen zuzena aukeratuko dugu. Intuitiboki, metodo horren helburua da puntu bakoitzetik zuzenera dagoen distantzia bertikala ahalik eta txikiena egitea.

Behin erregresio zuzeneko parametroak estimaturik, eredu estimatuaren adierazpena lortuko dugu:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

(Y = Osagai sistematikoa + Zorizko osagaia)

Eredu estimatuak eredu teorikoaren egitura berdina dauka, baina kasu honetan ezagunak dira $\hat{\beta}_0$ eta $\hat{\beta}_1$ balioak; X eta Y aldagaietarako behaturiko balioen bidez kalkulatzen dira, eta β_0 eta β_1 parametroen estimatzaileak dira.

Honako adierazpen honi X-ren gaineko Y-ren erregresio zuzen estimatua deritzo:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Zuzen hori erabiliz, X-ren balioen menpeko Y-ren batez besteko balioak edo itxarotako balioak kalkula ditzakegu, alegia $\hat{y}_i$. Errorearen estimazioak, $e_i = y_i - \hat{y}_i$, behaturiko puntuaren balioa eta erregresio zuzenaren arteko distantzia neurtzen du. Orduan, $y_i = \hat{y}_i + e_i$ betetzen da. Esan dugun bezala, eredu doituak eredu teorikoak zeukan egitura bera dauka, eta parametroen ordez estimatzaileak agertzen dira. $\hat{y}_i$ eta e_i **itxarotako balioak** eta **hondarrak** deritze, hurrenez hurren.

Itxarotako balioen propietateek baimenduko digute koaldagaiaren eta erantzun-aldagaiaren arteko erlazioari buruzko galdera zientifikoen erantzuna ematea. Hondarren propietateak, ordea, doitutako ereduaren egokitasuna eta eredu probabilistikoaren baldintzak epaitzeko erabiliko ditugu.

Aipatu dugun bezala, behaturiko puntuen eta doituriko zuzenaren arteko distantzia karratuen batura txikiena izatea da (hau da, hondar karratuen batura txikiena izatea) estimatzaileak aukeratzeko irizpide.

$$EKB = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

EKB **errorearen karratuen batura** deritzo, $\hat{\beta}_0$ eta $\hat{\beta}_1$ **estimatzaile minimo koadratikoak** deritze, eta EKB minimoa egiten duten parametroen balioak dira.

Karratu txikien estimazio metodoaren aplikazioaren ondorioz, $\hat{\beta}_0$ eta $\hat{\beta}_1$ estimatzaile minimo koadratikoak honako adierazpen honek definitzen ditu:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

non, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ eta $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ baitira.

Praktikan, funtsezkoa da parametroen estimazioen interpretazioa. Maldak edo $\hat{\beta}_1$-ek *X* unitate bat gehitzen denean *Y*-rako itxarotako gehikuntza adierazten du. Horrela, $\hat{\beta}_1$ positiboa izateak adierazten du *X* handitzerakoan *Y* ere handitu egiten dela; $\hat{\beta}_1$ negatiboa izateak, ordea, *X* handitzerakoan *Y* txikitu egiten dela esan nahi du; eta, azkenik, $\hat{\beta}_1 = 0$ izateak *X* eta *Y* artean erlazio linealik ez dagoela erakusten du. Gai askeak edo $\hat{\beta}_0$-k $X = 0$ denean *Y*-rako itxarotako balioa adierazten du. Egoeraren arabera, azken hori zentzurik gabeko interpretazio izan daiteke, beraz, $\hat{\beta}_0$-ren interpretazioa zuhurtasunez egin beharrekoa da.

Datu multzoak behaturiko datu horiek osatzen duten eremuan soilik eskaintzen digu linealtasun proba. Zer esanik ez, eremu horretatik kanpo dagoen balioentzat ez dagoela linealtasunaren nabaritasunik. Beraz, azpimarratu beharrekoa da oso arriskutsua izan daitekeela *X*-ren behaturiko balioetatik kanpo dagoen balioari dagokion *Y*-ren balioa aurresateko estimaturiko erregresio zuzena erabiltzea.

Estimaturiko erregresio zuzena idazterakoan, zuzenean dauden puntuak itxarotako batez besteko erantzuna dira, aldagai azaltzailearen balioa *x* denean. Erregresioaren bidez

lortutako emaitzak interpretatzerakoan, zuhurtasun ikaragarriarekin jokatu behar dugu, lortutako ereduaren ekuazioak, batez besteko erantzuna adierazteaz gainera, ale bakoitzaren itxarotako balioa ere adierazten baitu, *X*-ren balio bakoitzerako.

10.4.3 Parametroetarako konfiantza-tarteak

Praktikan, parametroen estimazioak parametroen benetako balioak aurresateko erabiliko dira. Beraz, aurreko gaietan definitu ditugun tekniketan egin dugun era berean, hemen ere inferentzia egitea izango da helburuetariko bat. Parametroetarako inferentzia estimazioa eta hipotesi-kontrastearen bidez egingo dugu, eta, horretarako, estimatzaileen desbideratze estandarrak kalkulatu behar ditugu. Zuzeneko malda eta gai askearen estimatzaileen desbideratze estandarrak honako adierazpen hauen bidez estimatzen dira:

$$DE(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]}$$

$$DE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

non $\hat{\sigma}^2$ **hondar bariantza** deritzon σ^2 parametroaren ondoko estimatzailea baita:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2} = \frac{EKB}{n-2}$$

Behin estimatzaileak eta haien desbideratze estandarrak kalkulatuta, beste parametro batzuetarako erabili ditugun metodo berdinak erabil ditzakegu β_0 eta β_1 parametroen %100·(1 – α) konfiantza-mailako konfiantza-tarteak kalkulatzeko.

$$I_{\beta_1}^{1-\alpha} = \hat{\beta}_1 \pm t_{\alpha/2;\,(n-2)} \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$I_{\beta_0}^{1-\alpha} = \hat{\beta}_0 \pm t_{\alpha/2;\,(n-2)} \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]}$$

Non $t_{\alpha/2;\,(n-2)}$ *Student*en t banaketaren balio kritikoa baita, $(n - 2)$ askatasun gradutarako. Alegia, dentsitate-funtzioaren eta ardatz horizontalaren artean eskuinerantz $\alpha/2$ azalera uzten duen $(n - 2)$ askatasun gradutako *Student*en t banaketaren balioa.

10.4.4 ANOVA taula eta erregresio-kontrastea

Erregresio zuzenak Y aldagaian behaturiko sakabanatzearen kantitate adierazgarria azaltzen duen kontrastatzeko teknikari bariantza-analisi deritzo. Hurrengo taulari ANOVA taula deritzo, eta erregresio lineal sinplerako informazioa biltzen du (10.4 taula). Honako hauek dira ANOVA taulan agertzen diren kantitateak:

KB(x) = *X* koaldagaiari dagokion karratuen batura; *X*-ren eta *Y*-ren arteko erlazio linealari lotzen zaion *Y*-ren aldakortasuna adierazten du.

EKB = Erroreen karratuen batura; *Y*-ren zorizko aldakortasuna adierazten du.

KBT = Karratuen batura totala; *Y*-ren aldakortasun osoa adierazten du.

Jatorria	Askatasun-graduak	Karratuen batura (*KB*)
Eredua	*1*	$KB(x) = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$
Errorea	$n - 2$	$EKB = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$
Totala	$n - 1$	$KBT = \sum_{i=1}^{n} (y_i - \bar{y})^2$

10.4 taula. ANOVA taularen adierazpen orokorra

Aurreko hiru kantitateek honako erlazio hau betetzen dute:

$$KBT = KB(X) + EKB$$

Beraz, erregresio linealaren erabilera egokia baldin bada, orduan $KB(X)$ osagaiak aldakortasun osoaren zatirik handiena osatuko du, eta zorizko aldakortasuna txikia izango da azaldutakoarekiko. Aldagai azaltzaileak erantzunean eraginik ez izatea bat dator $\beta_1 = 0$ hipotesiarekin. Beraz, erregresio lineal sinplearen ereduan hauxe da intereseko hipotesi-kontrastea:

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

Hipotesi-kontraste hori garatzeko, bi metodo azalduko ditugu. Funtsean, bi metodoak baliokideak dira erregresio lineal sinplerako.

Egiantz arrazoiaren testa: *F* estatistikoaren bidez gauzatzen da.

$$F_p = \frac{KB(x)/1}{EKB/(n-2)}$$

H_0 hipotesi nulua egia baldin bada, aurreko estatistikoaren banaketa *1* eta $n-2$ askatasun gradutako *F* banaketa da. Beraz, hipotesi-kontrastea egiteko, estatistikoaren balio kritikoa erabil dezakegu:

- $F_p > F_{\alpha,1,(n-2)} \Rightarrow H_0$ errefusatu
- $F_p \leq F_{\alpha,1,(n-2)} \Rightarrow H_0$ ez errefusatu

edo, bestela, *p*-balioa kalkula dezakegu, hots $p = P(F_{1,(n-2)} > F_p)$

Wald-en testa: *t* estatistikoaren bidez gauzatzen da.

$$t_{obs} = \frac{\hat{\beta}_1}{DE(\hat{\beta}_1)}$$

H_0 hipotesi nulua egia baldin bada, aurreko estatistikoaren banaketa $n-2$ askatasun gradutako *Student*en *t* banaketa da. Kasu horretan ere, hipotesi-kontrastea egiteko estatistikoaren balio kritikoa erabil dezakegu.

- $|t_p| > t_{\alpha/2,(n-2)} \Rightarrow H_0$ errefusatu
- $|t_p| \leq t_{\alpha/2,(n-2)} \Rightarrow H_0$ ez errefusatu

edo, bestela *p*-balioa kalkula dezakegu, hots $p = \begin{cases} 2P(t_{n-2} > t_p) & \text{baldin} \quad t_p > 0 \\ 2P(t_{n-2} < t_p) & \text{baldin} \quad t_p < 0 \end{cases}$

<u>Oharra</u>: esan dugun bezala, aurreko hipotesi-kontrastea gauzatzeko bi metodoak guztiz baliokideak dira, eta horren arrazoia da *F* eta *t* banaketen artean dagoen honako erlazio hau: $t^2_{(n-2)} = F_{1,(n-2)}$.

10.4.5 Determinazio-koefizientea

Azkenik, erregresio lineal sinplearen eta korrelazio linealaren arteko erlazioa aztertuko dugu. Erregresio linealaren eredua erabili ahal izateko, *X* gaineko *Y*-ren erregresio zuzena kalkulatu baino lehen, ontzat hartu dugu linealtasuna bi aldagaien artean. Beraz, beharrezkoa da behaturiko puntu multzora zuzenaren doikuntzaren egokitasuna neurtu ahal izateko metodo bat. Determinazio-koefizientea *X*-ren eta *Y*-ren arteko erlazio

linealaren indarra neurtzeko erabiltzen den estatistikoa da, bai erregresio linealaren ereduan, bai korrelazio-analisian.

X gaineko *Y*-ren erregresio lineal sinplea doitzen dugunean, honako adierazpen honek neurtuko du *X*-ren bidez azaldutako *Y*-ren aldakortasunaren proportzioa:

$$\frac{KB(x)}{KBT} = \frac{\text{Linealtasunari dagokion } Y - \text{ren aldakortasuna}}{Y - \text{ren aldakortasun osoa}} =$$

$$= \left\{ \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \right\}^2 = r^2$$

Non *r X*-ren eta *Y*-ren arteko *Pearson*en korrelazio linealaren koefizientea baita.

Ondorioz, *Pearson*en korrelazio linealaren koefizientearen karratuak *X*-ren bidez azaldutako *Y*-ren aldakortasun-proportzioaren estimazioa ematen digu. Kantitate horri **determinazio-koefiziente** deritzo eta $\boldsymbol{R^2}$ adierazten da. Erregresio lineal sinplerako, gauza bera dira *X* gaineko *Y*-ren erregresioan lortutako determinazio-koefizientea eta *X*-ren eta *Y*-ren arteko *Pearson*en korrelazio linealaren koefizientearen karratua.

