BIOESTATISTIKA

udako euskal unibertsitatea



IRUÑEA 1985

B I O E S T A T I S T I K A

Arantza Urkaregi Etxepare

EUSKO JAURLARITZAREN LAGUNTZAZ

Jabegoa: U.E.U.ko MATEMATIKA Saila
Lege-gordailua: BI-888-85
ISBN: 84-398-3695-3
Inprimategia: BOAN S.A.

HITZAURREA

Eskutartean daukazun liburua, Zientzi Fakultateko Bioestatistika ko irakasleok prestaturiko apunteetan oinarrituta dago. Beraz, ez da lan pertsonalik izan, kolektiboa baizik. Niri dagokidanez, apunte horien prestakuntzan harturiko partez gainera, euskaratze-lan izan da nerea.

Bioestatistika gaia irakasten hasi ginenean, kezka bat planteatu genion gure buruari: gai matematiko batekiko interesa, nola sortu Biologia ikasten ari direnengan?

Dudarik gabe, Estatistika edozein arazo biologiko aztertzeko pre miazko tresna da, honen frogapena tesina, tesia edota ikerketa-lan bat burutzen ari den edonork eman dezakeelarik. Baina, Biologiako bigarren kurtsoko ikasleek ez dute horrela kontsideratzen Bioestatistika, kanpoko gaia baizik.

Hau dela eta, apunteak prestatzerakoan, hauxe zen gure asmoa: frogapen matematikoak aldean utzirik, adibide biologikoen bidez, Estatisti karen mamia azaltzea; formula estatistikoak, zertarako, nola eta noiz erabili behar diren intuitiboki aditzera ematea.

Seguraski ez dugu oso ondo bete helburua; dena dela, hemen duzu gure lanaren emaitza.

Liburua, hamar gaiz osoturik dago. Lehenengoa, Estatistika Deskribatzaileari dagokio.

Bigarrenetik laugarrenera, probabilitatea eta probabilitate-banaketak azaltzen dira, geroko gaietan erabili ahal izateko.

Bostgarrenetik aurrera, Estatistikaren beste atal desberdina aztertzen hasten gara, Inferentzia Estatistikoa hain zuzen. Atal hau izango da erabilpenik handienekoa, beste gaiek Estatistika Induktiboa ulertzeko behar diren oinarri guztiak ematen dizkigutelarik.

Bukatu aurretik, nere eskerrik beroenak Martxel Ensunzari eman behar dizkiot, eskaini didan laguntza paregabeagatik.

Leioan, 1985.eko Martxoan

Arantza Urkaregi

A U R K I B I D E A

HITZAURREA 3

AURKIBIDEA 5

I. GAIA: ESTATISTIKA DESKRIBATZAILEA

I.1. Sarrera 11

I.2. Definizioak eta orokortasunak 12

I.3. Taula estatistikoak 15

I.4. Maiztasun-banaketen adierazpide grafikoak

I.5. Maiztasun-banaketak. Zenbait forma..... 24

I.6. Aldagai estatistikoaren zenbakizko deskribapena 27

II. GAIA: PROBABILITATEARI BURUZKO OINARRIZKO EZAGUMENDUAK.

II.1. Sarrera 55

II.2. Probabilitatearen kontzepturako hurbilketa intuikorra 57

II.3. Probabilitate-kalkuluaren axiomatika ... 61

II.4. Eredu probabilistikoak 65

II.5. Baldintzazko probabilitatea eta gertaeren arteko independentzia 70

I. ERANSKINA 81

II. ERANSKINA 85

III. GAIA: ZORIZKO ALDAGAI DISKRETUAK ETA BEREN PROBABILITATE-LEGEAK

III.1. Aldagai estatistikoa eta zorizko aldagaia 101

III.2. Zorizko aldagai diskretuaren probabilitate-legea eta banaketa-funtzioa. Adierazpide grafikoak 103

III.3. Zorizko aldagai diskretuaren batezbestekoa eta bariantza 107

III.4. Banaketa binomiala 118

III.5. Banaketa hipergeometrikoa 122

III.6. Banaketa multinomiala 126

III.7. Poisson-en banaketa 129

IV. GAIA: ZORIZKO ALDAGAI JARRAIAK

IV.1. Zorizko aldagai jarraiaren banaketa- eta dentsitate-funtzioak 147

IV.2. Zorizko aldagai jarraiaren batezbestekoa eta bariantza. Tchebycheff-en teorema 151

IV.3. Zenbait banaketa jarrai 156

V. GAIA: OINARRIZKO LAGIN-TEORIARAKO SARRERA

V.1. Oinarrizko lagin-teoriari buruzko oro kortasunak 185

V.2. $\bar{x}$ laginaren batezbestekoaren banaketa .. 188

V.3. Bi laginen batezbestekoen arteko kenduraren banaketa 194

V.4. Laginaren proportzioaren banaketa 198

V.5. Bi laginen proportzioen arteko kenduraren banaketa 200

V.6. χ^2, Pearson-en ji karratu banaketa 201

V.7. Student-en "t" banaketa 203

V.8. Fisher-Snedecor-en "F" banaketa 210

VI. GAIA: ESTIMAZIOA

VI.1. Inferentzia estatistikoaren beharra ... 229

VI.2. Puntu-estimazioa 230

VI.3. Konfidantza-tarteak 235

VI.4. Konfidantza-tarte arruntak 238

VI.5. Lagin-tamainuaren determinazioa 245

VII. GAIA: HIPOTESI-KONTRASTEA

VII.1. Sarrera eta adibideak 265

VII.2. Planteamendu orokorra 280

VII.3. Ohizko zenbait test 286

VIII. GAIA: JI KARRATU BANAKETAREN APLIKAZIOAK

VIII.1. Doikuntz egokitasunaren probak 323

VIII.2. Bi faktoreren arteko menpekotasun-ala independentzi proba 330

VIII.3. 2 x 2 kontingentzi taulak. Yates-en zuzenketa 336

VIII.4. Homogenotasun-probak 338

IX. GAIA: ERREGRESIOA ETA KORRELAZIOA

IX.1. Sarrera 357

IX.2. Erregresio-lerroa 358

IX.3. Aurresana eta minimo karratuen metodoa. 358

IX.4. Pearson-en korrelazio linealaren koefizientea 365

IX.5. Estimazio-errore standarda 369

IX.6. Erregresio eta korrelazio-arazoetarako konfidantza-tarteak 371

X. GAIA: BARIANTZ ANALISIA

X.1. Sarrera 387

X.2. Aldakuntz faktore bakar batekiko bariantz analisia 387

X.3. Bi aldakuntz faktore askerekiko bariantz analisia 393

X.4. Hiru aldakuntz faktore askerekiko bariantz analisia 400

TAULAK 415

HIZTEGIA 427

I. GAIA: ESTATISTIKA DESKRIBATZAILEA

I.1.- Sarrera.

I.2.- Definizioak eta orokortasunak.

I.3.- Taula estatistikoak.

I.4.- Maiztasun-banaketen adierazpide grafikoak.

I.5.- Maiztasun-banaketak. Zenbait forma.

I.6.- Aldagai estatistikoaren zenbakizko deskribapena:

I.6.1.- Joera zentraleko estatistikoak.

I.6.2.- Sakabanapen-estatistikoak.

I.6.3.- Posizio-estatistikoak.

I.6.4.- Forma-estatistikoak.

ESTATISTIKA DESKRIBATZAILEA

I.1.- SARRERA.

Arazo biologikoak ebazteko metodo estatistikoen aplikazioari "Bioestatistika" edo "Biometria" deitzen zaio.

Definizio hau ulertzeko, Estatistika terminoa definitu behar dugu: Natur fenomenoetan oinarrituta dauden zenbakizko datuak aztertzeko, eta baieztapen edo lege batzu onar daitezkeen ebazteko metodoak ematen dituen zientzia, Estatistika da.

Natur fenomenoetan ondoko fenomenoak sartuko ditugu: Al de batetik, natura bizidun eta bizigabean pertsonaren kontrolik gabe gertatzen diren gertaerak; eta beste aldetik, ikerlariak partzialki eragindakoak, bere kontrolaren menpean. Azken fenomeno hauei experimentu izena emango diegu.

Adibidez, kilkirren kantua, leka baten ilar-kopurua edo txita bat hazteko behar den denbora natur fenomenoak dira, eta bai adrenalinarekiko arratoien taupada ere, nahiz eta ikerlariak experimentaketen bidez fenomenoan eragin.

Fenomeno hauek, Estatistikaren bidez aztertu ahal izateko, bete behar dituzten baldintzak hurrengoak dira:

a) Baldintza egonkorren multzo baten menpean hainbat al diz errepika daitezke.

b) Proba bakoitzean ezin daiteke lortuko den emaitza aurresan.

c) Emaitza bakoitzaren maiztasun erlatiboak (emaitza gertatu deneko aldi-kopuruaren eta proba-kopuruaren arteko zatidurak) zenbaki batetan egonkortzera jotzen du.

Egindako proben kopurua emendatuz doan neurrian, egonkortasun hau gero eta handiagoa izango da.

Edozein azterketa estatistikoren oinarria, natur fenomenoetan gertatzen diren emaitzen erregulartasun edo egonkortasuna da.

Aipatutako baldintza hauek betetzen dituenean eta, gainera, ikerlariak eragindako experientzia baten emaitza denean, "zorizko experimentu" izena emango diogu fenomenoari.

Estatistika bi arlo nagusitan banatzen da, beraien artean oso desberdinak izanik, bai helburuetan zein kalkuluetarako erabiltzen diren lanabesetan:

- Estatistika deskribatzailea. Beregain, ondokoak hartzen dituelarik:

  a) Ikerketa burutzeko beharrezkoak diren datuak biltzea.

  b) Datuen antolakuntza, taula eta grafiko moduan aurkeztuz.

  c) Fenomenoaren ezaugarri garrantzitsuenak deskribatzen dituen zenbait kantitate kalkulatzea. Kantitate hauei estatistikoak deitzen zaie.

- Estatistika induktiboa. Dauzkan helburuak, hauxek dira:

  a) Emaitzen inferentziak edo orokorpenak burutzea eta beraien konfidantza-maila neurtzea.

  b) Aurreko etapan lortutako emaitzak interpretatzea.

I.2.- DEFINIZIOAK ETA OROKORTASUNAK.

1.D.- Azterketa estatistikoaren helburuak diren elemen-

tuen multzoari "populazioa" deituko diogu.

Populazioak finituak edo infinituak izan daitezke, baina, beti zehaztasunez definitu behar dira, zalantzagarritasunik gerta ez dadin.

2.D.- Lagina: Populazioaren azpimultzoa da.

3.D.- Unitate estatistikoa edo alea, populazioaren elementu bakoitza da.

4.D.- Laginaren (edo populazioaren) tamainua, laginaren (edo populazioaren) elementuen kopurua da.

Adibidea eta oharrak

50 urte baino zaharragoak diren pertsonen odolaren pH-ren balioa aztertu nahi badugu, hauxe izango da populazioa: 50 urte baino zaharragoak diren pertsona guztiek osotzen dutena.

Multzo honen elementu-kopurua infinitua dela kontsidera dezakegu.

Bere azterketa estatistikorako, 50 urte baino zaharragoak diren pertsona guztien pH-ren balioa aztertzea ezinezkoa denez, populazioaren azpimultzo finitu bat hartu beharko dugu. Baina, fenomenoaren azterketa estatistikoa burutzeko, laginek bete behar duten baldintza garrantzitsuenetariko bat ondokoa da: zorizkoak eta populazioaren errepresentagarriak izatea.

5.D.- Izaera, aztertzen den populazioaren propietatea da.

6.D.- Modalitatea, izaerak aurkez dezakeen era edo egoera da.
Aztertutako ale bakoitzak, izaeraren modalitate bat eta bakar bat aurkeztu behar du; hau da, modalitateek baztertzaile eta exhaustiboak izan behar dute.

Aurreko adibidean, odolaren pH-ren balioa izango da izaera, eta, modalitateak, pH honek har ditzakeen balio guztiak.

Izaeren sailkapena. Izaerak, mota honetakoak izan daitezke:

a) 7.D.- Kualitatiboak. Beren modalitateak ezin daitezke era numerikoan adieraz.

b) 8.D.- Kuantitatiboak. Beren modalitateak neurgarriak dira.

Adibideak: Sexua, profesioa edo begien kolorea, izaera kualitatiboak dira. Ordea, tenperatura, pisua, adina, izaera kuantitatiboak.

9.D.- Izaera kuantitatiboaren modalitateen neurketari dagokion zenbaki-multzoari aldagai estatistikoa deitzen zaio.

Aldagai estatistikoak modu honetan sailka daitezke:

a) 10.D.- Aldagai estatistiko diskretuak: Modalitateek hartzen dituzten balioak zenbaki diskretuak dira.

b) 11.D.- Aldagai estatistiko jarraiak: Tarte baten barnean, edozein izan daiteke modalitateak hartzen duen balioa.

Adibideak: Leka baten ilar-kopurua aldagai estatistiko diskretua da. Odolaren pH-ren balioa, aldiz, jarraia.

Oharra.- Aldagai estatistiko diskretu eta jarraien arteko desberdintasuna guztiz erlatiboa da; neurrien akatsak direla kausa azkenean, beti aldagaien balioak zenbaki diskretuak izango bait dira; baina, teorikoki, desberdintasuna argi dago eta, praktikoki, aldagaiak balio desberdin asko hartzen dituenean, aldagai estatistikoa jarraitzat hartuko da.

I.3.- TAULA ESTATISTIKOAK.

Estatistika deskribatzaileak egiten duen lehenengoa hau xe da: aztertu nahi den fenomenoari dagozkion datuak biltzea.

Lan hau arazo delikatua da, fase honetako erroreak datuen azterketa faltsutzen du eta. Kasu bakoitzean, lana burutzeko erabide eta teknikak, egiten den azterketaren eta ikertzailaren erizpiden menpean daude.

Datuak bildu eta gero, tauletan ordenatuta aurkeztea da bigarren lana.

Taula estatistikoetan, ondokoak adierazi behar dira:

a) Izaeraren modalitateak: c_i (Izaera kuantitatiboa bada, txikienetik handienera ordenaturik).

b) Maiztasun absolutuak: f_i, modalitate bakoitzari dagokion populazioaren edo laginaren ale-kopurua.

   Oharra: Populazioaren edo laginaren tamainua N bada, ondokoa bete behar da: $\sum_{i=1}^{N} f_i = N$

c) Izaera kuantitatiboa agertzen denean, F_i maiztasun metatuak adieraztea interesgarria da:

$$F_i = \sum_{j=1}^{i} f_j$$

   Batzutan, maiztasun erlatiboak $h_i = \frac{f_i}{N}$ eta maiztasun metatu erlatiboak $H_i = \frac{h_i}{N}$ seinalatzea interesa dakiguke.

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

a) Izaera kualitatiboa:

Adibidea: Intsektu-komunitate baten familiaka sailkaturiko "Heteroptero" ordena delakoaren ale-kopurua.

Familia	f_i
Alydidae	2
Anthocoridae	37
Coeridae	2
Ligacidae	318
Miridae	373
Pentatomidae	25
Tingidae	69
Heteropteroen totala	826

b) Izaera kuantitatiboa:

b-1) A.e. diskretua:

Adibidea: Landare-ekologiari buruzko experimentua. Landare mota bat nola banatzen den aztertzeko, area bat harturik, 500 karratutan zatitzen dugu.

Landare kopurua karratuko	f_i
0	181
1	118
2	97
3	54
4	32
5	9
6	5
7	3
8	1
Totala	500

b-2) A.e. jarraia:

Aldagaiak hartzen dituen balioen kopurua handia denean, maiztasun elkartuen izeneko metodoa erabiltzen da.

Beraren teknika ondokoa da:

i) Aldagai estatistikoaren heina edo aldakuntzaren tarte osoa determinatu. Horretarako, aldagaiak hartzen dituen baliorik handien eta txikien arteko kendura aurkitu behar dugu.

ii) Tartea, maiztasun-banaketa deskribatu nahi den bezain besteko azpitarteetan (klaseak deitutakoak) zatitzen da.

Zenbaki egokirik ez dago, baina muturreko kasuak ekidingo ditugu ahal den neurrian. Klase-kopuru txikiak ezin lezake neurrien arteko desberdintasuna erakuts eta, beraz, informazioa gal liteke. Gehiegizko klase-kopurua egoteak metodoa korapila lezake eta metodo honen bidez bilatu nahi duguna kalkuluak erraztea da, hain zuzen ere.

iii) Klaseak, beraien muturrak seinalatuz mugatzen dira. Mutur haiei beheko eta goiko klase-mugak deituko diegu, $C_i=(l_i,l_{i+1})$ adierazirik.

Klase muga hauek inolako zalantzagarritasunik gabe hautatuko dira. Hau da, datu bakoitzari dagokion klasea jakin beharko dugu.

Hau lortzeko bi teknika daude:

1) Datuek dituzten zifra adierazgarriak baino zifra gehiagorekin klase-mugak ematea:

2) Klase-mugak $(l_i,l_{i+1}]$ moduan eman. Honela, klase honi dagozkion datuak l_i baino handiago eta l_{i+1}-en berdin edo txikiagoak izango dira.

iv) Maiztasunak kontatu, eta aurreko kasuetan egin dugun antzera, taula eratu.

Oharra: Klaseak "muga-neurri" izenekoen bidez defini ditzakegu. Hauxek, klaseari dagozkion neurririk handiena eta txikiena dira, beraien prezisioa kontutan harturik. Kasu honetan, bi klase kontsekutiboren muga-neurriek ezin dezakete balio berbera har.

Adibidea: mmtan emandako 75 banana-orrien nerbiorik handienaren luzera:

113-126-139-171-119-134-170-144-153-175-126-180-139-126-149-117-154-122-137-140-142-168-152-149-129-148-175-168-104-144-134-145-122-137-140-153-149-169-132-147-150-152-118-161-153-177-146-152-112-140-145-152-151-145-112-162-188-156-160-170-165-156-157-161-155-162-155-170-160-172-158-155-182-132-131

Oharra: Adibide honetan aztertutako aldagaia, teorikoki jarraia da, baina praktikoki, neurriak hartzen dituguneko prezisioa dela kausa, balio etenak agertzen zaizkigu.

Dena dela, neurri desberdin edo modalitate asko daudenez, klaseetan elkartutako maiztasun-taula estatistikoaren bidez adierazi behar da aldagaia.

Klaseen muga-neurriak	Klaseen muga errealak	Erdiko puntuak	Maiztasunak
100-109	99,5-109,5	104,5	1
110-119	109,5-119,5	114,5	6
120-129	119,5-129,5	124,5	6
130-139	129,5-139,5	134,5	9
140-149	139,5-149,5	144,5	15
150-159	149,5-159,5	154,5	17
160-169	159,5-169,5	164,5	10
170-179	169,5-179,5	174,5	8
180-189	179,5-189,5	184,5	3
			75

I.4.- MAIZTASUN-BANAKETAREN ADIERAZPIDE GRAFIKOAK.

a) Izaera kualitatiboa

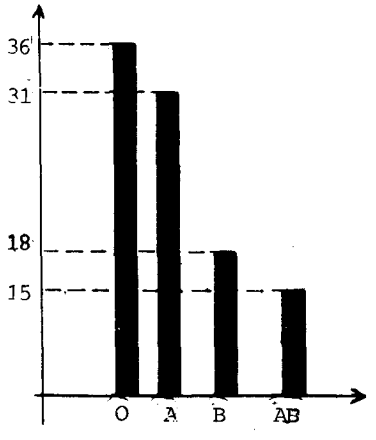
a-1) Barra-diagrama:

Modalitateak (abzisa-ardatzean) eta maiztasunak (ordenatu-ardatzean) ardatz kartesiarren gainean aurkeztuko dira, batzuen eta besteen arteko egokitasuna barra-sistema baten bidez deskribatuz.

Adibidea:

Odol-taldeka sailkaturiko 100 pertsonatako laginaren aleen banaketa.

Odol-taldea	f_i
O	36
A	31
B	18
AB	15
	100



a-2) Sektore-diagrama:

Zirkulu baten azalera osoa, modalitate desberdinen arauerako zabaltasunak dauzkaten sektore zirkularretan zatituz, lortzen da.

Adibide berberari jarraituz:

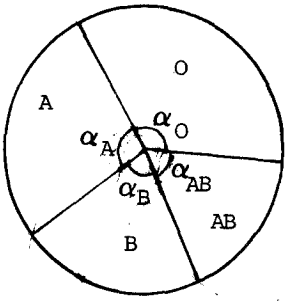
100 ------- 360°

36 ------- α_O ; α_O = 129,6°

31 ------- α_A ; α_A = 111,6°

18 ------- α_B ; α_B = 64,8°

15 ------- α_{AB} ; α_{AB} = 54°



b) Izaera kuantitatiboa. Aldagai estatistiko diskretua.

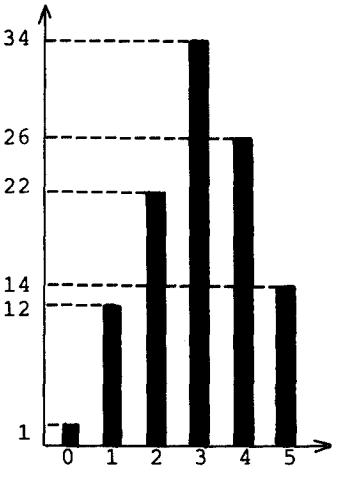
b-1) Barra-grafikoa:

Aurreko kasuaren analogoa da, baina, kasu honetan modalitateak neurgarriak izanez gero, beren ordenak garrantzia duela kontutan harturik.

Adibidea:

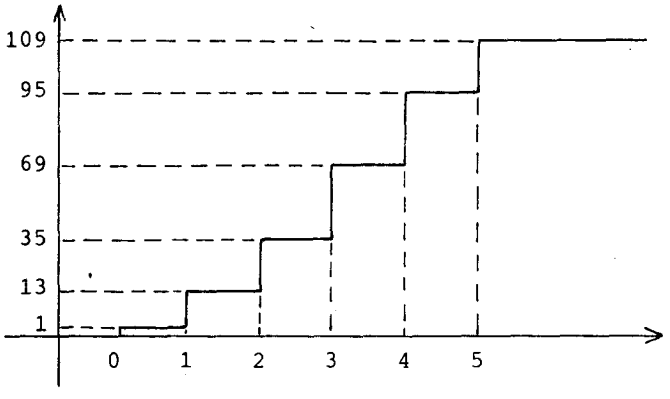
6 kumetako 109 kumaldi aztertuz, hurrengo "ar-kopuru/kumaldi" direlakoak lortu genituen:

Ar-kopurua	f_i	F_i	H_i
0	1	1	0,009
1	12	13	0,119
2	22	35	0,321
3	34	69	0,633
4	26	95	0,872
5	14	109	1



b-2) Grafiko metakorra:

$F(x)=F_i$ delako funtzioaren grafikoa da. Eskailera-itxura edukiko du beti, aldagairen bi balioren artean, (x_i, x_{i+1}), maiztasun metatua ez bait da aldatzen eta x_{i+1}puntua hartzen denean, grafikoak salto egiten bait du.



c) Izaera kuantitatiboa. Aldagai estatistiko jarraia.

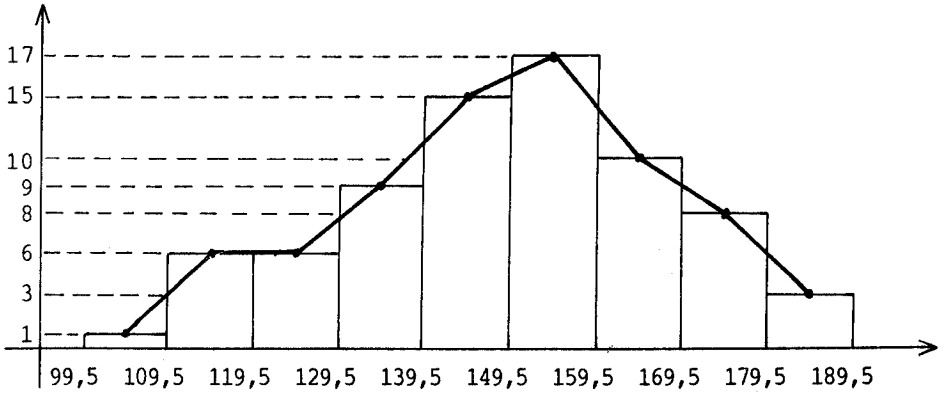
c-1) Histograma:

Ardatz kartesiarren gainean, hurrengo moduko errektangelu-sistema eraikitzen da: oinarriak, klaseen muga errealek definituak izango dira eta errektangelu bakoitzak, errepresentatzen duen klasearen maiztasun absolutuaren arauerako azalera edukiko du.

Oharra:

1) Klaseek tamainu berdina badaukate, maiztasunen arauerakoak edo berdinak izango dira altuerak.

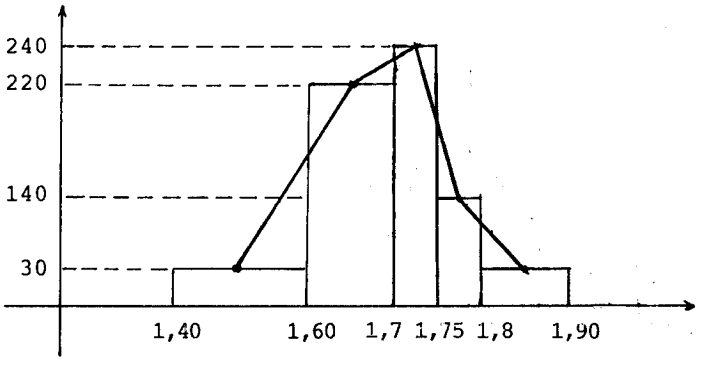
1. adibidea: 75 banana-orrien nerbiorik handienaren luzera.



2) Tamainu desberdinak badauzkate, errektangelu bakoitzaren altuera, maiztasun eta klasearen luzeren arteko zatidura (edo zatiduraren arauerakoa) izango da.

2. adibidea:

Altuerak	f_i	errektangeluen altuerak f_i/c_i
1,40-1,60	6	30
1,60-1,70	22	220
1,70-1,75	12	240
1,75-1,80	7	140
1,80-1,90	3	30



c-2) Maiztasun absolutuen poligonoa:

Histogramako errektangeluen goiko oinarrien erdiko puntuak lotzen dituen lerro poligonala da.

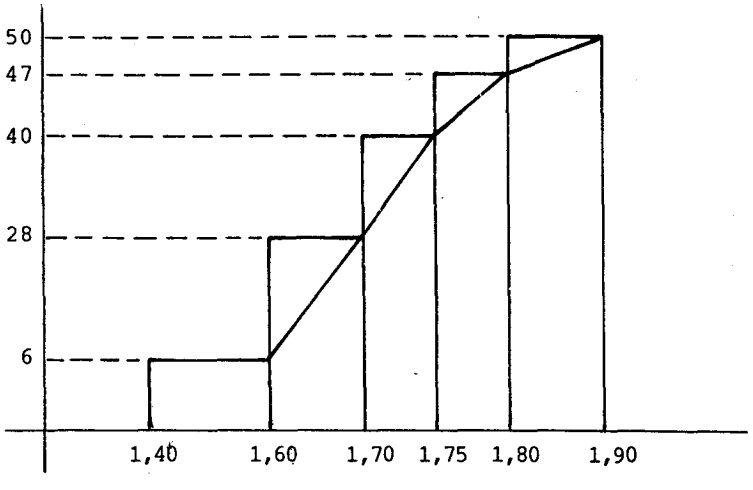
Begira itzazu aurreko adibideak.

c-3) Maiztasun metatuen histograma:

Maiztasun metatuen arauerako altuerak eta oinarritzat klase-tarteak dauzkan errektangelu-sistema da.

2.adibidea:

Altuerak	f_i	F_i
1,40-1,60	6	6
1,60-1,70	22	28
1,70-1,75	12	40
1,75-1,80	7	47
1,80-1,90	3	50



c-4) Maiztasun-poligono metakorra:

(l_{i+1}, F_i) puntuak lotuz eraikitako lerro poligonala da, $l_{i+1}=C_i$ klasearen goiko muga erreala eta $F_i=C_i$ klaseari dagozkion maiztasun metatua direlarik.

I.5.- MAIZTASUN-BANAKETAK. ZENBAIT FORMA.

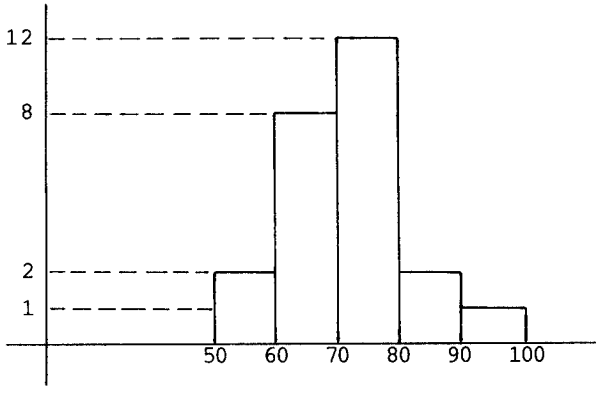
Adibidez, adin militarrean dauden gizonen pisu-populazioari dagozkion laginak atera nahiko bagenitu, aztertutako aldagai estatistikoaren adierazpen grafikoak egin genitzake.

Laginaren tamainua emendatuz eta klaseak txikiagotuz doan neurrian, aurkezpen grafiko hauek gero eta zehatzagoak izango dira. Mugan, histograma, aldagaiaren maiztasun-banaketa adieraziko duen kurba bihurtuko da.

Adibidea:

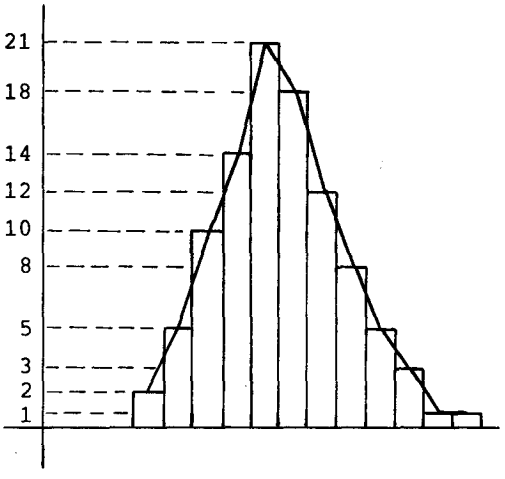
- 25 pertsonatako lagina:

Pisua	f_i
50- 60	2
60- 70	8
70- 80	12
80- 90	2
90-100	1

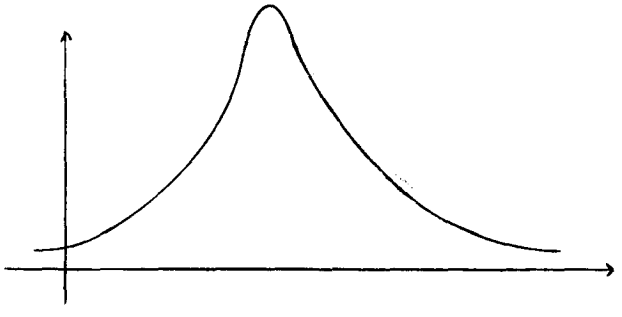


- 100 pertsonatako lagina:

Pisua	f_i
50- 55	2
55- 60	5
60- 65	10
65- 70	14
70- 75	21
75- 80	18
80- 85	12
85- 90	8
90- 95	5
95-100	5



- laginaren tamainua ∞-rantz doanean:



Maiztasun-banaketek adieraz ditzaketen zenbait forma hauxek dira:

a) Kanpai-forma.

Naharoena da, asimetria handiagoa edo txikiagoa aurkez dezakeelarik.

b) Bimodala (bi gailur dauzkana).

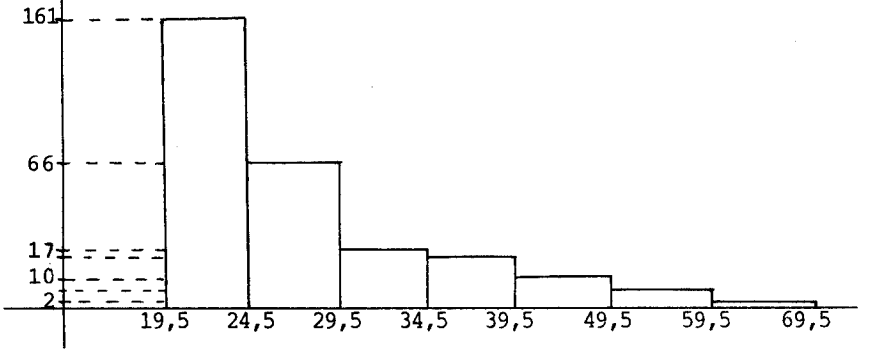
Maiztasun-banaketak bi maximo aurkezten ditu.

Ondorio garrantzitsuak atera daitezke. Adibidez, insektu-laginean neurriak bimodalki banatzen direla aurkituko bagenu, ondorioa ondokoa izan liteke: laginean, espezie edo arraza desberdinen presentzia edo sexu-desberdintasuna edo beste arrazoi partikularrak direla kausa, populazioa dimorfikoa dela.

c) L-ren edo U-ren formatako banaketak.

Adibidea: 1981.ean, emaztearen adinaren arauera sailkaturiko ezkontza-banaketa

Emaztearen adina	f_i
20tatik 24etara	161
25tatik 29tara	66
30etatik 34etara	17
35tatik 39tara	16
40tatik 49tara	10
50etatik 59tara	7
60 edo gehiago	2



I.6.1.- ALDAGAI ESTATISTIKOAREN ZENBAKIZKO DESKRIBAPENA.

Biologi azterketa estatistikoaren tratamendurako azter tzen dugun etapa deskribatzailean, laburpen-eredua behar izaten dugu. Eredu honek, datuen azterketa permitituko digu modu maneigarrian, bide batez, ondorio posibleak ateratzeko gaitasu na emango digularik.

Lehenago ikasitako diagramak, laburpen-mota bat dira. Hala ere, zenbakizko laburpena beharrezkoa da helburu gehienen tzat. Laburpen honek, laburki eta zehazki deskribatuko ditu az tertutako maiztasunen propietateak.

Zenbaki horiei, estatistiko deskribatzaileak deituko diegu.

Sailkapena:

a) Joera zentraleko estatistikoak
b) Sakabanapen-estatistikoak
c) Posizio-estatistikoak
d) Forma-estatistikoak

Estatistiko baterako, Yulek definitutako nahigarri diren zenbait propietateren deskribapena interesgarria da, estatistiko baten erabilera beste batena baino egokiagoa noiz den jakin ahal izateko.

Hauxek dira propietateak:

1) Modu objetiboz definitu behar da.
2) Datu guztiak erabili behar dira.
3) Esangura konkretua izan behar du.
4) Kalkulatzeko erraza izan behar du.
5) Kalkulu algebraikoaz egiteko modukoa izan behar du.
6) Fluktuazio muestralekiko ez du oso sentikorra izan behar.

I.6.1.- Joera zentraleko estatistikoak:

Balio hauek, neurriak beren inguruan biltzen direla adierazten dute nolabait.

Orain, estatistiko hauetariko batzu ikusiko ditugu:

1) Batezbesteko aritmetikoa.

Adierazkorrena eta erabiliena da.

Datu indibidualen baturaren eta laginaren elementu-kopuruaren arteko zatidura da. Hauxe da, n aletako lagin bat hartzen badugu eta bere ezaugarri estatistiko bat azterturik, $x_1, x_2, \ldots, x_n$ neurriak lortzen baditugu, $\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$ izango da.

Neurriak errepikatuko balira, maiztasunak sortuko lirateke eta batezbesteko aritmetikoaren adierazpena

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

izango litzateke, k modalitate edo neurri desberdinak $x_1, x_2, \ldots, x_k$ daudela eta berauek $f_1, f_2, \ldots, f_k$ maiztasunez banatzen direla suposaturik.

Analogoki, aldagai estatistikoaren banaketa klasez emango balitz, klase bakoitzaren zenbaki errepresentagarria bere erdiko puntua edo klase-marka, $x_i = \frac{l_i+l_{i+1}}{2}$, izango litzateke, batezbesteko aritmetikoa, $\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$ formula berberaz kalkulatuz.

Adibideak:

1) Igel-espezie batetan, 40 aleren luzera.