Gaur egungo pakete estatistiko gehienek, determinazio-koefiziente arruntaz gainera, determinazio-koefiziente zuzendua kalkulatzen dute. Azken hori aurrekoaren ordezkoa baizik ez da, baina ereduko parametro kopuruak zuzenduta. **Determinazio-koefiziente zuzendua** $\overline{R}^2$ adierazten da, eta honako formula honen bidez kalkulatzen da:

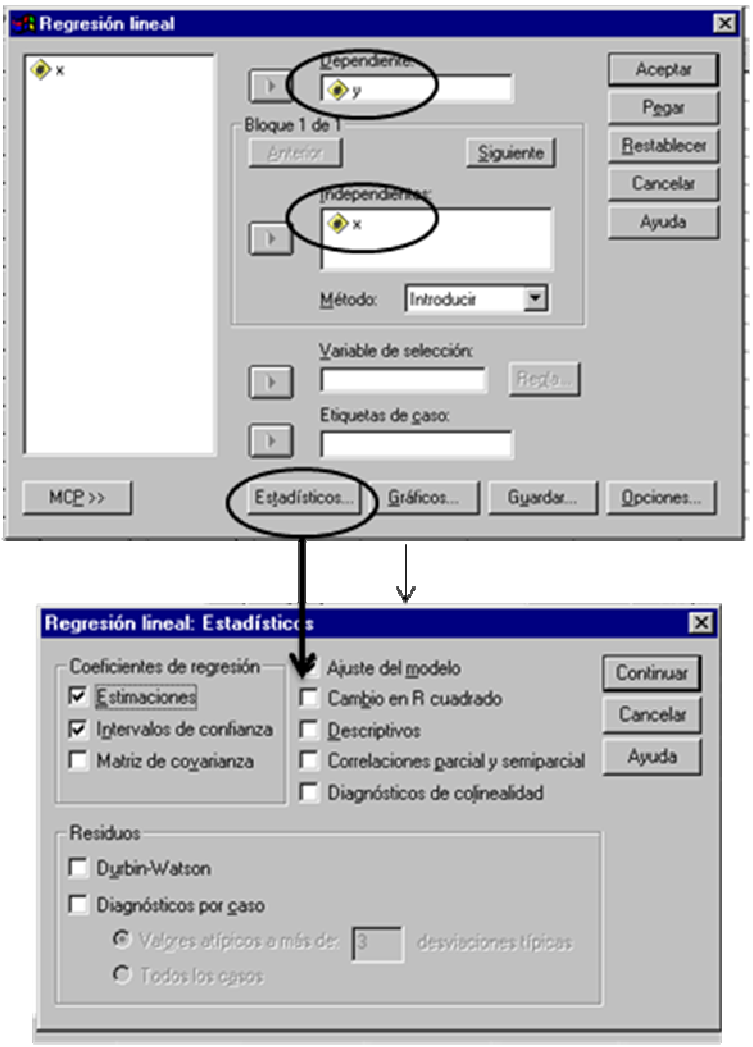
$$\overline{R}^2 = 1 - \frac{(n-i)(1-R^2)}{(n-2)}$$

non *n* lagineko ale kopurua baita eta *i* gai askeak ekuazioan egotearen aldagai adierazlea (1 balioa hartzen du β_0 ekuazioan baldin badago eta 0 ez badago).

10.5 ERREGRESIO LINEAL SINPLEA SPSS-REN BITARTEZ

Aurreko formuletan ikusi ahal izan dugunez, erregresio lineal sinplearen ereduaren aplikazioak eragiketa dezente egitea eskatzen du. Praktikan, eragiketa horiek guztiak ez dira eskuz egiten, baizik eta horretarako diseinaturiko *software*a erabiltzen da, esate baterako, *SPSS*. Beraz, gure helburua izango da ondorioak ateratzeko beharrezko emaitzak lokalizatzea eta interpretatzea.

*SPSS*ren bitartez erregresio lineala gauzatzeko erabili beharreko agindua hauxe da: *>Analizar >Regresión >Lineal*. Agertzen zaigun leihatilan, zein den menpeko aldagaia edo erantzuna eta zein den aldagai azaltzailea edo koaldagaia adierazi behar dugu (10.5 irudia).



10.5 irudia. Erregresio lineala *SPSS*ren bidez

Besterik adierazi ezean, agertzen diren emaitzak erregresio zuzena eta ANOVA taula dira. *Estadísticos* eta *Gráficos* bi botoiek aukera ematen digute kalkulatu nahi ditugun estatistiko gehigarriak eta irudikatu nahi ditugun grafikoak eskatzeko. *Guardar* botoiak eredua doitzerakoan kalkulatu diren beste aldagai batzuk gure datu-fitxategira biltzeko aukera eskaintzen digu, esate baterako itxarotako balioak edo hondarrak. *Opciones*

botoiak estimazio prozesua kudeatzeko aukera baliagarriak eskaintzen dizkigu; erregresio lineal sinplean gutxitan erabiltzen dira aukera horiek, baina erregresio lineal anizkoitzean oso beharrezkoak izango dira, hain zuzen. Azken botoi horretan, erregresio zuzena β_0 gai askea gabe estimatzea baimentzen digun aukera ere hauta dezakegu.

10.5 adibidea

10.1 adibidean azaldutako ikerketarako, 10.2 adibidean emandako datuak erabiliz erregresio lineal sinplea doituko dugu, datu horietarako *SPSS* erabiliz.

Estadísticos botoia erabiliz *Estimaciones* eta *Intervalos de confianza* aukerak hautatuko ditugu erregresio zuzeneko parametroen estimazioetarako. *Ajuste del modelo* aukera ere hautatuko dugu, determinazio-koefizientearen bidez ereduaren doikuntza neurtu ahal izateko.

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	762,149	1	762,149	520,568	,000[a]
	Residual	55,635	38	1,464		
	Total	817,784	39			

a. Variables predictoras: (Constante), X

b. Variable dependiente: Y

10.5 taula. 10.5 adibideko datuen ANOVA taula

Lorturiko emaitzak bi taulatan laburbiltzen dira. Lehena ANOVA taula da (10.5 taula); bertan erregresioaren hipotesi-kontrastea agertzen da, F estatistikoaren bidez egina. Kasu horretan, lortutako $p < 0{,}05$, beraz, odoleko glukosa-maila kalkulatzeko bi neurketen arteko erlazio lineala estatistikoki adierazgarria da.

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	,858	,352		2,440	,019	,146	1,570
	X	,912	,040	,965	22,816	,000	,832	,993

a. Variable dependiente: Y

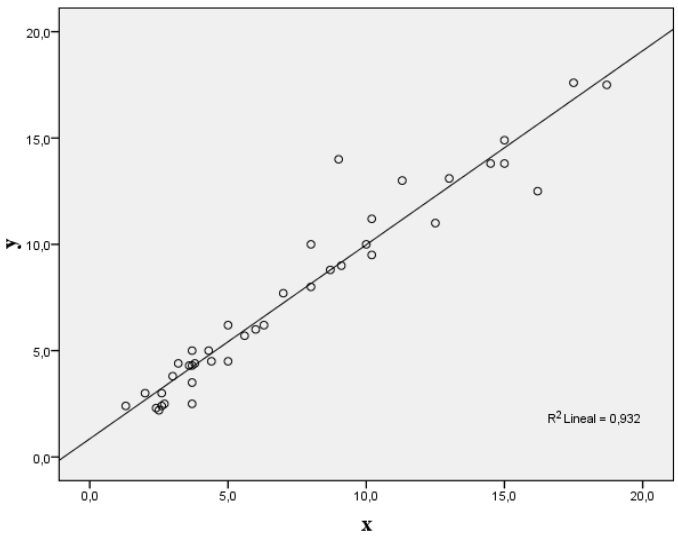
10.6 taula. 10.5 adibideko datuen erregresio zuzena eta ezaugarriak

Bigarren taulak (10.6 taula) parametroen estimazioak gehi haien desbideratze estandarra, konfiantza-tartea eta adierazgarritasunerako hipotesi-kontrastearen estatistikoa eta p-balioa erakusten dizkigu (Wald-en testa). Kasu horretan, hagatxoaren bidez egindako odoleko glukosa-mailaren neurketan unitate

bateko gehikuntza benetako odoleko glukosa-mailaren 0,912 unitateko gehikuntzari lotuta dago. Erlazio hori estatistikoki adierazgarria da ($p < 0,05$) eta, % 95eko konfiantzaz, 0,832 eta 0,993 balioen artean egongo da gehikuntza hori.

Balio horren interpretazioan jartzen badugu arreta, hagatxoak, laborategiko neurketaren ordez erabili ahal izateko, 1 balioa konfiantza-tartearen barruan egon beharko zukeen, horrek ziurtatuko baitzuen hagatxoaren bidez lortutako neurketaren fidagarritasuna benetako balioa aurresateko. Ikerketaren helburua da paziente diabetikoen odoleko glukosa-maila neurtzeko hagatxoaren zehaztasuna egiaztatzea, eta emaitza horren arabera hori ez da posible izango. Azkenik, gai askearen interpretazioa da hagatxoaren bidez lortutako neurketa zero denean itxarotako benetako odoleko glukosa-maila, eta kasu honetan hori ez da interpretagarria.

Erlazio hori grafikoki ikusi ahal izateko, aurretik egin dugun puntu-hodeiari erregresio zuzena gehitu ahal diogu, honako 10.6 irudi honetan adierazten den moduan.



10.6 irudia. 10.5 adibideko datuen puntu-hodeia eta erregresio zuzena.

Bukatzeko, gogora dezagun *Ajuste del modelo* aukera hautatu dugula ereduaren doikuntza neurtu ahal izateko. Agindu horren ondorioz, hirugarren taula bat lortu dugu doikuntzaren egokitasuna neurtzeko (10.7 taula). Taula horretan agertzen dira *Pearson*en korrelazio linealaren koefizientea, determinazio-koefiziente gordina, R^2, eta zuzendua, $\overline{R}^2$.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,965[a]	,932	,930	1,2100

a. Variables predictoras: (Constante), X

10.7 taula. 10.5 adibideko datuen determinazio koefizientea

Horrela, ikus dezakegu *Pearson*en korrelazio linealaren koefizientea, r = 0,965, aurreko adibidean kalkulatu dugunaren berdina dela. Determinazio-koefiziente gordina eta zuzendua R^2 = 0,932 eta $\overline{R}^2$ = 0,930 dira, hurrenez hurren. Azpimarratu beharrekoa da determinazio-koefiziente zuzendua beti dela gordina baino txikiagoa. Kontura gaitezke $R^2 = r^2$ betetzen dela ere.

Emaitza horren ondorioz, hagatxoaren bidez egindako odoleko glukosa-mailaren neurketak laborategian egindako benetako odoleko glukosa-mailaren aldakortasunaren % 93 biltzen du.

10.6 ERREGRESIO LINEAL SINPLEAREN EREDUAREN DIAGNOSIA

Erregresio lineal sinplea eredua definitu dugunean, erroreen baldintza teoriko batzuk finkatu ditugu. Erroreak zorizko aldagai askeak dira, eta beren banaketa normala, itxaropena zero eta bariantza σ^2. Hain zuzen,

$$E(Y_i) = \mu_i$$

$$Var(Y_i) = \sigma^2$$

Baldintza teoriko horiek egiaztatu beharko dira, erregresio lineal eredua erabili, emaitzak interpretatu eta ondorioak atera baino lehen. Laburbilduz, honako hauek dira bermatu beharko ditugun baldintzak: erlazioaren linealtasuna, behaketen independentzia, erroreen normaltasuna eta erroreen bariantza konstantea. Lehena, erlazioaren linealtasuna, aurretik egiaztatu beharrekoa da, esan dugunez, puntu-hodeiaren itxura grafikoa erabiliz. Behaketen independentzia laginketa-prozesuan bermatu beharrekoa da; zorizko lagin bakuna izateak ziurtatuko du, batez ere, behaketak askeak izatea, nahiz eta hori ez den bide bakarra independentzia lortzeko. Beste bi baldintzak erroreen banaketa eta bariantzarakoa dira. Erroreak zorizko aldagai behatu ezinak dira, beraz, haiengan egindako hipotesiak ezin izango ditugu zuzenean egiaztatu. Gai honen hasieran aipatu dugunez, erroreen estimazioak hondarrak dira, eta azken horietan egiaztatu beharko dugu ezarritako baldintzen betetze-maila. Horretarako, **hondar estandarizatuak** definituko dira honako era honetara:

$$\frac{e_i}{\sqrt{Var(e_i)}}$$

Ereduaren hipotesiak egokiak baldin badira, hondar estandarizatuak gutxi gorabehera banaketa normalari darraiote, itxaropena 0 eta bariantza 1 izanik. Ondorioz, hondar estandarizatuen histogramak banaketa simetrikoa erakutsi beharko luke, zeron zentratua, eta arretaz aztertu beharko genituzke ±3 baino handiagoak diren balioak.