Luzera (mm.tan)	f_i	x_i	$x_i f_i$
28-30	2	29	58
30-32	10	31	310
32-34	16	33	528
34-36	8	35	280
36-38	4	37	148
	40		1324

$$\bar{x} = \frac{1324}{40} = 33,1 \text{ mm.}$$

2) Intsektu zolduen kopurua, 5 intsektutako laginetan

Ints.zol.	lag-kop	$x_i f_i$
0	12	0
1	62	62
2	63	166
3	36	108
4	5	20
5	2	10
	200	366

$$\bar{x} = \frac{366}{200} = 1,83$$

Batezbestekoaren kalkulua errazteko metodoa.

Jatorri- edo eskala-aldaketa eginez, askoz errazagoak gertatzen dira batezbesteko aritmetikoa kalkulatzeko eragiketak sarritan.

Hau, $z = \frac{x-a}{c}$, a eta c edozein konstante izanik, aldagai-aldaketaz lortzen da. Baina errazpide egokia lortzeko, "a" aldagaiaren balio zentralen bat eta "c" klaseen tamainua (aldagaia klasez emanda badago) izatea komenigarria da.

Kasu honetan, $\bar{z} = \frac{\bar{x}-a}{c}$, hots, $\bar{x} = a + \bar{z}.c$ dela frogatzen da.

Adibidea:

Entrozitoen diametroen batezbestekoa determinatzeko, 2.000 neurtu ziren, hurrengo emaitzak lortu zirelarik:

Diametroa	f_i	x_i	$z_i = \frac{x-7,2}{0,4}$	$z_i f_i$
5,4 - 5,8	5	5,6	-4	-20
5,8 - 6,2	78	6	-3	-234
6,2 - 6,6	144	6,4	-2	-288
6,6 - 7	479	6,8	-1	-479
7 - 7,4	542	7,2	0	0
7,4 - 7,8	358	7,6	1	358
7,8 - 8,2	279	8	2	558
8,2 - 8,6	99	8,4	3	297
8,6 - 9	15	8,8	4	60
9 - 9,4	1	9,2	5	5
	2000			257

$\bar{z} = \frac{257}{2000} = 0,1285 = \frac{\bar{x}-7,2}{0,4}$; beraz, $\bar{x} = 0,1285 \cdot 0,4 + 7,2 = 7,2514$

Beste batezbesteko-mota batzu

Batezbesteko geometrikoa: $G = \sqrt[n]{(x_1)^{f_1}(x_2)^{f_2}\ldots(x_k)^{f_k}}$

Batezbesteko koadratikoa: $Q = \sqrt{\frac{x_1^2 f_1 + x_2^2 f_2 + \ldots + x_k^2 f_k}{n}}$

Batezbesteko harmonikoa: $H = \frac{n}{\sum_{x=1}^{k} \frac{f_i}{x_i}}$

2) Mediana

Mediana, Me, alde bakoitzean ale-kopurua berdina daukan aldagaiaren balioa da, txikienetik handienera ordenatuta daudela suposaturik; hau da, alde bakoitzean elementuen 50%a uzten duen aldagaiaren balioa.

Aldagai estatistikoa diskretua denean, hurrengo kasuak azter daitezke:

1) balio-kopuru bakoitia. Mediana, $\frac{n+1}{2}$ tokia betetzen duen balioa izango da.

2) balio-kopuru bikoitia. Kasu honetan, ez dago erdiko tokia betetzen duen baliorik eta, konbentzionalki, mediana, $\frac{n}{2}$ eta $\frac{n}{2}$ + 1 toki zentralak betetzen dituzten balioen arteko batezbesteko bezala kalkulatzen da.

Adibidea: 30 urtetako pertsonen pisuak.

1) 7 elementutako lagina: 62 - 65 - 70 - 71 - 77 - 79 - 86

Me = 71

2) 8 elementutako lagina: 62 - 65 - 70 - 71 - 77 - 84 - 86 - 88

$Me = \frac{71+77}{2} = 74$

Aldagai estatistikoa jarraia eta klasez bildua denean, mediana, interpolazioz kalkulatzen da, hurrengo formularen bidez:

$$Me = l_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} c_i$$

non l_i = mediana daukan klasearen beheko muga

N = laginaren tamainua

F_{i-1} = mediana daukanaren aurreko klasearen maiztasun metatua

f_i = mediana daukan klasearen maiztasuna

c_i = mediana daukan klasearen tamainua

diren.

Adibidea: Igel-mota baten 40 aleren luzera.

Luzera (mm.)	f_i	F_i
28 - 30	2	2
30 - 32	10	12
32 - 34	16	28
34 - 36	8	36
36 - 38	4	40

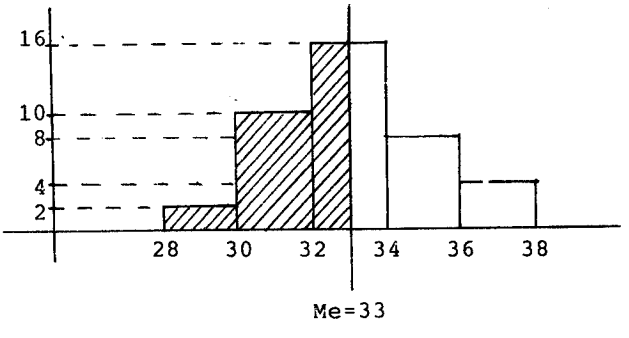
$\frac{N}{2} = \frac{40}{2} = 20$ dela eta, (32 - 34) klasean egon behar du medianak.

Beraz, $Me = 32 + \frac{20-12}{16} \cdot 2 = 32 + 1 = 33$

Grafikoki ikusirik, ondoko propietatea betetzen du medianak: aldagaiaren balio honetatik pasatzen den ordenatu-ardatzaren zuzen paraleloak, histogramaren azalera erdibitzen du.

Azalera marratua: $2 \cdot 2 + 2 \cdot 10 + 1 \cdot 16 = 40$

Marratu gabeko azalera: $1 \cdot 16 + 2 \cdot 8 + 2 \cdot 4 = 40$



3) Moda

Maiztasun handiena daukan aldagaiaren balioa da; hots, aldi-kopuru gehienetan agertzen den neurria. Maiztasun-banaketa batetan, grafikoaren maximoa izango litzateke.

Bi moda izaten dituenean, bimodala deitzen da banaketa.

Banaketak U forma baldin badauka, antimoda deritzo beheko puntuari.

Adibidea: Intsektu zolduen banaketa, 5 intsektutako 200 laginetan.

x_i	f_i
0	12
1	62
2	83
3	36
4	5
5	2

Mo = 2

Aldagai estatistikoa klasez elkartuta dagoen kasuan, maiztasun-kontzentrazio handiena daukan klasea ezagut dezakegu soilik definizio honen bidez. Klase honi, "klase modala" deituko diogu, eta kasu honetan, interpolazioz lortzen da modaren balioa, hurrengo adierazpenaren bidez, hain zuzen ere:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1+\Delta_2} \cdot c_i$$

non l_i = klase modalaren beheko muga

$$\Delta_1 = \frac{f_i}{c_i} - \frac{f_{i-1}}{c_{i-1}}$$

$$\Delta_2 = \frac{f_i}{c_i} - \frac{f_{i+1}}{c_{i+1}}$$

f_i = klase modalaren maiztasuna

c_i = klase modalaren tamainua

diren

Adibidea: Igel-mota baten 40 aleren luzera

Luzera (mm.)	f_i
28 - 30	2
30 - 32	10
32 - 34	16
34 - 36	8
36 - 38	4

Klase modala: (32 - 34)

$\Delta_1 = \frac{16}{2} - \frac{10}{2} = 8 - 5 = 3$

$\Delta_2 = \frac{16}{2} - \frac{8}{2} = 8 - 4 = 4$

$Mo = 32 + \frac{3}{3+4} \cdot 2 = 32,857$

Estatistiko baterako Yulenen propietateak gogoraturik, on dokoak frogatzen dira:

- batezbesteko aritmetikoak 1.,2.,3. eta 4. propietateak ondo egiaztatzen ditu, baina beraren kalkulua mediana edo modarena baino zailagoa da eta fluktuazio muestrale kiko sentikorra izan daiteke, muturretan jartzen diren balio estrainoek eragina daukatelako batezbestekoan.

- medianak 1.,3.,4. eta 6. propietateak betetzen ditu, 2. propietatea ukatzen duelarik, zeren datu guztiak beraien arteko ordena finkatzeko soilik erabiltzen bait dira. Honek 6. propietatea betetzen laguntzen du, medianan balio estrainoek eraginik ez daukate eta.

- modak Yulenen 1.,3. eta 4. propietateak betetzen ditu.

I.6.2.- Sakabanapen-estatistikoak.

Zenbakiz deskribatzea interesatzen zaigun lagin baten beste ezaugarri inportante bat, laginaren balioen kontzentrazioa da.

Esate baterako, zoriz aukeratuak izan diren 4 pertsonatako laginaren Sheldonp-indizea (altuera/$\sqrt[3]{\text{pisua}}$) kalkulatuko bagenu, 38,5 - 40 - 42,5 - 43 balioak lor genitzake. Beraien batezbestekoa 41 izango litzateke.

Ordea, 4 kirolarirentzako indize bera kalkulatuko bagenu, 40,5 - 40,7 - 41,2 - 41,6 balioak lor genitzake. Beraien batezbestekoa ere 41 izango litzateke, baina, bilduagoak izango lirateke balioak.

Bi laginen arteko ezberdintasun hau nabarerazteko, hurrengo sakabanapen-estatistikoak erabil ditzakegu:

a) Anplitudea edo aldakuntz tartea.

Laginaren neurri handien eta txikienaren arteko kendura da.

1. lagina: $R_1 = 43 - 38,5 = 4,5$

2. lagina: $R_2 = 41,6 - 40,5 = 1,1$

Interes gutxi dauka, balio estrainioek edo akatsdunek eragin handia bait daukate, baina, kalkulatzeko erraza denez gero, batzutan erabiltzen da.

b) Bariantza eta desbidazio standarda.

Sakabanapen-estatistikorik inportanteena desbidazio standarda da. Bariantzaren erro karratua da berau. Bariantza, batezbestekoarekiko desbidazioen karratuen batezbestekoa da. Hauxe da:

$$V(x) = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{N} \qquad \sigma(x) = \sqrt{V(x)}$$

Adibidea:

1.lagina: $V(x_1) = \frac{(38,5-41)^2+(40-41)^2+(42,5-41)^2+(43-41)^2}{4} = 3,375$

$\sigma(x_1) = \sqrt{3,375} = 1,837$

2.lagina: $V(x_2) = \frac{(40,5-41)^2+(40,7-41)^2+(41,2-41)^2+41,6-41)^2}{4} = 0,185$

$\sigma(x_2) = \sqrt{0,185} = 0,43$

Bariantzaren kalkulu praktikoa:

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

Frogapena: $V(x) = \frac{\sum_{i=1}^{k} (x_i-\bar{x})^2 f_i}{N} = \frac{\sum_{i=1}^{k} x_i^2 f_i - \sum_{i=1}^{k} 2\bar{x}x_i f_i + \sum_{i=1}^{k} \bar{x}^2 f_i}{N} =$

$$= \frac{\sum_{i=1}^{k} x_i^2 f_i}{N} - 2\bar{x}\frac{\sum_{i=1}^{k} x_i f_i}{N} + \bar{x}^2 \frac{\sum_{i=1}^{k} f_i}{N} = \frac{\sum_{i=1}^{k} x_i^2 f_i}{N} -$$

$$- 2\bar{x}^2 + \bar{x}^2 = \frac{\sum_{i=1}^{k} x_i^2 f_i}{N} - \bar{x}^2$$

Adibidea:

1.lagina: $V(x) = \frac{38,5^2+40^2+42,5^2+43^2}{4} - 41^2 = \frac{6737,5}{4} - 41^2 = 3,375$

Bariantzaren kalkuluak eta, beraz, desbidazio standardarenak, batezbestekoaren kalkulua errazteko erabilitako metodoa ere onartzen du.

Gogora dezagun, $z = \frac{x-a}{c}$ aldagai-aldaketan zetzala metodo hau, a eta c hautazko balioak izanik, nahiz eta eragiketak errazteko apropos aukeratuak izan.

$(z_i)_{i=1,2,\ldots,k}$ aldagaiaren balio berrien bidez, kalkulatuko dugu V(z), ondokoa, zailtasunik gabe frogatzen delarik:

$$V(z) = \frac{V(x)}{c^2} \text{ , hots, } V(x)=c^2 \cdot V(z) \text{ eta } \sigma(x)= c \cdot \sigma(z)$$

Adibidea: Eritrozitoen diametroen bariantza eta desbidazio standardaren kalkulua.

Diametroa	f_i	x_i	$z_i = \frac{x_i-7,2}{0,4}$	$z_i f_i$	$z_i^2 f_i$
5,4 - 5,8	5	5,6	-4	-20	80
5,8 - 6,2	78	6	-3	-234	702
6,2 - 6,6	144	6,4	-2	-288	576
6,6 - 7	479	6,8	-1	-479	479
7 - 7,4	542	7,2	0	0	0
7,4 - 7,8	358	7,6	1	358	358
7,8 - 8,2	279	8	2	558	1116
8,2 - 8,6	99	8,4	3	297	891
8,6 - 9	15	8,8	4	60	240
9 - 9,4	1	9,2	5	5	25
	2000			257	4467

Gogora dezagun kalkulaturiko batezbestekoa 7,2514 zela.
Beraz:

$$V(z) = \frac{\sum_{i=1}^{k} z_i^2 f_i}{N} - \bar{z}^2 = \frac{4467}{2000} - \frac{257}{2000} = 1,9765$$

$V(x) = 0,16 \cdot V(z) = 0,16 \cdot 1,9765 = 0,31624$

$\sigma(x) = 0,562$

c) Aldakuntz koefizientea.

Desbidazio standarda eta batezbestekoaren arteko zatidura da:

$$C.V. = \frac{\sigma_x}{\bar{x}}$$

Intuitiboki konparatzeko modukoak dira, baina, neurri-maila desberdinak dauzkaten populazioen aldakuntz maila konparatzeko erabiltzen da.

Dimentsio gabeko kantitatea da eta portzentaiatan, 100-ekin biderkatuz, adierazten dugu.

Beraren bidez, ondokoak konpara daitezke: bai laginaren beraren bi izaera desberdin, adibidez: kirolari-lagin batetan zein den aldakorragoa aztertzea, kapazitate torazikoa edo pisua; bai populazio desberdinen laginekiko antzeko izaerak, adibidez: zein den aldakorragoa aztertzea, pinu-mota baten edo haritz-mota baten altuera.

I.6.3.- Posizio-estatistikoak.

Mediana, alde bakoitzean balioen 50%a uzten duen aldagaiaren balioa izan arren, hots, balioen posizioarekiko balio zentrala izanik, zenbait posizio adierazten dizkiguten beste estatistiko batzu ere defini ditzakegu.

Hots, hauxe da posizio-estatistikoa: ezkerrean balio-portzentaia konkretua uzten duen aldagaiaren balioa.

Erabilgarrienak hurrengoak dira:

a) Koartilak: Q_1=Bere ezkerrean balioen 25%a uzten duen aldagaiaren balioa.
Q_3=Bere ezkerrean balioen 75%a uzten duen aldagaiaren balioa.
Q_2=Me

b) Dezilak: D_i=Bere ezkerrean balioen 10i%a uzten duen aldagaiaren balioa.
i=1,2,...,9
D_5=Me

c) Pertzentilak: P_i=Bere ezkerrean balioen i%a uzten duen aldagaiaren balioa.

Aldagaia,klasez bilduta dagoenean, eta medianaren kasuan egiten den moduan, interpolazioz kalkulatzen dira.

Hots, normalean, bere ezkerrean balioen i%a uzten duen P_i posizio-estatistikoa horrela kalkulatuko dugu:

$$P_i = l_i + \frac{\frac{i \cdot N}{100} - F_{i-1}}{f_i} \cdot c_i$$

, l_i=i portzentaia daukan klasearen beheko muga.
c_i=esandako klasearen anplitudea.
F_{i-1}=iN/100 baino txikiagoak diren maiztasun metatuetariko handiena.
f_i= i portzentaia daukan klasearen maiztasuna
izanik.

Adibidea:

200 esne-laginetako koipe-edukina.

Koipea	f_i	F_i
3,4-3,6	1	1
3,6-3,8	14	15
3,8-4	20	35
4 -4,2	52	87
4,2-4,4	56	143
4,4-4,6	35	178
4,6-4,8	18	196
4,8-5	4	200

$$\frac{50N}{100} = 100$$

$$Me = 4,2 + \frac{100-87}{56} \cdot 0,2 = 4,246$$

$$\frac{25N}{100} = 50$$

$$Q_1 = 4 + \frac{50-35}{52} \cdot 0,2 = 4,058$$

$$\frac{90N}{100} = 180$$

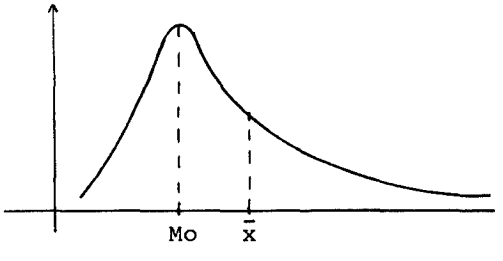
$$D_9 = 4,6 + \frac{180-178}{18} \cdot 0,2 = 4,622$$

I.6.4.- Forma-estatistikoak.

Estatistiko hauen kalkuluaren zailtasuna dela eta, "alborapena" ikusiko dugu soilik.

Beronek, banaketaren simetria neurtzen du, $\nu = \frac{\bar{x}-Mo}{\sigma_x}$ formularen bidez kalkulatzen delarik.

Banaketa alboratuen, hots, simetria gutxi daukatenen adibide argi bat, zoriz aukeratutako 100 langileren soldaten banaketa izan daiteke. Seguraski, honelako maiztasun-banaketa lortuko dugu:



Adibide honetan, $\bar{x} > Mo$ eta $\nu > 0$ izanik, banaketa eskuinerantz alboratua dagoela esango dugu.

$\nu < 0$ baldin bada, ezkerrerantz alboratua deituko diogu eta $\nu = 0$ baldin bada, alboragabea.

PROBLEMAK

1.- Sailka itzazu hurrengo aldagaiak:

a) Ospitale batetan sartzen diren eguneko gaixo-kopurua.

b) Pertsona-talde baten odolaren hemoglobina-edukina.

c) Behatoki batetan lortutako orduko tenperaturak.

d) 6 kumetako kumaldietan arratoiar zurien kopurua.

e) Pastila batzuren C bitaminaren edukina.

f) Birus baten latentzia-aldia.

g) Gaixo-talde baten minutuko taupada-kopurua.

2.- Birus baten latentzia-aldia ikertzeko, 100 txitei inokulatu zitzaien birusa. Bakoitzarentzako gaixotasunaren lehen sintomak agertu arte, pasatutako egun-kopurua aztertu zen. Lortutako datuak hurrengoak ziren:

8-10-8-14-16-9-12-13-9-12-5-9-11-13-5-9-12-13-8-14-7-8-12-12-8-6-8-9-9-15-9-12-8-6-9-14-13-8-12-9-10-13-6-5-14-12-14-6-11-12-6-15-7-9-15-9-7-10-7-10-12-10-15-8-6-8-5-14-6-13-7-11-8-16-10-6-11-8-7-12-9-9-10-12-11-12-10-12-6-7-10-8-11-11-14-15-8-9-8-13.

a) Ordena ezazu lagina, maiztasun-taula moduan aurkeztuz:

a-1) Aldagai estatistiko diskretua dela kontsideratuz.

a-2) Aldagai estatistiko jarraia dela kontsideratuz. Anplitudetzat unitatea duten klasez elkarturiko aldagaia suposatuz, klase-markak aldagaiaren balioak izanik.

b) Irudika itzazu ondokoak:

- a-1 kasuari dagokion barra-diagrama.
- a-2 kasuari dagozkion histograma eta poligono metakorra.

c) Zein da, 12 egun edo gutxiagoko gaixotasunaren latentzia-aldia izan zuten txiten portzentaia?

d) Zein da, 8 eta 12 egun (biak barne direlarik) arteko latentzia-aldia izan zuten txiten portzentaia?

3.- Ondoko maiztasun metatuen banaketa emanik, adieraz ezazue maiztasun absolutuen banaketa:

x_i	2	3	5	10	20	30
F_i	4	16	22	32	38	40

4.- Bete itzazu hurrengo taularen datuak:

x_i	1	2	3	4	5	6	7	8
f_i	4	4	-	7	-	5	-	-
F_i	-	-	-	-	33	-	45	-
h_i	-	-	0,16	0,14	-	-	-	-

5.- X eta Y aldagaiek ondoko balioak hartzen dituzte:

$x_1 = 2$; $x_2 = -5$; $x_3 = 4$; $x_4 = -8$

$y_1 = -3$; $y_2 = -8$; $y_3 = 10$; $y_4 = 6$

Kalkula itzazu:

a) $\left(\sum_{i=1}^{4} x_i\right)^2$ b) $\sum_{j=1}^{4} y_j$ c) $\sum_{i=1}^{4} x_i \sum_{j=i}^{4} y_j$

d) $\sum_{i=1}^{3} x_i \cdot y_{i+1}$ e) $\sum_{i=1}^{4} \sum_{j=1}^{2} x_i y_j$ f) $\sum_{i=1}^{4} x_i \sum_{j=1}^{i} y_j$

6.- Bigarren probleman agertzen den banaketarako, kalkula itzazu ondoko estatistikoak:

a) Batezbestekoa, mediana eta moda.

b) Aldagaiaren aldakuntz tartea, desbidazio standarda eta aldakuntz koefizientea.

c) Koartilak, koartilarteko heina eta 65 eta 85 pertzentilak.

d) Aldagaiak, banaketa normala duela suposatuz, kalkula itzazu birusaren latzentzia-egunen 68,7%, 95% eta 99%-a dauzkaten batezbestekoarekiko tarte simetrikoak.

7.- Pertsona normalen sero proteikoaren albumina-portzentaiari buruzko informazioa lortzeko, 10 eta 40 urteren arteko 210 pertsonen laginak aztertu ziren, hurrengo emaitzak lortu zirelarik:

Portzentaia	58-62	63-65	66-68	69-71	72-76	77-79
Pertsona-kopurua	15	30	42	51	65	7

Kalkula itzazu:

a) Maiztasun-histograma, batezbestekoa, mediana eta koartilak.

b) Alborapen-koefizientea.

c) Albuminaren portzentaia txikienak dauzkaten pertsonen 10%-a eta portzentaia handienak dauzkatenen 10%-a kenduz gero, zein portzentaia-tartetan egongo da laginaren aleen gaineratiko 80%-a?

8.- Baraurik dauden 120 pertsonengandik ateratako odolaren glukosa-edukinari dagokion banaketa emanik:

Glukosa	50-56	56-62	62-65	65-68	68-74	74-80
Pertsona-kopurua	5	39	31	30	12	3

Ondokoak kalkula itzazu:

a) Pertsonen glukosa-edukinen batezbestekoa.

b) Koartilak.

Zein da glukosa-edukina non 12 pertsonak bera baino glukosa-edukin handiagoa izango duten?

c) Zenbat dira, 67 baino glukosa-edukin txikiagoa daukaten pertsonak?

Zenbatek daukate, $(\bar{x}-2\sigma\ ,\ \bar{x}+2\sigma\ )$ tartean sartutako glukosa edukina?

9.- Airearen kutsadura aztertzeko, 57 hiritan aztertu zen esegidur materiaren partikula-edukina (mikrogramo/m^3-tan), hurrengo emaitzak lorturik:

Hiria	Balioa	Hiria	Balioa	Hiria	Balioa	Hiria	Balioa
1	68	16	65	31	12	46	22
2	63	17	43	32	32	47	43
3	42	18	25	33	49	48	27
4	27	19	74	34	38	49	49
5	30	20	51	35	42	50	28
6	36	21	36	36	27	51	23
7	28	22	42	37	31	52	19
8	32	23	28	38	50	53	46
9	79	24	31	39	38	54	30
10	27	25	28	40	21	55	43
11	22	26	25	41	16	56	49
12	23	27	45	42	24	57	12
13	24	28	12	43	69		
14	25	29	57	44	47		
15	44	30	51	45	23		

a) Histograma eta maiztasun-poligonoaren bidez, irudika ezazu aireango esegidur materiaren partikula-edukinen banaketa.

b) Kalkula itzazu oinarrizko estatigrafo guztiak (6.problemaren a, b eta c atalak).

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

10.- Ikerlariak, gazteen perimetro torazikoaren garapena aztertu nahi du. Honetarako hauxe da egiten duena: 15 urteko 90 gazte harturik, perimetro torazikoa neurtzen die, eta geroago, 18 urteak egitean, neurketa berbera egiten die, bakoitzaren lehengo eta bigarren neurrien arteko kendura idatzirik. Emaitzak hurrengoak dira:

Perimetro torazikoaren gehikuntza (cm.tan)	ale-kopurua
(-2,1;-1,5)	7
(-1,5;-0,9)	9
(-0,9;-0,3)	12
(-0,3; 0)	13
(0 ; 0,3)	17
(0,3; 0,9)	17
(0,9; 1,5)	12
(1,5; 2,1)	3

Aurki itzazue:

a) Bariantza eta aldakuntz koefizientea. Interpreta itzazu emaitzak.

b) Moda.

c) Aldagai estatistiko honen maiztasun-banaketa, simetrikoa al da?

11.- Gaixotasun kronikoetarako ospitale batetan gaixorik daudenen adinak (urtetan) aztertzen dira hilabetean, ondoko emaitzak lorturik:

Adinak	Gaixo-kopurua	Adinak	Gaixo-kopurua	Adinak	Gaixo-kopurua
1	2	25	1	56	1
2	3	27	2	58	2
3	1	28	2	60	1
5	2	29	1	61	2
6	3	30	1	63	2
7	5	32	2	69	2
8	1	33	4	70	1
9	2	36	1	71	1
10	3	37	1	72	1
11	1	39	1	74	1
12	1	40	1	76	1
14	2	42	2	77	1
15	1	46	2	81	2
16	1	48	3	85	4
17	1	49	1	87	3
18	1	51	3	88	1
21	1	52	3	89	1
22	3	53	4		
24	2	54	2		

a) Elkar itzazue datuak, anplitude bereko klaseetan.

b) Aurki ezazue, batezbestekoa ematen digun adierazpena, a (jatorria) eta c (klase-anplitudea) direlakoen funtzioan.

Batezbesteko adina lortzeko, aplika ezazue adierazpen hau, emandako probleman.

c) Lor ezazu mediana.

d) Banaketaren ezaugarrien artean, zein erabiliko zenukete:

d-1) gaixo ospitalizatuen maiztasun handieneko adina zein den adierazteko?

d-2) zoriz aukeraturiko adin bitariko batek, aipaturiko ezaugarriaren balioa gaindituko ez duela ziur egoteko?

12.- Pertsona normalen talde baten albumina-kantitate ibilkorraren batezbestekoa 130 gramo da. Gaixotasun bat jasatzen duten pertsona-talde baten batezbesteko hori, 100 gramo da. Osotasunean harturik, pertsona-talde bien albumina-kantitate ibilkorraren batezbestekoa 125 gramo da. Zein da pertsona gaixoen portzentaia?

13.- Galsoro batetan, A, B eta C, hiru ongarri-mota desberdin erabiltzen dira.

A ongarriz, 20 galbelar ongarriztatzen dira, beraren batezbesteko etekina 201 Kintale/Ha delarik.

B ongarriz, 40 galbelar ongarriztatzen dira, 228 Kintale/Ha balioko batezbesteko etekina lorturik.

C ongarriz, 30 galbelar ongarriztatzen dira, lortutako batezbesteko etekina 270 Kintale/Ha izanik.

Zein da galsoro osoaren batezbesteko etekina?

Ongarri-mota bakoitza galbelar-kopuru berdinerako erabiliko balitz, zein izango litzateke batezbesteko etekina?

14.- Pertsona-multzo baten batezbesteko pisua 65 kg. da, desbi dazio standarda 10 kg. izanik.

Halaber, beraren batezbesteko altuera 170 cm. da, desbida zio standarda 5 cm. izanik.

Pertsona batek, 71 kg.-tako pisua eta 175 cm.-tako altuera ditu.

Zein ezaugarrirekiko (pisua ala altuera) dauka posizio er latibo hoberik?

II. GAIA: PROBABILITATEARI BURUZKO OINARRIZKO EZAGUMENDUAK

II.1.- Sarrera.

II.1.1.- Probabilitatea, zertarako ikasi behar?

II.1.2.- Zorizko experimentuak.

II.2.- Probabilitatearen kontzepturako hurbilketa intuikorra.

II.2.1.- Zorizko experimentu bati elkartutako gertaerak.

II.2.2.- Multzoen bidezko gertaeren adierazpena.

II.3.- Probabilitate-kalkuluaren axiomatika.

II.3.1.- Zorizko experimentuen erregularta-sun estatistikoa. Maiztasunen egonkortasun-legea.

II.3.2.- Probabilitatearen axiomak.

II.4.- Eredu probabilistikoak.

II.4.1.- Zorizko eredu klasikoa.

II.4.2.- Eredu probabilistiko finitua.

II.4.3.- Eredu probabilistiko diskretua.

II.4.4.- Eredu probabilistiko orokorra.

II.5.- Baldintzazko probabilitatea eta gertaeren arteko independentzia.

II.5.1.- Baldintzazko probabilitatea.

II.5.2.- Gertaeren arteko independentzia.

II.5.3.- Bayes-en teorema.

PROBABILITATEARI BURUZKO OINARRIZKO EZAGUMENDUAK

II.1.- SARRERA

II.1.1.- Probabilitatea, zertarako ikasi behar?

Aurreko gaian egindako azterketa, emandako datuen multzoaren azterketara murriztua izan da. Metodo hau, Zientzi experimental gehienek sortzen dituzten eskakizunak betetzeko ez da aski.

Askotan, populazio osoaren azterketa ezinezkoa da. Orduan, azpipopulazio edo lagin bat edo batzuren azterketa burutu behar dugu, bertatik ateratako konklusioak populazio osora hedatzen ditugularik. Hau da Estatistika Deskribatzailetik Estatistika Induktiborako bidea. (*)

Bigarren kasu honetan, populaziotik ateratako laginetatik abiatuz, populazioaren ezaugarriren bat induzitu nahi dugunean, oinarrizko tresnaren erabilera ezagutu behar dugu: probabilitatearena, hain zuzen.

II.1.2.- Zorizko experimentuak

Naturan, oso izaera desberdina duten fenomenoak azter ditzakegu. Adibidez, kontsidera ditzagun ondoko bi fenomenoak: 1) Lurraren inguruko gorputzen erorketa askea; 2) Arrain-mota baten orduko oxigeno-kontsumoa.

Lehendabiziko kasuan, gorputzak erortzerakoan egindako

(*) Azterketa estatistikoa burutzerakoan, aztertuko dugun multzoa ea populazio osoa ala zati bat den hartu behar dugu kontutan.

espazioa, erorketa-denboraren menpean baizik ez dagoela adieraz ten digute datu experimentalek. (Ba dakigu higidura azaltzen duen ekuazioa, $s = \frac{1}{2} gt^2$ formulak emanik datorrela).

Bigarren kasuan, egindako behaketek datu ezberdinak ema ten dizkigute.

Lehendabiziko egoeran, baldintza definituen multzo baten pean, emaitza berbera lortzen da. Bigarrenean, ordea, nahiz eta arrain-mota, tamainua, inguruko baldintzak eta abar aurretik jarri, oxigeno-kontsumoaren neurriak ezberdinak izango dira.

Lehendabiziko motako fenomenoei, deterministak deituko diegu; bigarren motakoei, aldiz, zorizkoak.

Fisikan aztertzen diren gertaera gehienak (higidurak, elektrizitatea, uhinak,...) deterministak bailiren kontsidera daitezke. Baina, biologoentzat interesa daukaten arazo gehienek (heredentziaren transmisioa, giza-odolaren konposaketa, uren kutsadura-maila, animali kolonia batetan gaixotasunen transmisioa, eta abar) zorizko izaera daukate.

Esan dugunez, zorizko fenomenoek ondoko propietatea be te behar dute: Haiei elkartutako gertaera bakoitzerako, berau halabeharrez jaso behar deneko zirkunstantziak finkatzea, ezinezkoa da. Honi, "zori-baldintza" deritzogu, faktore ezezagunek edo aurrikustezinek parte hartzen dutela adierazten duelarik.

Baina, zorizko experimentuetako proba batetan lortuko den emaitza aurresan ezin bada, nola finka ditzakegu, aipaturi ko fenomenoak nolabait isladatzen dituzten legeak?

Izan ere, esate baterako, bakterio-mota baten bizitza-aldia kantitate aldakorra (zorizkoa) da eta badirudi bakterio baten bizitzaren iraupenari buruzko aurresanak egitea guztiz amu rragabea dela. Hala ere, perspektiba globaletik begiratuz, denbora-tarte mugatu baten bitartean, bakterioen heriotza "normalki" gertatzen dela ikusten dugu. Beste era batetara esanda, bakterio-populazio handi bat kontsideratzerakoan, zenbait uniformeta

sun egiaztatzen da; edo zehaztasun gehiagoz, baldintza egonkorren pean (tenperatura, ingurugiroa, presioa,...) bakterio bat aztertzeko eta bere heriotzaren unea notatzeko experientzia maiz errepikatzen badugu, lortuko ditugun neurriak, nahiz eta desberdinak izan, "gehienetan" denbora-balio konkretuen artean egoten dira. Erregulartasun-baldintza hau -geroago maiztasunen bidez zehaztuko duguna- zorizko experimentuez hitz egin ahal izateko premiazkoa da.

Jakina, erregulartasun-baldintzak, esentzialki konstanteak diren faktoreen pean, experimentua, idealki behintzat, etengabe errepikatzeko posibilitatea eskatzen du.

Gertaera biologiko askok, zorizko experimentuak definitzen dituzten zoriz, errepikatasun eta erregulartasunezko hiru baldintza hauek betetzen ditu. Baldintza hauek ezaugarri kuantitatiboak, matematikoki, eredu probabilistikoen bidez aztertzen dira. Horregatik, Estatistika eta Probabilitate-kalkuluak ezinbesteko tresnak izaten dira, beraien ondorioetan gutxienezko sakontasuna eskatzen duten ikerketa biologikoetarako.

II.2.- PROBABILITATEAREN KONTZEPTURAKO HURBILKETA INTUIKORRA

II.2.1.- Zorizko experimentu bati elkartutako gertaerak.

Gertaeraren nozioa, intuikorra, primarioa, sinplea, adibideen eta fenomeno errealen azterketaren bidez soilik transmiti daitekeena da. Zorizko experimentu bati elkartutako A gertaera ondo definituta dagoela esan dezakegu, proba partikular bat egiterakoan eta bere emaitza aztertzerakoan, ea A egiaztatzen denentz dakigunean. Erleen hegoen luzeraren neurketetan datzan experimentuan, "hegoen luzera lcm. baino luzeagoa da", elkartutako gertaera da, baina, ez da "erleak X gaixotasuna pairatzea". Beraz, gertaeraren kontzeptua, zorizko experimentuarenari lotuta dagoela ikusten dugu.

Nahiz eta zorizko experimentu bati elkartutako emaitzak gertaeren "forma" aditzera eman, ez dago emaitza eta gertaeraren

arteko identitaterik. Adibidez, gizabanako baten odol-presioa aztertzerakoan, 11 balioa lortzen badugu, "presioa 9 eta 14 artean dago" eta "presioa 10 baino handiagoa da" gertaerak jazo direla baiezta dezakegu.

Kontutan hartzen ditugun gertaerak zeintzu diren, ikertzaileak berak mugatu behar ditu, sinpletasunaren arrazoiek edo aztertzen den fenomenoaren barrutiko izaerarekiko komunztadura gehiago daukan beste argudio-mota batek gidaturik.

1. adibidea:

Kontsidera dezagun itsas-ekosistema baten barneko alelagin bat harrapatzean eta identifikatzean datzala experientzia. Azterketaren lehenengo fasean, ekoizleak diren populazioaren aleak eta kontsumatzaileak direnak bereizten dira. Aztertze-maila honetarako, ondokoak lirateke gertaerak: "Ekosistematik ateratako alea ekoizlea da", (E) adierazirik, eta "Ekosistematik ateratako alea kontsumitzailea da", (K) adierazirik.