*SPSS*k hondarrak gordetzea baimentzen du, *Guardar* aukera hautatuz, erregresio lineala gauzatzerakoan. Hondarrak datu-fitxategian ditugunean, *Gráficos* eta *Analizar* agindu orokorrak erabil daitezke, hondarren analisi grafikoa eta normaltasunerako doikuntza-egokitasun hipotesi-kontrastea egiteko. Halaber, erregresio lineala gauzatzerakoan aukera dago agindu berean diagnosi-analisi partziala egiteko.

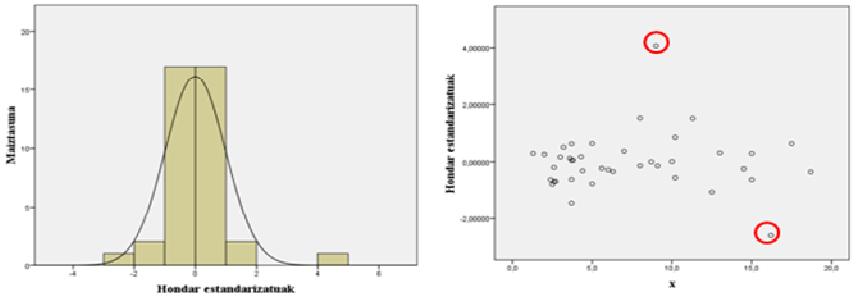
Kontrastatzeko zailena den baldintza hondarren homozedastizitatea da, alegia, bariantza konstantea izatea. Hala ere, hondarren bariantza konstantea ez den kasu gehienetan, *X*-ren balioen menpekoa izatea da arrazoia. Egoera hori grafikoki erraz ikusten da itxarotako balioak, $\hat{y}_i$, *X*-rekiko irudikatzen direnean.

Aipaturiko baldintza orokorrez gainera, erregresio lineala gauzatzerakoan kontuan hartu behar den beste faktore bat balio arraroena da. Bai *Y*-rako edo bai *X*-rako ale batzuen balioak oso arraroak izan daitezke, eta horiek eredua baldintzatzen badute balio eraginkorrak direla esaten da. Balio arraroek ez dute zertan eraginkorrak izan nahitaez, baina diagnosia egiterakoan balio arraroen miaketa egin behar da. Balio arraroren bat aurkituz gero, lehendabizi akatsak diren ala ez egiaztatu beharko dugu. Akatsak badira, zuzendu edo laginetik kanpo utziko ditugu. Benetako balioak baldin badira, ordea, ereduan mantentzearen egokitasuna eztabaidatu beharko dugu lortutako ereduan daukaten eraginkortasunaren arabera.

10.6 adibidea

10.1 adibidean azaldutako ikerketarako, 10.5 adibidean egindako erregresio linealaren hondar estandarizatuak gorde ditugu eta, lehendabizi, haien azterketa grafikoa egin dugu. Irudikatzen ditugun bi grafikoak hondarren histograma eta *X*-rekiko puntu-hodeia dira (10.7 irudia).

Bi grafiko horiek ez dute erakusten normaltasun faltarik, ezta bariantza konstantea izatearen hipotesia betetzen ez denik ere. Histogramak eskuinerantz alborapen txiki bat eduki arren, ez da 40 tamainako laginerako espero den baino handiagoa. Sakabanatze-irudia, balio absolutuan 3 baino handiagoa den hondar bat, eta 2 eta 3 bitartean beste bat ikus ditzakegu (gorriz markatuak irudian); baina, hori ere zorizkotasunari dagokion mugetan dago. Gainerako balioek bariantza nahiko konstante mantentzen dela erakusten digute. Azkenik, hondar estandarizatuen doikuntza-egokitasuna banaketa normalera kontrastatzen dugunean, *Kolmogorov-Smirnov* testa erabiliz, hipotesi nulua ez da errefusatzen (p = 0,278). Ondorioz, banaketa normalaren baldintza ere ez dugu zalantzan jartzen.



10.7 irudia. 10.1 adibideko datuentzat eraikitako erregresio zuzenaren egokitasun grafikoa

10.7 ARIKETAK

10.1. 1969. urtetik 1973. urterako 1.000 biztanleko jaiotza-tasa adierazten du 10.7 taulak:

Urteak	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
Jaiotza-tasa	56	59	64	65	73	78	81	94	93	95

10.7 taula. 10.1 ariketako datuen taula

a) Kalkula ezazu korrelazio linealaren koefizientea eta interpreta ezazu haren balioa.

b) Doi ezazu erregresio zuzena.

c) Kalkula itzazu erregresio linealaren koefizienteen % 95eko konfiantza-tarteak.

d) Kalkula ezazu ANOVA taula.

e) Zenbat jaiotza espero dugu 1979. urtean?

10.2. Bost pertsonaren altuerak eta pisuak agertzen dira 10.8 taulan:

Pisuak (kg-tan)	64	68	70	72	74
Altuerak (cm-tan)	160	70	180	190	195

10.8 taula. 10.2 ariketako datuen taula

a) Kalkula ezazu altueraren eta pisuaren arteko korrelazio linealaren koefizientea, eta interpreta ezazu balioa.

b) Doi ezazu altuerarekiko pisuaren erregresio zuzena.

c) Kalkula itzazu erregresio linealaren koefizienteen % 95eko konfiantza-tarteak.

d) Kalkula ezazu ANOVA taula.

e) Zein da itxarotako pisua 175 cm-ko pertsona batentzat?

10.3. Austriako Worther lakuko Y mg/l-ko oxigenoa neurtu da X metroko sakoneran, eta honako datu hauek lortu dira (10.9 taula):

Sakonera (m)	15	20	30	40	50	60	70
Oxigenoa (mg/l)	6,5	5,6	5,4	6	4,6	1,4	0,1

10.9 taula. 10.3 ariketako datuen taula

a) Doi iezaiozu erregresio zuzena lorturiko datuei.

b) Azter ezazu aldagai bien arteko korrelazioa.

c) Azter ezazu erregresio-koefizienteen adierazgarritasun-maila.

d) Kalkula ezazu ANOVA taula.

e) Zenbat oxigeno espero da ur-azaleran?

f) Zein da estimazioaren errore estandarra?

10.4. Honako datu hauek (10.10 taula) 18 urteko 10 gazteren altuera cm-tan (X) eta pisua kg-tan (Y) adierazten dute.

X	Y
169,6	71,2
166,8	58,2
157,1	56,0
181,1	64,5
158,4	53,0

X	Y
165,6	52,4
166,7	56,8
156,5	49,2
168,1	55,6
165,3	77,8

10.10 taula. 10.4 ariketako datuen taula

Erantzun itzazu honako galdera hauek:

a) Adieraz ezazu grafikoki *Y*-ren *X*-rekiko puntu-hodeia.

b) Froga itzazu: $\bar{x} = 165.52$, $\bar{y} = 59.47$, $S_{XX} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 472.076$, $S_{YY} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 731.961$ eta $S_{XY} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 274.786$. Kalkula itzazu malda eta lerro zuzenak *y* ardatza mozten duen puntuaren estimazioak. Adieraz ezazu erregresio zuzena.

c) Estima itzazu σ^2 eta $\hat{\beta}_0$ eta $\hat{\beta}_1$-en desbideratze estandarrak, eta $\hat{\beta}_0$ eta $\hat{\beta}_1$-en arteko kobariantza.

d) Egin itzazu honako hipotesi-kontraste hauek:

H_0: $\beta_1 = 0$ $\qquad$ H_0: $\beta_0 = 0$

H_1: $\beta_1 \neq 0$ $\qquad$ H_1: $\beta_0 \neq 0$

e) Kalkula ezazu ANOVA taula eta erregresioko F kontrastearen estatistikoaren balioa. Froga ezazu $t^2 = F$ erlazioa betetzen dela numerikoki.

f) Kalkula itzazu hondarrak eta Y-rako itxarotako balioak. Froga ezazu numerikoki hondarren batura zero dela. Egin ezazu hondarren adierazpen grafikoa itxarotako balioekiko.

g) Interpreta itzazu β_0 eta β_1 parametroen balioak. Zer unitatetan neurtzen dira parametro horiek? Eta σ^2?

10.5. Ikerketa ugari daude urtebetean behaturiko eguzkiko lohiune kopurua fenomeno lurtarrekin erlazionatzen dutenak. 10.11 taulan agertzen dira horretarako erabilitako datuen parte: Victoria lakuaren batez besteko maila (X), estandar finko batekiko neurri erlatibo modura neurtuta, eta 1902tik 1921era urteroko eguzkiko lohiune kopurua (Y).

Urtea	X	Y
1902	-10	5
1903	13	24
1904	18	42
1905	15	63
1906	29	54
1907	21	62
1908	10	49
1909	8	44
1910	1	19
1911	-7	6

Urtea	X	Y
1912	-11	4
1913	-3	1
1914	-2	10
1915	4	47
1916	15	57
1917	35	104
1918	27	81
1919	8	64
1920	3	38
1921	-5	25

10.11 taula. 10.5 ariketako datuen taula.

a) Adieraz ezazu grafikoki Y-ren X-rekiko puntu-hodeia.

b) Doi ezazu X-rekiko Y-ren erregresio zuzena. Kalkula itzazu:
   i. Erregresioko parametroen estimazioak.
   ii. Determinazio-koefizientea, R^2.
   iii. β_0 eta β_1-erako % 95eko konfiantza-tarteak.
   iv. Aipa itzazu iii. atalean erabilitako hipotesiak.

c) Egin ezazu hondarren adierazpen grafikoa itxarotako balioekiko.
   i. Doikuntza faltarik nabarmentzen al da?
   ii. Normaltasunaz aparte, ezin dugu adierazpen grafiko honetan egiaztatu aurreko b-iv) atalean aipaturiko hipotesietako bat. Zein da?
   iii. Proposa ezazu c-ii) atalean aipaturiko hipotesia egiaztatzeko metodoa. Egin ezazu proposaturiko grafikoa eta egiazta ezazu hipotesia.

10.6. Honako taula honetan (10.12 taula) agertzen dira apirilaren batean dagoen elur kantitatea (X), eta apiriletik ekainera ibaiaren arroak daraman ur kantitatea (Y), 17 urtean zehar.

X	*Y*
23,1	10,5
32,8	16,7
31,8	18,2
32,0	17,0
30,4	16,3
24,0	10,5
39,5	23,1
24,2	12,4
52,5	24,9

X	*Y*
37,9	22,8
30,5	14,1
25,1	12,9
12,4	8,8
35,1	17,4
31,5	14,9
21,1	10,5
27,6	16,1

10.12 taula. 10.6 ariketako datuen taula.

a) Emaitza horietatik abiatuz, kalkula itzazu:
   i. Erregresio zuzena.
   ii. Erregresioko parametroen estimatzaileen bariantzak eta bien arteko kobariantza.
   iii. Erregresioko parametroen estimatzaileen % 95eko konfiantza-tarteak.
   iv. Determinazio-koefizientea.
   v. Hondar-bariantza.

b) Interpreta itzazu erregresio-zuzena eta zuzena definitzen duten bi parametroen estimazioak.

c) Aurreko emaitzetan oinarrituz, apirilaren batean neurtutako elur kantitateak udaberrian zehar ibaiaren arroak daraman ur kantitatean eragin adierazgarria duela esango al zenuke? Azal ezazu zure erantzuna, % 5eko adierazgarritasun-mailaz.

10.7. Pertsona baten ezker eskuko bizitza-lerroaren luzeraren (X cm-tan) eta bizitzaren (Y urteetan) arteko erlazioa aztertzeko ikerketa egin da. Horretarako, 50 pertsonak osaturiko lagina aztertu da eta honako datu hauek lortu dira: $\sum yi = 3{,}333$; $\sum xi = 459{,}9$; $\sum y_i^2 = 231{,}933$; $\sum x_i^2 = 4308{,}57$ eta $\sum x_i y_i = 30{,}549$. Doi ezazu erregresio zuzena, kalkula itzazu erregresio-koefizienteen % 95eko konfiantza-tarteak, egin ezazu erregresioaren kontrastea, eta interpreta itzazu emaitzak, lorturiko datuen adierazgarritasun-mailaren arabera.