Bigarren fasean, hurrengo taldeak bereizten dira:

(E) Ekoizleak: - (Dt) Diatomeak
- (Df) Dinoflagelatuak
- (Z) Zooplanktona

(K) Kontsumitzaileak: - (Z) Zooplanktona

Kontutan hartzen diren gertaera-kopuruaren emendioa handiagotzen du honek.

Hirugarren fasean, hurrengo generoak bereiztuz, honela segi genezake:

Ekoizleak: - Diatomeak: - (Co) Corethona
- (Pl) Planktoniella
- (Ch) Chaetocerosa

- Dinoflagelatuak: - (Ce) Ceratiuma
- (Pe) Peridiniuna

- Zooplanktona: - (Ca) Calanusa
- (LT) Larba tunikatuak

Kontsumitzaileak: -Zooplanktona: - (Cl) Clupea
- (Sa) Sardinia

Eta honela, berriro, kontsideratzen diren gertaera-kopurua emendatzen dugu.

Gertaerari buruzko erlazioak eta definizioak sartzeko, A, B,..., letra majuskulez adieraziko ditugu gertaerak. Berauek, ξ zorizko experimentuari elkartuta daudela suposatuko dugu beti.

A gertaerak B izenekoa halabehartzen duela edo A parte B dela diogu, ondokoa agitzen denean: A gertatzen baldin bada, B gertatzen da. Eta $A \subset B$ edo $B \supset A$ moduan adierazten da.

$A \subset B$ eta $B \subset A$ betetzen direnean, berdinak direla esaten da, A=B adierazten delarik. Kasu honetan, proba baten emaitza bakoitzean, bai bata eta bai bestea gertatzen dira, ala bat ere ez da gertatzen.

A gertaera ez jazotzea, beste gertaera da. Honi, A-ren kontrako gertaera deritzogu eta $\bar{A}$ adierazten da.

A eta B gertaeren arteko bat jazotzea gutxienez, beste gertaera da. Gertaera berri honi, A eta B-ren arteko bildura deritzogu, $A \cup B$ adierazirik.

$A \cup \bar{A}$ gertaera, beti jazotzen dela begi bistakoa da. Hau dela eta, gertaera segurua deritzo, $E = A \cup \bar{A}$ adierazirik.

A eta B gertaerak batera gertatzen direnean, jazotzen den gertaerari A eta B-ren arteko ebakidura deituko diogu, $A \cap B$ moduan adieraziko dugularik.

A eta B, bi gertaera, ezin baldin badira batera jazo, bateraezinak dira. Kasu honetan, $A \cap B$ ez da inoiz gertatzen.

Gertaera honi, gertaera ezinezkoa deritzogu, $\emptyset = A \cap B$ moduan adierazten delarik.

Ondokoa betetzen baldin bada, C gertaerari, A eta B gertaeren arteko kendura deituko diogu: A gertatzen da eta B gertatzen ez denean, C gertatzen da. $C = A - B$ moduan adierazten da eta $A - B = A \cap \overline{B}$ berdintza kontutan hartu behar dugu.

Gertaera-familia batek, gertaera-sistema osoa osotzen du, baldin eta ondokoak betetzen baditu: binaka bateraezinak dira eta beren bildura gertaera segurua da. Beraz, proba bakoitzean, haien arteko bat eta bakar bat gertatzen da.

Azkenik, ondokoa betetzen baldin bada, A gertaera konposatua dela diogu: Bi gertaera, B eta C, existitzen dira, $A = B \cup C$ berdintza beterik, A-ren eta ezinezko gertaeraren desberdinak izanik. (A edozein jazoeratarako, $A = A \cup \emptyset$ berdintza beti betetzen dela kontutan hartu behar dugu). Konposatuak ez diren gertaerei, oinarrizko gertaerak deitzen diegu. (Gertaera ezinezkoa ez da oinarrizkotzat hartzen). Oinarrizko gertaerak, experimentu bati dagozkion gertaerarik sinpleenak dira.

II.2.2.- Multzoen bidezko gertaeren adierazpidea.

Aurreko atalean genioenez, experimentu bati elkartutako gertaeren eta emaitzen arteko identitaterik ez dago.

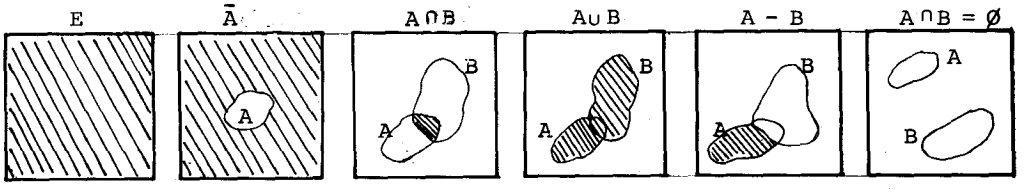
Zorizko experimentuaren emaitza posible guztien multzoari, lagin-espazioa edo multzo unibertsala deritzogu, Ω adierazirik.

Gertaera bakoitzari, bera definitzen duen emaitza-multzoa elkar diezaiokegu, hots, Ω -ren azpimultzo bat. Horrela, gertaeren eta emaitzen arteko erlazioa ezartzen dugu (aplikazio hau, gertaera-algebraren eta multzo-algebraren arteko isomorfismoa da, hain zuzen), zeinak gertaeren arteko eragiketak multzoteoriaren bidez azaltzen baimentzen bait digu. Hau interesgarria da, zeren multzoak Venn-en diagramen bidez aurkez bait daitezke,

nolabait ezagunak izanik. Beraz, dimentsio intuikorra galdu gabe, operatibotasun gehiago lortuko dugu.

Hona hemen, baliokidetasunen laburpena eta zenbait Venn-en diagrama:

Probabilitate-kalkulua	Multzo-teoria
Gertaera segurua	Multzo orokorra
Oinarrizko gertaera	Multzo orokorraren elementua
Gertaera	Multzo orokorraren azpimultzoa
Gertaeren bildura	Multzoen bildura
Gertaeren ebakidura	Multzoen ebakidura
Gertaeren kendura	Multzoen kendura
Gertaera bateraezinak	Multzo bateraezinak
Gertaera ezinezkoa	Multzo hutsa
Kontrako gertaera	Multzo osagarria
Gertaeren sistema osoa	Multzo orokorraren partiketa



II.1. Irudia

II.3.- PROBABILITATE-KALKULUAREN AXIOMATIKA

II.3.1.1.- Zorizko experimentuen erregulartasun estatistikoa. Maiztasunen egonkortasun-legea.

Atal honetan, zorizko experimentuen erregulartasunaren printzipioa zehatz-mehatz aztertu nahi dugu. Suposa dezagun zorizko experimentu baten n errepikapen egiten direla eta experi-

mentu honi elkartutako A gertaera, beraien arteko f aldiz gertatu dela. f zenbakiari, A gertatzeko maiztasun absolutua deritzogu, f eta n-ren arteko zatidura maiztasun erlatiboa deitua izanik. Bien artean, maiztasun erlatiboa interesatzen zaigu gehien, A gertatzeko "neurria" ematen bait du.

A eta B, bi gertaera, 5 aldiz gertatu direla esaten badigute, informazio honetatik ondorio gutxi atera dezakegu. Ordea, A gertaera 10 proben arteko 5 aldiz eta B delakoa 100 proben arteko 5 aldiz gertatu direla jakingo bagenu, A gertaera aldien erdian eta B-a aldien 5%-ean soilik gertatzen direla esango genuke. Maiztasun erlatiboak, proba-kopuruari buruzko informazioa barnean sarturik, gertatze-maila handia edo txikia konparatzea baimentzen digu. Errore abasolutu eta erlatiboari buruz egiten diren kontsiderazioen antzekoak dira.

ξ zorizko experimentuaren zenbait errepikapen-sekuentzia n', n", n"',..., egiten baditugu, A gertaerari dagozkion f'/n', f"/n", f"'/n"',..., maiztasun erlatiboak lortuko ditugu. n', n",..., kopuruak "handiak" baldin badira, A-ren maiztasun erlatiboak oso gutxi desberdintzen direla beraien artean, hots, zenbaki fixu baterantz jotzen dutela ohartzen da. Maiztasunen egonkortasun-lege honek, zorizko experimentuen erregulartasun-baldintzari ematen dio zehaztasuna.

Hala eta guztiz ere, antzirudiek ez gaitzatela engaina. Maiztasunen egonkortasun-legeak, nahiz eta zehaztasun inportantea suposatu, ez die hurrengo galderei erantzuten: A gertaeraren maiztasun erlatiboak, zein proba-kopurutatik daude hurbil beraien artean? Zorizko experimentuaren errepikapen-kopuru aski handia hartzeak, A-ren maiztasun erlatibotik zenbaki fixu hartara gutxi dagoela garantizatzen al du? Galdera delikatuak direla eta, oraingoz, ezin gaitezke haietan sar.

Ondoren, A gertaeraren maiztasun erlatiboaren zenbait propietate erraz aztertuko ditugu. Experimentua n aldiz errepikatzen dela eta A, f_1 aldiz gertatzen dela suposatuko dugu. A-ren maiztasun erlatiboa, fr(A) adierazirik, $fr(A)= f_1/n$ da.

1) $0 \leqslant f_1 \leqslant n$ dela eta, $0 \leqslant fr(A) \leqslant 1$

2) B gertaera A-rekiko bateraezina baldin bada eta B delakoa f_2 aldiz gertatzen baldin bada n probatan zehar, orduan, $A \cup B$ gertaera f_1+f_2 aldiz gertatzen da. Beraz, $fr(A \cup B) = fr(A)+fr(B)$.

3) E gertaera segurua, noski, n aldiz gertatzen denez gero, $fr(E) = 1$ da.

Material intuikor honez, Probabilitate-kalkuluaren axiometria formula dezakegu. Axiometria honek, hurrengo funtzio bikoitza dauka: Alde batetik, gertaeretarako eredu matematikoa, eta bestetik -eta beronen sostengaturik-, gertaeren probabilateetarako eredua.

II.3.2.- Probabilitatearen axiomak.

Kontsidera dezagun ξ zorizko experimentuari elkartutako A gertaera. Maiztasunen egonkortasun-legearen arauera, ξ experimentuaren n proba burutzen baditugu, n emendatzen doan neurrian, A gertaeraren maiztasun erlatiboak zenbaki fixu batetara jotzen du. Zenbaki honi, A gertaeraren probabilitatea deritzogu, p(A) adierazten dugularik.

Probabilitatearen ideia intuikor honetatik abiatuz, eta maiztasun erlatiboen propietateen orokorpenean oinarrituz, probabilitatearen definizio axiomatikora heltzen gara:

Bira, ξ experimentuari elkarturiko Ω lagin-espazioa eta $\mathcal{A}$ gertaera-algebra. $A \in \mathcal{A}$ gertaera bakoitzari, p(A) zenbaki erreala elkartzen diogu. p(A)-ri A-ren probabilitatea deritzogu, ondoko axiomak betetzen dituelarik:

i) $0 \leqslant p(A) \leqslant 1 \qquad \forall A \in \mathcal{A}$

ii) $A_1, A_2, \ldots, A_n \in \mathcal{A}$ binaka bateraezinak baldin badira (hots, $A_i \cap A_j = \emptyset$, $i \neq j$ baldin bada),

$$p(A_1 \cup A_2 \cup \ldots \cup A_n) = p(A_1) + p(A_2) + \ldots + p(A_n)$$

izango da.

iii) $p(E) = 1$

"Probabilitate" aplikazioak bete behar dituen hiru axioma hauek, Kolmogorov-en axiomak deitzen dira.

Ikus dezagun axiometatik eratortzen diren ondorioen bilduma (Algebraren gertaerak letra majuskulaz adierazirik):

1) $p(\bar{A}) = 1 - p(A)$. Partikularrean, $p(\emptyset) = 0$
Izan ere, gertaera segurua, $E = A \cup \bar{A}$ bateraezinezko bilduraren modura idatz dezakegu.
Beraz, $1 = p(E) = p(A \cup \bar{A}) = p(A) + p(\bar{A})$

2) $A \subset B$ baldin bada, $p(A) \leqslant p(B)$ izango da. Aurrekoa, edozein gertaera-bikotetan aplikatuz, ondokoa lortzen da:
$p(A \cap B) \leqslant p(A) \leqslant p(A \cup B)$
Izan ere, $B = A \cup (B-A)$, bildura bateraezinezkoa izanik. Beraz, $p(B) = p(\{A \cup (B-A)\}) = p(A) + p(B-A) \geqslant p(A)$ zeren $p(B-A) \geqslant 0$ bait da.

3) $p(A \cup B) = p(A) + p(B) - P(A \cap B)$, A eta B edozein gertaera izanik.
Formula hau, gertaera-kopuru handiagorako orokor daiteke. Adibidez, hiru gertaeratarako:
$p(A \cup B \cup C) = p(A)+p(B)+p(C)-p(A \cap B)-p(A \cap C)-p(B \cap C)+p(A \cap B \cap C)$
Izan ere:

$$\left.\begin{array}{l} A \cup B = A \cup (B-A) \\ B = (A \cap B) \cup (\bar{A} \cap B) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} p(A \cup B) = p(A) + p(B-A) = p(A) + p(\bar{A} \cap B) \\ p(B) = p(A \cap B) + p(\bar{A} \cap B) \end{array}\right\} \Longrightarrow$$

$$\left\{\begin{array}{l} p(A \cup B) = p(A) + p(\bar{A} \cap B) \\ p(\bar{A} \cap B) = p(B) - p(A \cap B) \end{array}\right\} \Longrightarrow p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

4) $p(A \cup B) \leqslant p(A) + p(B)$. Orokorrean,
$p(A_1 \cup A_2 \cup \ldots \cup A_n) \leqslant p(A_1) + p(A_2) + \ldots + p(A_n)$
3. atalean ikusi dugunaren ondorioa da.

Laburbilduz, eredu matematiko probabilistikoa ondoko hiru objetuek osotzen dute: Ω multzoa; $\mathcal{A}$, Ω-ren parteen familia; eta i), ii) eta iii) axiomak betetzen dituzten p(A) balio numerikoen edo gertaeren probabilitateen seriea. Hau dela eta, $(\Omega, \mathcal{A}, p)$ hirukoteari espazio probabilistiko orokorra deritzogu.

Eredu orokor honen zabaltasuna egiaztatzeko, bereizkeeta desberdinak ikusiko ditugu segidan.

II.4.- EREDU PROBABILISTIKOAK

II.4.1.- Zorizko eredu klasikoa

Bira ξ zorizko experimentua eta $\mathcal{A}$, berari elkartutako gertaera-algebra, zeinak $A_1, A_2, \ldots, A_N$ oinarrizko gertaerakopuru finitua daukan, gertaera guzti hauek ekiprobableak izanik.

Probabilitatea definitzen duten axiomen arauera, horrela izan behar da:

$$p(A_1) = p(A_2) = \ldots = p(A_N) = \frac{1}{N}$$

Algebraren A edozein gertaera konposatu, oinarrizko gertaeren (n, adibidez) bildura finitu bakar baten modura adieraz daitekeenez, p(A) = n/N dela bistakoa da. Horrela, edozein gertaeraren probabilitatea definituta dago.

Deskribapen-eskema honi "zorizko eredu klasikoa" deritzogu, historikoki, azareko jokoetatik, XVII. mendearen inguruan, sortu zen lehendabizikoa izan bait zen.

Aurreko berdintza, Laplace-ren hain definizio eztabaidatua da: gertaeraren probabilitatea, gertaeraren jazotzearen aldeko kasuen eta kasu posibleen kopuruen arteko zatidura da, guztiok ekiprobableak izanik. Beraz, eskema klasikoaren arauera, gertaeraren probabilitatearen kalkulua, aldeko kasuen eta kasu posibleen kalkulu bihurtzen da. Horretarako, Konbinato-

riak ematen dizkigu tresnak, zeinaren kontzeptu eta emaitza garrantzitsuenak eranskin moduan bildu ditugun.

Eredu klasikoa, bere sinpletasuna dela kausa, egoera errazetan aplikatzen da, non arrazoi batzurengatik (simetriagatik, adibidez) ekiprobabilitatea suposa bait dezakegu (edo beraiekiko informazio-eza nabaria den kasuetan eta emaitza batzu beste batzu baino probableagoak direla ezin dezakegu jakin). Honegatik, Laplace-ren legeari, batzutan, jakinezaren legea deitzen zaio.

2. adibidea:

Suposa dezagun, n makiletariko bakoitza bi zatitan hausten dela, bata luzea eta bestea laburra. 2n zati hauez, n bikote eratzen dira, berauek makila berriak osotzen dituztelarik. Ondoko probabilitateak aurki itzazu:

a) Zatiak, jatorrizko ordenan lotuta geratzen direnekoa.

b) Zati luze guztien bikotekideak laburrak izatekoa.

Lehenik, kasu posibleen kopurua kalkulatuko dugu. Horretarako, modu honetan arrazona dezakegu: 2n zatiak, 1-etik n-raino zenbatzen ditugu. Haien arteko bat aukeratzen dugu, la adibidez. Hau, 2n-1 gaineratiko zenbakien arteko edozeinekin pareka daiteke. Parekamendua eginez gero, gaineratiko bat aukeratzen dugu, zeina gelditzen diren 2n-3 zenbakiekin pareka bait daiteke, eta horrela, elkarren segidan. Beraz, kasu posibleen kopurua hauxe da: $(2n-1)\cdot(2n-3)\cdot(2n-5)\cdots 3.1$

a) atalerako, aldeko kasu bakar bat dago; b) atalerako, ordea, n! kasu. Beraz, eskaturiko probabilitateak ondokoak dira:

a) $\dfrac{1}{(2n-1)\cdot(2n-3)\cdots 3\cdot 1}$ b) $\dfrac{n!}{(2n-1)\cdot(2n-3)\cdots 3\cdot 1}$

Honen gutxi iradokorra den enuntziatu honek, hurrengo

interpretazioa onartzen du: Zelulak erradiazio kaltegarrien pean daudenean, kromosoma batzu apurtzen dira eta lehenago ikusitako makilen papera betetzen dute. Alde "luzea" zentromeroa daukana da. Zati "luzeak" edo "laburrak" elkarrekin lotzen badira, zelula hil egiten da.

II.4.2.- Eredu probabilistiko finitua

Lehen bezala, zorizko experimentuari elkartutako gertaeren algebrak, oinarrizko gertaera-kopuru finitua baldin badauka, eta honetaz gain, beraien ekiprobabilitatearen hipotesia suposatzen ez badugu, eredu finitura jotzen dugu, jarrera matematikotik aurreko eredua orokortzen duenera.

Oinarrizko gertaerak, A_1, A_2,..., A_N baldin badira, beren probabilitateak $p(A_1)= p_1$, $p(A_2) = p_2$,..., $p(A_N) = p_N$ funtsbaldintzak betetzen dituzten zenbakiak dira. Hauxe da:

i) $0 \leqslant p_i \leqslant 1 \qquad i = 1,2,\ldots,N$

ii) $\sum_{i=1}^{N} p_i = 1$

Adibidez, $A = A_1 \cup A_2 \cup \ldots \cup A_n$ gertaera konposatuak ondoko probabilitatea du:

$$p(A)=p(A_1 \cup A_2 \cup \ldots \cup A_n)=p(A_1)+p(A_2)+\ldots+p(A_n)=p_1+p_2+\ldots+p_n$$

Ezin dezakegu desberdintasun esentziala ahantz: eredu klasikoan, probabilitateen egokiera experientziaren "a priori" egiten da; eredu finituan, ordea, "a posteriori", gertaeren maiztasun erlatiboak aztertu eta gero.

3. adibidea:

Hereda daitezkeen karaktereak, "gen" deitutako zenbait transmisore bereziren menpekoak dira; hauek binaka agertzen dira, kasurik sinpleenean, bikote partikularraren gen bakoitzak, A eta a bi forma (alelo) har ditzakeelarik. Orduan, hiru bikote ezberdin era daitezke, eta, bikote partikular honekiko, organismoa AA, Aa, aa hiru genotipoetariko bat da (ez dago Aa eta aA-ren

arteko berezitasunik). Sarritan, bi genentzat, A_1, A_2,..., A_N, N mota desberdin daude, eta beraz, A_1A_1, A_1A_2,..., A_1A_N genotipoak.

Ugalketa-prozesuan, umearen gen-pare bakoitzean, aitarengandiko gen batek eta amarengandiko beste batek eragiten dute, eta hauek, transmitituak izateko probabilitate berdina daukate. Adibidez, Aa(aita)xAa(ama) parekatzearen genotipoak, AA, aa (genotipo hutsak edo homozigotikoak) eta Aa (hibridoa edo heterozigotikoa) dira, eta, kasu honetan, beren probabilitateak ez dira 1/3-takoak, experimentalki egiazta daitekeenez. Ordea, ondorengoen erdia hibridoak direla eta beste erdia AA eta aa genotipo hutsen artean erdibitzen dela adierazten digute AaxAa gurutzamenduaren emaitza experimentalek. Honek, aitarengandiko edo amarengandiko gen bakoitzak 1/2-tako transmititua izateko probabilitatea daukaneko; eta, beraz, AA, Aa eta aa genotipoekiko probabilitateak, 1/4, 1/2 eta 1/4 direneko eredua onartzera eramaten gaitu.

II.4.3.- Eredu probabilistiko diskretua

Kontutan hartutako oinarrizko gertaera-kopurua infinitu zenbakigarria, hots, $\mathbb{N}$ zenbaki arrunten multzoaren kardinal berberekoa denean, eredu diskretua daukagu. Oinarrizko gertaeren "zenbakigarritasuneko" hipotesia dela eta, berauek, $A_1, A_2, .., A_N$-ren bidez adieraz ditzakegu.

A_i bakoitzaren gertatzeari dagozkion p_i zenbakiak, ez-negatiboak dira eta beren batura 1 da, beti bezala. Matematikoki esanik:

$$0 \leqslant p_i \leqslant 1 \qquad \forall i \qquad ; \qquad \sum_{i=1}^{\infty} p_i = 1$$

Zenbaki finitua ezik, p_i guztiak zero direnean, eredu finituaren kasuan gaudela kontutan hartu berhar dugu.

Ikuspuntu biologikotik, eskema edo eredu diskretua ez da apetatsu den zerbait. Ondoko fenomenoetan aplikatzen da: zeluletan, X izpien irradiazioaren bidezko kromosoma-elkartrukean;

likido-bolumen batetako bakterio-banaketan; odolaren analisi kuantitatiboan; desintegrazio irradiaktiboan ; jaiotze anizkoitzetan;... Adibide hauen eta aipa litezkeen beste askoren azterketa sistematikoa, aurrerago egingo dugu, Poisson-en banaketa delakoa edo gertaera "arraroen" legea ikustean.

4. adibidea:

Suposa dezagun, laborategiko partzela batetan n emaitza ekiprobable lor daitezkeela, eta arteko bati begiematen diogula. Experientzia, aurretik jarritako emaitza lortzeko, ahal den hainbat aldiz proba errepikatzean datza. Zein da "arrakasta" r-garren errepikapenean lortzeko probabilitatea?

j_1 emaitza lehen proban, j_2 bigarren proban,..., jr r-garrenean lortzen direla adierazteko, $(j_1, j_2,\ldots,j_r)$ moduko ikurrak erabiltzea komeni da. $(j_1, j_2,\ldots, j_r)$ "aldeko" segidak, $j_1, j_2,\ldots,j_{r-1}$ emaitzak "a" emaitza aurridatziaren ezberdinak direneko, baina, j_r berdin "a" deneko baldintzaren menpekoak dira. Beraien kopurua $(n-1)^{r-1}$ da; segida posibleen kopurua, aldiz, n^r. Beraz, eskatutako probabilitatea hauxe da:

$$q_r = \frac{(n-1)^{r-1}}{n^r} = \left(\frac{n-1}{n}\right)^{r-1} \cdot \frac{1}{n}$$

Inolako proban "a" lortzen ez bada, prozesua inoiz ez bukatzea teorikoki posiblea izan arren, $\{q_r\}$ serie geometrikoaren batura 1 izateak (erraz egiaztatzen denez), gertaera honek probabilitate nulua daukala ondorioztatzen du.

II.4.4.- Eredu probabilistiko orokorra

Oraindik, aurreko ereduak betetzen ez dituzten arazo biologikoak, ugariak -eta oso interesgarriak- dira. Adibidez, nola deskriba dezakegu arrain-mota baten oxigeno-kontsumoa, landareen hazierarako tratamenduaren emaitzak, odol-tentsioa, eta abar? Eta orokorrean, nola deskriba dezakegu balio "jarraiak" har ditzakeen kantitatea (zorizkoa)?

Bistakoa da, kantitate horri X baderitzogu, kontutan

hartzen ditugun gertaerak, $\{a \leqslant X \leqslant b\}$ formakoak direla, non a eta b muturrak zenbaki errealak diren. Baina, honek, elkartuta ko gertaera-kopurua infinitu "zenbakiezina" dela halabehartzen du. Aurreko atalean emandako espazio probabilistikoaren definizioa hain orokorra izatea, horrelako argudioek justifikatzen dute.

Behintzat, oraindik, gaian sakontzea nahi izan gabe, soilik seinalatzen dugu ezen aurrekoak bezalako zorizko kantitateen deskribapena ezin daitekeela $\{X=a\}$ moduko gertaeretara ko probabilitateen egokiera oinarri, beste arazo-mota batetan baizik. (Izatez, tratamendu pean dagoen landareak 2,619 cm-tako, doi-doi, edo beste edozein balio finkotako haziera jasotzen duena ez da inoiz gertatzen; hau dela eta, gertaera honi probabilitate nulua eman behar zaio. Honek, ezinezko gertaeraren eta probabilitate zero duen gertaeraren artean bereiztera eramaten gaitu).

II.5.- BALDINTZAZKO PROBABILITATEA ETA GERTAEREN ARTEKO INDEPENDENTZIA.

II.5.1.- Baldintzazko probabilitatea

Batzutan, ondokoa gertatzen da: Zorizko experimentuari buruzko informazio-mota zehatza lor dezakegu gutxi gora behera, eta honek, berari buruzko ikuspen orokorra aldarazten digu. Horrela adibidez, A gertaera burutu dela ezagutzen badugu, beste zenbait gauzaren artean, $\bar{A}$ gertaera ez dela burutu esan dezakegu, edo $B \subset A$ gertaerari, gertatzeko probabilitate handiagoa eman diezaiokegu.

Aurreko informazioaren nozio intuikorra, gure eredu probabilistikoari eransteko arazoa agertzen da. Ideia, maiztasunen bidez, ondoko oharpenek azaltzen dute:

Suposa dezagun zorizko experimentua N aldiz errepikatzen dela, non A gertaera f_A aldiz jazo bait da, eta berauen artean, f_{AB} aldiz jazo da B gertaera. Orduan, B-ren maiztasunak,

A jazotzen diren kasuak soilik kontutan harturik, edo A-ren jazotzeak baldintzatuak, zeina fr(B/A) moduan adierazten dugun, f_{AB}/f_A balio du.

$A \cap B$ gertaeraren maiztasunak $\frac{f_{AB}}{N}$ balio du.

Beraz,

$$fr(B/A) = \frac{f_{AB}}{f_A} = \frac{f_{AB}/N}{f_A/N} = \frac{fr(A \cap B)}{fr(A)}$$

Definizio berberaz, fr(B/A) zenbakiak, A gertatu denekо informazioak, B gertatzeko probabilitatea aldatzen dueneko neurri bat ematen digu.

Ideia hauek abstradituz, hurrengo definizio formala ematen dugu:

Biz $(\Omega, \mathcal{A}, p)$ espazio probabilistikoa eta $p(A) > 0$ baldintza betetzen duen A gertaera. A-rekiko baldintzazko B gertaeraren probabilitatea, hurrengo p(B/A) zenbakiari deituko diogu:

$$p(B/A) = \frac{p(B \cap A)}{p(A)}$$

Kontutan hartu behar dugu, p(B/A) balioa edozein gertaeratarako definitua geratzen dela, ondoko propietateak betetzen dituelarik:

i) $0 \leqslant p(B/A) \leqslant 1 \qquad \forall B \in \mathcal{A}$

ii) $p(E/A) = 1$

iii) $\{B_i\}_{i=1}^{\infty}$ gertaera-familia binaka bateraezinak baldin badira,
$p(B_1 \cup B_2 \cup \ldots B_n \cup \ldots /A) = p(B_j/A) + \ldots + p(B_n/A) + \ldots$
izango da.

Beraz, p-tik abiatuz, p(·/A) probabilitate berria definitzen da, baldintzazko probabilitatea izendatua eta gertaera desberdinetarako probabilitateen egokiera, A gertatu dela jakinik, adierazten duena.

Interpretazio hau indartzen da, ondoko propietateak egiaztatzerakoan:

a) $p(\bar{A}/A) = 0$

Orokorrean, $B \subset \bar{A}$ baldin bada, $p(B/A) = 0$ izango da

b) $p(A/A) = 1$

Halaber, $B \supset A$ baldin bada, $p(B/A) = 1$ izango da

c) Edozein B gertaerak, non $B \subset A$ den, $p(B/A) \geq p(B)$ betetzen du.

$p(A) > 0$ baldin bada, $p(B/A) = \frac{p(B \cap A)}{p(A)}$ modura definitu dugu. Aurreko baldintza, $p(A \cap B) = p(A)p(B/A)$ erara jarrita, probabilitate konposatuaren teorema izenaz ezagutzen da.

Ondokoa, erraz ikusten da:

$p(A \cap B \cap C) = p(A \cap B)p(C/A \cap B) = p(A)p(B/A)p(C/A \cap B)$
baldin $p(A \cap B) > 0$

Indukzioz, n gertaeratarako orokor daiteke aurreko formula:

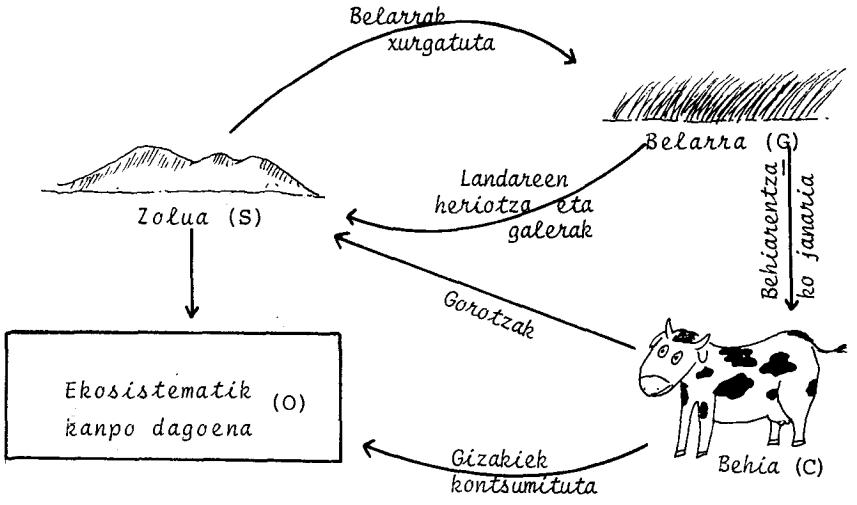
$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1)p(A_2/A_1)\ldots p(A_n/A_1 \cap \ldots \cap A_{n-1})$
baldin $p(A_1 \cap \ldots \cap A_n) > 0$

Baldintzazko probabilitate desberdinak ezaguturik, n gertaera aldi berean gertatzeko probabilitatea kalkulatzeko metodoa ematen digu formula honek.

5. adibidea:

Ekosistema batetan, gai-ibilbidearen eredua.

Kontutan har ezazu ekosistema batetako (artzantzarako zelaian) fosforo-atomo baten ibilbidearen ondoko eredua:



Atomoa, zoluan (S) baldin badago, hurrengo egunean, zoluan irauteko probabilitatea 3/5 da; belarrean (G) egotekoa (landareek asimilatuak izatearren), 3/10; edo, ekosisteman gertu ez egotekoa (O) (adibidez, erosioa dela kausa), 1/10.

Atomoa, belarretan (G) baldin badago, hurrengo egunean, zoluan (S) egoteko probabilitatea 1/10 da; belarretan (G) irautekoa, 4/10; edo, behi batek asimilatua izatekoa, 1/2.

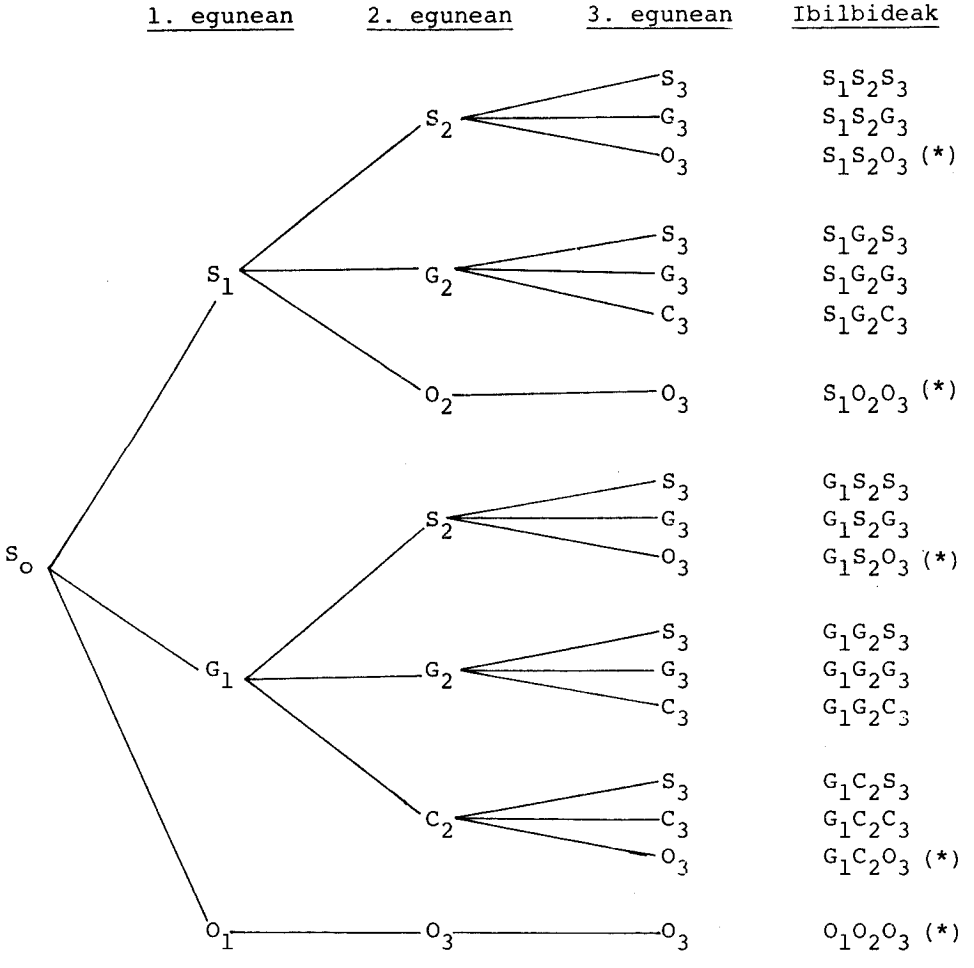
Atomoa, behi batetan (C) baldin badago, berton (C) irauteko probabilitatea 1/5 da; zolura (S) pasatzekoa, 3/4; edo, ekosistematik kanpo(O) galtzekoa, 1/20.

Ekosistematik kanpo (O) baldin badago, hara itzultzeko probabilitatea nulua da.

Atomo bat zoluan baldin badago (S_0, zero unean), zein da, hiru egunen buruan, ekosistematik kanpo (0_3 gertaera) egoteko probabilitatea?

Kalkulu hau burutzeko, 0_3 gertaera, 0_3-ra daramaten aurreko uneei dagozkien gertaera-kateen funtzioan deskonposa daiteke. Konkretuki, ekosisteman ibil ditzakeen "istorio" edo "bide" guztien multzoa eta beraiei elkartutako probabilitateak,

ondoko grafikoan emanak daude.



O_3 gertaera, hurrengo gertaera disjuntuen (ibilbideen) bildurak osotuta egongo da:

$S_1 \text{——} O_2 \text{——} O_3$, $\frac{3}{5} \cdot \frac{1}{10} \cdot 1 = \frac{3}{50}$ probabilitatearekin

$S_1 \text{——} S_2 \text{——} O_3$, $\frac{3}{5} \cdot \frac{3}{5} \cdot \frac{1}{10} = \frac{9}{250}$ probabilitatearekin

$G_1 \text{——} S_2 \text{——} O_3$, $\frac{3}{10} \cdot \frac{1}{10} \cdot \frac{1}{10} = \frac{3}{1000}$ probabilitatearekin

$G_1 \text{——} C_2 \text{——} O_3$, $\frac{3}{10} \cdot \frac{1}{2} \cdot \frac{1}{20} = \frac{3}{400}$ probabilitatearekin

$O_1 \text{——} O_2 \text{——} O_3$, $\frac{1}{10} .1.1 = \frac{1}{10}$ probabilitatearekin

Hau dela eta, eskatutako probabilitatea hauxe izango da:

$$p(O_3) = \frac{3}{50} + \frac{9}{250} + \frac{3}{1000} + \frac{3}{400} + \frac{1}{10} = \frac{413}{2000}$$

II.5.2.- Gertaeren arteko independentzia

Intuikorki, gertaeren arteko independentziaren nozioak, beren gertatzeen arteko eraginik ez dagoela erakusten du. Adibide erraz batez baliatuz, bola zuri eta beltzak dauzkan kutxontzi batetatik bi ateraldi itzuleradunen emaitzak askeak dira.

Nahiz eta gertaera askeen esanahiari buruzko ideia hartzea erraza izan, bi gertaera edo bi edo faktore gehiago askeak direneko arazo praktikoa oso problematikoa gerta daiteke. Gizakian pisua eta altuera, edo pultsua eta tenperatura, zein puntu taraino dira askeak? Zenbait aldagai biologikoren arteko menpekotasun-maila, baliogabeko mailetatik elkarrekiko eragin handietara oszila daiteke. Alegia, menpekotasunez ala independentziaz ezin da absolutuki hitz egin, menpekotasun-maila ezberdinez baizik.(*)

A-rekiko baldintzazko B-ren probabilitatearen definizioak, B gertatzean, A gertatzearen eragin-maila eransten duela eta, ondokoa definitzea logikoa da:

(*) Geroago, aldagaien arteko erregresio eta korrelazioaz hitz egiterakoan, eragin hoiek zenbakitzeko ereduak ezarriko ditugu. Ideia-maila honetan, tentagarria da kausa eta efektuen kontzeptuetan sartzea, era zabalean azterturik. Baina, plan teamendu hauek, Estatistikatik kanpo dauden "sasi-galderetara" (lehen aipatuak) jotzen dute. Hots, biztanlego batetarako pisu eta altueraren arteko erlazioa aztertzen badugu, esate baterako, metodo estatistikoek erlazio hau kuantifikatzen dute, baina ez dute ea altuera pisuaren "kausa" denentz adierazten.

Bira A eta B gertaera, $p(A)>0$ izanik. $p(B/A) = p(B)$ baldin bada, B gertaera A gertaerarekiko askea dela esaten dugu.

Kasu bakar honetan, A gertatzearen informazioak ez digu ezer esaten B gertatzeari buruz.

Ondoren, erraz egiazta daitekeen enunziatua azaltzen dugu:

Proposamendua: Bira A eta B gertaerak, $p(A)>0$ eta $p(B)>0$ izanik. Orduan, ondokoak baliokideak dira:

i) B gertaera, A gertaerarekiko askea da.
ii) A gertaera, B gertaerarekiko askea da.
iii) $p(A \cap B) = p(A)p(B)$

Lehen bi atalek, independentziaren nozioa simetrikoa dela erakusten dute, eta, beraz, "gertaera askeez" hitz egin dezakegula justifikatzen dute. Hirugarrenari dagokionez, egile askok, $p(A \cap B) = p(A)p(B)$ formula gertaera askeen definiziotzat hartzen du, zeren $p(A)=0$ edo $p(B)=0$ diren kasuak barnekoak izatearen abantaila eskeintzen bait du.

Are gehiago, gertaera batzutarako independentziaren kontzeptua orokortzeko era aditzera ematen du. Honela, ondokoa definitzen dugu:

$A_1,\ldots,A_n$ gertaerak "elkar askeak" deitzen dira, $1 \leqslant i < j < \ldots \leqslant n$ konbinazio guztietarako, hurrengo arauak betetzen baldin badira:

$$p(A_i \cap A_j) = p(A_i)p(A_j) \qquad \forall i \neq j$$
$$p(A_i \cap A_j \cap A_k) = p(A_i)p(A_j)p(A_k) \qquad \forall i \neq j \neq k$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1)p(A_2) \ldots p(A_n)$$

Lehendabiziko lerroan jarritako $\binom{n}{2}$ baldintzak soilik betetzen baldin badira, gertaerak binaka askeak direla esaten da.

6. adibidea: Sexu-banaketa.

Kontsidera ditzagun hiru umetako familiak, aaa,aae,.., eee zortzi posibilitateen arteko bakoitzak 1/8-tako probabilitatea daukala suposatzen dugularik. Bira H, "familiak sexu bitako umeak edukitzeari" dagokion gertaera eta A, "gehien jota, eme bat egoteko" gertaera. Orduan, $p(H) = 6/8$ eta $p(A) = 4/8$ dira. A eta H batera gertatzeak, aae, aea, eaa hiru posibilitateetariko bat adierazten du, $p(A \cap H) = 3/8 = p(A)p(H)$ delarik.

Horrela, hiru umetako familietan, gertaera hauek askeak dira. Baina, hau, bi edo lau umetako familietan, egia ez dela kontutan hartu behar dugu. Honek, independentzia egotea bidezkoa ez dela frogatzen du.

II.5.3.- Bayes-en teorema.

Probabilitate osoaren izeneko teorema frogatzen hasiko gara, laster tresna gisara erabiltzeko.

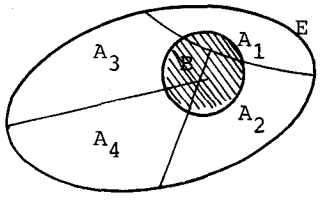
Probabilitate osoaren teorema:

Bira $A_1,\ldots,A_n$ gertaera-sistema osoa, hots, $\bigcup_{i=1}^{n} A_i = E$ eta $A_i \cap A_j = \emptyset \ \forall i \neq j$ betetzen dituen n gertaeratako familia. Biz B edozein gertaera. Orduan, $p(B) = p(B/A_1)p(A_1)+\ldots+p(B/A_n)p(A_n)$ da, $p(A_1)>0,\ldots,p(A_n)>0$ suposaturik.

Frogapena: Kontutan har dezagun, B, gertaera-bildura disjuntu bezala adieraz daitekeela, alegia:

$$B=B \cap E=B \cap (\bigcup_{i=1}^{n} A_i) =$$

$$= (B \cap A_1) \cup \ldots \cup (B \cap A_n)$$



Beraz, $p(B) = p(B \cap A_1) + \ldots + p(B \cap A_n)$

$p(A_i \cap B) = p(B/A_i)p(A_i)$ dela gogoratuz, azkenean hauxe daukagu:

$$p(B) = p(B/A_1)p(A_1)+\ldots+p(B/A_n)p(A_n)$$

Jadanik, ondokoa aipa dezakegu:

Bayes-en teorema:

Bira $A_1,\ldots, A_n$ gertaera-sistema osoa eta B beste gertaera bat, guztien probabilitateak hertsiki positiboak izanik. Familiaren edozein gertaeraren (A_1-en, esate baterako) B gertaeraren baldintzazko probabilitatea, modu honetan ematen da:

$$p(A_1/B) = \frac{p(B/A_1)p(A_1)}{p(B/A_1)p(A_1)+\ldots+p(B/A_n)p(A_n)}$$

Frogapena:

Definizioz, $p(A_1/B) = \dfrac{p(A_1 \cap B)}{p(B)}$

Baina, zenbakitzailea $p(A_1 \cap B) = p(B/A_1)p(A_1)$ erara jar daiteke eta probabilitate osoaren teorema erabiliz p(B) adierazteko, teoremaren adierazpena lortzen da berehala.

Bayes-en teoremaren interpretazioaren azalpena egitea interesgarria da: Suposa dezagun B gertaera, $A_1,\ldots, A_n$ "kausa" ezberdinek (elkar-ukatzaileak) sorterazia izan daitekeela. Halaber, suposa ditzagun A_i kausaren presentzian B gertatzeko probabilitatea $p(B/A_i)$ 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" sorterazi duen probabilitateaz galdetzen diogu gure buruari. Probabilitate hau, Bayes-en teoremak ematen duena da. $p(A_i)$ zenbakiak "a priori" probabilitateak deitzen dira, $p(B/A_i)$-ak, egiantzak eta $p(A_i/B)$-ak, "a posteriori" probabilitateak.

Bayes-en teoremari buruzko eztabaida handia izan dela, eta dela, aipatu behar da. Teknikoki, praktikan jartzerakoan, "a priori" probabilitateak kalkulatzeko oztopoak daude. Baina, batez ere, bere interpretazioan dagoen filosofia da eztabaidaren zergatia.

Xehetasunak sakonki aztertu nahi gabe, esan dezakegu na hauxe da: Bayes-en teorema aplikatzerakoan, kontu handiz ibili behar dugu. Interpretaziotik kanpo, dudarik gabeko egia matematikoa dauka aipaturiko teoremak.

7. adibidea:

A, B eta C, hiru bakteriotako preparazioa daukagu, 10%, 30% eta 60%-ean, hurrenez hurren. A motakoek, 80%-an sulfatoekin erreakzionatzen dute; B motakoek, 60%-an, eta C motakoek, 40%-an.

Bakterio bat isolatzen da eta sulfatoekin erreakzionatzen duela egiaztatzen da. Hemendik aurrera, A motakoa edo B-koa edo C-koa izateko probabilitateak kalkula daitezke.

Izan ere, $p(A) = 0,1$; $p(B) = 0,3$; $p(C) = 0,6$ dira.

Biz S, "bakterio isolatuak sulfatoekin erreakzionatzen dueneko" gertaera. Ondokoak dakizkigu:

$$p(S/A) = 0,8 \; ; \; p(S/B) = 0,6 \; ; \; p(S/C) = 0,4$$

Bayes-en teorema aplikatuz, hauxe lortzen da:

$$p(A/S) = \frac{p(S/A)p(A)}{p(S/A)p(A)+p(S/B)p(B)+p(S/C)p(C)} =$$

$$= \frac{0,8 \cdot 0,1}{0,8 \cdot 0,1+0,6 \cdot 0,3+0,4 \cdot 0,6} = \frac{8}{50} = 0,16$$

Era berean:

$$p(B/S) = \frac{18}{50} = 0,36$$

$$p(C/S) = \frac{24}{50} = 0,48$$

Bakterioa C motakoa izateko probabilitatea, beste motatakoa baino handiagoa dela egiaztatzen da lorturiko emaitzetatik.

I. ERANSKINA

MULTZO-TEORIARI BURUZKO IRAZKIN LABURRAK

Nahiz eta multzo-teoria bere forma axiometrikoan aurkeztea ri uko egin (ukoa bera, bestalde, baliagarria gertatzen zaigu hemen), multzo, elementu eta elementu bat multzo baten barnekoa izatea edo ez izatearen nozioak intuikorrak direla aipatu behar dugu; hots, ezin daitezke defini. Gogora dezagun, multzo bat determinatuta dagoela, osotzen duten elementuak azaltzen dituzunean edo bere elementuek eta beraiek soilik egiaztatzen duten propietateren bat ematen dugunean.

Elementuak, letra minuskulaz adierazten dira; multzoak, letra majuskulaz edo beraien elementuak giltzen artean itxiz, eta barnekoa izatea edo ez izatearen ikurrak, $\in$, $\notin$, hurrenez hurren.

Errepasa ditzagun propietate nagusiak:

- A multzoa B multzoan edukia dagoela edo B multzoak A daukala edo A multzoa B-ren azpimultzoa dela esaten dugu, A-ren elementu guztiak B-ren barnekoak baldin badira, $A \subset B$ edo $B \supset A$ adierazirik.

- $A \subset B$ eta $B \subset A$ baldin badira, A=B dela esaten da. Horrela, multzoak berdinak direla frogatzeko, edukin bikoitzez egin behar da.

- A eta B, bi multzotatik abiatuz, multzo berriak eraiki daitezke: $A \cup B$ bildura, non elementuak A barnekoak edo B barnekoak bait dira; $A \cap B$ ebakidura, non elementuak bien elementu komunak bait dira; A-B kendura, B barnekoak ez diren A barnekoek osotua.
  Multzo-bilduma zenbakigarri baterako (finitua edo infinitu zenbakigarria), kendura, bildura eta ebakidura analogoki definitzen dira, $\bigcup_{i=1}^{n} A_i$, $\bigcap_{i=1}^{n} A_i$ (n multzotako bildura finiturako), $\bigcup_{i=1}^{\infty} A_i$, $\bigcap_{i=1}^{\infty} A_i$ (bilduma zenbakigarrirako) adierazten direlarik. A-k eta B-k elementu ko

munik ez badaukate, $A \cap B$ multzoak ez dauka elementurik, $A \cap B = \emptyset$ adierazirik.

- A, E-ren azpimultzoa baldin bada, A barnekoak ez diren E-ren elementu-multzoari E-rekiko A-ren multzo osagarria ($\bar{A}$ adierazia) deritzogu.

- Multzo-bildura eta ebakiduraren trukatze-, elkartze-, eta banatze- propietateak oso ezagunak dira; halaber, osagarrietarako Morgan-en legeak:

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \quad ; \quad \overline{A \cap B} = \bar{A} \cup \bar{B}$$

Lege hauek, $\{A_i\}_{i=1}^{\infty}$ edozein bildumatarako orokortzen dira:

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \bar{A}_i \quad ; \quad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \bar{A}_i$$

- E multzoa emanik, E-ren azpimultzo guztiek osotzen duten multzoari E-ren zatien multzoa deritzogu, $\mathcal{P}(E)$ adierazten delarik. E-ren n elementu-kopurua finitua denean, $\mathcal{P}(E)$ delakoak 2^n elementu ditu; hauetan, 0 elementutako azpimultzoak (multzo hutsa, edozein multzoren azpimultzotzat hartzen dena), elementu batetako $\binom{n}{1}$ azpimultzoak, 2 elementutako $\binom{n}{2}$ azpimultzoak,..., n elementutakoak (hots, E bera) sartzen direlarik.
Denetara:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} = (1+1)^n = 2^n$$

Newton-en binomioaren formula kontutan hartuz.

- Ikus dezagun pare bat "maina", probabilitate-teorian interesgarriak direnak: A multzoa, multzo-bilduma baten bildura, hots, $A = \bigcup_{i=1}^{\infty} A_i$, multzo disjuntuen bildura modura (bi multzo disjuntuak dira, beraien ebakidura hutsa denean), edo multzo-familia "gorakor" baten bildura modura adieraz daiteke beti.

Lehenengo kasuan, hauxe da eraiketa:

$B_1 = A_1$

$B_2 = A_2 \cap \overline{A}_1$

$B_3 = A_3 \cap (\overline{A_1 \cup A_2})$

.....

$B_n = A_n \cap (\overline{A_1 \cup \ldots \cup A_{n-1}})$

.....

$B_1, B_2, \ldots, B_n, \ldots$ multzoak disjuntuak dira binaka eta

$A = \bigcup_{i=1}^{\infty} B_i$

Bigarrenari dagokionez, ondokoa kontsideratzen dugu:

$C_1 = A_1$

$C_2 = A_1 \cup A_2$

.....

$C_n = A_1 \cup A_2 \cup \ldots \cup A_n$

.....

Orduan, $C_1 \subset C_2 \subset \ldots \subset C_n \subset \ldots$ eta $A = \bigcup_{i=1}^{\infty} C_i$

- E multzoa emanik, $A_1, \ldots, A_n$ beraren azpimultzo-bilduma (zenbakigarria izan daitekeena) bati E-ren partiketa deritzogu, $\bigcup_{i=1}^{n} A_i = E$ eta $A_i \cap A_j = \emptyset \ \forall i \neq j$ baldin badira.
  Erraz ikusten denez, E multzoa, era askotan parti daitekeke (kasu nabariak ahazturik).

- Azkenik, gogora dezagun aplikazioaren kontzeptua:
  A eta B, bi multzo emanik, A-tik B-rako aplikazioa, A-ren elementu bakoitzari B-ren bat eta bakar bat elkartzen dion "legea" da, $f: A \longrightarrow B$ adierazirik.
  A multzoari jatorrizko multzoa deitzen diogu, eta B, irudi multzoa da.
  f aplikazioak, $a \in A$ elementuari $b \in B$ elementua elkartzen badio, b, f-ren bidezko a-ren irudia dela esaten da, $f(a) = b$ idazten delarik.

II. ERANSKINA

KONBINATORIAREN ELEMENTUAK

Konbinazio-analisiak ebatzi nahi duen arazo generikoa, ondoko terminoez formula liteke: m elementutako multzoa emanik, zenbat n elementutako azpimultzo "desberdin" era daiteke, emandako multzoaren elementuak erabiliz? Erantzuna ez da bakarra, eta, kasu bakoitzean, azpimultzo "desberdinez" ulertzen denaren menpekoa da, frogatzeko aukera izango dugunez.

Aldez aurretik, sar dezagun izaera teknikoaren zenbait nozio:

n zenbaki arrunt bakoitzerako, n faktoriala, n lehen zenbaki arrunten arteko biderkadura da, $n!=n(n-1)\cdot(n-2)\ldots\ldots 3\cdot 2\cdot 1$ idazten dugularik.

$n!=n(n-1)! = n(n-1)\cdot(n-2)!= \ldots$ erraz ikusten da. Komenientziazko arrazoiak direla eta, $0! = 1$ definitzen da.

Halaber, m, n zenbaki arruntak emanik, $m \geqslant n$ izanik (noizean behin, $m=n=0$ izan daitezke), "m gain n" konbinazio-zenbakia honela definitzen dugu:

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}$$

Frogatzeko propietate errazak, hauxek dira:

1) $\binom{m}{0} = \binom{m}{m} = 1 \quad ; \quad \binom{m}{1} = m$

2) $\binom{m}{n} = \binom{m}{m-n}$

3) $\binom{m}{n} + \binom{m}{n+1} = \binom{m+1}{n+1}$

Egingo ditugun baieztapenak frogatu nahi gabe, hel diezaiogun gaiari:

$\{a_1, a_2, \ldots, a_m\}$ m elementutako multzoa kontsideratzen dugu. Bertatik atera daitezkeen n elementutako ($n \leqslant m$) taldeei honela deituko diegu:

- Aldakuntza arruntak: taldeak, desberdintzat hartzen ditugu, elementuren batez diferitzen direnean, edo, elementu berberak izanik, beraien ezarrera-ordenaz diferitzen direnean. Beraien kopurua, $V_m^n = m(m-1)\ldots(m-n+1)$ da.

- Konbinazio arruntak: bi talde, desberdintzat hartzen ditugu, elementuren batez diferitzen direnean. Beraien kopurua, $C_m^n = \binom{m}{n}$ da.

- Permutazio arruntak: bi talde bereizteko erizpidea, elementuen ordenan datza soilik. (Kasu honetan, jakina, n=m izan behar du). Beraien kopurua, $P_n = n!$ da.

Kontutan har dezagun, $V_m^n = C_m^n \cdot P_n$ dela.

$\{a_1, a_2, \ldots, a_m\}$ m elementutako multzoa emanik, elementuak errepikatu ahal izanik, beraietatik atera daitezkeen n elementutako (n, hautazkoa) taldeei honela deituko diegu:

- Errepikatuzko aldakuntzak: bi talde bereizteko erizpidea, aldakuntza arruntena da eta beraien kopurua $V'^n_m = m^n$ da.

- Errepikatuzko konbinazioak: bi talde bereizteko erizpidea, konbinazio arruntena da, beraien kopurua $C'^n_m = \binom{m+n-1}{n}$ delarik.

Azkenik, $\underbrace{a,\ldots,a}_{n_1},\underbrace{b,\ldots,b}_{n_2},\ldots,\underbrace{k,\ldots,k}_{n_k}$ multzoa emanik, $n_1+\ldots+n_k = n$ izanik, hots, n elementutakoa, beraien arteko n_1 lehenengo klasekoak, n_2 bigarren klasekoak,..., n_k k.klasekoak izanik, berari eman dakizkiokeen ordenamendu desberdinei, multzo honen errepikatuzko permutazioak deritzegu. Beraien kopurua, hauxe da:

$$P_n^{n_1,n_2,\ldots,n_k} = \frac{n!}{n_1!n_2!\ldots n_k!}$$

Konbinazio-analisiaren hurbil dagoen formula bat, Newton-en binomioarena da. Hain zuzen ere:

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b+\ldots+\binom{n}{n-1}\, a\, b^{n-1}+\binom{n}{n}b^n$$

$a,b\in\mathbb{R}$, $n\in\mathbb{N}$-etarako baliagarria izanik.

Partikularrean, $\binom{n}{0} + \binom{n}{1} +\ldots+ \binom{n}{n} = (1+1)^n = 2^n$

Binomioaren formularen orokorpen garrantzitsua, ondokoa da:

$$(p_1 + p_2 +\ldots+ p_k)^n = \sum \frac{n!}{x_1!x_2!\ldots x_k!}\, p_1^{x_1}\, p_2^{x_2}\ldots p_k^{x_k}$$

non batukaria $0\leqslant x_1<x_2<\ldots<x_k\leqslant n$ zenbaki arrunteta-ra hedatua bait dago, $x_1+x_2+\ldots+x_k = n$ baldintza bete-rik.

PROBLEMAK

1.- Bira A, B eta C, hautazko hiru gertaera. Aurki itzazu hurrengo gertaeretarako adierazpenak (zeintzu A, B eta C-ren menpekoak bait dira, jakina):

I) A soilik gertatzen da.

II) A nahiz B gertatzen dira, C ez da gertatzen ordea.

III) Hirurak gertatzen dira.

IV) Gutxienez bat gertatzen da.

V) Gutxienez bi gertatzen dira.

VI) Bat eta bakar bat gertatzen da.

VII) Bi gertatzen dira soilik.

VIII) Bat ere ez da gertatzen.

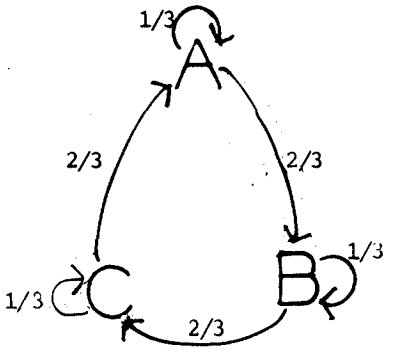
IX) Bi baino gehiago ez dira gertatzen.

2.- Jokalari italiarrak Galileori bere harridura adierazi zion, hiru datoz jokatzerakoan, 10 batura 9a baino maizago agertzen zela kontutan hartzean.

Jokalariak zioenez, 9aren aldeko kasuak hauxek ziren: 126 - 135 - 144 - 225 - 234 - 333; eta 10aren aldekoak: 136 - 145 - 226 - 235 - 244 - 334.

Baina Galileok, konbinazio hauek ezin zitezkeela ekipro babletzat har ikusi zuen. Adiera ezazu zergatia eta kalkula itzazu dagozkien probabilitateak.

3.- A, B, C, hiru egoerek ekosistema osotzen dute. Bertan, elementu bat aurki daiteke, esate baterako, nitrogenoa. Nitrogeno atomo baten ibilbidea, honela eskematiza daiteke:



Une batetan, atomoa A (edo B edo C) egoeran baldin badago, hurrengo etapan, ondoko bi posibilitateak ditu:

1) B (edo C edo A) ondoz ondoko egoerara pasatzea, probabilitatea 2/3 izanik

2) hasierako egoeran bertan irautea, probabilitatea 1/3 izanik.

Atomoa, A egoeran dagoela suposatzen baldin badugu, aurki itzazu:

I) bi etaparen buruan, B-an egoteko probabilitatea

II) Hiru etaparen buruan, C-an egotekoa

III) Lau etaparen buruan, A-ra, lehen aldiz, itzultzekoa

IV) Lau etaparen buruan, A-n egotekoa.

4.- Egoteko kasuan, testak T motako bakterioen presentzia detektatzen du uretan, probabilitatea 0,9 izanik. Bakteriorik ez badago, egoneza detektatzen du, probabilitatea 0,8 izanik.

Ur laginak, benetan, T motako bakterioak edukitzeko probabilitatea 0,2 dela jakinik, kalkula itzazu:

a) Testa positiboa izan denean, ur laginak T motako bakterioak edukitzeko probabilitatea.

b) Testa negatiboa izan denean, ur laginak T motako bakterioak edukitzekoa.

c) Ur laginak T motako bakterioak edukitzeko, eta, gainera, testak positiboa izatekoa.

5.- Tratamendu berriaren eraginaz, hurrengo datuak ditu osagileak:

	(C) Hirian bizi direnak		($\bar{C}$) Hirian bizi ez direnak	
	(B) tratat.	($\bar{B}$) ez trat.	(B) trat.	($\bar{B}$) ez tratat.
(A) sendatuak	1000	50	95	5000
($\bar{A}$) ez sendatuak	9000	950	5	5000

A = "sendatuak", B = "tratatuak", C = "hiriko biztanleak" izanik, kalkula itzazu: $p(A/B)$; $p(A/\bar{B})$; $p(A/B \cap C)$; $p(A/\bar{B} \cap C)$; $p(A/\bar{B} \cap \bar{C})$. Interpreta itzazu emaitzak.

6.- h elementutako populaziotik, r elementutako lagina hartzen da. Aurki ezazu, N elementu aurrefinkaturen arteko bat ere laginaren barnekoa ez izateko probabilitatea, laginketa:

a) itzuleragabea

b) itzuleraduna

dela suposatzen baldin badugu.

Konpara itzazu bi metodoen zenbakizko balioak:

I) $h = 100$, $r = N = 3$

II) $h = 100$, $r = N = 10$

izanik.

7.- Biz E, "Drosophila melanogaster"-ko π populazioan jaiotako ale bat harrapatzean eta aztertzean datzan experientzia. Experimentatzaileak, A = "Bar begiak" eta B = "hego bestigialak" karaktere genetikoak aztertu nahi ditu.

i) Deskriba ezazu behaketa-maila honi elkartutako gertaera-algebra (zenbat eta zeintzu diren oinarrizko gertaerak

eta zenbat eta zeintzu diren gertaera konposatuak).

ii) Suposa dezagun, fenomenoaren etengabeko behaketatik, $p(A) = 1/4$ eta $p(B) = 1/2$ direla ondorioztatu dugula. Datu hauetatik abiatuz, algebraren gertaera guztien probabilitateak kalkula al daitezke?

iii) Suposa dezagun A eta B karaktereak askeak direla. Kalkula itzazu algebraren gertaera guztien probabilitateak.

iv) Aurreko experientzian, A, B eta C hiru karaktere genetiko kontutan hartzen baldin baditugu, deskriba ezazu behaketa-maila horri elkartutako gertaera-algebra.

Are gehiago, A_1, A_2,..., A_k, k karaktere genetiko kontutan hartuko balira, deskriba ezazu, halaber, elkartutako gertaera-algebra.

8.- Zurrumurru-hedapena.

h+1 biztanletako herrian, pertsona batek beste bigarren pertsona bati zerbait esaten dio zurrumurru modura; bigarrenak beste hirugarren bati errepikatzen dio, eta abar. Pauso bakoitzean, zurrumurruaren hartzailea, zoriz aukeratzen da h pertsona gertuen artean.

Aurki ezazu, zurrumurrua r aldiz igarotzeko probabilitatea:

a) sortu zuenari itzuli gabe

b) pertsona bati errepikatu gabe.

9.- h bola, h gelatxotan zoriz ezartzen baldin badira, gelatxo batek gutxienez bola bat edukitzeko probabilitatea hauxe da:

$$\frac{h!}{h^h}$$ (Egiazta ezazu)

Balio hau ustegabeko txikia da; esate baterako, h = 7 denean, 0,00612 balio du. Honek, ondokoa adierazten du:

Hiri batetan, 7 istripu astero gertatzen baldin badira, orduan, (banaketa guztiak ekiprobableak direla suposatuz), aste guztietan bi istripu edo bi baino istripu gehiago dau den egunak egongo dira, eta, batezbestekoan, 165 astetako aste bakar batek, "istripu bat eguneko" banaketa uniformea aurkeztuko du.

10.- Demagun OBAC errektangelua, non OB eta OC aldeak h eta p neurrikoak bait dira, hurrenez hurren, hp laukitan zatitu ta dagoela. O-tik A-raino doan edozein bideri L deitzen badugu, eskuineranzko desplazamendu horizontalak edo goranzko bertikalak egon daitezkeela suposaturik, froga eza zu bide guztiek luzera berdina daukatela eta kalkula ezazu dagoen bide-kopurua.

Bide guztiak ekiprobableak baldin badira, zein da O-tik A-raino joateko probabilitatea, aurrefinkatutako "diagonal" baten azpian pasatuz?

11.- Zentsuaren datuetatik abiatuz, begi beltzetako aitek eta begi beltzetako umeek ($A \cap B$), zentsuan dauden pertsonen 5%-a osotzen dute; begi beltzetako aitek eta begi argieta ko umeek ($A \cap \overline{B}$), 7,9%-a; begi argietako aitek eta begi bel tzetako umeek ($\overline{A} \cap B$), 8,9%-a; eta begi argietako aitek eta umeek ($\overline{A} \cap \overline{B}$), 78,2%-a. Aurki ezazu aita eta umeen begien koloreen artean dagoen lokarria.

12.- Suposa dezagun minbizirako proba diagnostikoa dagoela, $p(\overline{A}/\overline{C}) = 0,95$ eta $p(A/C) = 0,95$ propietateak betetzen ditue la, C = "pertsona batek minbizia dauka" eta A = "testa po sitiboa da" gertaerak izanik.

p(C) = 0,005 suposatuz, kalkula ezazu, testak positiboa emanez gero, pertsona batek minbizia edukitzeko probabilitatea.

13.- Superbizipen-eredua.

Kontsidera dezagun populazio baten aleen bizitza-aldia (zorizkoa). Suposa ditzagun, sinplifikatzeko, hurrengo hipotesiak:

a) Ale guztiak aldiberean jaiotzen dira.

b) Aldia, iraupen berdineko tarteetan banatzen da.

c) Bizitza-aldiak ez du A bornea gainditzen.

Ondoko experientzia diseinatzen dugu: (0,A) tartea, iraupen berdineko azpitarteetan zatitzen dugu eta l_0 jaioberri-kopurua kontsideratzen dugu.

t = 0,1,2,....,k uneetan bizirik irauten duteneko kopuruak idazten ditugu, l_0, l_1,..., l_{k-1}, $l_k = 0$ lorturik, hurrenez hurren.

t unean, ale bat bizirik irauteko probabilitatea $p_t = \frac{l_t}{l_0}$ den hipotesia egiten dugu, t = 0,1,...,k izanik. (Eztabaida ezazu zein puntutaraino bidezkoa den).

Froga itzazu ale bat, t eta t+1-en arteko tartean, hila izateko probabilitatea,

$$q_t = \frac{l_t - l_{t+1}}{l_0} = P_t - P_{t+1}$$

dela; eta t-1 unera bizirik heldu deneko baldintzarekin, t unean hila izateko probabilitatea,

$$r_t = \frac{l_{t-1} - l_t}{l_{t-1}} = \frac{P_{t-1} - P_t}{P_{t-1}} \quad \text{dela.}$$

Orain, demagun 1000 ale jaioberritako populazioa. Ere duaren hipotesiak zuzenak baldin badira, zenbat alek iraun go dute bizirik, t une arte? Zenbat heriotza gertatuko di ra, t eta t+1-en arteko tartean?

14.- "Sitophilus"-ko populazioaren sexuarteko arrazoia.

Experimentatzaileak, "Sitophilus Zea-mais" delakoaren 500 ale-bikote isolatzen ditu zoriz. Berauetan, sexu ber-dineko eta desberdineko ale-bikoteak sartzen omen dira.

500 bikote hauen artean, 25-ek soilik izaten duten on dorengoa. Ondorengorik ez daukaten (adibidez, beraien se-xujotzeko aparailuak ondo ez gardainatzeagatik) benetako (emea eta arra) bikoteen proportzioa 5%-a da.

Datu hauetatik abiatuz, 500 bikoteetan dauden benetako bikoteen proportzioa (eme-proportzioa arrena baino handia-goa dela suposatzeko arrazoiak ba daude) eta π populazioa-ren sexu-proportzioa kalkulatzea eskatzen da. (Sexu biak morfologikoki bereizteak, zenbait oztopo aurkez ditzakeela eta, kalkulu hauek zeharkako bidez burutzea interesgarria izan daiteke).

Honetarako, bira hurrengo gertaerak:

D = "sexu desberdineko bikotea" eta

V = "ondorengoa izaten duen bikotea".

Froga ezazu $p(D) = 1/19$ (*) dela ($p(V/\overline{D})=0$ suposaturik, nabaria dirudienez).

π populazioaren sexu bataren proportzioari p, eta bes tearenari q deiturik, ondokoa arrazona ezazu:

Aleak zoriz aukeratzen baldin badira, sexu baten p^2 se xubakarreko bikote, 2pq bikote heterosexual (**) eta beste sexuaren q^2 sexubakarreko bikote egongo direla.

(*) eta (**)-tik abiatuz, ondoriozta itzazu:

p(f) = 0,97 eta p(m) = 0,03 ala

p(m) = 0,97 eta p(f) = 0,03 direla (f=emea eta m=arra)

15.- Gautxori batek hiru giltza dauzka. Giltzak guztiz bereizgaitzak dira eta beraien arteko bakar batek irikitzen du beraren etxebizitzako atea. Behar duen giltza aurkitzeko, hurrengo metodoetakoren batez baliatzen da:

1) giltzak, bata bestearen ondoren erabiltzen ditu, giltza berbera berriro ez erabiltzeko kontuz ibilirik.

2) giltza bat erabiltzen du, eta balio ez badu, giltzatakoa astindu eta berriro saiatzen da, horrela, berriro erabiltzeko arriskuan egonik.

I) 1. metodoari jarraituz, zein da hirugarren aldiz irikitzeko probabilitatea? Eta, 2. metodoari jarraituz?

II) Gehiegi edan eta gero, etxeratzerakoan (hau, hiru egunetatik behin gertatzen da), gautxoriak 2. metodoa erabiltzen duela, eta edan gabe datorrenean, 1. metodoa erabiltzen duela ba dakigu.

Lehenengo bi aldiz huts egin duela jakinez gero, zein da gautxoria mozkorra izateko probabilitatea?

16.- Mendelen experientzietariko bat aztertuko dugu: lore hori eta berdeak dauzkaten ilar homozigotikoen arteko gurutzamendua.

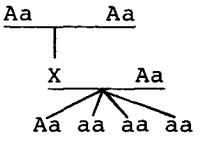
Lehenengo belaunaldian, ilar guztiek lore horiak izango dituzte, zeren A (horia) genea meneratzailea bait da eta a (berdea), ordez, errezesiboa. Ilar hauetariko bi gurutzatzean, lori hori eta berdeak dauzkaten ilarra, 3:1 proportzioan agertzen dira bigarren belaunaldian; hots,

agertzen diren lore horiko ilar-kopurua, berdearenaren hirukoitza da.

I) Bigarren belaunaldi honen lore horiko ilar bat hartzen dugu. Zein da homozigotikoa izateko probabilitatea? Eta heterozigotikoa izatekoa?

II) Lore horiko ilar hau, beste lore berdeko batekin gurutzatzen dugu, 5 ilar hori lortzen ditugularik. Mendelen experientziaz dakigunez, ilarra homozigotikoa baldin bada, gurutzamenduz lorturiko ilar guztiak horiak izango dira, eta heterozigotikoa baldin bada, ondorengo bakoitza lore horikoa izateko probabilitatea 1/2 da. Zein da hartutako landareak daukan homozigotikoa izateko probabilitatea?

17.- Biz ondoko zuhaitz genealogikoa:



Aurki itzazu, X alea: I) AA; II) Aa; III) aa izateko probabilitateak

18.- Familia batek seme bat edo alaba bat izateko probabilitatea 1/2 dela ba dakigu.

I) Bi umetako familia bat aukeratzen dugu, beraien arteko bat gizonezkoa dela ikusirik. Zein dá biak gizonezkoak izateko probabilitatea?

II) Biz hiru umetako familia bat. "Familia sexu bitako umeak dauzka" gertaerari H deritzogu eta "ez dago emakume bat baino gehiagorik" gertaerari, A.

A eta H gertaerak askeak al dira?