10.8. Soziologoek AEBko hilketen intzidentziaren gehikuntza aztertu nahi dute. Horretarako, honako datu hauek bildu dituzte hogei hiritan:

Y: Hilketa-tasa 100.000, mila biztanleko.
X_1: Urtero $5000 baino gutxiago irabazten duten familien proportzioa.
X_2: Langabezia-tasa.

a) Hiri batean urtero $5000 baino gutxiago irabazten duten familien proportzioak hiri horretako hilketa-tasan daukan eragina aztertzeko,

erregresio lineala erabiliz, honako taula hauek lortzen ditugu *SPSS* programatik (10.13 taula).

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,840[a]	,705	,689	5,51192

a. Variables predictoras: (Constante), Irabazi txikiko familia-proportzioa

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1308,339	1	1308,339	43,064	,000[a]
	Residual	546,863	18	30,381		
	Total	1855,202	19			

a. Variables predictoras: (Constante), Irabazi txikiko familia-proportzioa

b. Variable dependiente: Hilketa-tasa

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-29,901	7,789		-3,839	,001
	Irabazi txikiko familia-proportzioa	2,559	,390	,840	6,562	,000

a. Variable dependiente: Hilketa-tasa

Correlaciones de los coeficientes[a]

Modelo			Irabazi txikiko familia-proportzioa
1	Correlaciones	Irabazi txikiko familia-proportzioa	1,000
	Covarianzas	Irabazi txikiko familia-proportzioa	,152

a. Variable dependiente: Hilketa-tasa

10.13 taula. 10.8 ariketako datuen lehen *SPSS*ko irteera.

Emaitza horietatik abiatuz, kalkula itzazu:

i. Erregresio zuzena.
ii. Erregresioko parametroen estimatzaileen bariantzak eta bien arteko kobariantza.
iii. Erregresioko parametroen estimatzaileen % 95eko konfiantza-tarteak.
iv. Determinazio-koefizientea.
v. Hondar-bariantza.

b) Hiri bateko langabezia-tasak hiri horretako hilketa-tasan daukan eragina aztertzeko erregresio lineala erabiliz, honako taula hauek lortuko ditugu *SPSS* programatik (10.14 taula).

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,865[a]	,748	,734	5,09685

a. Variables predictoras: (Constante), Langabezi-tasa

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1387,600	1	1387,600	53,415	,000[a]
	Residual	467,602	18	25,978		
	Total	1855,202	19			

a. Variables predictoras: (Constante), Langabezi-tasa

b. Variable dependiente: Hilketa-tasa

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-28,527	6,814		-4,187	,001
	Langabezi-tasa	7,080	,969	,865	7,309	,000

a. Variable dependiente: Hilketa-tasa

Correlaciones de los coeficientes[a]

Modelo			Langabezi-tasa
1	Correlaciones	Langabezi-tasa	1,000
	Covarianzas	Langabezi-tasa	,938

a. Variable dependiente: Hilketa-tasa

10.14 taula. 10.8 ariketako datuen bigarren *SPSS*ko irteera.

Emaitza horietatik abiatuz, kalkula itzazu:

i. Erregresio zuzena.
ii. Erregresioko parametroen estimatzaileen bariantzak eta bien arteko kobariantza.
iii. Erregresioko parametroen estimatzaileen % 95eko konfiantza-tarteak.
iv. Determinazio-koefizientea.
v. Hondar-bariantza.

10.9. Indiako Ozeanoko eta Pazifikoko itsasertz babestuetan zehar, haitzetan eta maskorretan erantsita egoten den Patelloida pygmaea deritzon laparen altueraren (X) eta luzeraren (Y) arteko erlazioa aztertzeko egindako esperimentuan, honako taula honetako datuak lortu dira (10.15 taula).

x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
0,9	3,1	1,9	5,0	2,1	5,6	2,3	5,8
1,5	3,6	1,9	5,3	2,1	5,7	2,3	6,2
1,6	4,3	1,9	5,7	2,1	5,8	2,3	6,3
1,7	4,7	2,0	4,4	2,2	5,2	2,3	6,4
1,7	5,5	2,0	5,2	2,2	5,3	2,4	6,4
1,8	5,7	2,0	5,3	2,2	5,6	2,4	6,3
1,8	5,2	2,1	5,4	2,2	5,8	2,7	6,3

10.15 taula. 10.9 ariketako datuen taula.

Emaitza horietatik abiatuz, kalkula itzazu:

a) Erregresio zuzena.

b) Erregresioko parametroen estimatzaileen bariantzak eta bien arteko kobariantza.

c) Erregresioko parametroen estimatzaileen % 95eko konfiantza-tarteak.

d) Determinazio-koefizientea.

e) Hondar-bariantza.

10.10. Denboran zehar, preparatu hormonalaren aktibitatea zehaztu da, eta 10.16 taulan agertzen diren emaitzak lortu dira.

Denbora (hilabeteetan)	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Aktibitatea	90	85	9	72	69	73	50	42	47	30	34	32	23	21	25

a) Kalkula ezazu denborarekiko aktibitatearen erregresio zuzena.

b) Interpreta itzazu erregresio-koefizienteak.

c) Esango al zenuke hormonaren batez besteko aktibitatea denboraren menpekoa dela, % 5eko adierazgarritasun-mailaz? Arrazoitu ezazu zure erantzuna.

d) Estima ezazu, erregresio-zuzenaren arabera, hiru hilabeteri dagokion batez besteko hormonaren aktibitatea.

e) Noiz izango da nulua aktibitatea? Zentzuzkoa al da galdera hori egitea eta emandako erantzuna?

10.11. Eguneko kaloria kopuruari buruz eta asteko pisu-galerari buruz, 10.17 taulako datu hauek bildu ditu medikuak.

Kaloria kopuruak	Pisu-galerak
2000	1,0
1200	3,0
1500	2,0
2000	1,5
1100	1,5
1600	2,5
1800	3,1
2100	3,5
1200	1,5

Kaloria kopuruak	Pisu-galerak
800	5,0
500	4,0
1000	3,5
800	4,5
1100	4,0
900	3,5
1000	2,5
1200	2,0
800	4,5

10.17 taula. 10.11 ariketako datuen taula.

a) Aurki ezazu eguneko kaloria kopuruaren eta asteko pisu-galeraren arteko korrelazioa.

b) Kalkula ezazu kaloria-kontsumoarekiko pisu-galeraren erregresio zuzena.

c) Interpreta itzazu erregresio-koefizienteak.

d) Esango al zenuke astean izandako pisu-galera kaloria kontsumoaren menpekoa dela, % 5eko adierazgarritasun-mailaz? Arrazoitu ezazu zure erantzuna.

e) Kalkula eta interpreta ezazu determinazio-koefizientea.

f) Eguneko 700 kaloriako jan-neurria egiten duen pertsonarentzat, zer pisu-galera espero izango dugu?

10.12. Glukogeno kantitate finkoko hogei lagin hartu dira. X glukogenasa kantitatea (mmo1/1-tan) aplikatu zaie, kasu bakoitzaren Y erreakzio-abiadura neurturik (mol/ml-tan). Lortutako datuak 10.18 taulan adierazi dira.

x	y
1,0	18
2,0	35
3,0	60
0,2	8
0,5	10
1,0	20
3,0	57
0,8	17
2,2	38
1,5	25

X	y
1,2	20
1,8	31
0,2	11
3,1	64
0,7	16
1,5	27
2,8	55
0,9	19
2,0	31
1,7	29

10.18 taula. 10.12 ariketako datuen taula.

a) Kalkula ezazu glukogenasa kantitatearen eta erreakzio-abiaduraren arteko korrelazioa.

b) Datu horietatik, ondorioztatzen ote da erreakzio-abiadura glukogenasa kantitatearekin batera emendatzen dela? Justifika ezazu erantzuna, erregresio zuzena eta % 5eko adierazgarritasun-maila erabiliz.

c) Kalkula eta interpreta ezazu determinazio-koefizientea.

d) Lagin bati 5 mmol/1-ko glukogenasa kantitatea aplikatuko bagenio, zein izango litzateke itxarotako erreakzio-abiadura?

e) Kalkula ezazu erregresio zuzenari dagokion hondar-bariantza.

10.13. Inguruko tenperatura ezberdinetan neurturik, txoriek kontsumituriko oxigeno kantitateak dira honako datu hauek (10.19 taula).

0_2-ren kontsumoa (ml./g.h.-tan)	Tenperatura (ºC-tan)
5,2	-18
4,7	-15
4,5	-12
3,9	-10
3,6	-8
3,5	-5
2,9	-2

0_2-ren kontsumoa (ml./g.h.-tan)	Tenperatura (ºC-tan)
3,4	0
3,1	1
2,8	5
2,5	7
2,5	10
2,2	14
1,8	19

10.19 taula. 10.13 ariketako datuen taula.

Tenperatura *X* aldagai asketzat harturik, honako hau eskatzen da:

a) Estima ezazu erregresio zuzena.

b) Interpreta itzazu erregresio-koefizienteak eta haien % 95eko konfiantza-tarteak.

c) Kalkula ezazu estimazioaren hondar-bariantza.

d) Kalkula eta interpreta ezazu korrelazio-koefizientea.