III. GAIA: ZORIZKO ALDAGAI DISKRETUAK ETA BEREN PROBABILITATE-LEGEAK.

III.1.- Aldagai estatistikoa eta zorizko aldagaia.

III.2.- Zorizko aldagai diskretuaren probabilitate-legea eta banaketa-funtzioa. Adierazpide grafikoak.

III.2.1.- Zorizko aldagaiaren probabilitate-legea.

III.2.2.- Banaketa-funtzioa. Propietateak.

III.2.3.- Adierazpide grafikoak.

III.3.- Zorizko aldagai diskretuaren batezbestekoa eta bariantza.

III.3.1.- Zorizko aldagai diskretuaren batezbestekoa edo itxaropena.

III.3.2.- Itxaropen matematikoaren propietateak.

III.3.3.- Zorizko aldagai diskretuaren bariantza eta desbidazio standarda.

III.3.4.- Bariantzaren propietateak.

III.3.5.- Sakabanatzea eta Tchebycheff-en teorema.

III.4.- Banaketa binomiala.

III.5.- Banaketa hipergeometrikoa .

III.6.- Banaketa multinomiala.

III.7.- Poisson-en banaketa.

ZORIZKO ALDAGAI DISKRETUAK ETA BEREN PROBABILITATE-BANAKETAK

III.1.- ALDAGAI ESTATISTIKOA ETA ZORIZKO ALDAGAIA.

ξ zorizko experimentua kontsideratzen badugu eta berari dagozkion n proba egiten baditugu, lorturiko behaketen multzoari, n hedadurako zorizko lagina deituko diogu. Emaitzen multzo honek taula estatistikoa ematen du eta bertan aldagaiaren balioei funtzio batzu dagozkie. ξ zorizko experimentuaren n burutzapenen n emaitzak aurkezten dituen aldagai honi, aldagai estatistikoa deritzogu.

1.adibidea:

π populazioaren aleek, lokus batetarako, A eta a alelo meneratzailekideak aurkez ditzakete, fenotipikoki, AA, Aa eta aa genotipoko aleak bereizirik. Ale bakoitzerako, daraman A alelo-kopurua kontutan hartzen dugu. 10 ale aztertzerakoan, ondoko balioak eta maiztasunak hartzen ditu aldagai honek:

X	0	1	2
fr.	6/10	3/10	1/10

ξ experimentuari dagokion kontaezinezko proba egin dugula suposatzen baldin badugu, kontaezinezko emaitza posibleek, ξ experimentuari elkartutako zorizko aldagaiaren nozioa sorterazten dute. Kasu honetan, zorizko aldagaiak, aipaturiko experimentuaren oinarrizko gertaera posibleak adierazten dituzten balioak hartzen ditu, dagozkien probabilitateekin batera. Aurreko kasuan,

$$X: \Omega \longrightarrow \mathbb{R}$$

$$\omega \rightsquigarrow X(\omega) = \text{"}\omega\text{-ren A alelo-kopurua"}$$

zorizko aldagaiak ondoko balioak hartzen ditu:

$$X(\omega)=\begin{cases}0 \text{ baldin } \omega \in aa\\ 1 \text{ baldin } \omega \in Aa\\ 2 \text{ baldin } \omega \in AA\end{cases}$$

$p(AA)=0,49$, $p(Aa)=0,42$, $p(aa)=0,09$ maiztasun genotipikoek emandako probabilitateak izanik.

Taula modura adierazirik:

X	0	1	2
p	0,09	0,42	0,49

Horrela, lagin, maiztasun eta aldagai estatistikoen kontzeptu konkretuak lortzen ditugu. Berauetatik, abstrakzio-prozesuaren bidez, populazio edo unibertso, probabilitate eta zorizko aldagaien kontzeptu teorikoak ondorioztatzen dira.

Beraz, zorizko experimentu bati elkartutako zorizko aldagaia, experimentuaren emaitza posibleen unibertso edo osotasunaren parteen gainean definituriko funtzio erreala da. Experimentuaren unibertsoa, zorizko aldagaiaren eremua da, aldagaiaren balioen multzoa bere heina izanik.

Zorizko aldagaiak, diskretu eta jarraietan sailkatzen dira, hartzen duten balio-kopuruaren arauera: kopuru hau, finitua edo infinitu zenbakigarria (zenbaki arruntena bezala) baldin bada, aldagaia diskretua izango da; aldagaiak, ordea, segmento edo zuzenerdi baten balio guztiak (adibidez, zenbaki erreal positibo guztiak) har baldin baditzake, jarraia izango da.

2.adibidea:

Azterketa hematologikoan, mm^3 odol bakoitzari linfozito-kopurua egoki dakioke, horrela zorizko aldagaia definiturik. Pertsona normalen linfozito-kopururik handiena mm^3 odoleko 3.500 eta txikiena 900 direla suposatuz, aldagai honen heina hauxe da:

X: 900, 901, 902, ..., 3.500

III.2.- ZORIZKO ALDAGAI DISKRETUAREN PROBABILITATE-LEGEA ETA BANAKETA-FUNTZIOA. ADIERAZPIDE GRAFIKOAK.

III.2.1.- Zorizko aldagaiaren probabilitate-legea.

Zorizko aldagaiaren balio bakoitzari, beraren gertatze-mailaren neurria ematen digun probabilitatea dagokio. X aldagaiaren heina $\{x_1,\ldots,x_n\}$ baldin bada, $f(x_1)=p(X=x_1),\ldots,$ $f(x_n)=p(X=x_n)$ adieraziko ditugu, $\{f(x_1),\ldots,f(x_n)\}$ n zenbakitako multzoari X zorizko aldagaiaren probabilitate-legea deiturik.

Probabilitate-legea ezagutuz gero, edozein zorizko aldagaitarako planteiatzen den ondoko arazo nagusia ebatzita dago: Zorizko aldagaia, bere heinaren edozein bi balioren artean egoteko probabilitatea kalkulatzea. Esate baterako, $p(x_2 \leqslant X \leqslant x_5) = f(x_2)+f(x_3)+f(x_4)+f(x_5)$.

Baina,beti, ez da erraza f(x) determinatzea. Batzutan, arazoak baimentzen duenean, konbinazio-metodoak erabiltzen dira.

3.adibidea:

Pertsona heldu normalaren mm^3 odol batetan, 7.000 leukozito zenbatu dira, beraien arteko 4.500 neutrofiloak eta 2.500 beste motatakoak izanik. 20 leukozito zoriz aukeratzen baldin badira, zein da x neutrofilo eta beste motatako 20-x izateko probabilitatea?

4.500 artean, x neutrofilo aukeratzeko era-kopurua $\binom{4.500}{x}$ da, eta 2.500 artean, 20-x aukeratzekoa, $\binom{2.500}{20-x}$.

Eskatutako gertaeraren aldeko kasu-kopurua hauxe da: 20 leukozito talde-kopurua, non x neutrofiloak eta 20-x beste motatakoak bait dira; hots $\binom{4.500}{x} \cdot \binom{2.500}{20-x}$.

Kasu posibleen kopurua, $\binom{7.000}{20}$ da; beraz, eskaturiko probabilitatea hauxe da:

$$f(x) = p(X=x) = \frac{\binom{4.500}{x} \cdot \binom{2.500}{20-x}}{\binom{7.000}{20}} \qquad x=0,1,\ldots,20$$

Fenomeno naturaletan, beren arteko batzu geroago aztertuko ditugu, gertatzen diren zorizko aldagaien legeak, oso ondo bete dituzten probabilitate-legeen zenbait eredu tekniko aztertzen ditu Estatistika matematikoak.

Nabarerazi ditzagun probabilitate-legearen propietateak:

X, $\{x_1,\ldots,x_n\}$ heineko zorizko aldagaia baldin bada, orduan:

i) $f(x_i) \geqslant 0 \qquad i=1,2,\ldots,n$

ii) $\sum_{i=1}^{n} f(x_i)=1$

III.2.2.- Banaketa-funtzioa. Propietateak.

Zorizko aldagai diskretua aztertzean, probabilitateak sartzeko era alternatibo bat, beraren banaketa-funtzioaren bidez egiten dena da.

X, $\{x_1,\ldots,x_n\}$ balioak hartzen dituen zorizko aldagaia baldin bada, X-en banaketa-funtzioa da, $\mathbb{R}$ gainean definiturik funtzioa, non $F(x)=p(X\leqslant x)$ den. Hau da, zorizko aldagaiak x-en berdin edo x baino txikiagoko balioa hartzeko probabilitatea da F(x).

f(x) ezagutzeak, F(x) ezagutzeko aukera ematen digu, eta alderantziz. Izan ere:

$$F(x) = \sum_{x_i \leqslant x} f(x_i) \; ; \; f(x_i)=F(x_i)-F(x_{i-1}) \quad i=1,2,\ldots,n$$

Erraz egiazta daitekeenez, banaketa-funtzioak ondoko propietateak betetzen ditu:

i) $F(+\infty)=1$; $F(-\infty)=0$

ii) F(x) ez-beherakorra da.

iii) Edozein puntutan, F(x) eskuinetik jarraia da.

iv) F(x)-ek, $x_1,\ldots, x_n$ puntuetan eta beraietan soilik, etenak aurkezten ditu, jauziak $f(x_1),\ldots, f(x_n)$-koak izanik, hurrenez hurren.

4.adibidea:

Demagun 1.adibidearen zorizko aldagaia. Bira:

$f(0) = p(X=0) = 0,09$

$f(1) = p(X=1) = 0,42$

$f(2) = p(X=2) = 0,49$

X-en banaketa-funtzioa kalkulatzeko, $F(x)=\sum_{x_i \leqslant x} f(x_i)$ formula aplikatuko dugu, $F(x)=0$ baldin $x<0$

$F(x)=0,09$ baldin $0 \leqslant x<1$

$F(x)=0,51$ baldin $1 \leqslant x<2$

$F(x)=1$ baldin $x \geqslant 2$ lorturik.

III.2.3.- Adierazpide grafikoak.

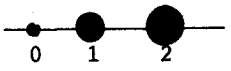
f(x), masa-banaketaren modura adieraz daiteke: Aldagai diskretuaren balioak eskala baten gainean adierazten baldin badira, balio bakoitzaren gainean bere "probabilitate-masaren" zirkulu adierazgarri bat seinalatuz, probabilitate-legearen adierazpide grafiko zuzen eta iradokorra daukagu.

Probabilitateen barra-diagrama, aldagaiaren balioak eskala horizontal baten gainean adieraztean datza, balio bakoitzean, probabilitatearen berdina den altuera daukan barra bat eraikirik. Grafikoan, probabilitateen eskala bertikala agertzea komeni da.

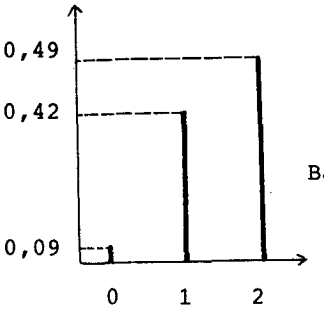
Diagrama metakorrari dagokionez, banaketa-funtzioaren adierazpidea besterik ez da, non zorizko aldagai diskretuetarako grafikoak, eskaileraren forma hartzen bait du.

5.adibidea:

X="Ale baten A alelo-kopurua" zorizko aldagaira berriro itzuliz, dagozkion grafikoak ondokoak izango dira:



Masa-banaketa



Barra- diagrama

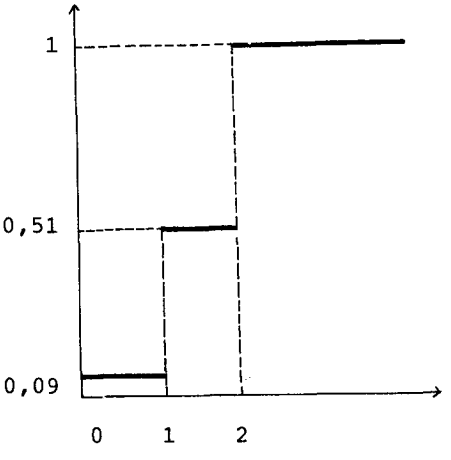


Diagrama metakorra

III.3.- ZORIZKO ALDAGAI DISKRETUAREN BATEZBESTEKOA ETA BARIANTZA.

III.3.1.- Zorizko aldagai diskretuaren batezbestekoa edo itxaropena.

Probabilitate-legeak, zorizko aldagaiaren jokabidea zeharo determinatzen du. Baina, aplikazioetan, zorizko aldagaitzat jokatzen duten neurketetarako, funtzio hau ezagutzea ezinezkoa gerta daiteke (experientziari buruzko datu edo informazioaren falta dela kausa). Orduan, zorizko aldagaia, zenbait parametroren bidez (esate baterako, batezbesteko balioa edo itxaropen matematikoa eta bariantzaren bidez) karakterizatzea baliagarria da. Batezbestekoa, aldagaiaren zentralizapen-neurria da, eta bariantza, aldiz, sakabanatzearena.

X, $\{x_1,\ldots,x_n\}$ heineko zorizko aldagaia baldin bada, X-en batezbestekoa, batezbesteko balioa edo itxaropen matematikoaren definizioa hauxe da:

$$\mu = E(X) = \sum_{i=1}^{n} x_i p(X=x_i) = \sum_{i=1}^{n} x_i f(x_i)$$

non f, X-en probabilitate-legea den.

Zorizko aldagai diskretuaren batezbestekoa, beraren n baliotako laginaren batezbestekoaren limitetzat, n-k ∞-rantz jotzen duenean, har daiteke. Izan ere, zorizko fenomenoa emanik, zeini X zorizko aldagaia elkartuko bait diogu, proba R aldiz errepikatzerakoan, aldagaiaren balio desberdinak, beren maiztasunekin batera lortzen dira:

X	x_i	x_2	...	x_n
f	f_1	f_2	...	f_n

Balio hauen batezbesteko aritmetikoa hauxe da:

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

$p(X=x_i) = \lim_{R\to\infty} \frac{f_i}{R}$, probabilitateen interpretazio frekuentziala kontutan hartuz, ondokoa lortzen da:

$$\mu = \sum_{i=1}^{n} x_i p(X=x_i) = \lim_{R\to\infty} \sum_{i=1}^{n} x_i \frac{f_i}{R}$$

X zorizko aldagaiaren itxaropen matematikoa, aldagaiaren puntuen grabitate-zentrutzat ere interpreta daiteke, bertan, bakoitzari bere probabilitatearen proportzionala den masa elkartzen zaiolarik.



E(X)

Beste esangura bat, probabilitate-teoriaren azareko jokoetango aplikaziotik dator. Azareko jokoa emanik, X zorizko aldagaiaren definizioa, emaitza edo gertaera bakoitzak halabeharr tzen duen irabazia edo galera baldin bada, E(X), X-en itxaropen matematikoa, batezbesteko irabazia jokaldikoa da. Jokuak, itxaropen nulua baldin badauka, orekatua da; bere itxaropena 0 baino handiagoa baldin bada, aldekoa, eta osterantzean, kontrakoa.

6.adibidea:

Aurreko adibidearen populazioan, A motako genen batezbesteko kopurua, horrela definitutako zorizko aldagaiaren itxaropen matematikoa da:

$$X(\omega) = \begin{cases} 0 & \text{baldin } \omega \in aa \\ 1 & \text{" } \omega \in Aa \\ 2 & \text{" } \omega \in AA \end{cases}$$

Orain $p(aa)=p(AA)=\frac{1}{4}$ eta $p(Aa)=\frac{1}{2}$ suposatuz, ondokoa kalkulatzen da:

$$E(X)=0\cdot p(X=0)+1\cdot p(X=1)+2\cdot p(X=2)=0\cdot\frac{1}{4}+1\cdot\frac{1}{2}+2\cdot\frac{1}{4}=1$$

Kasu honetan, E(X), X-ek hartzen dituen balioetariko baten berdina da, baina hau ez da egoera orokorra. Probabilitateak ondokoak izango balira:

$$p(AA)=\frac{3}{8} \qquad p(Aa)=\frac{1}{2} \qquad p(aa)=\frac{1}{8}$$

X-en itxaropen matematikoa hauxe izango litzateke:

$$E(X)=0\cdot\frac{1}{8}+1\cdot\frac{1}{2}+2\cdot\frac{3}{8}=\frac{5}{4}=1,25$$

III.3.2.- Itxaropen matematikoaren propietateak.

1) Konstante baten itxaropen matematikoa, bera da; hots, $E(c)=c$.

   Izan ere, c konstantea, c balio posible bakarra daukan zorizko aldagai diskretutzat hartzen da, beraren probabilitatea 1 izanik. Beraz, $E(c)=c.1=c$

2) Bi zorizko aldagairen arteko baturaren itxaropen matematikoa, batugaien itxaropen matematikoen arteko batura da.

   Matematikoki: $E(X+Y)=E(X)+E(Y)$

   Sinplifikatzeko, demagun X eta Y-ren probabilitate-banaketak ondokoak direla:

X	x_1	x_2
p	p_1	p_2

Y	y_1	y_2
p'	p'_1	p'_2

Orduan, X+Y-ren probabilitate-legea hauxe izango da:

X+Y	x_1+y_1	x_1+y_2	x_2+y_1	x_2+y_2
p*	p_{11}	p_{12}	p_{21}	p_{22}

eta X+Y-ren itxaropena horrela idatziko dugu:

$$E(X+Y)=(x_1+y_1)p_{11}+(x_1+y_2)p_{12}+(x_2+y_1)p_{21}+(x_2+y_2)p_{22}=$$

$$= x_1(p_{11}+p_{12})+x_2(p_{21}+p_{22})+y_1(p_{11}+p_{21})+y_2(p_{12}+p_{22})=$$

$$= x_1p_1+x_2p_2+y_1p'_1+y_2p'_2 = E(X)+E(Y)$$

$$\left.\begin{array}{l} p_{11}+p_{12}=p_1 \\ p_{21}+p_{22}=p_2 \\ p_{11}+p_{21}=p'_1 \\ p_{12}+p_{22}=p'_2 \end{array}\right\} \text{direla eta.}$$

Zorizko aldagaiek balio-kopuru handiagoa har baldin badezakete, aurreko frogapena orokor daiteke; eta, halaber, zorizko aldagai-kopuru handiagorako. Beraz,

$$E(X_1+X_2+\ldots+X_n)=E(X_1)+E(X_2)+\ldots+E(X_n)$$

berdintza ere egiazkoa da.

3) Konstante eta zorizko aldagaiaren arteko biderkaduraren itxaropen matematikoa, konstantearen eta zorizko aldagaiaren itxaropen matematikoaren arteko biderkadura da.

Hauxe da, $E(cX) = c.E(X)$

Izan ere:

$$E(cX) = \sum_{i=1}^{n} (cx_i)p_i = c \sum_{i=1}^{n} x_i p_i = c.E(X)$$

4) Itxaropen matematikoa eragile lineala da; hots,

$$E(c_1X_1+c_2X_2+\ldots+c_nX_n)=c_1E(X_1)+c_2E(X_2)+\ldots+c_nE(X_n)$$

Frogapena, aurreko propietateetatik ondorioztatzen da berehala.

5) Bi zorizko aldagai askeren (*) arteko biderkaduraren itxaropen matematikoa, berauen itxaropen matematikoen arteko biderkadura da. Hauxe da:

$E(X\cdot Y) = E(X)\cdot E(Y)$

Izan ere, demagun X eta Y-ren probabilitate-legeak hauxek direla:

X	x_1	x_2
p	p_1	p_2

Y	y_1	y_2
p'	p'_1	p'_2

X·Y aldagai berriaren probabilitate-legea ondokoa izango da:

(*) Ikus hurrengo orrialde.

X·Y	x_1y_1	x_1y_2	x_2y_1	x_2y_2
p^*	p_1p_1'	p_1p_2'	p_2p_1'	p_2p_2'

Beraz,

$$E(X\cdot Y) = x_1y_1p_1p_1' + x_1y_2p_1p_2' + x_2y_1p_2p_1' + x_2y_2p_2p_2' =$$

$$= y_1p_1' \cdot (x_1p_1 + x_2p_2) + y_2p_2' \cdot (x_1p_1 + x_2p_2) =$$

$$= (x_1p_1 + x_2p_2)(y_1p_1' + y_2p_2') = E(X)\cdot E(Y)$$

Zorizko aldagai askeen edozein kopuru finitutarako, hauxe izango da:

$$E(X_1X_2\ldots X_n) = E(X_1)\cdot E(X_2)\cdot \ldots \cdot E(X_n)$$

III.3.3.- Zorizko aldagai diskretuaren bariantza eta desbidazio standarda.

Zorizko aldagaiaren itxaropen matematiko edo batezbestekoak, aldagaia fluktuatzen deneko inguruaren balioa adierazten du, baina ez du fluktuazio honen tamainuari buruzko informaziorik ematen.

(*) $\{x_1,\ldots,x_n\}$, $\{y_1,\ldots,y_m\}$ heinetako, hurrenez hurren, X eta Y bi aldagai askeak dira, baldin

$$p(X=x_i\ ,\ Y=y_j) = p(X=x_i)p(Y=y_j) \quad i=1,\ldots,n \quad j=1,\ldots,m$$

badira.

Kontzeptu honen interpretazioa, gertaera askez hitz egiterakoan egin zenaren guztiz antzekoa da. (Propietate hau, zorizko aldagai askeetarako soilik dela baliozkoa kontutan hartu behar da).

Probabilitate-masa, zorizko aldagaiaren inguruan, nola banatzen den adieraziko lukeen balioa hauxe izango litzateke:

$$E(|X-\mu|) \quad \text{non} \quad \mu = E(X)$$

Hala ere, kalkuluaren erraztasuna dela kausa, gehiago erabiltzen da ondoko adierazpena:

$$\sigma_x^2 = \text{Var}(X) = E\{(X-\mu)^2\}$$

X-en bariantza deitua izanik.

X-ek, $\{x_1,\ldots,x_n\}$ balioak hartzen baditu, itxaropen matematikoaren definiziotik, bariantzaren adierazpen hau ondorioztatzen da:

$$\text{Var}(X) = \sum_{i=1}^{n}(x_i-\mu)^2 p_i \quad \text{non} \quad p_i = p(X=x_i) \quad \text{den.}$$

Bariantzaren erro karratu positiboari, desbidazio standarda deritzogu; hots,

$$\sigma_x = \sqrt{\text{Var}(X)} = \sqrt{E\{(X-\mu)^2\}}$$

7.adibidea:

Aurreko adibideko zorizko aldagaiaren bariantza hauxe da:

$$\text{Var}(X) = \sum_{i=1}^{2}(x_i-\mu)^2 p_i = (0-1)^2\frac{1}{4} + (1-1)^2\frac{1}{2} + (2-1)^2\frac{1}{4} =$$

$$= \frac{1}{2}$$

Beraz, $\sigma_X = \sqrt{Var(X)} = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}}$

π-ren aleak, batez beste, 1-en berdina den genen batezbesteko balioarekiko, $\frac{1}{\sqrt{2}}$ genetan diferitzen direla esan daiteke.

III.3.4.- Bariantzaren propietateak.

1) Bariantza, beti positiboa edo nulua da, zeren $(X - \mu)^2 \geqslant 0$ bait da.

2) Bariantza nulua da, baldin eta soilik baldin, zorizko aldagaia konstantea bada.

Izan ere:

Baldin X=c bada, $Var(X)=Var(c)=E(\{c-E(c)\}^2) =$

$= E\{(c-c)^2\} = E(0) = 0.$

Eta alderantziz, baldin $Var(X)=E\{(x-\mu)^2\}=0$ bada, orduan, $X=\mu$

Hots, "sakabanatze" nulua daukaten zorizko aldagai bakarrak, konstanteak dira.

3) $Var(X) = E(X^2) - \{E(X)\}^2 = E(X^2) - \mu^2$

Izan ere:

$Var(X)=E\{(X-\mu)^2\}=E(X^2-2\cdot\mu\cdot X+\mu^2)=$

$E(X^2)-2\mu E(X)+E(\mu^2)=E(X^2)-2\mu\cdot\mu+\mu^2=E(X^2)-\mu^2$

non itxaropen eragilearen linealitatea erabili bait da.

4) X eta Y askeak baldin badira, $Var(X+Y)=Var(X)+Var(Y)$.

Izan ere:

(Gogora dezagun, X eta Y askeak baldin badira, orduan, $E(X\cdot Y)=E(X)E(Y)$ dela)

$$Var(X+Y)=E\{(X+Y)^2\}-\{E(X+Y)\}^2=E(X^2+2XY+Y^2)-$$
$$-\{E(X)+E(Y)\}^2=E(X^2)+2E(XY)+E(Y^2)-$$
$$-\{E(X)\}^2-2E(X)E(Y)-\{E(Y)\}^2=Var(X)+Var(Y)$$

Propietate hau, zorizko aldagai askeen edozein kopuru finitutarako orokortzen da.

5) $Var(cX)=c^2Var(X)$

Ondoko azalpenaz froga dezakegu:

$$Var(cX)=E\{(cX)^2\}-\{E(cX)\}^2=E(c^2X^2)-\{cE(X)\}^2=$$
$$=c^2E(X^2)-c^2\{E(X)\}^2=c^2[E(X^2)-\{E(X)\}^2]=$$
$$=c^2Var(X)$$

6) $Var(X+c)=Var(X)$ (Zorizko aldagaiaren bariantza, translazioekiko "inbariantea" da).

Izan ere:

$$Var(X+c)=E[\{X+c-E(X+c)\}^2]=E[\{X+c-E(X)-c\}^2]=$$
$$=Var(X)$$

III.3.5.- Sakabanatzea eta Tchebycheff-en teorema.

Bariantza, batezbestekoarekiko sakabanatzearen neurria deneko zergatiaren justifikapen ona, hurrengo teoremak ematen digu. Oraingoz, ez dugu beraren frogapena ikusiko.

Tchebycheff-en teorema:

Biz X, itxaropen eta bariantza finituko zorizko aldagaia. Orduan, ondokoa egiaztatzen da:

$$p(|X - E(X)| > h) \leqslant \frac{Var(X)}{h^2} \qquad \forall h > 0$$

Partikularrean, $h=k\sigma$ hartuz, non σ X-en desbidazio standarda den, hauxe daukagu:

$$p(|X - E(X)| > k\sigma) \leqslant \frac{1}{k^2}$$

Esate baterako, k=4 denean, ondoko ondorioa ateratzen dugu:

$$p(|X - E(X)| > 4\sigma) \leqslant 0,0625$$

Hots, X-en balio bat zoriz hartzerakoan, berau eta $E(X)=\mu$ -ren arteko distantzia 4σ baino luzeagoa izateko probabilitatea 0,0625 baino txikiago edo berdina da.

8.adibidea:

X, zorizko aldagai diskretua bada, non $E(X)=0$ eta $\sigma(X)=1$ diren, Tchebycheff-en teoremak hauxe azaltzen digu:

$$p(|X| > 4) \leqslant 0,0625 \Longleftrightarrow p(-4 \leqslant X \leqslant 4) \geqslant 0,9375$$

Emaitza honek, ondoko interpretazioa onartzen du: Nahiz eta X zorizko aldagaiak balio-kopuru handia hartu ahal izan, proba-kopuru handia egiten badugu, bakoitzean X-en balioa behaturik, haien arteko 90%-a baino gehiago -4 eta 4-ren artean dagoela ikusiko dugu.

III.4.- BANAKETA BINOMIALA.

Probabilitate-kalkuluaren helburuetariko bat, zientzi experimentaletan -beraien artean, Biologian- agertzen diren zorizko fenomeno ugariren eredutzat har daitezkeen banaketak aurkitzea da. Atal honetan, eredu binomiala azaltzen hasiko gara.

Biz ξ zorizko ezperimentua. Berarekin erlazionatuta dauden A gertaera (arrakasta) eta $\bar{A}$ kontrakoa (porrota) kontsideratzen ditugu. Suposa dezagun $p(A)=p$ eta $p(\bar{A})=1-p=q$ direla.

Demagun, askeak eta baldintza berberen pean egiten diren ondoz-ondoko hiru proba. Emaitza posibleak hauxek dira:

$$AAA \ ; \ \bar{A}AA \ ; \ A\bar{A}A \ ; \ AA\bar{A} \ ; \ \bar{A}\bar{A}A \ ; \ \bar{A}A\bar{A} \ ; \ A\bar{A}\bar{A} \ ; \ \bar{A}\bar{A}\bar{A}$$

Gertaera-multzo honek, hiru proba errepikatuek egindako experimentu konposatuari dagokion lagin-espazioa osotzen du. Independentzia kontutan hartuz, gertaera hauen probabilitateak ondokoak izango dira, hurrenez hurren:

$$ppp \ ; \ qpp \ ; \ pqp \ ; \ ppq \ ; \ qqp \ ; \ qpq \ ; \ pqq \ ; \ qqq$$

Lagin-espazioaren gainean, zorizko aldagai binomiala honela definitzen dugu:

X = "arrakasta-kopurua, proba konposatu batetan"

Lagin-espazioaren puntuetan, ondoko balioak hartzen ditu aldagai honek: 3, 2, 2, 2, 1, 1, 1, 0.

Batukortasuna kontutan harturik, balio hauei ondoko probabiliteak dagozkie:

$$p(X=3)=p^3 \ ; \ p(X=2)=3p^2q \ ; \ p(X=1)=3pq^2 \ ; \ p(X=0)=q^3$$

Honela, n=3 probatarako banaketa binomiala, zeharo definituta dago.

Kasu orokorrean, n proba egiten ditugu, ondoko baldintzen pean:

1) Proba bakoitzean, elkar-ukatzaileak diren emaitza posibleetariko bat lortzen da. Emaitza posibleen bati, arrakasta, eta besteari, porrota deritzogu.

2) Arrakastaren probabilitateak, 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 $\bar{A}$ balioak, era posible guztietara egokituz, lortzen dira lagin-espazioaren elementuak. A eta $\bar{A}$ elementuen n ordeneko errepikatuzko aldakuntzak diren 2^n gertaera izango dira.

Lagin-espazio honen gainean, X="Kontutan harturiko gertaeraren arrakasta-kopurua" zorizko aldagai binomiala definitzen dugu.

r arrakasta eta n-r porroteko gertaeraren probabilitatea, $p^r q^{n-r}$ da.

Baina, $A\overbrace{\ldots A}^{r}\bar{A}\overbrace{\ldots\bar{A}}^{n-r}$ n letren ordenamendu posibleak, A eta $\bar{A}$ bi letren n ordeneko errepikatuzko permutazioak dira, non A letra-kopurua r eta $\bar{A}$ letrarena n-r diren; hots,

$$\frac{n!}{r!(n-r)!}$$

Beraz, bilatu nahi dugun probabilitatea, hauxe izango da:

$$p(X=r) = \frac{n!}{r!(n-r)!}\, p^r q^{n-r} \qquad r=0,1,\ldots,n$$

Hau da "banaketa binomial" izenekoa, beraren izena, probabilitate hauek ondoko garapen binomikoaren gaiak izatetik datorrelarik:

$$1 = 1^n = (p+q)^n = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} = \sum_{r=0}^{n} p(X=r)$$

Banaketa-funtzioa ondokoa izango da:

$$F(x) = p(X \leqslant x) = \sum_{r \leqslant x} \binom{n}{r} p^r q^{n-r}$$

Aldagai binomialaren batezbestekoa eta bariantza, hurrenez hurren, hauxek direla erraz froga daiteke:

$$\mu = E(X) = np \qquad \sigma_x^2 = Var(X) = npq$$

Dakusagunez, banaketa binomiala, n proba-kopuruak eta proba bakoitzeko p arrakasta-probabilitateak karakterizaturik dago. Kantitate hauek, banaketa binomialaren <u>parametroak</u> dira, n eta p parametroetako X zorizko aldagai binomiala B(n,p) adierazirik.

<u>9.adibidea</u>:

Zoldura-prozesu baten azterketa, banaketa binomialari jarraitzen zaion zorizko aldagaiaren bidez.

Intsektu-populazio bat aztertzen dugu. Beronen aleen arteko 40%-ak zoldura birikoa aurkezten duela ba dakigu, eta dagokion prozesu-mota kutsakorra ezagutu nahi dugu. Horretarako, populazioaren 5 aletako (intsektutako) 2423 lagin hartzen ditugu, beraietako ale zolduen kopurua behaturik eta 0, 1, 2, 3, 4 eta 5 zoldu, hurrenez hurren, aurkezten dituzten laginen maiztasunak edo proportzio experimentalak lorturik.

Intsektu baten zoldura, laginaren gaineratiko osagaiena naren askea dela suposaturik, X="Populazioaren 5 intsektutako laginaren ale zolduen kopurua" zorizko aldagaiak betetzen duen banaketa teorikoa, n=5 eta p=0,4 parametroetako binomiala da.

X-en ibilbidea {0,1,2,3,4,5} da, beraren dentsitate-funtzioa ondokoa izanik:

$$f(0) = \binom{5}{0} \cdot 0,4^0 \cdot 0,6^5 = 0,07776$$
$$f(1) = \binom{5}{1} \cdot 0,4^1 \cdot 0,6^4 = 0,25920$$
$$f(2) = \binom{5}{2} \cdot 0,4^2 \cdot 0,6^3 = 0,34560$$
$$f(3) = \binom{5}{3} \cdot 0,4^3 \cdot 0,6^2 = 0,23040$$
$$f(4) = \binom{5}{4} \cdot 0,4^4 \cdot 0,6^1 = 0,07680$$
$$f(5) = \binom{5}{5} \cdot 0,4^5 \cdot 0,6^0 = 0,01024$$

Balio hauek, harturiko laginak, 0,1,2,3,4,5 ale zoldu aurkezteko probabilitateei (teorikoei) dagozkie, hurrenez hurren.

Praktikan, emaitza teoriko hauek eta aurkituriko proportzio experimentalak konparatzea interesatzen zaigu; horrela, ea experientziak betetzen dituen suposaturiko baldintzak ontzat eman daitezkeen kontrastatzeko.

Horretarako, 5 intsektutako 2423 laginak aztertuz gero, hurrengo taula eraikitzen dugu:

Intsektu zolduen kopurua lagineko (x_i)	Probabilitateak edo itxaroten ditugun maiztasun erlatiboak	Itxaroten ditugun maiztasun absolutuak ($2423p_i$)	Behaturiko maiztasun absolutuak (f_i)
0	0,07776	184,4	202
1	0,25920	628,0	643
2	0,34560	837,0	817
3	0,23040	558,3	535
4	0,07680	186,1	197
5	0,01024	24,8	29
	1,00000	2423,0	2423

$$E(X) = np = 5 \cdot 0,4 = 2 \qquad \bar{x} = 1,98721$$

$$\sigma_x = \sqrt{npq} = \sqrt{5 \cdot 0,4 \cdot 0,6} = 1,09545 \qquad s_x = 1,11934$$

Azterturiko zorizko aldagaiak banaketa binomiala betetzen badu, itxaroten ditugun maiztasunak eta lehendabiziko azterketa honetan lorturiko maiztasun experimentalak oso antzekoak direla ikus daiteke. Ea benetan, maiztasun hauek nabariki desberdinak direnentz erabakitzen baimentzen diguten metodoak, Inferentzia Estatistikoan azaltzen dira.

III.5.- BANAKETA HIPERGEOMETRIKOA.

Demagun N tamainuko populazio finitua dikotomikoa dela; hots, beraren elementuak bi klase baztertzaile eta exhaustiboren barnekoak direla (esate baterako, sexuaren arauera sailkaturiko gizakien populazioa dikotomikoa da).

Klaseetariko bati "klase nagusia" deritzogu, A adierazirik; besteari, "kontrako klasea" deituko diogu, $\bar{A}$ adierazten delarik.

Suposa dezagun, populaziotik, n elementu aldiberean edo elkarren segidan itzulerarik gabe ateratzen direla, A klase nagusiaren x elementu-kopurua zenbatzen delarik. Elementu-kopuru hau zorizko aldagaia da, hipergeometrikoa deitua izanik.

Populazioaren elementu nagusien proportzioa, $p=\frac{M}{N}$ baldin bada, non M="populazioaren elementu nagusien kopurua" den, orduan, $1-p = \frac{N-M}{N}$ da kontrako elementuen proportzioa.

n tamainuko laginean, x elementu nagusi eta n-x kontrako elementu aurkitzeko probabilitatea lortu nahi dugu.

M-en arteko x elementu nagusi lortzeko era-kopurua, $\binom{M}{x}=\binom{Np}{x}$ da.