A. ERANSKINA: TAULAK

A.1 taula. Bin(n,p) banaketa. $P(X=k)=\binom{n}{k}p^k q^{n-k}$

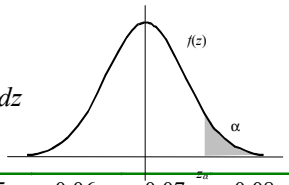
n	k	p	0.01	0.05	0.10	0.15	0.20	0.25	0.30	1 / 3	0.35	0.40	0.45	0.49	0.50
2	0		0.9801	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4444	0.4225	0.3600	0.3025	0.2601	0.2500
	1		0.0198	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4444	0.4550	0.4800	0.4950	0.4998	0.5000
	2		0.0001	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1111	0.1225	0.1600	0.2025	0.2401	0.2500
3	0		0.9703	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2963	0.2746	0.2160	0.1664	0.1327	0.1250
	1		0.0294	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4444	0.4436	0.4320	0.4084	0.3823	0.3750
	2		0.0003	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2222	0.2389	0.2880	0.3341	0.3674	0.3750
	3		0.0000	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0370	0.0429	0.0640	0.0911	0.1176	0.1250
4	0		0.9606	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1975	0.1785	0.1296	0.0915	0.0677	0.0625
	1		0.0388	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3951	0.3845	0.3456	0.2995	0.2600	0.2500
	2		0.0006	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.2963	0.3105	0.3456	0.3675	0.3747	0.3750
	3		0.0000	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.0988	0.1115	0.1536	0.2005	0.2400	0.2500
	4		0.0000	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0123	0.0150	0.0256	0.0410	0.0576	0.0625
5	0		0.9510	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1317	0.1160	0.0778	0.0503	0.0345	0.0313
	1		0.0480	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3292	0.3124	0.2592	0.2059	0.1657	0.1563
	2		0.0010	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3292	0.3364	0.3456	0.3369	0.3185	0.3125
	3		0.0000	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1646	0.1811	0.2304	0.2757	0.3060	0.3125
	4		0.0000	0.0000	0.0005	0.0022	0.0064	0.0146	0.0284	0.0412	0.0488	0.0768	0.1128	0.1470	0.1563
	5		0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0041	0.0053	0.0102	0.0185	0.0282	0.0313
6	0		0.9415	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0878	0.0754	0.0467	0.0277	0.0176	0.0156
	1		0.0571	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2634	0.2437	0.1866	0.1359	0.1014	0.0938
	2		0.0014	0.0305	0.0984	0.1762	0.2458	0.2966	0.3241	0.3292	0.3280	0.3110	0.2780	0.2436	0.2344
	3		0.0000	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2195	0.2355	0.2765	0.3032	0.3121	0.3125
	4		0.0000	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0823	0.0951	0.1382	0.1861	0.2249	0.2344
	5		0.0000	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0165	0.0205	0.0369	0.0609	0.0864	0.0938
	6		0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0014	0.0018	0.0041	0.0083	0.0138	0.0156
7	0		0.9321	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0585	0.0490	0.0280	0.0152	0.0090	0.0078
	1		0.0659	0.2573	0.3720	0.3960	0.3670	0.3115	0.2471	0.2048	0.1848	0.1306	0.0872	0.0604	0.0547
	2		0.0020	0.0406	0.1240	0.2097	0.2753	0.3115	0.3177	0.3073	0.2985	0.2613	0.2140	0.1740	0.1641
	3		0.0000	0.0036	0.0230	0.0617	0.1147	0.1730	0.2269	0.2561	0.2679	0.2903	0.2918	0.2786	0.2734
	4		0.0000	0.0002	0.0026	0.0109	0.0287	0.0577	0.0972	0.1280	0.1442	0.1935	0.2388	0.2676	0.2734
	5		0.0000	0.0000	0.0002	0.0012	0.0043	0.0115	0.0250	0.0384	0.0466	0.0774	0.1172	0.1543	0.1641
	6		0.0000	0.0000	0.0000	0.0001	0.0004	0.0013	0.0036	0.0064	0.0084	0.0172	0.0320	0.0494	0.0547
	7		0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0005	0.0006	0.0016	0.0037	0.0068	0.0078
8	0		0.9227	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0390	0.0319	0.0168	0.0084	0.0046	0.0039
	1		0.0746	0.2793	0.3826	0.3847	0.3355	0.2670	0.1977	0.1561	0.1373	0.0896	0.0548	0.0352	0.0313
	2		0.0026	0.0515	0.1488	0.2376	0.2936	0.3115	0.2965	0.2731	0.2587	0.2090	0.1569	0.1183	0.1094
	3		0.0001	0.0054	0.0331	0.0839	0.1468	0.2076	0.2541	0.2731	0.2786	0.2787	0.2568	0.2273	0.2188
	4		0.0000	0.0004	0.0046	0.0185	0.0459	0.0865	0.1361	0.1707	0.1875	0.2322	0.2627	0.2730	0.2734
	5		0.0000	0.0000	0.0004	0.0026	0.0092	0.0231	0.0467	0.0683	0.0808	0.1239	0.1719	0.2098	0.2188
	6		0.0000	0.0000	0.0000	0.0002	0.0011	0.0038	0.0100	0.0171	0.0217	0.0413	0.0703	0.1008	0.1094
	7		0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0024	0.0033	0.0079	0.0164	0.0277	0.0313
	8		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0002	0.0007	0.0017	0.0033	0.0039
9	0		0.9135	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0260	0.0207	0.0101	0.0046	0.0023	0.0020
	1		0.0830	0.2985	0.3874	0.3679	0.3020	0.2253	0.1556	0.1171	0.1004	0.0605	0.0339	0.0202	0.0176
	2		0.0034	0.0629	0.1722	0.2597	0.3020	0.3003	0.2668	0.2341	0.2162	0.1612	0.1110	0.0776	0.0703
	3		0.0001	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2731	0.2716	0.2508	0.2119	0.1739	0.1641
	4		0.0000	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2048	0.2194	0.2508	0.2600	0.2506	0.2461
	5		0.0000	0.0000	0.0008	0.0050	0.0165	0.0389	0.0735	0.1024	0.1181	0.1672	0.2128	0.2408	0.2461
	6		0.0000	0.0000	0.0001	0.0006	0.0028	0.0087	0.0210	0.0341	0.0424	0.0743	0.1160	0.1542	0.1641
	7		0.0000	0.0000	0.0000	0.0000	0.0003	0.0012	0.0039	0.0073	0.0098	0.0212	0.0407	0.0635	0.0703
	8		0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0009	0.0013	0.0035	0.0083	0.0153	0.0176
	9		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0008	0.0016	0.0020
10	0		0.9044	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0173	0.0135	0.0060	0.0025	0.0012	0.0010
	1		0.0914	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0867	0.0725	0.0403	0.0207	0.0114	0.0098
	2		0.0042	0.0746	0.1937	0.2759	0.3020	0.2816	0.2335	0.1951	0.1757	0.1209	0.0763	0.0494	0.0439
	3		0.0001	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2601	0.2522	0.2150	0.1665	0.1267	0.1172
	4		0.0000	0.0010	0.0112	0.0401	0.0881	0.1460	0.2001	0.2276	0.2377	0.2508	0.2384	0.2130	0.2051
	5		0.0000	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1366	0.1536	0.2007	0.2340	0.2456	0.2461
	6		0.0000	0.0000	0.0001	0.0012	0.0055	0.0162	0.0368	0.0569	0.0689	0.1115	0.1596	0.1966	0.2051
	7		0.0000	0.0000	0.0000	0.0001	0.0008	0.0031	0.0090	0.0163	0.0212	0.0425	0.0746	0.1080	0.1172
	8		0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0030	0.0043	0.0106	0.0229	0.0389	0.0439
	9		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0005	0.0016	0.0042	0.0083	0.0098
	10		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0010

A.2 taula. $\mathcal{P}(\lambda)$ banaketa. $P(X=k)=e^{-\lambda}\dfrac{\lambda^k}{k!}$

λ	*k*	0	1	2	3	4	5	6	7	8	9	10	11	12
0.1		0.9048	0.0905	0.0045	0.0002									
0.2		0.8187	0.1637	0.0164	0.0011	0.0001								
0.3		0.7408	0.2222	0.0333	0.0033	0.0003								
0.4		0.6703	0.2681	0.0536	0.0072	0.0007	0.0001							
0.5		0.6065	0.3033	0.0758	0.0126	0.0016	0.0002							
0.6		0.5488	0.3293	0.0988	0.0198	0.0030	0.0004							
0.7		0.4966	0.3476	0.1217	0.0284	0.0050	0.0007	0.0001						
0.8		0.4493	0.3595	0.1438	0.0383	0.0077	0.0012	0.0002						
0.9		0.4066	0.3659	0.1647	0.0494	0.0111	0.0020	0.0003						
1.0		0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001					
1.1		0.3329	0.3662	0.2014	0.0738	0.0203	0.0045	0.0008	0.0001					
1.2		0.3012	0.3614	0.2169	0.0867	0.0260	0.0062	0.0012	0.0002					
1.3		0.2725	0.3543	0.2303	0.0998	0.0324	0.0084	0.0018	0.0003	0.0001				
1.4		0.2466	0.3452	0.2417	0.1128	0.0395	0.0111	0.0026	0.0005	0.0001				
1.5		0.2231	0.3347	0.2510	0.1255	0.0471	0.0141	0.0035	0.0008	0.0001				
1.6		0.2019	0.3230	0.2584	0.1378	0.0551	0.0176	0.0047	0.0011	0.0002				
1.7		0.1827	0.3106	0.2640	0.1496	0.0636	0.0216	0.0061	0.0015	0.0003	0.0001			
1.8		0.1653	0.2975	0.2678	0.1607	0.0723	0.0260	0.0078	0.0020	0.0005	0.0001			
1.9		0.1496	0.2842	0.2700	0.1710	0.0812	0.0309	0.0098	0.0027	0.0006	0.0001			
2.0		0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0034	0.0009	0.0002			
2.2		0.1108	0.2438	0.2681	0.1966	0.1082	0.0476	0.0174	0.0055	0.0015	0.0004	0.0001		
2.4		0.0907	0.2177	0.2613	0.2090	0.1254	0.0602	0.0241	0.0083	0.0025	0.0007	0.0002		
2.6		0.0743	0.1931	0.2510	0.2176	0.1414	0.0735	0.0319	0.0118	0.0038	0.0011	0.0003	0.0001	
2.8		0.0608	0.1703	0.2384	0.2225	0.1557	0.0872	0.0407	0.0163	0.0057	0.0018	0.0005	0.0001	
3.0		0.0498	0.1494	0.2240	0.2240	0.1680	0.1008	0.0504	0.0216	0.0081	0.0027	0.0008	0.0002	0.0001
3.2		0.0408	0.1304	0.2087	0.2226	0.1781	0.1140	0.0608	0.0278	0.0111	0.0040	0.0013	0.0004	0.0001
3.4		0.0334	0.1135	0.1929	0.2186	0.1858	0.1264	0.0716	0.0348	0.0148	0.0056	0.0019	0.0006	0.0002
3.6		0.0273	0.0984	0.1771	0.2125	0.1912	0.1377	0.0826	0.0425	0.0191	0.0076	0.0028	0.0009	0.0003
3.8		0.0224	0.0850	0.1615	0.2046	0.1944	0.1477	0.0936	0.0508	0.0241	0.0102	0.0039	0.0013	0.0004
4.0		0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.1042	0.0595	0.0298	0.0132	0.0053	0.0019	0.0006
5.0		0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462	0.1044	0.0653	0.0363	0.0181	0.0082	0.0034
6.0		0.0025	0.0149	0.0446	0.0892	0.1339	0.1606	0.1606	0.1377	0.1033	0.0688	0.0413	0.0225	0.0113
7.0		0.0009	0.0064	0.0223	0.0521	0.0912	0.1277	0.1490	0.1490	0.1304	0.1014	0.0710	0.0452	0.0263
8.0		0.0003	0.0027	0.0107	0.0286	0.0573	0.0916	0.1221	0.1396	0.1396	0.1241	0.0993	0.0722	0.0481
9.0		0.0001	0.0011	0.0050	0.0150	0.0337	0.0607	0.0911	0.1171	0.1318	0.1318	0.1186	0.0970	0.0728
10.0		0.0000	0.0005	0.0023	0.0076	0.0189	0.0378	0.0631	0.0901	0.1126	0.1251	0.1251	0.1137	0.0948

λ	*k*	13	14	15	16	17	18	19	20	21	22	23	24
3.6		0.0001											
3.8		0.0001											
4.0		0.0002	0.0001										
5.0		0.0013	0.0005	0.0002									
6.0		0.0052	0.0022	0.0009	0.0003	0.0001							
7.0		0.0142	0.0071	0.0033	0.0014	0.0006	0.0002	0.0001					
8.0		0.0296	0.0169	0.0090	0.0045	0.0021	0.0009	0.0004	0.0002	0.0001			
9.0		0.0504	0.0324	0.0194	0.0109	0.0058	0.0029	0.0014	0.0006	0.0003	0.0001		
10.0		0.0729	0.0521	0.0347	0.0217	0.0128	0.0071	0.0037	0.0019	0.0009	0.0004	0.0002	0.0001

A.3 taula. Z: $N(0.1)$ banaketa $\int_{z_\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$

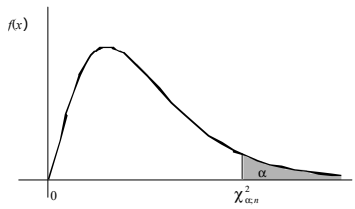


z_α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

A.4 taula. 2.000 zorizko zenbaki

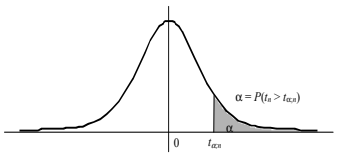
1	57785	73283	69160	81349	79421	47737	42582	02361
2	88277	69862	95341	38066	82338	06389	24003	70989
3	36441	29730	61465	90214	02553	42307	57809	26642
4	80368	41646	84621	10875	63047	65203	46013	68655
5	71367	03723	86816	74692	29153	78248	63551	91770
6	10511	70271	31458	57863	77123	13433	15739	07962
7	10834	73567	64471	15570	55527	38359	27866	96232
8	37934	71135	44309	41684	53770	41193	62388	74618
9	32054	22506	75736	74065	81980	09818	99791	98985
10	21437	02995	34214	13720	80796	95173	43211	63532
11	93805	98465	54917	00674	99352	89713	15572	19930
12	53696	40935	29482	84933	18057	66999	22698	32718
13	16811	49263	30992	80266	48599	82930	58513	99031
14	09258	76799	24010	57686	09280	81704	24077	58731
15	28037	52868	16856	57978	40523	08422	00716	83232
16	22803	75989	46045	18954	14277	75111	46912	25052
17	44440	24080	60454	69663	72268	21104	51045	72919
18	96521	85985	82310	44278	42622	75550	96454	13519
19	49567	13498	74207	45878	31943	53141	53152	58262
20	72159	51759	95487	17169	48598	14743	84575	12465
21	68975	00270	32078	00811	60302	77363	84025	00858
22	38317	90080	37210	31327	43275	08610	11518	98849
23	39679	77948	37600	13231	73244	33168	15331	25671
24	57356	65346	12452	64044	81941	75598	10165	47535
25	26468	56159	22184	80449	84861	50532	87888	62234
26	21075	88805	07640	79009	31181	93754	84426	42891
27	63163	75976	24141	54636	02688	86890	69601	87394
28	24175	21032	27538	67491	78864	54881	71884	72063
29	27224	95169	25152	98278	58098	09740	95852	85645
30	02000	18648	60053	33644	89770	99106	22657	58692
31	47626	29085	45903	74235	27718	23903	45254	76083
32	53405	33689	65142	97562	27289	78409	49568	12674
33	35853	17030	10191	32365	57848	44713	98493	92393
34	03286	47110	01560	66691	09132	95980	10541	98948
35	56871	30356	82922	93673	63669	11648	54564	28469
36	90472	15523	86006	19141	52122	07891	47246	96469
37	51748	31725	71004	72525	84341	72319	16164	10211
38	74879	81933	68167	51643	39749	14602	77125	28567
39	01960	32332	51096	41275	82887	17621	00675	87077
40	34775	56876	06016	01137	79857	34654	39889	53364
41	35437	00031	83336	23357	48071	93353	03333	45773
42	83079	54769	00945	17857	91613	53966	24199	86093
43	14643	67519	53663	26847	32672	71349	15495	25728
44	03039	47198	00582	49813	59310	69965	40242	87953
45	50218	06603	20384	98391	75596	77257	00152	56204
46	30075	07250	71997	32094	95802	68409	02057	71111
47	60370	46422	15627	46352	76250	80575	58903	61648
48	98718	70909	51389	34055	90449	17829	87171	92984
49	37232	03902	20096	36438	32948	02382	52669	62507
50	41994	53782	75705	53873	28053	94560	83082	86762