(N-M)-ren arteko n-x kontrako elementu lortzeko era desberdinen kopurua, hauxe izango da: $\binom{N-M}{n-x} = \binom{N(1-p)}{n-x}$

x elementu nagusitako talde bakoitza eta n-x kontrako elementutako talde bakoitza parekatuz, x nagusi eta n-x kontrako elementutako laginaren aldeko kasu-kopurua lortuko dugu, hain zuzen ere: $\binom{M}{x}\cdot\binom{N-M}{n-x}$

Kasu posibleak, N tamainuko populaziotik atera daitezkeen n tamainuko lagin desberdin guztiak dira; hots, $\binom{N}{n}$.

Beraz, "n tamainuko laginaren elementu nagusien kopurua" X deituriko zorizko aldagaiak, ondoko probabilitate-legea dauka:

$$p(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{N.p}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} \qquad 0\leqslant x\leqslant \min(n,M)$$

X zorizko aldagai hipergeometrikoaren batezbestekoa eta bariantza hauxek direla frogatzen da:

$$\mu = E(X) = np = n\frac{M}{N} \quad ; \quad \sigma_x^2 = \frac{N-n}{N-1}\cdot n.p.(1-p)$$

10.adibidea:

Animali populazio baten tamainuaren estimazioa, berrarrapaketaren datuen bidez.

Demagun, laku batetan harrapaturiko 1000 arrain, gorriz seinalatzen direla, eta, ondoren, askatzen direla. Handik denbora apur bat, berriro 1000 arrain harrapatzen dira eta beraien arteko 100-ek makula gorriak dauzkatela behatzen da.

Lakuko arrain-kopuruari buruz, zer ondorioztatuko dugu? Hau da estimazio estatistikoaren arazo karakteristikoa. Estatistikaren metodo berriak deskribatu nahi izango bagenitu, gure helburutik urrunduko ginateke, eta, horregatik, banaketa hipergeometrikoak ematen digun arazoaren soluzioa aztertuko dugu soilik.

Bi harrapaketak, lakuko arrain guztiek osotzen duten populazioaren zorizko laginak direla suposatuko dugu. (Praktikan, bi harrapaketak leku berberean eta oso denbora-tarte laburrean egiten diren egoerak baztertzen ditu suposamendu honek).

Halaber, lehendabiziko harrapaketatik bigarrenera doan aldian, arrain-kopurua ez dela aldatzen suposatuko dugu.

Hautazko lagin-tamainuak onartuz, orokortzen dugu arazoa. Beraz, bira:

N = lakuko arrain-kopurua (ezezaguna).

M = lehendabiziko harrapaketako arrain-kopurua, zeinak "elementu nagusien" papera betetzen bait du.

n = bigarren harrapaketako arrain-kopurua.

x = bigarren harrapaketako arrain gorrien kopurua.

$q_x(N)$ = bigarren harrapaketak x arrain gorri edukitzeko probabilitatea.

Formulazio honetan, $q_x(N)$, ondoko formularen bidez lortzen dela bistakoa da:

$$q_x(N) = \frac{\binom{N}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad (1)$$

Praktikan, M, n eta x-en balioak ezagutzen dira; N-arena, ordea, ezezaguna da. Kontutan har ezazu, N, azarearen pean ez dagoen zenbaki fixua dela. Beraz, adibidez, N, 6000 baino handiagoa izateko probabilitateaz galdetzea ez dauka zentzurik. M+n-x arrain desberdin harrapatu direla badakigu, eta, beraz, $N \geqslant M+n-x$. Hau da segurtasunez baiezta dezakegun gauza bakarra.

Gure adibidean, M=n=1000 eta x=100 geneuzkan. Lakuak 1900 arrain soilik dauzkala pentsa daiteke, baina hipotesi honetatik abiatzen bagara, oso probabilitate txikia daukan gertaera jazo dela ondorioztatuko dugu. Izan ere, denetara N=1900 arrain dagoela suposatzen baldin badugu, jazotako gertaeraren probabilitatea, (1)-en arauera, hauxe izango da:

$$\frac{\binom{1000}{100}\binom{900}{900}}{\binom{1900}{1000}} = \frac{(1000\,!)^2}{100\,!\;1900\,!}$$

Stirling-en formulak(*), probabilitate hau 10^{-430} ordenakoa dela erakusten digu, eta, egoera honetan, zentzuak hipotesia errefusatzen du, arrazonagarria ez dela eta.

Antzeko arrazonamenduak, n oso handia (esate baterako, miloi bat) deneko hipotesia errefusatzera eramango gintuzke.

Erreflexio hauek N-ren balio partikularra, non $q_x(N)$ delakoak balio maximoa hartzen bait du, kontutan hartzera garamatza, zeren eta, N honetaz, gure behaketak probabilitaterik handiena izango bait du.

Horretarako, kontsidera dezagun ondoko arrazoia:

$$\frac{q_x(N)}{q_x(N-1)} = \frac{(N-M)\,(N-n)}{N\,(N-M-n+x)}$$

(*) $n\,! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$ non $\sim$ ikurrak, bi gaien arteko zatidurak, n-k ∞-ra jotzen duenean, 1-era jotzen duela adierazten duen.

Arrazoi hau 1 baino handiago ala txikiagoa, $Nx < Mn$ ala $Nx > Mn$ denaren arauera izango dela erakusten digu kalkulu errazak. Honek esan nahi du, N emendatzean, $q_x(N)$ segida, lehenik handitzen dela eta gero gutxitzen dela, N, $\frac{Mn}{x}$ baino txikiagoa den zenbaki osoaren hurbilena denean, M_α balio maximoa lorturik. Gure adibide honetan, N=10000.

N-ren benetako balioa, batzutan handia izango da eta beste batzutan, txikia; haien artean N dagoela espero izatea arrazonagarria deneko mugak hauteman ahal izanik. Baina, oraindik, ezin diogu arazo honi heldu.

III.6.- BANAKETA MULTINOMIALA.

Banaketa binomialaren orokorpena, banaketa multinomiala da. Binomialean, populazio dikotomiko eta infinituaren laginketatik abiatzen gara, k zelula baztertzaile eta exhaustiboetan zatituriko populazio infinituaren laginketa aztertzea naturala dirudielarik.

Bektore hitza, zenbaki-multzo ordenatutzat ulertuko dugu. Adibidez. (2, 5, 7) hiru osagaitako bektorea da eta (9, 4, 6, 8, 3), bost osagaitakoa.

Beraren osagaiak zorizko aldagaiak direnean, zorizkoa da bektorea. Beraz, zorizko bektorea, zorizko aldagai-multzo ordenatua izango da.

Suposa dezagun, A_1, A_2, ..., A_k zeluletan, $A_i \cap A_j = \emptyset \ \forall i \neq j$ eta $A_1 \cup A_2 \cup \ldots \cup A_k = E$ izanik, zatituriko populaziotik n ateralditako segida egiten dela eta laginean dauden zelula bakoitzeko elementuak zenbatzen direla. x_1 A_1-ekoak, x_2 A_2-koak, ..., x_k A_k-koak baldin badira, $(x_1, x_2, \ldots, x_k)$ multzo ordenatua k osagaitako bektorea da, zorizko bektore multinomiala deitua izanik.

Deskribaturiko populazioari dagokionez, n ateraldi as ketako segida, honela eskematiza daiteke:

$$A_2A_5A_1A_k\overset{n}{\ldots}A_3A_2A_1 \qquad (2)$$

Suposa dezagun, aurreko segidaren elementuen artean, A_1-eko x_1, A_2-ko x_2, ..., A_k-ko x_k daudela, $x_1+x_2+\ldots x_k=n$ izanik eta zelula desberdinen probabilitateak $p(A_1)=p_1$, $p(A_2)=p_2,\ldots,p(A_k)=p_k$ direla, $p_1+p_2+\ldots+p_k=1$ izanik.

Gertaera askeen arteko ebakiduraren probabilitatea kalkulatzeko legea aplikatuz, (2) segida gertatzeko probabilitatea lortzen da. Hauxe, hain zuzen ere:

$$p_2p_5p_1p_k\overset{n}{\ldots}p_3p_2p_1 = (p_1)^{x_1}(p_2)^{x_2}\ldots(p_k)^{x_k}$$

Baina, (2) segida ez da A_1-eko x_1, A_2-ko x_2, ..., A_k-ko x_k dauzkan segida bakarra. (2)-ko elementuak, era posible guztietara permutatzen baldin badira, A_1-eko x_1, A_2-ko x_2, ..., A_k-ko x_k dauzkaten n ateralditako segida guztiak lortzen dira. Baina, (2)-ko permutazioen kopurua, n elementutakoa da, non balio bakoitza x_1, x_2, ..., x_k aldiz errepikatzen den; hots,

$$\frac{n!}{x_1!x_2!\ldots x_k!}$$

Beraz, (2)-ren ordenamendu guztien (baztertzaileak direnak) probabilitatea hauxe izango da:

$$p(X_1=x_1,X_2=x_2,\ldots,X_k=x_k) = \frac{n!}{x_1!x_2!\ldots x_k!}(p_1)^{x_1}(p_2)^{x_2}\ldots(p_k)^{x_k} \qquad (3)$$

non $x_1+x_2+\ldots+x_k=n$ eta $p_1+p_2+\ldots+p_k=1$ diren.

(3) adierazpena, zorizko bektore multinomialaren probabilitate-funtzioari dagokio. Binomiala, kasu partikular bezala, k=2 denean, lortzen da.

Azkenik, kontutan har dezagun (3)-ren gaiak, $(p_1+p_2+\ldots+p_k)^n$ garapenaren koefizienteak direla.

11.adibidea:

Sistema baten energi egoeraren karakterizazio estatistikoa.

Sistemaren partikula baten egoera, n dimentsiotako espazio euklidearraren puntu baten bidez adieraz daiteke (posizioko, spineko,..., eta abarreko koordenatuak emanez). Egoera hauetariko bakoitzari, E energi balioa dagokio, non E, partikula batek aurkez ditzakeen $E_1,\ldots,E_s$ energi balioen multzo finituaren elementua den. Partikularen egoera posibleetako multzoaren gaineko $E_1,\ldots,E_s$ partiketa definitzen du honek eta $\omega \in \Omega$ egoera E_i-ren barnekoa da, E_i energia erantsia badarama.

Bira $p_1,\ldots,p_s$ energi balio hauei atxekitako probabilitateak.

N partikulak eratutako sistemaren energi egoera, energi balio bakoitzak aurkezten duen partikula-kopuruaren bidez karakterizatzea logikoa da. Hau, $(X_1,\ldots,X_s)$ zorizko aldagai bektorialaren bidez adieraz daiteke, non X_i="E_i energia daukan N_i partikula-kopurua" i=1,...,s diren.

Edozein partikularen egoera, gaineratikoenaren independentea baldin bada, $(X_1,\ldots,X_s)$ zorizko bektore honek banaketa multinomiala beteko du, eta, beraz:

$$p(X_1=N_1,\ldots,X_s=N_s) = \frac{N!}{N_1!\ldots N_s!} (p_1)^{N_1}\ldots(p_s)^{N_s}$$

non $N=N_1+\ldots+N_s$ den.

III.7.- POISSON-EN BANAKETA

Banaketa diskretu guztien artean, interesa daukaten egoera biologiko (eta ez biologikoàk soilik) askori deskribapen egokia eskaintzen diona, Poisson-ena da. mm^3 odoleko agente patogenoen kopurua, substantzia irradiaktiboak igorritako partikula-kopurua denboraldi batetan, ur-bolumenean dagoen bakterio-kopurua, eta abar, Poisson-en eredua gutxi gora behera segitzen duten zorizko aldagai diskretuak dira.

Poisson-en banaketa bereziki sartzeko era argigarria ondokoa da:

Suposa dezagun, Geiger kontagailu baten instalazioaren bidez, substantzia irradiaktiboak, 5 minututan igorritako α partikula-kopurua aztertu nahi dugula. Denbora-tartea (5 minutu), n oso une-kopuru handitan, esate baterako, milisegundotan (*) zatitzen baldin badugu, partikularik igortzen ez den une asko eta igortzen den oso gutxi egongo dira.

Partikula bat, milisegundo batetan igortzeko probabilitatea "p" deitua izanik, eta igorpenak independenteak direla suposatuz, ondoko adierazpenera garamatza eredu binomialak:

$$p(X=x) = \binom{n}{x} p^x q^{n-x} \qquad x=0,1,\ldots,n \qquad (4)$$

Hala ere, probabilitate hauen kalkuluak, ondoko arazoak aurkitzen dìtu: Lehenik, p(X=x) ebaluatzea, n handia eta p oso txikia direnean; eta bigarrena, lehena baino garrantzitsuagoa, p-ren determinazioa.

Kontutan har dezagun, kasu honetan, p oso kantitate eskuraezina izan arren, E(X)=np="5 minututan igorritako batezbesteko partikula-kopurua"-rekin ez dela gauza bera gertatzen.

(*) Hain partiketa fina egitearen helburua, 5 minututako tarteko zatiketa bakoitzean, igorpen bat gehienez sortzen dela seguratzea da.

Beraz, (4)-aren muga-balioak, n-k ∞-ra jotzen duenean, aztertu behar ditugu.

Proposamendua.-

x zenbaki oso ez-negatibo bakoitzerako, $\binom{n}{x}p^x q^{n-x}$ kantitateek $\frac{\lambda^x e^{-\lambda}}{x!}$ -era jotzen dute, baldin $n\to\infty$, $p\to 0$ eta $np\to\lambda$ badira.

Izan ere:

$$\binom{n}{x} p^x q^{n-x} = \frac{n(n-1)\dots(n-x+1)}{x!} p^x (1-p)^{n-x}$$

Eta limiteak hartuz:

$$\lim_{n\to\infty} \frac{n(n-1)\dots(n-x+1)}{x!} p^x (1-p)^{n-x} = \frac{1}{x!} \lim_{n\to\infty} n(n-1)\dots(n-x+1)\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-x+1}{n}\cdot\frac{(1-\lambda/n)^n}{(1-\lambda/n)^x} = \frac{\lambda^x}{x!} 1\cdot 1\dots 1\cdot\frac{e^{-\lambda}}{1} = \frac{\lambda^x e^{-\lambda}}{x!}$$

$\left\{\frac{\lambda^x e^{-\lambda}}{x!}\right\}$ zenbakien segidak probabilitate-banaketa osotzen duela froga daiteke, zeren eta ez-negatiboak bait dira eta beraien batura 1 bait da:

$$\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

X zorizko aldagaia, non bere heina zenbaki oso ez-negatiboak diren eta bere probabilitate-legea

$$p(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad x=0,1,\dots$$

den, λ parametroko Poisson-en banaketari jarraitzen zaio, $X = P(\lambda)$ adierazirik.

Poisson-en banaketaren batezbestekoa eta bariantza, λ parametroaren berdinak direla erraz froga daiteke.

Gertaerak oso probabilitate txikia daukanean, eta gertaera, behaketa-kopuru handiko k aldiz gertatzeko p_k probabilitateak lortu nahi dugunean, Poisson-en banaketa aplikatzen da maiz. Horregatik, batzutan, gertaera arraroen legea deitzen zaio. $p<0,1$ eta $np<5$ baldin badira, ontzat hartzen da hurbilketa.

λ parametroak, zeina batezbesteko eta bariantzaren berdina bait da, Poisson-en banaketa zeharo determinatzen duela ikusi dugu. Praktikan, λ parametroaren balioa ondoko erez lor daiteke:

1) n eta p ezagunetatik abiatuz, $\lambda = np$.
2) Zorizko aldagaiaren itxaropenetik abiatuz, $\lambda = E(X)$.
3) Experientzia batetan, aldagaiak harturiko balioen laginaren batezbestekotik abiatuz, ondoko estimazioaren bidez:

$$\lambda \simeq \frac{1}{s} \sum_{i=1}^{s} x_i$$

4) X=0 gertaeraren probabilitatetik abiatuz, zeren $p(X=0)=e^{-\lambda}.\frac{\lambda^0}{0!} = e^{-\lambda}$. Beraz, $\lambda = -\ln\{p(X=0)\}$

12.adibidea:

Kretazikoan, hamar narrazti-motaren suntsidura-prozesua azaltzen duten hipotesiei buruzko kontsiderazioak.

Aro kretazikoaren amaieran, existitzen ziren hamar narrazti-motetariko bost desagertu ziren: dinosauro lurtarren bi mota, pterosauroen bat (hegalariak) eta ur-narraztien bi mota, iktiosauro eta pleisiosauro izenez ezaguturik. Bost mota

hauek, 35 familia zeuzkaten, berauetariko bakoitza espezie - kopuru handiz osotuta zegoelarik.

Kretazikoa bukatzerakoan, kontutan harturiko narrazti-mota baten batezbesteko bizitza-aldia 100 miloi urtekoa zela estimatzen baldin bada, zein den gertaera honen probabilitatea, suntsiduran, aparteko giro-faktorerik parte hartu ez balu eta mota bakoitzaren desagerpena gaineratikoaren independentea izan balitz, kalkulatzea interesgarria zitekeen.

Kasu honetan, hots, iraupen honetako aldietan, suntsidurak "zoriz" agertu balira, X="$5\cdot10^6$ urteko aldiko suntsidura-kopurua" zorizko aldagaia, Poisson-en banaketari jarraitu litzaiokeen, non λ parametroak, X-en itxaropena adierazten bait du, hots, 5 miloi urteko aldiko batezbesteko suntsidura-kopurua.

Izan ere, mota baten "batezbesteko bizitza" 100 miloi urtekoa dela estimatzen dela eta, suntsidura bat, batez beste, 100 miloi urteko agertuko da. Honek, 5 miloi urteko λ batezbesteko suntsidura-kopurua hauxe dela adierazi nahi du:

$$\lambda = 5\cdot10^6 \frac{1}{100\cdot10^6} = \frac{5}{100}$$

Beraz, 5 miloi urteko aldian, 5 suntsidura agertzeko probabilitatea ondokoa da:

$$e^{-\frac{5}{100}} \frac{\left(\frac{5}{100}\right)^5}{5!} = \frac{5^5}{e^{\frac{5}{100}}\cdot120\cdot100^5} = 2{,}477\cdot10^{-9}$$

Ikusten denez, oso txikia den probabilitate honek, desoreka ekologikoekin loturiko inguruneko baldintzen existentzia iradokitzen digu. Guzti honek, seguraski, espezie-kopuru handiaren suntsidura sorterazi zuen, berauekin erlazionaturiko beste batzurena ere ekarririk. Aipaturikoak bide zitezkeen narrazti handien erabateko desagerpenaren behin betiko zergatia.

PROBLEMAK

1.- Demagun, ume-kopuru berbereko familien sexuaren arauerako banaketan, semea edukitzeko probabilitatea 0,6 dela. Halaber, ume baten sexua, gaineratikoen sexu-banaketaren independentea dela suposaturik, zoriz aukeraturiko 8 umetako familia batetan, aurki itzazu:

i) Guztiok, sexu berberekoak izateko probabilitatea.

ii) Lehenengo hirurak, sexu batekoak, eta gaineratikoa, bestekoa izateko probabilitatea.

iii) Bost seme eta hiru alaba izateko probabilitatea.

iv) Gutxienez, bost seme izatekoa.

v) Lehenengo biak, semeak; eta, gutxienez, bost seme izatekoa.

2.- Herri batetan, hiru arraza-mota daude. 50%-a "A" arrazakoa da; 30%-a, "B" arrazakoa; eta, 20%-a, "C"-koa.

10 pertsona zoriz aukeraturik, zein da:

i) "A" arrazako hiru izateko probabilitatea?

ii) "A" arrazako bost, "B" arrazako hiru eta "C" arrazako bi izatekoa?

iii) Arraza batetako hiru, beste batetako hiru eta beste batetako lau izatekoa?

iv) "C" arrazakorik ez izatekoa?

3.- Egindako azterketak direla medio, intsektu-mota baten 80%-ak, gutxi gora behera, "G" karaktere genetikoa aurkezten duela ba dakigu. 160 intsektutako multzotik, 4 aletako talde bat

aukeratzen da zoriz.

i) Zein da, hauetariko hiruk "G" karakterea aurkezteko probabilitatea?
Eta, "G" karaktererik gabeko bat aurkitzekoa?
Gertaera hauek askeak al dira?

ii) Zein da, gutxienez, "G" karakteredun bat egoteko probabilitatea?
Hiru gertaerak askeak al dira?
Eta, lehendabizikoa eta hirugarrena?

4.- Demagun, hilabete batetan, egoera txarrean emandako gripearen aurkako txerto-kopurua, $P(3)$ banaketari jarraitzen zaiola.

i) Hilabetean, zein da egoera txarrean emandako 3 edo 3 baino txerto gehiago izateko probabilitatea?

ii) 4 hilabetetako aldia aztertzen badugu, zein da, hilabete batetan gutxienez, egoera txarrean dauden 3 edo 3 baino txerto gehiago emateko probabilitatea?

5.- Arrainontziak 3 arrain exotiko potolo eta 7 deselikatu dauzka. Felix katuak, goseti, arrainontzitik 3 arrain ateratzen ditu batera. Hirurak potoloak dira eta Felix katua jateko prest dago. Une horretan, bere jabea agertzen da eta honela esaten dio: "Felix, egin duzuna, saiaturiko 3-tatik birritan errepikatzeko gai bazara, arrainak jatea utziko dizut. Lortzen ez baduzu, zure betiko barraskilo zatituak jango dituzu".

Zein da Felix-ek arrainak jateko probabilitatea?

6.- Odol-lagin batetako globulu gorrien kopuruaren zenbaketa, Poisson-en banaketaren bidez.

Lagin-odola mikroskopioz aztertuz, eta koadrikula erregular egokiaren karratuetan, hematien banaketa behatuz, globulu gorrien urritasuna determina daiteke. Koadrikula erabilgarrienetariko bat Neubaner-ena da, bere dimentsioak 1mm x 1mm = $1mm^2$-etako azalerakoa eta 0,1 mm-tako sakonerakoa izanik eta 400 karratuetan zatiturik.

Honetarako, odol-lagina 0,25%-ez, adibidez, diluitzen eta homogenotzen da, koadrikulan, hematiak zoriz banatzen direla kontsideratu ahal izanik. Baldintza hauen pean, X = "Karratu batetako hemati-kopurua" zorizko aldagaia, λ parametroko Poisson-en banaketari jarraitzen zaiola esan daiteke.

Suposa dezagun, experientzia egiten dela eta karratu hutsen kopurua zenbatzean, 25 lortu ditugula.

Froga itzazu:

i) λ parametroko Poisson-en banaketarako, $\lambda = -\ln\{p(X=0)\}$ dela.

ii) Kasu honetan, p(X=0) eta "X=0" gertaeraren maiztasun erlatiboa identifikatuz, λ = 2,77 dela

iii) Batez beste, mm^3 odoleko $4,44.10^6$ hemati daudela

7.- Desintegrazio irradiaktiboak

Substantzia irradiaktibo batek α partikulak igortzen ditu. t denboraldian, espazioko zati batetara heltzen den α partikula-kopurua, Poisson-en banaketa betetzen duten zorizko aldagaien adibiderik onena da. Jakina, substantziak deskonposatzen dirau, epe luzetara, α partikula-dentsitatea gutxituko delarik. Hala eta guztiz ere, urteak pasatuko dira giroko materiaren gutximenik detektatu gabe. Aldi

laburretan, baldintzak konstanteak kontsidera daitezkeela eta, Poisson-en banaketara eraman zuen hipotesiaren burutzapen ideala daukagu.

Experimentu famatuan, 7,5 segundotako 2608 aldi-tarteren bitartean behatu zen substantzia irradiaktiboa, tarte bakoitzean, kontagailura heltzen zen α partikula-kopurua lorturik.

K partikula dauzkan N_k tarte-kopurua adierazten du ondoko taulak:

K	0	1	2	3	4	5	6	7	8	9	$k \geqslant 10$
N_k	57	203	383	525	532	408	273	139	45	27	16

Laginaren batezbestekoaren bidez, λ parametroa estimatuz, konpara itzazu neurturiko maiztasunak eta Poisson-en ereduaren arauera lortzen diren probabilitate teorikoak.

8.- Sero eta txertoen probak

Demagun abere-gaixotasunaren zoldura-maila arrunta 25%-a dela. Sero berria probatzeko, "h" animalia osasuntsuri xiringatzen diogu. Seroa guztiz desegokia izango balitz, X= "zoldura hartzen ez duen animali-kopurua", Bin(h, 0,75) zorizko aldagaiaren arauera banatuko litzateke.

Lau experientzia egin ziren, ondoko emaitzak lorturik:

	Animalia xiringatuak	Zoldura hartzen dutenak	Zoldura hartzen ez dutenak
1. experientzia	10	0	10
2. experientzia	12	0	12
3. experientzia	17	1	16
4. experientzia	23	2	21

Seroaren eraginkortasuna, zein experientziak sostengatuko du gehien?

Erizpide honen arauera, ordena itzazu experientziak.

9.- Hegazti batek harrapaturiko Lepidopteroen batezbesteko kopurua. Itxaropen matematikoa, oreka ekologiko posiblearen adierazle bezala.

Mota baten hegaztiak, egun batetan, populazio berbereko (oso handia) 40 tximeleta jan ditzake gehien jota. Tximeleta hauek landare pozoitsuak jaten dituzte, eta, hegaztiak tximeleta pozoindu bat jaten duenean, populazio honen tximeletak harrapatzeari uzten dio, egun horretan behintzat.

Tximeleta-populazioaren 40%-ak landare pozoitsuak jan dituela suposatuz, egun batetan, hegaztiak jan gabe uzten duen batezbesteko tximeleta-portzentaia kalkula daiteke.

X = "Hegaztiak, egun batetan, jaten duen tximeleta-kopurua", zorizko aldagaitzat hartzen dugu.

Froga itzazu:

i) X-en probabilitate-legea, banaketa geometrikoaren bidez deskriba daitekeela; hots, $p(X=k) = 0,6^{k-1}0,4$

ii) Orokorrean, $p(X=k) = q^{k-1}p$ probabilitate-legea betetzen duen zorizko aldagai geometrikorako, $E(X)= \frac{1}{p}$ dela. Ondorio bezala, adibidearen zorizko aldagairako, $E(X) = 2,5$ dela.

iii) Jan litzakeen tximeleta guztien 93,75%-a, batez beste, ez duela jaten hegaztiak.

iv) Tximeleta guztiak pozoitsuak izango balira, orduan, $E(X)= 1$ izango litzatekeela, eta, egun batetan hegaztiak jaten ez duen tximeleta-portzentaia 97,5%-era igoko litzateke.

v) Aurreko datuen arauera, egin itzazu "hegazti-lepidoptero-landare pozoitsu" sistemaren egoera ekologikoari buruzko kontsiderazioak.

10.- Suposa dezagun, intsektu batek "x" arraultze errutzeko probabilitatea, λ parametroko Poisson-en banaketak ematen duela, arraultze bat hazteko probabilitatea "p" izanik. Arraultzeen hazkundean, elkarrekiko independentzia suposatzen bal din bada, froga ezazu, "r" arraultzek bizirik irauteko probabilitatea, λp parametroko Poisson-en banaketak emanda dagoela.

Oharra: Egoera berberaren beste adibide bat hauxe da: "k" kromosoma apurtzeko probabilitatea, λ parametroko Poisson-en banaketak emanda dago, haustura bat sendatzeko probabilitatea "p" izanik.

11.- Seroa xiringatu eta gero, gizabanakoa positiboki erreakzionatzeko probabilitatea 0,991-etakoa dela jakinik, determina ezazu, txertatuak dauden 2000 pertsonatako taldean, gehien jota, negatiboki erreakzionatzen diren hiru egoteko probabilitatea.

12.- "K" zorizko aldagaiak, laborategiko tresna elektroniko baten matxura-kopurua adierazten du. "K" aldagaia, Poisson-en banaketaren arauera banatzen da. 8 tresnaz experimentatzen da, eta, tresna bakoitzerako, hilabeteko matxura-kopurua neurtzen da, 6, 3, 1, 3, 1, 4, 0, 2 lorturik.

Ondokoa eskatzen da:

i) λ estimatzea

ii) K_o muga aurkitzea, non $p(K \leqslant K_o) = 0,95$ den.

iii) Lehendabiziko tresnarako, matxuraren kanpo-zergatirik egotea onar al daiteke?

iv) Hau egia baldin bada, ez dugu lehendabiziko tresna kontutan hartzen. Zein da batezbestekoaren estimazio berria?

13.- Animali kolonia baten alearen ondorengo-kopurua, 6 parametroko Poisson-en banaketaren bidez ondo deskribatuta dagoen zorizko aldagaia da.

Baina, aleari medikamendua ematen baldin bazaio, orduan, beraren ondorengo-kopurua, 3 parametroko Poisson-en banaketari jarraitzen zaion zorizko aldagaia da.

Koloniaren 30%-a medikaturik dagoela suposatzen da.

i) Zein da, azarez aukeraturiko koloniaren aleak 5 ondorengo, gehienez, izateko probabilitatea?

ii) Aleak 5 ondorengo izan dituela jakinik, zein da medikamenduaren pean egoteko probabilitatea?

iii) Kolonian, zein da aleko batezbesteko ondorengo-kopurua?
Aleko batezbesteko ondorengo-kopurua 4 izatea nahi baldin badugu, koloniaren zein portzentaia medikatu behar dugu?

14.- Biz, dato bat jaurkitzean datzan E experimentua. X zorizko aldagaia, lehen aldiz 6 bat lortzeko behar den jaurtikipen-kopurua da. Jakina, X aldagaiak 1, 2, 3,... balioak hartuko ditu, 6 bat agertzen zaigun lehendabiziko aldiz, lehendabiziko jokaldian ala bigarrenean ala hirugarrenean,...., denaren arauera.

Ondokoa eskatzen da:

i) X zorizko aldagaiaren probabilitate-legea ematea; hots, p(X=h) h= 1,2,... kalkulatzea.

ii) Benetan, probabilitate-legea dela frogatzea.

iii) Edozein ordena bikoititako jaurtikipen edo jokalditan, lehendabiziko aldiz, sei bat ateratzeko probabilitatea.

15.- Artikulu akasdunen kopurua kontrolatzeko, 20 egunetan, 10 artikulu aztertzen dira, ondoko banaketa lorturik:

Eguna	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Art. akasdunen kop.	0	1	0	0	1	2	1	0	3	1	2	2	3	1	2	4	0	1	1	2

i) Artikuluak akasdunak izateko probabilitatea fixua dela suposaturik, doi iezaiezu banaketa binomiala datuei.

ii) Aurkitutako banaketa, prozesuaren benetako legea baldin bada, zein da, egun batetan azterturiko 10 artikulutariko bi baino akasdun gehiago egoteko probabilitatea?

16.- X zorizko aldagaiak, ondoko erara definituriko probabilitate-legea du:

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

i) Determina ezazu "K"

ii) Kalkula itzazu $p(X\leqslant 2)$ eta $p(0<X<2)$

iii) Zein da x, X-en baliorik txikiena, non $p(X\leqslant x)>0,5$ den?

iv) Determina itzazu X-en banaketa funtzioa, E(X) eta Var(X).

17.- Ondoko taulak, banaketa estatistikoa aurkezten du, non, hiri handi batetan, 0, 1, 2, 3, 4, 5 istripu gertatzen diren egun-kopurua adierazten bait da:

Istripu-kopurua	0	1	2	3	4	5
Egun-kopurua	24	17	8	7	2	2

i) 60 egunetako aldian, kalkula ezazu eguneko batezbesteko istripu-kopurua.

ii) Poisson-en banaketaren bidez, doi ezazu aurreko banaketa.

iii) Banaketa doituaren arauera, kalkula itzazu 0, 1, 2, 3, 4, 5 istripu izango direneko egun-kopuru teorikoak.

18.- N elementutako populazioa, $Np_1,\ldots, Np_r$ tamainuko $E_1,..,E_r$ azpiklaseetan zatitzen da. Zoriz eta itzuleraz aukeratzen dira populaziotiko n elementu. Azal ezazu experimentu hau deskribatzen duen eredu probabilistikoa.

19.- Estimazio-arazo bat.

Kutxontzi zenbatuak, 1-etatik N-raino zenbaturiko bolak dauzka. Biz X, itzuleraz eginiko h ateralditan lorturiko zenbakirik handiena.

i) Froga ezazu,

$$p(X=k) = \frac{k^h-(k-1)^h}{N^h} \qquad k= 1,2,\ldots,N \text{ dela}$$

Horretarako, kontutan har ezazu, "X=k" gertaerak ondokoa adierazten duela: "h" ateratako zenbaki guztiak k baino txikiagoak edo berdinak direla. Halaber, $p(X\leqslant k) = k^h/N^h$ eta $p(X=k)=p(X\leqslant K) - p(X\leqslant k-1)$ direla.

ii) Froga ezazu,

$$E(X) = \frac{N^{h+1} - \sum_{k=1}^{N} (k-1)^h}{N^h} \quad (*)$$

Kontutan har ezazu, (*) batukaria, N handitarako, $y=x^h$ kurbaren pean, 0-tik N-raino dagoen azalerara hurbiltzen dela.
Ondoriozta ezazu,

$$E(X) = \frac{h}{h+1} N \quad \text{dela.}$$

iii) Herri batetan, N = 1000 txakur baldin badago eta h=10 aletako lagina aztertzen bada, behaturiko plaken itxarotako zenbakirik handiena (aleatoriotasuna suposaturik) 910 da gutxi gora behera.

(Hala ere, estatistikatzaile praktikoa alderantziz aritzen da; hots, laginean neurturiko maximoa erabiltzen du, N zenbaki ezezaguna kalkulatzeko. Metodo hau, azken gerratean erabili zen, etsaiaren arma-ekoizpena kalkulatzeko).

20.- Probabilitate aldakorretako Bernouilliren probak.

Kontsidera ditzagun h proba aske, non bakoitzak E eta F (arrakasta eta porrota) bi emaitza posible bait dauzka. K. probako E-ren probabilitatea p_k da, F-rena $q_k = 1-p_k$ izanik. (Edozein k-tarako, $p_k=p$ baldin bada, orduan, planteamendu hau Bernouilliren proba bihurtzen da).

Deskribatzeko erarik sinpleena, 0 eta 1 balioak E eta F-ri egokieraztean datza. Orduan, h kopuruko X_k zorizko aldagai askeK $p(X_k=1)=p_k$ eta $p(X_k=0)=q_k$ beraren probabilitate-legea izanik, zeharo deskribatzen dute eredua.

i) Froga itzazu $E(X_k) = p_k$ eta $Var(X_k) = p_k q_k$

ii) $S_n = X_1+X_2+\ldots+X_n$ zorizko aldagaia kontutan hartu rik, egiazta itzazu $E(S_n) = p_1+p_2+\ldots+p_n$ eta

$Var(S_n) = p_1q_1+p_2q_2+\ldots+p_nq_n$. S_n aldagaia, n proba asketako arrakasta-kopuru osotzat interpreta daiteke.

iii) $p = \frac{p_1+\ldots+p_n}{n}$ kantitatea, batezbesteko arrakasta-probabilitatetzat har daiteke. Hau dela eta, egoera hau eta p arrakasta-probabilitate konstantea daukaten Bernouilliren probak konparatzea normala dirudi.

Egiazta ezazu ere $Var(S_n) = np - \sum_{k=1}^{n} p_k^2$ dela.

$\sum_{k=1}^{n} p_k=np$ betetzen duten $\{p_k\}_{k=1}^{n}$ segida guztien artean, p_k guztiak berdinak direnean, $\sum_{k=1}^{n} p_k^2$ baturak bere balio minimoa hartzen duela froga daiteke.

Guztiau kontutan harturik, ondoriozta ezazu ondokoa:

p batezbesteko arrakasta-probabilitatea konstantea baldin bada, hots, $p_1=p_2=\ldots=p_n=p$ direnean, $Var(S_n)$ delakoak bere balio maximoa hartzen duela.

Hona hemen emaitza harrigarria: p_k-ren aldakortasunak, hots, uniformetasun-galerak, bariantzak neurturiko zorizko fluktuazioen magnitutea gutxitu erazten du. Adibidez, n makinen p batezbesteko kalitatea emanik, makinak berdinak baldin badira, aldakorragoa izango da ekoizpena.

IV. GAIA: ZORIZKO ALDAGAI JARRAIAK

IV.1.- Zorizko aldagai jarraiaren banaketa- eta dentsitate-funtzioak.

IV.1.1.- Aurretiko ideiak.

IV.1.2.- Banaketa-funtzioa.