A.5 taula. χ^2_n banaketa



α / n	0.995	0.99	0.98	0.975	0.95	0.90	0.10	0.05	0.025	0.02	0.01
1	0.000	0.000	0.001	0.001	0.004	0.016	2.706	3.841	5.024	5.412	6.635
2	0.010	0.020	0.040	0.051	0.103	0.211	4.605	5.991	7.378	7.824	9.210
3	0.072	0.115	0.185	0.216	0.352	0.584	6.251	7.815	9.348	9.837	11.345
4	0.207	0.297	0.429	0.484	0.711	1.064	7.779	9.488	11.143	11.668	13.277
5	0.412	0.554	0.752	0.831	1.145	1.610	9.236	11.070	12.832	13.388	15.086
6	0.676	0.872	1.134	1.237	1.635	2.204	10.645	12.592	14.449	15.033	16.812
7	0.989	1.239	1.564	1.690	2.167	2.833	12.017	14.067	16.013	16.622	18.475
8	1.344	1.647	2.032	2.180	2.733	3.490	13.362	15.507	17.535	18.168	20.090
9	1.735	2.088	2.532	2.700	3.325	4.168	14.684	16.919	19.023	19.679	21.666
10	2.156	2.558	3.059	3.247	3.940	4.865	15.987	18.307	20.483	21.161	23.209
11	2.603	3.053	3.609	3.816	4.575	5.578	17.275	19.675	21.920	22.618	24.725
12	3.074	3.571	4.178	4.404	5.226	6.304	18.549	21.026	23.337	24.054	26.217
13	3.565	4.107	4.765	5.009	5.892	7.041	19.812	22.362	24.736	25.471	27.688
14	4.075	4.660	5.368	5.629	6.571	7.790	21.064	23.685	26.119	26.873	29.141
15	4.601	5.229	5.985	6.262	7.261	8.547	22.307	24.996	27.488	28.259	30.578
16	5.142	5.812	6.614	6.908	7.962	9.312	23.542	26.296	28.845	29.633	32.000
17	5.697	6.408	7.255	7.564	8.672	10.085	24.769	27.587	30.191	30.995	33.409
18	6.265	7.015	7.906	8.231	9.390	10.865	25.989	28.869	31.526	32.346	34.805
19	6.844	7.633	8.567	8.907	10.117	11.651	27.204	30.144	32.852	33.687	36.191
20	7.434	8.260	9.237	9.591	10.851	12.443	28.412	31.410	34.170	35.020	37.566
21	8.034	8.897	9.915	10.283	11.591	13.240	29.615	32.671	35.479	36.343	38.932
22	8.643	9.542	10.600	10.982	12.338	14.041	30.813	33.924	36.781	37.659	40.289
23	9.260	10.196	11.293	11.689	13.091	14.848	32.007	35.172	38.076	38.968	41.638
24	9.886	10.856	11.992	12.401	13.848	15.659	33.196	36.415	39.364	40.270	42.980
25	10.520	11.524	12.697	13.120	14.611	16.473	34.382	37.652	40.646	41.566	44.314
26	11.160	12.198	13.409	13.844	15.379	17.292	35.563	38.885	41.923	42.856	45.642
27	11.808	12.878	14.125	14.573	16.151	18.114	36.741	40.113	43.195	44.140	46.963
28	12.461	13.565	14.847	15.308	16.928	18.939	37.916	41.337	44.461	45.419	48.278
29	13.121	14.256	15.574	16.047	17.708	19.768	39.087	42.557	45.722	46.693	49.588
30	13.787	14.953	16.306	16.791	18.493	20.599	40.256	43.773	46.979	47.962	50.892

$$n > 30 \Rightarrow \chi^2_{\alpha,n} = \frac{1}{2}\left(z_\alpha + \sqrt{2n-1}\right)^2$$



A.6 taula. t_n banaketa.

n \ α	0.40	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756	3.396	3.660
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626	3.174	3.390
200	0.254	0.525	0.843	1.286	1.653	1.972	2.345	2.601	3.131	3.340
500	0.253	0.525	0.842	1.283	1.648	1.965	2.334	2.586	3.107	3.310
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.327	2.576	3.091	3.291

A.7 taula. F_{n_1,n_2} banaketa.

n_1		1	2	3	4	5	6	7	8	9	10
n_2	α										
1	0.005	16213	19997	21614	22501	23056	23440	23715	23924	24092	24222
1	0.010	4052.2	4999.3	5403.5	5624.3	5764.0	5859.0	5928.3	5981.0	6022.4	6055.9
1	0.025	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28	968.63
1	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
1	0.100	39.864	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.857	60.195
2	0.005	198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39	199.39
2	0.010	98.502	99.000	99.164	99.251	99.302	99.331	99.357	99.375	99.390	99.397
2	0.025	38.506	39.000	39.166	39.248	39.298	39.331	39.356	39.373	39.387	39.398
2	0.050	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396
2	0.100	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392
3	0.005	55.552	49.800	47.468	46.195	45.391	44.838	44.434	44.125	43.881	43.685
3	0.010	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228
3	0.025	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419
3	0.050	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785
3	0.100	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230
4	0.005	31.332	26.284	24.260	23.154	22.456	21.975	21.622	21.352	21.138	20.967
4	0.010	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546
4	0.025	12.218	10.649	9.979	9.604	9.364	9.197	9.074	8.980	8.905	8.844
4	0.050	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
4	0.100	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920
5	0.005	22.785	18.314	16.530	15.556	14.939	14.513	14.200	13.961	13.772	13.618
5	0.010	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
5	0.025	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619
5	0.050	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
5	0.100	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297
6	0.005	18.635	14.544	12.917	12.028	11.464	11.073	10.786	10.566	10.391	10.250
6	0.010	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874
6	0.025	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461
6	0.050	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
6	0.100	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937
7	0.005	16.235	12.404	10.883	10.050	9.522	9.155	8.885	8.678	8.514	8.380
7	0.010	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620
7	0.025	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761
7	0.050	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
7	0.100	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703
8	0.005	14.688	11.043	9.597	8.805	8.302	7.952	7.694	7.496	7.339	7.211
8	0.010	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814
8	0.025	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295
8	0.050	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347
8	0.100	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538
9	0.005	13.614	10.107	8.717	7.956	7.471	7.134	6.885	6.693	6.541	6.417
9	0.010	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
9	0.025	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964
9	0.050	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
9	0.100	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416

n_1		11	12	15	20	24	30	40	60	120	∞
n_2	α										
1	0.005	24334	24427	24632	24837	24937	25041	25146	25254	25358	25466
1	0.010	6083.4	6106.7	6157.0	6208.7	6234.3	6260.4	6286.4	6313.0	6339.5	6365.6
1	0.025	973.03	976.72	984.87	993.08	997.27	1001.4	1005.6	1009.8	1014.0	1018.3
1	0.050	242.98	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.32
1	0.100	60.473	60.705	61.220	61.740	62.002	62.265	62.529	62.794	63.061	63.328
2	0.005	199.42	199.42	199.43	199.45	199.45	199.48	199.48	199.48	199.49	199.51
2	0.010	99.408	99.419	99.433	99.448	99.455	99.466	99.477	99.484	99.491	99.499
2	0.025	39.407	39.415	39.431	39.448	39.457	39.465	39.473	39.481	39.489	39.498
2	0.050	19.405	19.412	19.429	19.446	19.454	19.463	19.471	19.479	19.487	19.496
2	0.100	9.401	9.408	9.425	9.441	9.450	9.458	9.466	9.475	9.483	9.491
3	0.005	43.525	43.387	43.085	42.779	42.623	42.466	42.310	42.150	41.990	41.829
3	0.010	27.132	27.052	26.872	26.690	26.597	26.504	26.411	26.316	26.221	26.125
3	0.025	14.374	14.337	14.253	14.167	14.124	14.081	14.036	13.992	13.947	13.902
3	0.050	8.763	8.745	8.703	8.660	8.638	8.617	8.594	8.572	8.549	8.526
3	0.100	5.222	5.216	5.200	5.184	5.176	5.168	5.160	5.151	5.143	5.134
4	0.005	20.824	20.705	20.438	20.167	20.030	19.892	19.751	19.611	19.469	19.325
4	0.010	14.452	14.374	14.198	14.019	13.929	13.838	13.745	13.652	13.558	13.463
4	0.025	8.794	8.751	8.657	8.560	8.511	8.461	8.411	8.360	8.309	8.257
4	0.050	5.936	5.912	5.858	5.803	5.774	5.746	5.717	5.688	5.658	5.628
4	0.100	3.907	3.896	3.870	3.844	3.831	3.817	3.804	3.790	3.775	3.761
5	0.005	13.491	13.385	13.146	12.903	12.780	12.656	12.530	12.402	12.274	12.144
5	0.010	9.963	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
5	0.025	6.568	6.525	6.428	6.329	6.278	6.227	6.175	6.123	6.069	6.015
5	0.050	4.704	4.678	4.619	4.558	4.527	4.496	4.464	4.431	4.398	4.365
5	0.100	3.282	3.268	3.238	3.207	3.191	3.174	3.157	3.140	3.123	3.105
6	0.005	10.133	10.034	9.814	9.589	9.474	9.358	9.241	9.122	9.001	8.879
6	0.010	7.790	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
6	0.025	5.410	5.366	5.269	5.168	5.117	5.065	5.012	4.959	4.904	4.849
6	0.050	4.027	4.000	3.938	3.874	3.841	3.808	3.774	3.740	3.705	3.669
6	0.100	2.920	2.905	2.871	2.836	2.818	2.800	2.781	2.762	2.742	2.722
7	0.005	8.270	8.176	7.968	7.754	7.645	7.534	7.422	7.309	7.193	7.076
7	0.010	6.538	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
7	0.025	4.709	4.666	4.568	4.467	4.415	4.362	4.309	4.254	4.199	4.142
7	0.050	3.603	3.575	3.511	3.445	3.410	3.376	3.340	3.304	3.267	3.230
7	0.100	2.684	2.668	2.632	2.595	2.575	2.555	2.535	2.514	2.493	2.471
8	0.005	7.105	7.015	6.814	6.608	6.503	6.396	6.288	6.177	6.065	5.951
8	0.010	5.734	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
8	0.025	4.243	4.200	4.101	3.999	3.947	3.894	3.840	3.784	3.728	3.670
8	0.050	3.313	3.284	3.218	3.150	3.115	3.079	3.043	3.005	2.967	2.928
8	0.100	2.519	2.502	2.464	2.425	2.404	2.383	2.361	2.339	2.316	2.293
9	0.005	6.314	6.227	6.032	5.832	5.729	5.625	5.519	5.410	5.300	5.188
9	0.010	5.178	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
9	0.025	3.912	3.868	3.769	3.667	3.614	3.560	3.505	3.449	3.392	3.333
9	0.050	3.102	3.073	3.006	2.936	2.900	2.864	2.826	2.787	2.748	2.707
9	0.100	2.396	2.379	2.340	2.298	2.277	2.255	2.232	2.208	2.184	2.159