IV.1.3.- Dentsitate-funtzioa.

IV.2.- Zorizko aldagai jarraiaren batezbestekoa eta bariantza. Tchebycheff-en teorema.

IV.2.1.- Batezbestekoa edo itxaropen matematikoa.

IV.2.2.- Bariantza eta desbidazio standarda.

IV.2.3.- Sakabanatzea eta Tchebycheff-en teorema.

IV.3.- Zenbait banaketa jarrai.

IV.3.1.- Banaketa exponentziala.

IV.3.2.- Banaketa normala.

IV.3.3.- Limitearen teorema zentrala.

ZORIZKO ALDAGAI JARRAIAK

IV.1.- ZORIZKO ALDAGAI JARRAIAREN BANAKETA-ETA DENTSITATE-FUNTZIOAK.

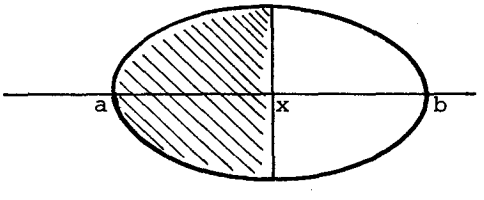
IV.1.1.- Aurretiko ideiak.

X zorizko aldagaia jarraia da, baldin eta edozein tarte finitu edo infinituren puntu guztiak har baditzake. Tarte hori, (a,b), $(-\infty,b)$, (a,∞) edo $(-\infty,\infty)$ motakoa izan daiteke.

Gogora dezagun, $\{x_1,\ldots,x_n\}$ heineko X zorizko aldagai diskretua, $f(x_i)=p(X=x_i)$ $i=1,\ldots,n$ zehaztuz deskribatuta dagoela, probabilitate-legearen bidez. Tratamendu hau ezin daiteke aldagai jarraietarako erabil, zeren aldagai jarraiak balio fixu bat hartzeko probabilitatea nulua bait da.

Probabilitate-banaketa jarraia deskribatzeko, zorizko aldagai jarraiaren heinaren edozein zatiri dagokion probabilitatea kalkulatzeko arazoa ebaztea baimentzen diguten tresna berriak behar ditugu.

Ikerketa honetan, hurrengo interpretazio intuikorrak lagunduko digu: probabilitate-banaketa jarraia, aldagaiaren heinean zeharrezko masa unitarioaren zatiketa jarraitzat interpreta daiteke. (IV.1.irudia).



IV.1.irudia

Zati marratuaren masa, $p(a\leqslant X\leqslant x)$ probabilitateari dagokio.

IV.1.2.- Banaketa-funtzioa.

X zorizko aldagai jarraiaren heineko x balio bakoitzerako banaketa-funtzioa ondoko modura definitzen dugu:

$$F(x) = p(X \leqslant x)$$

F(x), (a,x) segmentuan, edo, heinaren ezkerreko muturra $-\infty$ baldin bada, $(-\infty, x)$ tartean dagoen probabilitate-masatzat kontsidera daiteke.

$F(x)=0$ baldin $x<a$ eta $F(x)=1$ baldin $x>b$ direneko hitzarmena hartuz, F(x), edozein balio errealetarako definituta dago.

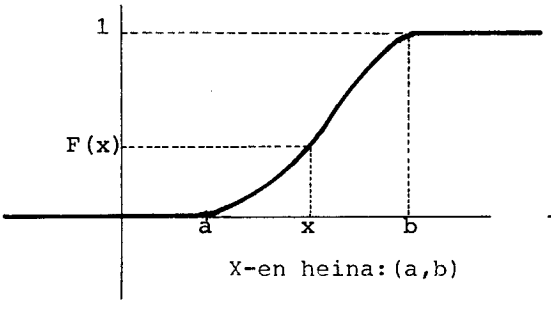
F(x)-en ondoko propietateak, egiaztatzeko errazak dira:

1) $F(-\infty) = 0$; $F(+\infty) = 1$.

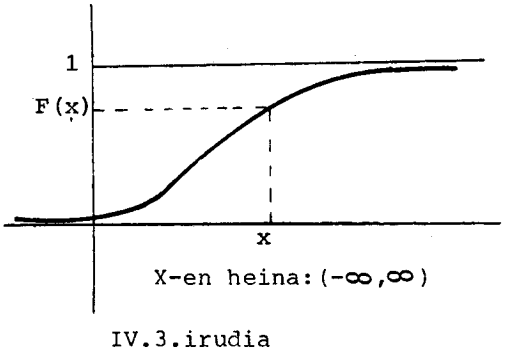
2) F(x) ez-beherakorra, ez-negatiboa eta 0 eta 1-en artean bornatua da.

3) Probabilitate-banaketa jarraia baldin bada, F(x) jarraia da.

F(x)-en grafiko kartesiarrak, "ojiba-forma" delakoa hartzen du. (IV.2. eta IV.3.irudiak).



IV.2.irudia



IV.3.irudia

x balio bakoitzerako kurbaren ordenatuak $p(X \leqslant x)$ ematen digu.

Beraz, F(x) ezagutzeak, x_1, x_2 X-en heineko edozein bi puntu izanik, $p(x_1 \leqslant X \leqslant x_2)$ kalkulatzea baimentzen digu, zeren

$p(x_1 \leqslant X \leqslant x_2) = F(x_2) - F(x_1)$ bait da.

IV.1.3.- Dentsitate-funtzioa.

Ikusi dugunez, F(x+h)-F(x) kendurak, h luzerako (x,x+h) segmentuan dagoen probabilitate-masa adierazten du. Hau dela eta, $\frac{F(x+h)-F(x)}{h}$ zatidura, segmentuaren batezbesteko probabilitate-dentsitatetzat har daiteke.

F(x) banaketa-funtzioa deribagarria dela suposatzen baldin badugu,

$$\lim_{h \to 0} \frac{F(x+h)-F(x)}{h} = F'(x) \quad \text{izango da.}$$

F'(x)-ek, probabilitate-dentsitatea, x puntuan, adierazten du, f(x) = F'(x) funtzioa dentsitate-funtzioa deitua izanik.

Kalkulu integralaren zenbait oinarrizko emaitza direla eta, banaketa- eta dentsitate-funtzioek elkar determinatzen dutela bistakoa da, zeren F(x) ezaguturik, f(x) = F'(x) bait da; eta, alderantziz, f(x) ezaguturik,

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{bait da.}$$

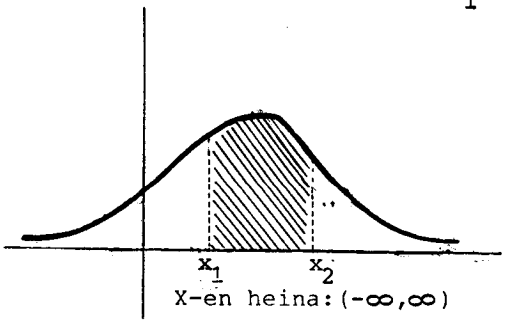
f(x) dentsitate-funtzioa ardatz erreal osoan definituta dago. X zorizko aldagai jarraiak (a,b) hein finitua daukan kasuan, F(x)=0 baldin $x<a$, eta F(x)=1 baldin $x>b$ direnez, orduan f(x)=0 baldin $x<a$ edo $x>b$.

Dentsitate-funtzioaren propietate inportanteenak honela adieraz daitezke:

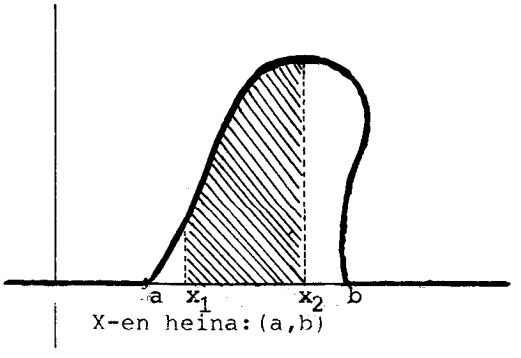
1) F(x) ez-beherakorra dela eta, f(x) ez-negatiboa da.

2) $\int_{-\infty}^{+\infty} f(x)dx=1$, edo, X-en heina (a,b) den kasuan, $\int_{a}^{b} f(x)dx=1$.

3) $p(x_1 \leqslant X \leqslant x_2) = \int_{x_1}^{x_2} f(x)dx$.



IV.4.irudia



IV.5.irudia

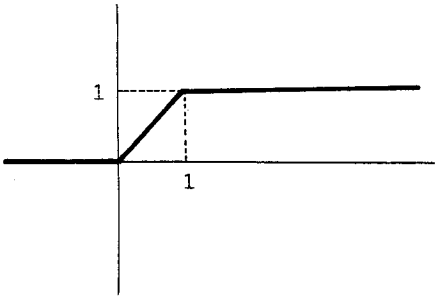
Azken batez, X aldagai jarrairako, x_1 eta x_2 $(x_1 < x_2)$ edozein zenbaki izanik, $p(x_1 \leqslant X \leqslant x_2)$ kalkulatzeko funtsezko arazoa, banaketa-funtzioa ezagutuz ebazten da, zeren kasu honetan, $p(x_1 \leqslant X \leqslant x_2)=F(x_2)-F(x_1)$ bait da, edo, dentsitate-funtzioaren bidez,

$$p(x_1 \leqslant X \leqslant x_2) = \int_{x_1}^{x_2} f(x)dx \quad \text{bait da.}$$

Horrela, X zorizko aldagai jarraia, bere banaketa- edo dentsitate-funtzioen bidez determinaturik dago.

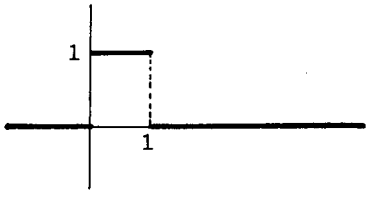
1.adibidea:

$$F(x) = \begin{cases} x & \text{baldin } 0 \leqslant x \leqslant 1 \\ 0 & \text{baldin } x < 0 \\ 1 & \text{baldin } x > 1 \end{cases}$$



Bere dentsitate-funtzioa ondokoa izango da:

$$f(x) = F'(x) = \begin{cases} 1 & \text{baldin } 0 \leqslant x \leqslant 1 \\ 0 & \text{baldin } x < 0 \\ 0 & \text{baldin } x > 1 \end{cases}$$



Banaketa hau, uniforme edo errektangularra deitzen da.

$$p(0,2 \leqslant X \leqslant 0,6) = F(0,6)-F(0,2) = 0,6-0,2 = 0,4.$$

IV.2.- ZORIZKO ALDAGAI JARRAIAREN BATEZBESTEKOA ETA BARIANTZA. TCHEBYCHEFF-en TEOREMA.

IV.2.1.- Batezbestekoa edo itxaropen matematikoa.

X, f(x) dentsitate-funtzioa daukan zorizko aldagai jarraia baldin bada, batezbestekoa edo itxaropen matematikoa honela definitzen da:

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

integrala, finitua baldin bada.

Aldagai jarraiaren batezbestekoa, aldagai diskretuaren antzera, probabilitate-masa bere inguruan kontzentratzen den balioa adierazten duen zentralizazio-neurria da. Izan ere, X-en heina, "oso luzera laburreko" tarte-kopuru "handian" zatitzen baldin bada, tarte bakoitzaren klase-marka beraren probabilitate-masaz biderkatzen bada, eta lorturiko biderkadurak batzen badira, $\sum_{i=1}^{n} x_i p_i$ batura, tarteen anplitudeak zerorantz doazenean, $\int_{-\infty}^{+\infty} x\, f(x)dx$ bihurtzen da.

E(X)-en propietateak, zorizko aldagai diskretuaren itxaropenerako ikusi zirenen guztiz antzekoak dira, kalkulu integralaren propietateetatik ondorioztaturik:

1) $E(X+a) = E(X) + a \quad \forall a \in \mathbb{R}$

Partikularrean, $E(X-\mu) = 0$.

Izan ere:

$$E(X+a) = \int_{-\infty}^{+\infty} (x+a)f(x)dx = \int_{-\infty}^{+\infty} xf(x)dx + \int_{-\infty}^{+\infty} af(x)dx =$$
$$= E(X) + a\int_{-\infty}^{+\infty} f(x)dx = E(X) + a$$

2) $E(k\cdot X) = k\cdot E(X) \quad \forall k \in \mathbb{R}$

Izan ere:

$$E(k\cdot X) = \int_{-\infty}^{+\infty} kxf(x)dx = k\int_{-\infty}^{+\infty} xf(x)dx = k\cdot E(X)$$

3) X eta Y bi zorizko aldagai jarrai baldin badira, orduan, $E(X+Y) = E(X) + E(Y)$ da.
(Propietatea, zorizko aldagai jarraien kopuru finitura hedatzen da).
Egiten ez dugun frogapena ez da batere erraza.

4) $E(aX+bY) = aE(X) + bE(Y) \quad \forall a,b \in \mathbb{R}$
Itxaropena, "eragile lineala" dela adierazten du propietate honek.
Frogapenari dagokionez, 2) eta 3) propietateen korolarioa besterik ez da.

5) X eta Y bi aldagai jarrai baldin badira, orduan,

$$E[(X\cdot Y)^2] \leqslant E(X^2)E(Y^2).$$

6) X eta Y aldagai jarrai askeak(*) baldin badira, orduan, $E(XY) = E(X)E(Y)$
(Aldagai jarrai askeen kopuru finiturako, gauza bera gertatzen da).
Hemen, frogapena ez da batere nabaria.

(*) X eta Y bi aldagai jarrai, askeak dira baldin:
$p\{(X \leqslant a) \cap (Y \leqslant b)\} = p(X \leqslant a)p(Y \leqslant b) \quad \forall a,b \in \mathbb{R}$
Intuikorki, batak hartzen dituen balioak ezagutzeak, besteak hartzen dituen balioetan eraginik ez daukala adierazten du.

2.adibidea:

Banaketa errektangularreko kasuan, hauxe da bere batez bestekoa:

$$\mu = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} x \cdot 1 \, dx = \frac{x^2}{2}\Big]_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$$

IV.2.2.- Bariantza eta desbidazio standarda.

X, μ batezbestekoa daukan aldagai jarraia baldin bada, X-en bariantza ondoko integrala da:

$$\text{Var}(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

berau finitua dela suposaturik.

X-en desbidazio standarda bariantzaren erro karratu positiboa da, hots,

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Bariantzak (edo desbidazio standardak), probabilitate-masaren batezbestekoaren inguruko sakabanatze-maila ematen digu.

Ondoren, bere propietateak aipatzen ditugu, irakurlea, ariketa gisa, haietariko batzu frogatzen saia dadin:

1) $\text{Var}(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu_X^2 = E(X^2) - E(X)^2$
2) $\text{Var}(X+k) = \text{Var}(X) \quad \forall k \in \mathbb{R}$
3) $\text{Var}(k.X) = k^2 \text{Var}(X) \quad \forall k \in \mathbb{R}$
4) $\text{Var}(X) = 0 \Longleftrightarrow X = \text{kte.}$ ia seguraski
5) X eta Y askeak baldin badira, orduan,
   $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Aurreko propietateek eta batezbestekoarenek, maiz erabiliko dugun proposamendua frogatzea baimentzen digute:

Proposamendua:

X, μ batezbesteko eta σ desbidazio standardeko aldagai jarraia baldin bada, orduan, "tipifikatua" izeneko $Z=\frac{X-\mu}{\sigma}$ aldagai berriak, 0 batezbestekoa eta 1 bariantza dauzka.

Frogapena:

$$E(Z) = \frac{1}{\sigma} E(X-\mu) = \frac{1}{\sigma} \{E(X)-\mu\} = \frac{1}{\sigma} \cdot 0 = 0$$

$$Var(Z) = Var(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2} Var(X-\mu) = \frac{1}{\sigma^2} Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

3.adibidea:

Banaketa errektangularraren bariantza eta desbidazio standarda ondoko modura lortzen dira:

$$\sigma_x^2 = \int_0^1 (x - \frac{1}{2})^2 f(x)\,dx = \int_0^1 (x- \frac{1}{2})^2 dx =$$

$$= \frac{(x- \frac{1}{2})^3}{3} \Big]_0^1 = \frac{1}{24} - (- \frac{1}{24}) = \frac{1}{12}$$

$$\sigma_x = \sqrt{1/12} = 0,2887$$

IV.2.3.- Sakabanatzea eta Tchebycheff-en teorema.

Bariantza, sakabanatze-neurria dela azaldu dugu lehenago. Erraza izan arren, orokortasun eta erabilpen handiak dauzkan ondoko teoremak zehazten du baieztapen hau.

Tchebycheff-en teorema:

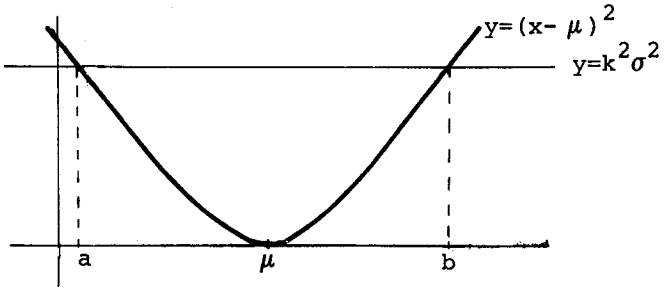
Biz f(x) dentsitate-funtzioko X aldagai jarraia, non μ batezbestekoa eta σ^2 bariantza finituak diren. k konstante positiboa baldin bada, orduan:

$$p(|X-\mu| \leqslant k\sigma) \geqslant 1 - \frac{1}{k^2}$$

Frogapena:

Demagun $y=(x-\mu)^2$ parabola, eta $y=k^2\sigma^2 > 0$ zuzena; a eta b, bien arteko ebaki-puntuen abzisak izanik.

$x<a$ edo $x>b \Longleftrightarrow (x-\mu)^2 > k^2\sigma^2$
dela bistakoa da.



Orduan:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx \geqslant \int_{-\infty}^{a} (x-\mu)^2 f(x)dx +$$

$$+ \int_{b}^{+\infty} (x-\mu)^2 f(x)dx \geqslant k^2 \cdot \sigma^2 \int_{-\infty}^{a} f(x)dx +$$

$$+ k^2\sigma^2 \int_{b}^{+\infty} f(x)dx = k^2\sigma^2 \{p(X<a) + p(X>b)\} =$$

$$= k^2\sigma^2 p\{(X<a) \cup (X>b)\} = k^2\sigma^2 p\{(X-\mu)^2 > k^2\sigma^2\}$$

Hemendik abiatuz:

$$\frac{1}{k^2} \geqslant p\{(X-\mu)^2 > k^2\sigma^2\} \Longrightarrow 1 - \frac{1}{k^2} \leqslant p\{(X-\mu)^2 \leqslant k^2\sigma^2\}$$

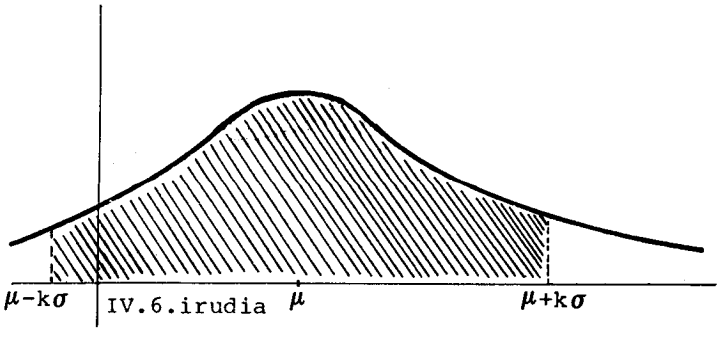
Beraz:

$$p(-k\sigma \leqslant X-\mu \leqslant k\sigma) \geqslant 1 - \frac{1}{k^2}$$

Edo beste era batetara:

$$p(\mu - k\sigma \leqslant X \leqslant \mu + k\sigma) \geqslant 1 - \frac{1}{k^2}$$

Grafikoki, Tchebycheff-en teoremak, IV.6. irudiko azalera marraztua, $1 - \frac{1}{k^2}$ baino handiagoa edo berdina dela adierazten digu, edo beste hitz batzutan, "isatsen" azalerak ez du $\frac{1}{k^2}$ gainditzen.



IV.6.irudia

k=2 kasuan: $p(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \geqslant \frac{3}{4}$

k=4 kasuan: $p(\mu - 4\sigma \leqslant X \leqslant \mu + 4\sigma) \geqslant \frac{15}{16} = 0,9375$

Hots, X-ek 4σ baino laburragoa den μ-tiko distantzia daukan balioa hartzeko probabilitatea 0,9 baino handiagoa da.

IV.3.- ZENBAIT BANAKETA JARRAI.

IV.3.1.- Banaketa exponentziala.

Zenbait egoeratan, "gertaera bat jazo arte iraganiko aldia", zorizko aldagaitzat jotzea interesatzen zaigu. Zorizko aldagai hau, baldintza batzuren pean, banaketa exponentzialaren bidez deskriba daiteke egokiro.

X zorizko aldagai jarraiak, <u>banaketa exponentzial negatiboa</u> dauka, baldin bere dentsitate-funtzioak ondoko itxura badu:

$$f(x) = \begin{cases} \beta e^{-\beta x} & \text{baldin } x > 0 \\ 0 & \text{baldin } x \leqslant 0 \end{cases}$$

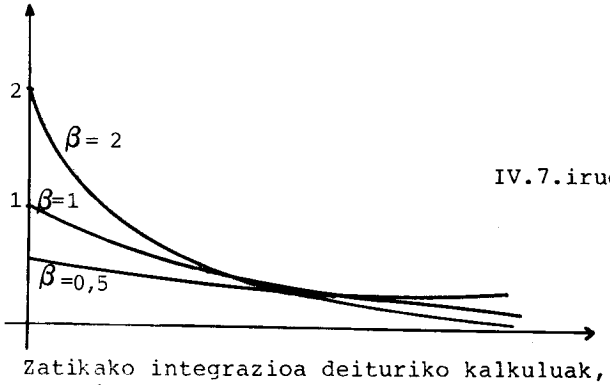
β konstante positiboa izanik ($\beta > 0$).

Ondokoa ikusiz, f(x) dentsitate-funtzioa dela froga daiteke:

i) $f(x) \geqslant 0 \quad \forall x \in \mathbb{R}$

ii) $$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} \beta e^{-\beta x} dx = -e^{-\beta x}\Big]_0^{+\infty} = 0 + 1 = 1$$

β-ren balio desberdinetarako f(x)-en grafikoak, IV.7. irudian ikus daitezke:



IV.7.irudia

Zatikako integrazioa deituriko kalkuluak, $E(X) = \frac{1}{\beta}$ eta $Var(X) = \frac{1}{\beta^2}$ direla erakusten digu.

Banaketa-funtzioa hauxe da:

$$F(x) = \begin{cases} 0 & \text{baldin } x \leqslant 0 \\ 1 - e^{-\beta x} & \text{baldin } x > 0 \end{cases}$$

Banaketa exponentzialak hurrengo propietatea betetzen du:

"$x>0$ eta $t>0$ baldin badira, $p\,(X>t+x \,/\, X>t\,) = p(X>x)$"

Izan ere:

$$p\,(X>t+x \,/\, X>t\,) = \frac{p\{(X>t+x)\cap(X>t)\}}{p\,(X>t)} = \frac{p(X>t+x)}{p(X>t)} =$$

$$= \frac{1-F(t+x)}{1-F(t)} = \frac{e^{-\beta\,(t+x)}}{e^{-\beta\,t}} = e^{-\beta\,x} = p(X>x)$$

Propietate honek ondoko interpretazioa onartzen du: X-ek, itxarote-aldia (gertaera jazo arte) adierazten baldin badu, t aldia itxaron eta gero, x+t baino itxarote-aldi luzeagoa izateko probabilitatea ez dago lehenago iraganiko t aldiaren pean. Aldiarekiko homogenotasun honek, banaketa exponentziala unibokoki karakterizatzen duela frogatzen da.

4.adibidea:

Demagun, "bakterioa zatitu arte iraganiko aldia" zorizko aldagaia, β=2,5h. parametroko banaketa exponentzialari jarraitzen zaiola. Bakterioak, 2,7h. eta 3h.-ren arteko bizitza-aldia izateko probabilitatea hauxe da:

$$p(2,7\leqslant X\leqslant 3) = F(3)-F(2,7) = 1-e^{-2,5\cdot 3}-(1-e^{-2,5\cdot 2,7}) =$$

$$= 1,17\cdot 10^{-3}-5,53\cdot 10^{-4} = 6,18\cdot 10^{-4}$$

Analogoki, 2,7h. baino gehiago bizi izan dela jakinik, bakterioa 3h. baino gutxiago bizitzeko probabilitatea hauxe izango da:

$$p(X\leqslant 3/X\geqslant 2,7) = \frac{p\{(X\leqslant 3)\cap(X\geqslant 2,7)\}}{p(X\geqslant 2,7)} = \frac{p(2,7\leqslant X\leqslant 3)}{p(X\geqslant 2,7)} =$$

$$= \frac{6,18\cdot 10^{-4}}{1,17\cdot 10^{-3}} = 0,528$$

IV.3.2.- Banaketa normala.

Banaketa normala, Laplacek eta Gaussek eriden zuten, behaketen akats-banaketaren azterketatik abiatuz. Probabilitate-banaketa jarrairik garrantzitsuena da, beraren aplikazio praktiko anitz eta interes handiko beste zenbait banaketaren sortzailea izatea direla kausa.

$(-\infty,\infty)$ ardatz erreal osoa heintzat daukan X zorizko aldagai jarraia, μ batezbesteko eta σ^2 bariantzako banaketa normalari jarraitzen zaio, baldin bere dentsitate-funtzioa ondoko adierazpenak emanik badago:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < +\infty$$

non μ eta σ zenbaki errealak eta $\sigma > 0$ diren.

X aldagaiak, μ batezbesteko eta σ^2 bariantzako banaketa normala betetzen duela adierazteko, honela idatzen da: $N(\mu, \sigma^2)$.

Nahiz eta nabaria ez izan, ondokoa froga daiteke:

i) $f(x)$ dentsitate-funtzioa da; hots, $f(x) \geq 0 \ \forall x \in \mathbb{R}$ eta $\int_{-\infty}^{+\infty} f(x)dx = 1$.

ii) $E(X) = \mu$, $Var(X) = \sigma^2$.

Banaketa normalaren kurbaren forma aztertzeko, $f(x)$-en zenbait propietate nabareraziko ditugu, haiek guztiak, egiaztatzeko izanik:

1) $f(\mu - k) = f(\mu + k) \quad \forall k \in \mathbb{R}$; hots, kurba normala, ordenatu-ardatzaren $x=\mu$ zuzen paraleloarekiko simetrikoa da.

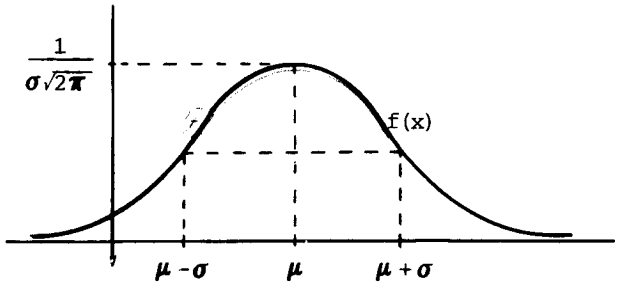
2) x, μ baliotik ∞-raino emendatzerakoan, f(x) beheratzen da; ∞-rantz doanean, zerorantz jotzen duela rik.

3) f'(x)=0 ekuazioa ebatziz, kurbak, $x=\mu$ abzisan, bere ordenatu maximoa hartzen duela aurkitzen da, ordenatu hau $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$ izanik.

4) f"(x)=0 ekuazioa ebatziz, $\mu-\sigma$ eta $\mu+\sigma$ abzisako kurbaren puntuak, kurbaren inflexio-puntuak direla lortzen da.

5) Azkenik, $(\mu-\sigma\ ,\ \mu+\sigma\ )$ tartean, f"(x) negatiboa da, eta, bertatik kanpo, ordea, positiboa. $(\mu-\sigma\ ,\ \mu+\sigma\ )$ tartean, ahurtasun negatibokoa (-0,4-takoa, gutxi gora behera) eta tartetik kanpo, ahurtasun positibokoa (+0,4-takoa, gutxi gora behera) dela adierazten digu honek.

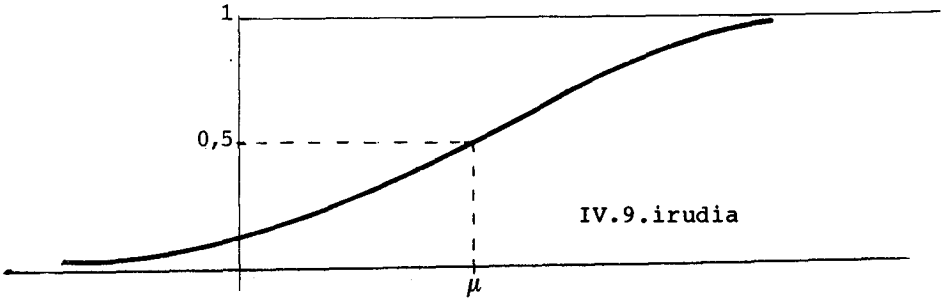
Propietate hauek ikusiz, IV.8.irudiko forma (Gauss-en kanpai famatua) daukan f(x)-en grafikoa eraiki daiteke:



ıdia

Banaketa-funtzioa, $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2} dt$ da.

IV.9.irudian agertzen den kurba bere grafikoa izanik:



IV.9.irudia

f(x) dentsitate-funtzioa, balio-kopuru infinitua har dezaketen μ eta σ parametroen pean dagoenez, "Gauss-en kanpai"-kopurua "bi bider infinitua" dela ondorioztatzen da. μ aldatzen denean (σ -k konstante dirauelarik), kanpaia OX ardatzaren paraleloki desplazatzen da, desitxuratu gabe; eta σ aldatzen denean (μ -k konstante dirauelarik), kanpaia, aitzitik, hestuagoa edo zapalduagoa egiten da.

"Gauss-en kanpai" guztien artean, $\mu=0$ eta $\sigma=1$ balioetakoari "kanpai tipikoa" deritzogu, ordenatu-ardatza simetri-ardatztzat daukana, aparteko garrantzia izanik.

Haatik, honela adieraz daitekeen ondoko arazo praktikoa ebatzi gabe gelditzen da: Zenbat balio du $p(x_1 \leqslant X \leqslant x_2)$ probabilitateak, non x_1 eta x_2 $(x_1 < x_2)$ edozein zenbaki erreal diren?

Teorikoki, hauxe da:

$$p(x_1 \leqslant X \leqslant x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \, dx$$

Hala ere, integral honen kalkulua ez da eskurakoia, eta, horrez gainera, μ eta σ-ren balio bakoitzerako kalkulu berriak egin beharko genituzke.

Baina, dakigunez, $X=N(\mu,\sigma)$ emanik, $Z = \frac{X-\mu}{\sigma}$ aldagai tipifikatua kontutan hartzen badugu, orduan, Z-k, 0 batezbestekoa eta 1 desbidazio standarda dauzka.

Z aldagaia normala dela frogatzea ez da zaila. Beraz, $X=N(\mu,\sigma)$ emanik, Z aldagai tipifikatua N(0,1) da.

Ondorioz, hauxe daukagu:

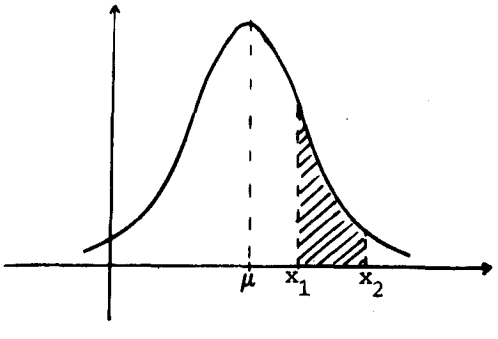
$$p(x_1<X<x_2) = p(x_1-\mu< X-\mu < x_2-\mu) =$$

$$= p(\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}) =$$

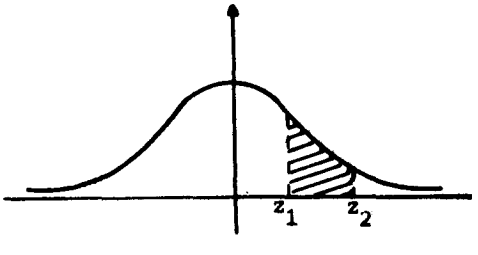
$$= p(z_1< Z <z_2)$$

$z_1 = \frac{x_1-\mu}{\sigma}$ eta $z_2 = \frac{x_2-\mu}{\sigma}$ izanik.

Horrela, $N(\mu,\sigma^2)$ dentsitate-kurbaren pean, μ eta σ^2 edozein izanik, IV.10.irudiko barruti marratuaren azalera aurkitzeko arazoa, N(0,1) kurbaren pean, IV.11.irudiko barruti marratuaren azalera aurkitzeko arazo bihurtzen da. Praktikan, banaketa normalaren taulak erabiliz ebazten da azken arazo hau.



IV.10.irudia



IV.11.irudia

Problema askotan, $p(|X-\mu| > \gamma_p\sigma)$ (IV.12.irudian) kalkulatzea interesatzen zaigu; hots, batezbestekoarekiko desbidazio absolutua, γ_p edozein faktore eta σ-ren arteko biderkadura baino luzeagoa izateko probabilitatea. Hauxe daukagu:

$$p(|X-\mu| > \gamma_p\sigma) = p\left(\left|\frac{X-\mu}{\sigma}\right| > \gamma_p\right) =$$

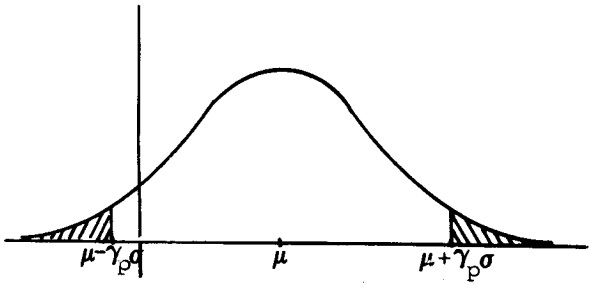
$$= p(|Z| > \gamma_p) =$$

$$= p(Z < -\gamma_p) + p(Z > \gamma_p) =$$

$$= 2p(Z > \gamma_p) =$$

$$= 2\{1 - p(Z < \gamma_p)\}$$

non $Z = \frac{X-\mu}{\sigma}$ $N(0,1)$ den.



IV.12.irudia

Ondoren adierazten ditugun p-ren zenbait baliori dagozkien γ_p-ren balioak, N(0,1) banaketaren taulatik abiatuz lorturikoak, bereziki garrantzitsuak dira aplikazioetan:

Behaketen 50%-a, $\mu \pm 0{,}68\sigma$ mugen artean dago

Behaketen 68,2%-a, $\mu \pm \sigma$ mugen artean dago

Behaketen 95%-a, $\mu \pm 1{,}96\sigma$ mugen artean dago

Behaketen 95,4%-a, $\mu \pm 2\sigma$ mugen artean dago

Behaketen 99%-a, $\mu \pm 2{,}58\sigma$ mugen artean dago

Behaketen 99,7%-a, $\mu \pm 3\sigma$ mugen artean dago

Behaketen 99,9%-a, $\mu \pm 3{,}29\sigma$ mugen artean dago

5.adibidea:

Bihotz-gaixoen mota baten odol-presioa aztertzerakoan, diastole-presioen batezbestekoa 90mmHg-takoa eta desbidazio standarda 12mmHg-takoa zirela lortu zen, presioen populazioa gutxi gora behera normala izanik.

Aurki itzazu, taldeko gaixoren baten diastole-presioak ondoko baldintzak betetzeko probabilitateak:

a) 115mmHg. baino handiagoa izateko.

b) Batezbestekotik, 20 unitate (ezkerretik ala eskuinetik) baino gutxiagotan diferitzeko.

c) Determina ezazu presioaren balioa, non gaixoen 5%-ak soilik, balio hori baino diastole-presio handiagoa bait dauka.

a) $p(X>115) = p\left(\frac{X-90}{12} > \frac{115-90}{12}\right) = p(Z>2{,}08) = 0{,}0188$

b) $p(|X-90|<20) = p\left(\left|\frac{X-90}{12}\right| < \frac{20}{12}\right) = p(|Z|<1{,}67) =$

$= p(-1{,}67<Z<1{,}67) = 1-2p(Z>1{,}67) =$

$= 1-2\cdot 0{,}0475 = 0{,}905$

c) Eskatutako balioa, P_{95}(95 pertzentila) adieraziko dugu.

$$p(X > P_{95}) = p\left(\frac{X-90}{12} > \frac{P_{95}-90}{12}\right) = p\left(Z > \frac{P_{95}-90}{12}\right) = 0,05$$

Banaketa normalaren taulen bidez, hauxe lortuko dugu gutxi gora behera:

$$\frac{P_{95}-90}{12} \simeq 1,64$$

Hau dela eta, $P_{95} = 12 \cdot 1,64+90 = 109,68$mmHg.