n_1		1	2	3	4	5	6	7	8	9	10
n_2	α										
10	0.005	12.827	9.427	8.081	7.343	6.872	6.545	6.303	6.116	5.968	5.847
10	0.010	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
10	0.025	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717
10	0.050	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
10	0.100	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323
11	0.005	12.226	8.912	7.600	6.881	6.422	6.102	5.865	5.682	5.537	5.418
11	0.010	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539
11	0.025	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526
11	0.050	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
11	0.100	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248
12	0.005	11.754	8.510	7.226	6.521	6.071	5.757	5.524	5.345	5.202	5.085
12	0.010	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296
12	0.025	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374
12	0.050	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
12	0.100	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188
13	0.005	11.374	8.186	6.926	6.233	5.791	5.482	5.253	5.076	4.935	4.820
13	0.010	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100
13	0.025	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250
13	0.050	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
13	0.100	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138
14	0.005	11.060	7.922	6.680	5.998	5.562	5.257	5.031	4.857	4.717	4.603
14	0.010	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939
14	0.025	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147
14	0.050	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
14	0.100	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095
15	0.005	10.798	7.701	6.476	5.803	5.372	5.071	4.847	4.674	4.536	4.424
15	0.010	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805
15	0.025	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060
15	0.050	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
15	0.100	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059
16	0.005	10.576	7.514	6.303	5.638	5.212	4.913	4.692	4.521	4.384	4.272
16	0.010	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691
16	0.025	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986
16	0.050	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
16	0.100	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028
17	0.005	10.384	7.354	6.156	5.497	5.075	4.779	4.559	4.389	4.254	4.142
17	0.010	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593
17	0.025	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922
17	0.050	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
17	0.100	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001
18	0.005	10.218	7.215	6.028	5.375	4.956	4.663	4.445	4.276	4.141	4.030
18	0.010	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508
18	0.025	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866
18	0.050	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
18	0.100	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977

n_1		11	12	15	20	24	30	40	60	120	∞
n_2	α										
10	0.005	5.746	5.661	5.471	5.274	5.173	5.071	4.966	4.859	4.750	4.639
10	0.010	4.772	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
10	0.025	3.665	3.621	3.522	3.419	3.365	3.311	3.255	3.198	3.140	3.080
10	0.050	2.943	2.913	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
10	0.100	2.302	2.284	2.244	2.201	2.178	2.155	2.132	2.107	2.082	2.055
11	0.005	5.320	5.236	5.049	4.855	4.756	4.654	4.551	4.445	4.337	4.226
11	0.010	4.462	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.603
11	0.025	3.474	3.430	3.330	3.226	3.173	3.118	3.061	3.004	2.944	2.883
11	0.050	2.818	2.788	2.719	2.646	2.609	2.570	2.531	2.490	2.448	2.405
11	0.100	2.227	2.209	2.167	2.123	2.100	2.076	2.052	2.026	2.000	1.972
12	0.005	4.988	4.906	4.721	4.530	4.431	4.331	4.228	4.123	4.015	3.904
12	0.010	4.220	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
12	0.025	3.321	3.277	3.177	3.073	3.019	2.963	2.906	2.848	2.787	2.725
12	0.050	2.717	2.687	2.617	2.544	2.505	2.466	2.426	2.384	2.341	2.296
12	0.100	2.166	2.147	2.105	2.060	2.036	2.011	1.986	1.960	1.932	1.904
13	0.005	4.724	4.643	4.460	4.270	4.173	4.073	3.970	3.866	3.758	3.647
13	0.010	4.025	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
13	0.025	3.197	3.153	3.053	2.948	2.893	2.837	2.780	2.720	2.659	2.596
13	0.050	2.635	2.604	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
13	0.100	2.116	2.097	2.053	2.007	1.983	1.958	1.931	1.904	1.876	1.846
14	0.005	4.508	4.428	4.247	4.059	3.961	3.862	3.760	3.655	3.547	3.436
14	0.010	3.864	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
14	0.025	3.095	3.050	2.949	2.844	2.789	2.732	2.674	2.614	2.552	2.487
14	0.050	2.565	2.534	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
14	0.100	2.073	2.054	2.010	1.962	1.938	1.912	1.885	1.857	1.828	1.797
15	0.005	4.329	4.250	4.070	3.883	3.786	3.687	3.585	3.480	3.372	3.260
15	0.010	3.730	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.869
15	0.025	3.008	2.963	2.862	2.756	2.701	2.644	2.585	2.524	2.461	2.395
15	0.050	2.507	2.475	2.403	2.328	2.288	2.247	2.204	2.160	2.114	2.066
15	0.100	2.037	2.017	1.972	1.924	1.899	1.873	1.845	1.817	1.787	1.755
16	0.005	4.179	4.099	3.920	3.734	3.638	3.539	3.437	3.332	3.224	3.112
16	0.010	3.616	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
16	0.025	2.934	2.889	2.788	2.681	2.625	2.568	2.509	2.447	2.383	2.316
16	0.050	2.456	2.425	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
16	0.100	2.005	1.985	1.940	1.891	1.866	1.839	1.811	1.782	1.751	1.718
17	0.005	4.050	3.971	3.793	3.607	3.511	3.412	3.311	3.206	3.097	2.984
17	0.010	3.518	3.455	3.312	3.162	3.083	3.003	2.920	2.835	2.746	2.653
17	0.025	2.870	2.825	2.723	2.616	2.560	2.502	2.442	2.380	2.315	2.248
17	0.050	2.413	2.381	2.308	2.230	2.190	2.148	2.104	2.058	2.011	1.960
17	0.100	1.978	1.958	1.912	1.862	1.836	1.809	1.781	1.751	1.719	1.686
18	0.005	3.938	3.860	3.683	3.498	3.402	3.303	3.201	3.096	2.987	2.873
18	0.010	3.434	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
18	0.025	2.814	2.769	2.667	2.559	2.503	2.445	2.384	2.321	2.256	2.187
18	0.050	2.374	2.342	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
18	0.100	1.954	1.933	1.887	1.837	1.810	1.783	1.754	1.723	1.691	1.657

n_1		1	2	3	4	5	6	7	8	9	10
n_2	α										
19	0.005	10.073	7.093	5.916	5.268	4.853	4.561	4.345	4.177	4.043	3.933
19	0.010	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434
19	0.025	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817
19	0.050	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
19	0.100	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956
20	0.005	9.944	6.987	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847
20	0.010	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
20	0.025	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774
20	0.050	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
20	0.100	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937
21	0.005	9.829	6.891	5.730	5.091	4.681	4.393	4.179	4.013	3.880	3.771
21	0.010	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310
21	0.025	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735
21	0.050	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
21	0.100	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920
22	0.005	9.727	6.806	5.652	5.017	4.609	4.322	4.109	3.944	3.812	3.703
22	0.010	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258
22	0.025	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700
22	0.050	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
22	0.100	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904
23	0.005	9.635	6.730	5.582	4.950	4.544	4.259	4.047	3.882	3.750	3.642
23	0.010	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211
23	0.025	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668
23	0.050	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
23	0.100	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890
24	0.005	9.551	6.661	5.519	4.890	4.486	4.202	3.991	3.826	3.695	3.587
24	0.010	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168
24	0.025	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640
24	0.050	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
24	0.100	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877
25	0.005	9.475	6.598	5.462	4.835	4.433	4.150	3.939	3.776	3.645	3.537
25	0.010	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129
25	0.025	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613
25	0.050	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
25	0.100	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866
26	0.005	9.406	6.541	5.409	4.785	4.384	4.103	3.893	3.730	3.599	3.492
26	0.010	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094
26	0.025	5.659	4.265	3.670	3.329	3.105	2.945	2.824	2.729	2.653	2.590
26	0.050	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
26	0.100	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855
27	0.005	9.342	6.489	5.361	4.740	4.340	4.059	3.850	3.687	3.557	3.450
27	0.010	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062
27	0.025	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568
27	0.050	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
27	0.100	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845

n_1		11	12	15	20	24	30	40	60	120	∞
n_2	α										
19	0.005	3.841	3.763	3.587	3.402	3.306	3.208	3.106	3.000	2.891	2.776
19	0.010	3.360	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
19	0.025	2.765	2.720	2.617	2.509	2.452	2.394	2.333	2.270	2.203	2.133
19	0.050	2.340	2.308	2.234	2.155	2.114	2.071	2.026	1.980	1.930	1.878
19	0.100	1.932	1.912	1.865	1.814	1.787	1.759	1.730	1.699	1.666	1.631
20	0.005	3.756	3.678	3.502	3.318	3.222	3.123	3.022	2.916	2.806	2.691
20	0.010	3.294	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
20	0.025	2.721	2.676	2.573	2.464	2.408	2.349	2.287	2.223	2.156	2.085
20	0.050	2.310	2.278	2.203	2.124	2.082	2.039	1.994	1.946	1.896	1.843
20	0.100	1.913	1.892	1.845	1.794	1.767	1.738	1.708	1.677	1.643	1.607
21	0.005	3.680	3.602	3.427	3.243	3.147	3.049	2.947	2.841	2.730	2.614
21	0.010	3.236	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
21	0.025	2.682	2.637	2.534	2.425	2.368	2.308	2.246	2.182	2.114	2.042
21	0.050	2.283	2.250	2.176	2.096	2.054	2.010	1.965	1.916	1.866	1.812
21	0.100	1.896	1.875	1.827	1.776	1.748	1.719	1.689	1.657	1.623	1.586
22	0.005	3.612	3.535	3.360	3.176	3.081	2.982	2.880	2.774	2.663	2.546
22	0.010	3.184	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.306
22	0.025	2.647	2.602	2.498	2.389	2.332	2.272	2.210	2.145	2.076	2.003
22	0.050	2.259	2.226	2.151	2.071	2.028	1.984	1.938	1.889	1.838	1.783
22	0.100	1.880	1.859	1.811	1.759	1.731	1.702	1.671	1.639	1.604	1.567
23	0.005	3.551	3.474	3.300	3.116	3.021	2.922	2.820	2.713	2.602	2.484
23	0.010	3.137	3.074	2.931	2.780	2.702	2.620	2.536	2.447	2.354	2.256
23	0.025	2.615	2.570	2.466	2.357	2.299	2.239	2.176	2.111	2.041	1.968
23	0.050	2.236	2.204	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
23	0.100	1.866	1.845	1.796	1.744	1.716	1.686	1.655	1.622	1.587	1.549
24	0.005	3.497	3.420	3.246	3.062	2.967	2.868	2.765	2.658	2.546	2.428
24	0.010	3.094	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
24	0.025	2.586	2.541	2.437	2.327	2.269	2.209	2.146	2.080	2.010	1.935
24	0.050	2.216	2.183	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
24	0.100	1.853	1.832	1.783	1.730	1.702	1.672	1.641	1.607	1.571	1.533
25	0.005	3.447	3.370	3.196	3.013	2.918	2.819	2.716	2.609	2.496	2.377
25	0.010	3.056	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.170
25	0.025	2.560	2.515	2.411	2.300	2.242	2.182	2.118	2.052	1.981	1.906
25	0.050	2.198	2.165	2.089	2.007	1.964	1.919	1.872	1.822	1.768	1.711
25	0.100	1.841	1.820	1.771	1.718	1.689	1.659	1.627	1.593	1.557	1.518
26	0.005	3.402	3.325	3.151	2.968	2.873	2.774	2.671	2.563	2.450	2.330
26	0.010	3.021	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.132
26	0.025	2.536	2.491	2.387	2.276	2.217	2.157	2.093	2.026	1.954	1.878
26	0.050	2.181	2.148	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
26	0.100	1.830	1.809	1.760	1.706	1.677	1.647	1.615	1.581	1.544	1.504
27	0.005	3.360	3.284	3.110	2.927	2.832	2.733	2.630	2.522	2.408	2.287
27	0.010	2.988	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
27	0.025	2.514	2.469	2.364	2.253	2.195	2.133	2.069	2.002	1.930	1.853
27	0.050	2.166	2.132	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
27	0.100	1.820	1.799	1.749	1.695	1.666	1.636	1.603	1.569	1.531	1.491

n_1		1	2	3	4	5	6	7	8	9	10
n_2	α										
28	0.005	9.284	6.440	5.317	4.698	4.300	4.020	3.811	3.649	3.519	3.412
28	0.010	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032
28	0.025	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.547
28	0.050	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
28	0.100	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836
29	0.005	9.230	6.396	5.276	4.659	4.262	3.983	3.775	3.613	3.483	3.376
29	0.010	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005
29	0.025	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.669	2.592	2.529
29	0.050	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
29	0.100	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827
30	0.005	9.180	6.355	5.239	4.623	4.228	3.949	3.742	3.580	3.451	3.344
30	0.010	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979
30	0.025	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511
30	0.050	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
30	0.100	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819
40	0.005	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117
40	0.010	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801
40	0.025	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388
40	0.050	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
40	0.100	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763
60	0.005	8.495	5.795	4.729	4.140	3.760	3.492	3.291	3.134	3.008	2.904
60	0.010	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632
60	0.025	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270
60	0.050	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
60	0.100	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707
120	0.005	8.179	5.539	4.497	3.921	3.548	3.285	3.087	2.933	2.808	2.705
120	0.010	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472
120	0.025	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157
120	0.050	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910
120	0.100	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652
∞	0.005	7.880	5.299	4.280	3.715	3.350	3.091	2.897	2.745	2.621	2.519
∞	0.010	6.635	4.605	3.782	3.319	3.017	2.802	2.640	2.511	2.408	2.321
∞	0.025	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114	2.048
∞	0.050	3.842	2.996	2.605	2.372	2.214	2.099	2.010	1.939	1.880	1.831
∞	0.100	2.706	2.303	2.084	1.945	1.847	1.774	1.717	1.670	1.632	1.599

n_1		11	12	15	20	24	30	40	60	120	∞
n_2	α										
28	0.005	3.322	3.246	3.073	2.890	2.794	2.695	2.592	2.483	2.369	2.247
28	0.010	2.959	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
28	0.025	2.494	2.448	2.344	2.232	2.174	2.112	2.048	1.980	1.907	1.829
28	0.050	2.151	2.118	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
28	0.100	1.811	1.790	1.740	1.685	1.656	1.625	1.592	1.558	1.520	1.478
29	0.005	3.287	3.211	3.038	2.855	2.759	2.660	2.557	2.448	2.333	2.210
29	0.010	2.931	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
29	0.025	2.475	2.430	2.325	2.213	2.154	2.092	2.028	1.959	1.886	1.807
29	0.050	2.138	2.104	2.027	1.945	1.901	1.854	1.806	1.754	1.698	1.638
29	0.100	1.802	1.781	1.731	1.676	1.647	1.616	1.583	1.547	1.509	1.467
30	0.005	3.255	3.179	3.006	2.823	2.727	2.628	2.524	2.415	2.300	2.176
30	0.010	2.906	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
30	0.025	2.458	2.412	2.307	2.195	2.136	2.074	2.009	1.940	1.866	1.787
30	0.050	2.126	2.092	2.015	1.932	1.887	1.841	1.792	1.740	1.683	1.622
30	0.100	1.794	1.773	1.722	1.667	1.638	1.606	1.573	1.538	1.499	1.456
40	0.005	3.028	2.953	2.781	2.598	2.502	2.401	2.296	2.184	2.064	1.932
40	0.010	2.727	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
40	0.025	2.334	2.288	2.182	2.068	2.007	1.943	1.875	1.803	1.724	1.637
40	0.050	2.038	2.003	1.924	1.839	1.793	1.744	1.693	1.637	1.577	1.509
40	0.100	1.737	1.715	1.662	1.605	1.574	1.541	1.506	1.467	1.425	1.377
60	0.005	2.817	2.742	2.570	2.387	2.290	2.187	2.079	1.962	1.834	1.689
60	0.010	2.559	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
60	0.025	2.216	2.169	2.061	1.944	1.882	1.815	1.744	1.667	1.581	1.482
60	0.050	1.952	1.917	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
60	0.100	1.680	1.657	1.603	1.543	1.511	1.476	1.437	1.395	1.348	1.292
120	0.005	2.618	2.544	2.373	2.188	2.089	1.984	1.871	1.747	1.606	1.431
120	0.010	2.399	2.336	2.191	2.035	1.950	1.860	1.763	1.656	1.533	1.381
120	0.025	2.102	2.055	1.945	1.825	1.760	1.690	1.614	1.530	1.433	1.311
120	0.050	1.869	1.834	1.750	1.659	1.608	1.554	1.495	1.429	1.352	1.254
120	0.100	1.625	1.601	1.545	1.482	1.447	1.409	1.368	1.320	1.265	1.193
∞	0.005	2.433	2.359	2.187	2.000	1.898	1.789	1.669	1.533	1.364	1.016
∞	0.010	2.248	2.185	2.039	1.878	1.791	1.697	1.592	1.473	1.325	1.015
∞	0.025	1.993	1.945	1.833	1.709	1.640	1.566	1.484	1.388	1.269	1.012
∞	0.050	1.789	1.752	1.666	1.571	1.517	1.459	1.394	1.318	1.222	1.010
∞	0.100	1.571	1.546	1.487	1.421	1.383	1.342	1.295	1.240	1.169	1.008

A.8 taula. Estimazioaren laburpena

Parametroa	Baldintzak	Lagin-Banaketa	%(1-α)100 mailako Konfiantza-Tartea	Estatistikoa
μ	Populazio normala σ ezaguna	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} : N(0,1)$	$\overline{X} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$
μ	σ ezaguna $n>30$	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$	$\overline{X} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$
μ	Populazio normala σ ezezaguna	$\frac{\overline{X}-\mu}{S/\sqrt{n}} : t_{n-1}$	$\overline{X} \mp t_{\alpha/2;n-1}\frac{S}{\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$
μ_1 - μ_2	Populazio normal askeak σ_1 eta σ_2 ezagunak	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} : N(0,1)$	$(\overline{X}_1-\overline{X}_2) \mp z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
μ_1 - μ_2	Populazio askeak σ_1 eta σ_2 ezagunak $n_1>30$ eta $n_2>30$	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \approx N(0,1)$	$(\overline{X}_1-\overline{X}_2) \mp z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
μ_1 - μ_2	Populazio normal askeak σ_1 eta σ_2 ezezagunak $\sigma_1=\sigma_2$	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \approx t_{n_1+n_2-2}$	$(\overline{X}_1-\overline{X}_2) \mp t_{\alpha/2;(n_1+n_2-2)}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	$\frac{\overline{X}_1-\overline{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$
μ_1 - μ_2	Populazio normal askeak σ_1 eta σ_2 ezezagunak $\sigma_1 \neq \sigma_2$	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \approx t_g$	$(\overline{X}_1-\overline{X}_2) \mp t_{\alpha/2;g}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$
p	Populazio binomiala $n>30$	$\frac{\hat{p}-p}{\sqrt{pq/n}} \approx N(0,1)$	$\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$	$\frac{\hat{p}-p_0}{\sqrt{p_0q_0/n}}$
p_1 - p_2	Populazio binomial askeak $n_i>30$ eta $n_2>30$	$\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} \approx N(0,1)$	$\hat{p}_1-\hat{p}_2 \mp z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$	$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}}$
σ^2	Populazio normal	$\frac{(n-1)S^2}{\sigma^2} : \chi^2_{n-1}$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}\right)$	$\frac{(n-1)S^2}{\sigma_0^2}$
$\frac{\sigma_1^2}{\sigma_2^2}$	Populazio normal askeak	$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} : F_{(n_1-1),(n_2-1)}$	$\left(\frac{S_1^2/S_2^2}{F_{\alpha/2;n_1-1;n_2-1}}, \frac{S_1^2/S_2^2}{F_{1-\frac{\alpha}{2};n_1-1;n_2-1}}\right)$	$\frac{S_1^2}{S_2^2}$

A.9 taula. Hipotesi-kontrastearen laburpena

Param.	Baldintzak	Estatistikoa	Aldebiko Kontrastea *p*-balioa	Eskuin Kontrastea *p*-balioa	Ezker Kontrastea *p*-balioa
μ	Populazio normala σ ezaguna	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$p=2P(Z>\lvert z_p\rvert)$	$p=P(Z>z_p)$	$p=P(Z<z_p)$
μ	σ ezaguna $n>30$				
μ	Populazio normala σ ezezaguna	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$p=2P(t_{n-1}>\lvert t_p\rvert)$	$p=P(t_{n-1}>t_p)$	$p=P(t_{n-1}<t_p)$
μ_1 - μ_2	Populazio normal askeak σ_1 eta σ_2 ezagunak	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$p=2P(Z>\lvert z_p\rvert)$	$p=P(Z>z_p)$	$p=P(Z<z_p)$
μ_1 - μ_2	Populazio askeak σ_1 eta σ_2 ezagunak $n_1>30$ eta $n_2>30$				
μ_1 - μ_2	Populazio normal askeak σ_1 eta σ_2 ezezagunak $\sigma_1=\sigma_2$	$\frac{\overline{X}_1-\overline{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$p=2P(t_{n_1+n_2-2}>\lvert t_p\rvert)$	$p=P(t_{n_1+n_2-2}>t_p)$	$p=P(t_{n_1+n_2-2}<t_p)$
μ_1 - μ_2	Populazio normal askeak σ_1 eta σ_2 ezezagunak $\sigma_1\neq\sigma_2$	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	$p=2P(t_g>\lvert t_p\rvert)$	$p=P(t_g>t_p)$	$p=P(t_g<t_p)$
p	Populazio binomiala $n>30$	$\frac{\hat{p}-p_0}{\sqrt{p_0q_0/n}}$	$p=2P(Z>\lvert z_p\rvert)$	$p=P(Z>z_p)$	$p=P(Z<z_p)$
p_1 - p_2	Populazio binomial askeak $n_1>30$ eta $n_2>30$	$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}_1\hat{q}_1/n_1+\hat{p}_2\hat{q}_2/n_2}}$			
σ^2	Populazio normal	$\frac{(n-1)S^2}{\sigma_0^2}$	$s^2\leq\sigma_0^2\rightarrow$ $p=2P(\chi^2_{n-1}<\chi^2_p)$ $s^2>\sigma_0^2\rightarrow$ $p=2P(\chi^2_{n-1}>\chi^2_p)$	$p=P(\chi^2_{n-1}>\chi^2_p)$	$p=P(\chi^2_{n-1}<\chi^2_p)$
$\frac{\sigma_1^2}{\sigma_2^2}$	Populazio normal askeak	$\frac{S_1^2}{S_2^2}$	$F_p\geq 1\rightarrow$ $p=2P(F_{n_1-1,n_2-1}>F_p)$ $F_p<1\rightarrow$ $p=2P(F_{n_1-1,n_2-1}<F_p)$	$p=P(F_{n_1-1,n_2-1}>F_p)$	$p=P(F_{n_1-1,n_2-1}<F_p)$

$$s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$g=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1}+\frac{\left(s_2^2/n_2\right)^2}{n_2}}$$

B. ERANSKINA: ERREFERENTZIAK

Hodges, J.L.; Lehmann, E.L.. *Basic Concepts of Probability and Statistics*. Holden Day, San Francisco, 1970.

Ingelfinger, J.A.; Mosteller, F.; Thibodean, L.A.; Ware, J.H.. *Biostatistics in Clinical Medicine*. MacMillan Publishimg Co., Inc. New York, 1987.

Martinez de Lizardui, A.. *Estatistika Deskribatzailea*. Elhuyar Matematika Taldea, 1983.

Milton, J.S.. *Estadística para Biología y Ciencias de la Salud*. 3. argitalpen zabaldua. Interamericana-McGraw-Hill, 2007.

Quesada, V.; Isidoro, A.; López, L.A.. *Curso y Ejercicios de Estadística.* Alhambra Universidad, 1982.

Rosner, B.. *Fundamentals of Biostatistics*. 7. argitalpena. Wadsworth Cengage Learning, 2010.

Urkaregi, A.. *Bioestatistika*. UEU, 1991.

Zar, J.H.. *Biostatistical Analysis*. Fifth Edition. Pearson Prentice Hall, 2010.