IV.3.3.- Limitearen teorema zentrala.

Banaketa normalari (doi-doi ala gutxi gora behera) jarraitzen zaion zorizko aldagai jarraien kopurua, izugarri handia da. Honen azalpen matematikoa, limitearen teorema zentrala delakoa da, teoria estatistikoaren garrantzitsuenetariko bat. Nahiz eta hemen beraren frogapenean ezin sartu, bere esanguraren deskribapen intuikorra azalduko dugu, inferentzia estatistikoan daukan aplikazio anitz dela kausa.

Zorizko aldagai ugariak, batzen diren oso antzeko bariantza finituko faktore txiki independenteen konkurrentziaz sortzen dira. Adibidez, uzta baten laranja baten pisua, beste faktore batzuren artean, lurraren ezaugarri, erabilitako ongarri, hezetasun-maila, tenperatura eta abarren pean dagoen zorizko aldagaia da. Eta, bariantza txiki eta desberdin samar duten aldagai txiki askeen kopuru handia batuz eratzen diren zorizko aldagai guztiak, Gauss-en lege normalari darrazkio.

Aldagai biologiko anitz horrela sortzen direnez gero, banaketa normala, hain maiztasun handiaz agertzeak ez gaitu harritu behar.

Hona hemen limitearen teorema zentralaren enuntziatu baten aurkezpen zehatza:

Levy-Lindeberg-en teorema:

$X_1, X_2, \ldots, X_n$, μ batezbesteko eta $\sigma^2 \neq 0$ bariantza finituko zorizko aldagai aske eta berdinki banatuen segida baldin bada, orduan,

$$Z_n = \frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}} \text{ , } N(0,1) \text{ da asintotikoki;}$$

hots,

$$p(z_1 < Z_n < z_2) \simeq \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

n zenbat eta handiagoa izan, hurbilketa gero eta hobeagoa delarik. Edo, era baliokide batetara esanda, $S_n = X_1+X_2+\ldots+X_n$ zorizko aldagaia, $N(n\cdot\mu\ ,\ \sigma\cdot\sqrt{n})$ da gutxi gora behera.

Kontutan har dezagun, $E(X_1+X_2+\ldots+X_n) = E(X_1)+E(X_2)\ldots+E(X_n) = n\mu$ eta $Var(X_1+X_2+\ldots+X_n) = Var(X_1)+Var(X_2)+\ldots+Var(X_n) = n\cdot\sigma^2$ direla, zeren X_i aldagaiak askeak eta banaketa berberekoak bait dira.

6.adibidea:

Demagun, populazioaren altuera, $E(X) = \mu$ batezbesteko eta $Var(X) = \sigma^2$ bariantza finituko X zorizko aldagaia dela. Halaber, suposa dezagun, n pertsonaren altuerak neurtzen direla, ondoren, beren batezbestekoa kalkulaturik:

$$\bar{X} = \frac{X_1+X_2+\ldots+X_n}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n$$

$\bar{X}$, zorizko aldagaia da, zeren $\frac{1}{n} X_1,\ldots,\frac{1}{n} X_n$ banaketa berbereko zorizko aldagai askeen batura bait da,

$$E(\frac{1}{n} X_i) = \frac{1}{n} E(X_i) = \frac{\mu}{n} \text{ eta}$$

$$Var(\frac{1}{n} X_i) = \frac{1}{n^2} Var(X_i) = \frac{1}{n^2} \cdot \sigma^2 \text{ izanik.}$$

$\bar{X}$, asintotikoki, $N(\mu, \frac{\sigma}{\sqrt{n}})$ dela, aurreko teoremaren berehalako ondorioa da, X_i zorizko aldagaien banaketaren independentea izanik.

Partikularra den oso kasu garrantzitsua hauxe da:

Moivre-ren teorema:

X, Bin(n,p), $p \neq 0$ eta $p \neq 1$ izanik, zorizko aldagaia baldin bada, orduan,

$\frac{X-np}{\sqrt{npq}}$ zorizko aldagaia, N(0,1) da asintotikoki.

Oharrak:

1) Aurreko teoremarekiko lotura, honela adieraz daiteke: populazioaren elementu nagusiak, 1-ez, eta kontrakoak, 0-z seinalatzen baditugu, X aldagai binomiala, 1+0+0+1+1+...+1+0=X baturaren bidez adieraz daiteke, hots, $X=X_1+X_2+\ldots+X_n$ baturaren bidez, non $X_1,\ldots,X_n$, Bin(1,p) banaketari jarraitzen zaizkion zorizko aldagai askeak bait dira.

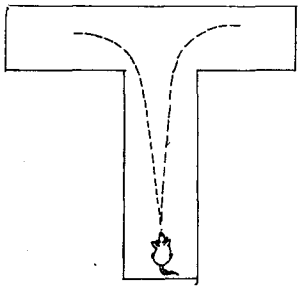
2) Gogora dezagun, X=Bin(n,p) baldin bada, orduan, E(X)=np eta Var(X)=npq direla. Beraz, $p \neq 0$ eta 1 baldintza eta $Var(X) \neq 0$ baldintza baliokideak dira.

3) n zenbat eta handiagoa izan eta p, $\frac{1}{2}$ -ren hurbilago egon, hurbilketa gero eta hobeagoa da.

($p=\frac{1}{2}$ -rako, banaketa binomiala guztiz simetrikoa da).

Praktikan, $n>30$ eta $0,1<p<0,9$ baldin badira, hurbilketa hau erabiltzen da. p-ren balio txiki zein handietarako ($p\leqslant 0,1$ ala $p\geqslant 0,9$), hobe da Poisson-en hurbilketa erabiltzea normalarena baino.

7.adibidea:



Ondoko jokabide-experientzia diseinatu da: irudiak adierazten duenez, labirintoa daukagu, non T-aren bi adarrak guztiz berdinak bait dira. Proba, arratoi batek ibilaldia egitean datza, T-aren adar bien artean, zein hartzen duen ikusirik.

Hasieran, bi adarrak berdinak direnez, arratoiak adarren bat bereziki nahiago izateko, inolako motibapenik ez duela suposatzen da, hots, "a priori", adarren batetara jotzeko probabilitatea $\frac{1}{2}$ dela kontsideratzen da.

Experientzia 1000 aldiz errepikatuko da, era independentean eta zirkunstantzia berberetan, bai 1000 arratoi erabiliz, zein probak behar den neurrian urrunduz, horrela arratoi berberak aurreko proba ahantz dezan.

"A priori" formulaturiko galdera teorikoa hauxe da: Zein da, T-aren eskuineko adarretik ibiltzen den aldi-kopurua, 490 eta 510 aldiren artean egoteko probabilitatea?

X="eskuineko adarretiko ibilaldi-kopurua" zorizko aldagaia, Bin(1000, $\frac{1}{2}$) banaketari darraio teorikoki. Baina, n aski handia eta $0,1<p<0,9$ direnez, X, $N(np, \sqrt{npq}) = N(500, 15,81)$ dela kontsidera daiteke.

Orain, ondoko tarteari dagokion probabilitatea bilatuz, arazoa ebatz dezakegu:

$$p(490<X<510) = p\left(\frac{490-500}{15,81} < \frac{X-500}{15,81} < \frac{510-500}{15,81}\right) =$$

$$= p(-0,6324<Z<0,6324) = 0,48$$

Alegia, zera espero dugu: arratoia eskuinerantz doaneko aldi-kopurua, 490 eta 510-artean egoteko probabilitatea 48%-takoa da.

Posible da espero dugun emaitza ez lortzea, pentsa bait liteke, ezen diseinuaren baldintzetan kontutan hartu ez dugun faktoreren batek, bai kanpokoak zein barnekoak, arratoiaren hobespena, bi adarretariko batetarantz joatekoa sorterazten duela (aurreko experientziak, usainak, zarataren batek, animaliaren ezaugarri optikoek, jaiotzetiko joerak eta abarrek).

Baina, planteiamenduaren interesa, hain zuzen ere, mota bereko proba desberdinetan lorturiko emaitzak eta itxaroten ditugunak konparatzeko ahalmenean datza, experimentatzailearen suposamenduak egokiak direnentz erabaki daitekeelarik.

PROBLEMAK

1.- Birusaren bizitza, ondoko dentsitate-funtzioa daukan zoriz ko aldagaia da:

$$f(x) = \begin{cases} 0 & \text{baldin } x < 1 \\ k/x^4 & \text{baldin } x \geq 1 \end{cases}$$

i) Kalkula ezazu, f(x) dentsitate-funtzioa izateko k-ren balioa.

ii) Zein da birusaren batezbesteko bizitza?

iii) Aurki ezazu banaketa-funtzioa.

iv) Kalkula itzazu mediana eta koartilak.

v) Kalkula ezazu desbidazio standarda.

vi) Aurki ezazu, zoriz aukeraturiko birusa, 5 ordu baino gehiago bizitzeko probabilitatea.

2.- X zorizko aldagaia, [a,b] tartean, banaketa uniformeari darraio.
Kalkula itzazu bere itxaropena eta desbidazio standarda.
Kalkula ezazu $p(m<X<n)$, probabilitate hau (n-m)-ren menpean soilik dagoela ikusirik.

3.- Zorizko aldagai jarraiaren dentsitate-funtzioa hauxe da:

$$f(x) = \begin{cases} ax^2+b & \text{baldin } x \in (0,2) \\ 0 & \text{baldin } x \notin (0,2) \end{cases}$$

$p(\frac{1}{2}<x<1) = 0,1357$ dela jakinez, kalkula itzazu a eta b.
Kalkula itzazu moda eta mediana, ea haien artekoren bat batezbestekoaren berdina den azterturik.

4.- Herri baten biztanlegoaren 40%-a, A odol-taldearen barnekoak direla aurkitu da. Lor ezazu, 200 pertsonatako komunitatean, hauetariko 90 baino gehiago, A taldearen barnekoak izateko probabilitatea.

5.- Istripua pairatzen duten 5 pertsonen artean, orokorki, emakume bat dagoela jakinik, kalkula itzazu, 200 kotxe-istripu tan:

a) Gizonak, 30%-a baino gutxiago izateko probabilitatea.

b) Gizonak, 80%-a baino gehiago izatekoa.

c) Gizonak, 40% eta 60%-ren artean egotekoa.

6.- Produktu baten lurrunketako pisu-galeren erregistroek, 6,45 gr.-tako batezbesteko galera eta 1,30 gr.-tako desbidazio standarda aurkezten dituzte. Erlo batetatik, bi pakete zoriz aukeratzen badira:

i) Zein da, biek 8gr. baino galera handiagoa aurkezteko probabilitatea?

ii) Bost pakete aukeratzerakoan, zein da haietariko batek, gutxienez, 8gr. baino galera handiagoa aurkezteko probabilitatea?

7.- Biztanlego baten pertsonen indize zefalikoa (beraren luzeraren portzentaia modura adierazitako burmuinaren zabalera), banaketa normalari darraiola suposatzen da. 58%-a, dolikozefaloak ($i<75$) dira; 38%-a, mesozefaloak ($75\leqslant i\leqslant 80$), eta, 4%-a, ekizefaloak ($i>80$).

Aurki itzazu, i-ren batezbestekoa eta desbidazio standarda.

8.- Biztanlego baten gizonen altueren banaketa, 170 cm.-tako batezbesteko eta 7 cm.-tako desbidazio standarda dituen banaketa normala da; emakumeen altuerena, 160 cm.-tako batezbesteko eta 6 cm.-tako desbidazio standarda dituen beste bana-

keta normala izanik. Emakume bat eta gizon bat aukeratzen dira zoriz. Zein da, gizona emakumea baino altuagoa izateko probabilitatea?

9.- Demagun, X = "animali mikroorganismoaren bizitza-aldia" zorizko aldagaia, ondoko dentsitate-funtzioak determinatuta dagoela:

$$f(x) = \begin{cases} 2e^{-2x} & \text{baldin } x \geq 0 \\ 0 & \text{baldin } x < 0 \end{cases}$$

Ondokoak eskatzen dira:

i) Kalkula ezazu X-en banaketa-funtzioa; eta aurki ezazu mikroorganismoa ordu bete baino gehiago bizitzeko probabilitatea.

ii) Zein da mikroorganismoarentzako itxarotako bizitza-aldia?

iii) Aurki ezazu x unea, non mikroorganismoaren bizitza-aldia x baino laburragoa izateko probabilitatea 0,99 den.

10.- Gaixotasuna sendatzeko behar den denbora, N(33,9) banaketari jarraitzen zaion zorizko aldagaia da.

i) Zein da, pertsona baten gaixotasunaren iraupena, desbidazio standarda baino laburragoa den balioan, batezbestekotik (edozein norantzatan) desbidatzeko probabilitatea?

ii) Gaixotasun honen iraupenari, X deitzen badiogu, kalkula ezazu, $p(X \leq t) = 0,2$ betetzen duen t denbora.

iii) Orain, gaixotasunaren C kostua (zorizkoa) kontsideratzen dugu, non

$$C = \begin{cases} 1000 & \text{baldin } X \leq t \\ 2000 & \text{" } \quad t < X \leq 42 \\ 5000 & \text{" } \quad X > 42 \end{cases}$$

den.

Kalkula ezazu, gaixotasunerako itxarotako kostua.

11.- Tchebycheff-en teoremaren aplikazioa, biztanlego baten gizabanakoen altueren bornapenean. Zehaztasunaren arazoa.

Herri baten adin militarreko gizonezkoen populazioaren gizabanakoen altuerari dagokion X zorizko aldagaiak, $E(X)=$ = 175 cm.-tako itxaropen matematikoa eta σ_x = 7,5 cm.-tako desbidazio standarda dauzka.

Behin batez, premia militarrek, gizonezko gertuen 95%-aren bilketa eskatzen dute. Portzentaia hau lortzeko erizpide bat, batezbestekotik $\delta > 0$ kantitatean diferitzen ez den altuera daukatenen mobilizapenean datza, $E(X)-\delta$ eta $E(X)+\delta$ -en artean dauden altueretako gizabanakoen proportzioa 95%-a delarik.

Tchebycheff-en teoremaren bidez, egiazta itzazu δ =33,5 dela, eta, beraz, bilatzen dugun tartea (141,5, 208,5) dela.

Horrela, 95%-aren mobilizapena aseguratuko litzateke gutxienez, baina, seguraski, ez litzateke errentagarria izango, zeren behar direnak baino gizonezko gehiago mobilizatuko bait lirateke.

Orain, X zorizko aldagaia banaketa normalari jarraitzen zaiola suposatuz, egiazta ezazu, zuzenki, bilatzen dugun tartea (160,3, 189,7) dela.

Konpara itzazu emaitzak.

12.- Komunikapen-arazo bat.

A telefonetxeak, B telefonetxeko 2000 abonaturi ematen die zerbitzua. A-tik B-rako 2000 hari tronkal jartzea, kostu handikoa eta premiagabekoa izango litzateke. Baldintza arrunten pean, 100 deiren arteko batek soilik, hari tronkal gerturik ez aurkitzeko N hari-kopuru bezain handia jartzea nahikoa da.

Demagun, eguneko ordurik beteenean (baldintzarik kalte garrienetan, behintzat), abonatu bakoitzak B-ren hari tronkal bat eskatzen duela, 2 minututako batezbestekoaz.

Beteriko orduaren une fixu batetan, egoera hau eta 2000 proba, non bakoitzak haria eskatzeko $p = \frac{1}{30}$ probabilitatea bait dauka konparatzen ditugu. Baldintza arrunten pean, proba hauek askeak direla suposa daiteke. (Nahiz eta honek bat bateko zaparradak edo lurrikarak bezalako gertaeretarako balio ez izan, zeren eta honelako gertaerek pertsona askori telefonoa jo erazten bait diote; kasu honetan, ezin daiteke teoria aplika, "hariak" asetzen dira eta).

Orduan, $p = \frac{1}{30}$ -eko Bernouilliren 2000 proba dauzkagu eta N zenbakirik txikiena eskatzen da, non N arrakasta baino gehiago izateko probabilitatea 0,01 baino txikiagoa bait da.

Alegia, X = "arrakasta-kopurua" zorizko aldagaia, $Bin(2000, \frac{1}{30})$ banaketari jarraitzen zaio. Matematikoki, horrela formulatzen da galdera:

$$p(X > N) < 0{,}01$$

Hurbilketa normalak N = 86 balioa ematen digula eta, A eta B lotzen duten 86 hari tronkal jartzea aski izango da.

13.- Lore arrosak dauzkaten "gau-dama" izeneko bi landare (heterozigotikoak) gurutzatzen dira. Biz A ondoko gertaera: lorturiko landareak lore zuriak ditu.

Kolore zuria errezesiboa dela jakinik, kalkula itzazu:

i) 10 gurutzamendutan, lore zuriko bi landare gehienez lortzeko probabilitatea.

ii) 768 gurutzamendutan, lore zuriko 204 landare gutxienez lortzekoa.

14.- Fase epidemiologikoan, 2 urte baino ume gazteagoengan, sepsi meningokozikoaren eraginak, 0,2%-tako balioa dauka.

2 urte baino gazteagoak diren 200 umetako biztanlegoan, zeintzu dira:

a) gaixotasuna, 4 umek pairatzeko probabilitatea.

b) 4 edo 4 umek baino gutxiago pairatzekoa.

15.- Intsektu-mota baten luzera, ondoko dentsitate-funtzioa daukan zorizko aldagaia da:

$$f(x) = \begin{cases} 20x^3(1-x) & \text{baldin } 0 \leqslant x \leqslant 1 \\ 0 & \text{gaineratikoetan} \end{cases}$$

i) Kalkula itzazu banaketa-funtzioa eta $p(X \leqslant 2)$

ii) Kalkula ezazu $p(X \leqslant \frac{1}{2} \ / \ \frac{1}{4} < X < \frac{3}{4})$

iii) Mota honetako 25 intsektutako lagina daukagu. 5 intsektuk, gutxienez, 0,9 baino luzera handiagoa baldin badaukate, lagina aztertzea interesatuko zaigu. Zein da lagina aztertzeko probabilitatea?

16.- Telefonetxe batek 450 lotura jar ditzake minutu batetan. Minutuko eskaturiko lotura-kopurua, Poisson-en banaketari darraio, $\lambda = 400$ izanik.

Minutu batetan, zein da telefonetxea guztiz beteta egoteko probabilitatea?

17.- Balantzaren pisaldietan gertatzen diren zorizko erroreak, $N(0, \frac{1}{2})$ banaketaren arauera banatzen dira. Ondokoak eskatzen dira:

i) Pisaldi batetan egindako errore maximoa, 0,99-tako probabilitatearekin.

ii) Pisaldi-kopuru maximoa, batezbesteko errorea 0,1 paino txikiagoa dela aseguratzeko, 0,95-etako probabilitatea rekin.

18.- Toxikotasun-azterketetan, aztertzen den alearen heriotza eragiten duen substantzia baten kantitate minimoari, dosi hilgarri minimoa deritzogu.

Suposa dezagun, arratoi-populazio batetan, substantziaren dosi hilgarri minimoa $N(\mu, \sigma)$ banaketari jarraitzen zaiola.

Determina ezazu, $\mu \pm \gamma\sigma$ motako tartea (hots, determina ezazu γ), non substantziak eragindako arratoien 50%-ak, dosi hilgarri minimoa bait dauka beraren barruan).

19.- Banaketa moztuak.

Biz X, R populazioaren A ezaugarria deskribatzen duen zorizko aldagai jarraia, $F(x) = p(X \leqslant x)$ bere banaketa-funtzioa eta f(x) bere dentsitate-funtzioa izanik.

Suposa dezagun, A-ren balioak x_o kantitatea gainditzen duteneko elementuek osoturiko R-ren azpipopulazioan, A-ren jokabidea deskribatu nahi dugula.

Azpimultzoari dagokion zorizko aldagai berriaren banaketa-funtzioa hauxe izango da:

$$G(x) = p(X \leqslant x / X > x_o)$$

Froga ezazu G(x) delakoak, $G(x) = \frac{F(x)-F(x_o)}{1-F(x_o)}$ adieraz pena onartzen duela, eta, beraz, zorizko aldagai honen (zeinaren heina, $[x_o, +\infty)$ tartea den) dentsitate-funtzioa

$$g(x) = f(x) \frac{1}{1-F(x_o)} \quad \text{dela.}$$

$F(x_o)$ balioari, mozketa-maila deritzogu.

Aurreko metodoa, beraren modukoak diren ondoko kasuetan aplikatzen da:

i) Demagun, giza-organismoaren albumina-maila N(130gr., $15gr^2$.) banaketa normalari jarraitzen zaiola eta 125 gr. baino albumina-maila handiagoa daukaten pertsonen taldea aztertu nahi dugula.

ii) Bakterio-mota baten bizitza-aldia, $\beta = \frac{1}{30}$ parametro banaketa exponentzialari jarraitzen zaio, eta, orain dela 10 egun, beraien kultibo bat daukagu.

Halaber, goi-mozketa edo mozketa bikoitzeko kasuak kontsidera daitezke.

20.- Arazo irradiaktiboa

Substantzia irradiaktiboaren X bizitza-aldia, β parametroko banaketa exponentzialaren bidez ondo deskribatzen den zorizko aldagaia da. Substantzia-mota bakoitzerako, β-ren balio berezia dago, eta deskribapen-arazoa (iraupen denboralaren ikuspuntutik), β-ren determinazioan datza hain zuzen ere.

X aldagai jarrairako, mediana, $p(X \leqslant x) = \frac{1}{2}$ deneko x balioa da, alegia, $F(x) = \frac{1}{2}$ ekuazioaren soluzioa. Froga ezazu, banaketa exponentzialerako medianaren balioa $t = \frac{\ln 2}{\beta}$ dela.

Fisikariek, erdidesintegrazio-aldia ("batezbesteko bizitza" izenekoa ere), atomoen erdia desintegratzeko behar den alditzat definitzen dute. Kontutan har ezazu, "batezbesteko bizitza" medianaren interpretazio frekuentziala dela. Beraz, experimentalki, substantziaren "batezbesteko bizitza" determinatzen baldin badugu, orduan, β parametroaren balioa, $\frac{\ln 2}{\text{"batezbesteko bizitza"}}$ izango da gutxi gora behera.

Hala ere, ez du ematen erabide honi jarraitzea praktikoa denik (esate baterako $_6C^{14}$ karbono irradiaktiboaren desintegrazio-aldia 5770 urtekoa da eta $_{92}U^{238} \longrightarrow {}_{92}Th^{234} + {}_2He^4$ uranioaren desintegrazio-prozesuak $4{,}5 \cdot 10^9$ urteko "batez-besteko bizitza" dauka).

Soluziora hurbiltzeko era, desintegrazio-abiaduraren kontzeptua sartzea da:

Bira N_o hasieran dagoen atomo-kopurua, eta N(t), t unean dagoen atomo-kopurua. Probabilitatearen interpretazio frekuentzialaren arauera,

$\frac{N(t)}{N_o} \simeq p(X > t) = 1 - F(t)$ da.

$\frac{N(t+\Delta t)-N(t)}{\Delta t} = N_o \frac{F(t)-F(t+\Delta t)}{\Delta t}$ dela erraz ikusten da.

Eta limitera pasatuz:

$\frac{dN(t)}{dt} = -N_o f(t)$, non $f(t) = \beta e^{-\beta t}$ X-en dentsitate-funtzioa den.

$\frac{dN(t)}{dt}$ balioa, t uneko desintegrazio-abiadura da, eta, halaber, ondokoa erraz froga daiteke:

$$\frac{dN(t)}{dt} = -N_o \beta p(X > t) = -\beta N(t) \quad (*)$$

Alegia, une bakoitzeko desintegrazio-abiadura, dagoen atomo-kopuruaren proportzionala da, proportzionaltasun-konstantea β izanik hain zuzen.

(*) ekuazio diferentziala, edo beste era batetara, $\frac{dN(t)}{N(t)} = -\beta\, dt$ integra daiteke, $\ln\{N(t)\} = -\beta t + k$ lorturik.

Beraz, $\ln\{N(t)\}$ delakoaren grafikoa, t-ren funtzioan, zuzena da, non bere maldak, β ezagutzea baimentzen bait digu.

21.- Gamma banaketa

a) Alde teorikoak:

Analisian, $\Gamma(p)$ gamma funtzioa, $\Gamma(p) = \int_0^\infty e^{-x}x^{p-1}dx$ integralaren bidez definitzen da, $p>0$-rako integrala konbergentea izanik.

$p \geqslant 1$ suposaturik, zatikako integrazioaren bidez,
$\Gamma(p) = (p-1).\Gamma(p-1)$ lortzen da.

p osoa bada, prozesua errepika daiteke ondokoa lortu arte:

$\Gamma(p) = (p-1)(p-2)...2\cdot\Gamma(1) = (p-1)!$ zeren $\Gamma(1)= \int_0^\infty e^{-x}dx = 1$ bait da.

$\Gamma(p) = \int_0^\infty e^{-x}x^{p-1}dx$ integralean, $x = ay$ $(a>0)$ motako aldagai-aldaketa eginez, $\Gamma(p) = a^p \int_0^\infty e^{-ay}y^{p-1}dy$ lortzen da.

Beraz,
$f_1(x) = \frac{e^{-x}\,x^{p-1}}{\Gamma(p)}$ $(p>0)$ eta $f_2(x) = \frac{a^p}{\Gamma(p)}\,e^{-ax}\,x^{p-1}$ $(a>0, p>0)$ funtzioak, balio positiboetarako definiturik, dentsitate-funtzioak dira.

Zorizko aldagaia, beraren dentsitate-funtzioa $f_1(x)$ baldin bada, $\gamma(p)$ p parametroko gamma banaketari jarraitzen zaio; eta, $f_2(x)$ baldin bada, $\gamma(a,p)$ a eta p parametroetako gamma banaketari.

Kontutan har dezagun, $p = r$ (zenbaki arrunta) baldin bada,

$f_1(x) = \frac{e^{-x}\,x^{r-1}}{(r-1)!}$ eta $f_2(x) = \frac{a^r}{(r-1)!}\,e^{-ax}\,x^{r-1}$ direla, beraz, $\gamma(a,1)$ banaketa, banaketa exponentziala izanik.

b) Gamma banaketaren interpretazioa:

Demagun, denboran zehar ematen diren $t_1, t_2, \ldots$ uneetan jazotzen diren zorizko gertaeren segida. Adibidez, substan

tzia irradiaktiboen desintegrazioak, jaiotzak, hilketak, Geiger-Muller kontagailuaren gaineko partikulen kontaketa eta abar.

Edozein t denbora-tarte fixutan, k gertaera (esate baterako, jaiotza) jazotzeko probabilitateaz galdetzen diogu gure buruari.

Probabilitate honek, $p_k(t) = \frac{(\omega t)^k}{k!} e^{-\omega t}$ (*) balio du, prozesuak ondoko baldintzak betetzen baldin baditu:

1) Δt denbora-tarte laburrean, gertaera bakar bat jazotzeko probabilitatea, $\omega \Delta t$ ordenakoa da, hots, tartearen luzeraren proportzionala.

2) Δt denbora-tarte laburrean, bi gertaera edo bi baino gehiago jazotzeko probabilitatea, nulua da praktikoki.

3) bi denbora-tarte baztertzailetan jazotzen diren gertaera-kopuruak, zorizko aldagai askeak dira.

4) denboran zehar, $p_k(t)$ probabilitateak berdinak dira; hots, (t_1,t_2), (t_3,t_4) bi tarteen probabilitateak berdinak dira, baldin $t_2-t_1 = t_4-t_3$ bada.

Frogapena nabaria ez denez, lehen irakurketan utziko dugu, baina, (*) formula berez garrantzitsua dela bistakoa da.

r.gertaera jazo arte iraganiko itxarote-denboraren probabilitate-legeaz galdetzen diogu gure buruari.

Biz $F_r(t)$, t.gertaera jazo arte iraganiko denbora, t baino laburragoa edo berdina izateko probabilitatea. Horrela, $1-F_r(t)$, itxarote-denbora t baino luzeagoa izateko probabilitatea izango da; baina, hau da t denboran agertzen den gertaera-kopurua, r baino txikiagoa izateko probabilitatea; hots,

$$1 - F_r(t) = \sum_{k=0}^{r-1} \frac{(\omega t)^k}{k!} e^{-\omega t}$$

Adierazpen hau deribatuz, ondoko dentsitate-funtzioa lortzen dugu:

$$f_r(t) = \frac{\omega^r}{(r-1)!} e^{-\omega t} t^{r-1}$$

Alegia, $\gamma(\omega,r)$ banaketaren dentsitate-funtzioa.

Partikularrean, r = 1-erako, gertaera bakar bat jazo arte itxarote-denbora, banaketa exponentzialari jarraitzen zaion zorizko aldagaia da.

V. GAIA: OINARRIZKO LAGIN-TEORIARAKO SARRERA.

V.1.- Oinarrizko lagin-teoriari buruzko orokortasunak.

V.1.1.- Laginketa-motak.

V.1.2.- Lagin-banaketaren kontzeptua.

V.2.- $\bar{x}$ laginaren batezbestekoaren banaketa.

V.3.- Bi laginen batezbestekoen arteko kenduraren banaketa.

V.3.1.- Normalki banaturiko populazioan egindako laginketa.

V.3.2.- Normalki ez banaturiko populazioan egindako laginketa.

V.4.- Laginaren proportzioaren banaketa.

V.5.- Bi laginen proportzioen arteko kenduraren banaketa.

V.6.- χ^2, Pearson-en ji karratu banaketa.

V.6.1.- Definizioa.

V.6.2.- Baturaren teorema.

V.6.3.- Kuasibariantzaren banaketa.

V.6.4.- Ji karratu banaketaren beste aplikazio bat.

V.7.- Student-en "t" banaketa.

V.7.1.- Definizioa.

V.7.2.- "t" banaketa eta batezbestekoen arteko kendura.

V.8.- Fisher-Snedecor-en "F" banaketa.

V.8.1.- Definizioa.

V.8.2.- Kuasibariantzen arteko zatiduraren banaketa.

OINARRIZKO LAGIN-TEORIARAKO SARRERA

V.1.- OINARRIZKO LAGIN-TEORIARI BURUZKO OROKORTASUNAK.

Esan genuenez, Estatistika induktiboa, Probabilitate-teorian oinarriturik, ondoko helburuak dauzkan metodo-multzoa da:

a) Emaitzetako inferentziak edo orokorpenak egitea, beraien konfidantza-maila neurturik.

b) Emaitzak interpretatzea.

Orokortu, haien konfidantza-maila neurtu eta interpretatu nahi ditugun emaitzak dira populaziotik ateratako lagin baten edo batzuren azterketaren bidez lortzen direnak.

1.adibidea:

Suposa dezagun, medikamendu baten kalitatea kontrolatu nahi dugula, A osagaiaren edukina aztertuz.

Medikamenduaren eguneroko ekoizpena 10000 unitatekoa da.

Kontrolatzeko metodorik seguruena, haiek guztiak aztertzea, eta bakoitzaren A osagaiaren edukina neurtzea izango litzateke. Baina, operazio hau materialki ezinezkoa denez, hauxe da egingo duguna: eguneroko ekoizpenetik lagin bat aukeratuko dugu zoriz, eta A osagaiaren edukina neurtzeko behar diren kontrolak, bere gainean burutuko ditugu.

Geroago, metodo estatistikoen bidez, emaitza hauek orokortzen, eta, behar den konfidantza-mailaz, zein den medikamenduaren A osagaiaren edukina finkatzen saiatuko gara.

V.1.1.- Laginketa-motak.

Orokorrean, bi laginketa-mota daude: laginketa probabilistikoa eta ez-probabilistikoa. Inferentziak egitea baimentzen diguten metodo estatistiko seguruak lehen laginketa-motarako soilik daudenez, beraren azterketa egingo dugu.

Definizioa: Populaziotik ateratako lagina, lagin probabilistikoa da, baldin eta laginean sartzeko populazioaren elementu guztien probabilitatea ezaguna bada.

Gehiago sinplifikatuz, laginketa probabilistikoaren kasu partikular bakar bat aztertuko dugu: Zorizko laginketa bakuna.

Definizioa: Populazioaren elementu guztiek, aukeratuak izateko probabilitate berdina baldin badaukate, zorizkoa da lagina. Gainera, laginketa baliagarria izateko, populazio osoaren adierazgarria izan behar du laginak.

2.adibidea:

18 eta 24 urteren artean dauden pertsonen altuerari buruzko azterketa egin nahi badugu, lagin egokia hartu beharrean gaude.

Esate baterako, ikasten ari diren pertsonek eta soilik haiek osoturiko lagina hartzen badugu, errealitatetik urrun gertatuko da azterketa, laginak inolako adierazgarritasunik ez bait du.

Laginaren aleak aukeratzeko beste metodo bat, 6 letratako eta P-tik hasitako lehendabiziko deitura duten pertsonak soilik aukeratzea izango litzateke; baina, lagina ez litzateke adierazgarria izango, zeren eta deiturak luzeak izateko edo P-tik hasteko joera, seguraski, herrialde batzuetan besteetan baino handiagoa izango bait da.

Zorizko lagin adierazgarria lortzeko metodo seguruenetariko bat, zorizko zenbakien taulen erabilpenean datza.

3.adibidea.

Suposa dezagun, 1.adibideko medikamenduaren kontrolean, 25 aletako lagin adierazgarria hartu nahi dugula. Aukerapen-metodoa ondokoan datza:

Ekoitzitako 10.000 medikamenduak zenbatzen dira (jeneralki, beren fabrikapen-zenbakia ba daukate) eta zorizko zenbakien taularen bidez, bertatik, 25 zenbaki ateratzen dira, zenbaki haiei dagozkien medikamenduak harturik.

V.1.2.- Lagin-banaketaren kontzeptua.

Gai honetan erabiliko dugun kontzeptu nagusia, lagin-banaketarena da.

Definizioa: Populaziotik zoriz ateratako tamainu berdineko laginetatik abiatuz, estatigrafoak har ditzakeen balio posible guztien banaketa, estatigrafoaren lagin-banaketa deitzen da.

4.adibidea:

Berriro ere, suposa dezagun medikamenduaren A osagaiaren edukina kontrolatu nahi dugula, laginaren elementuak aukeratzeko metodoa ba daukagularik.

A osagaiaren edukina, maiztasun-banaketa izango duen aldagaia da.

Baina, $\bar{x} = \frac{x_1+\ldots+x_{25}}{25}$ laginaren batezbestekoa ere, aldagaia izango da, zeren eta 25 elementutako lehendabiziko lagina hartzen badugu, batezbestekoaren $\bar{x}_1$ balio konkretua lortuko bait dugu; baina, beste 25 elementutako lagin bat hartuz, batezbestekoaren $\bar{x}_2$ beste balio desberdina lortuko dugu; lagin

posible guztiak ($C_{10.000,25}$ lagin desberdin) atereaz eta beraiei dagozkien batezbestekoak aurkituz, batezbestekoaren lagin-banaketa deskriba dezakegularik.

Populazioaren tamainua txikia ez denean, estatigrafoaren lagin-banaketaren eraiketa honi dagokion lana, itzelezkoa da, eta, infinitua denean, ezinezkoa. Baina, lagin-banaketak matematikoki ondoriozta daitezke, gure helburua beraiei buruzko on doko hiru ezaugarriak ezagutzea izanik: batezbestekoa, bariantza eta forma funtzionala.

V.2.- $\bar{x}$ LAGINAREN BATEZBESTEKOAREN BANAKETA.

5.adibidea:

Demagun, (6,8,10,12,14) ospitale batetan dauden umeen adinek osoturiko 5 tamainuko populazioa dugula. Hurrengo metodoen bidez, azter dezagun 2 tamainuko laginen batezbestekoen lagin-banaketa:

a) Itzuleraz:

Lagin posibleak

(6,6)	(6,8)	(6,10)	(6,12)	(6,14)
(8,6)	(8,8)	(8,10)	(8,12)	(8,14)
(10,6)	(10,8)	(10,10)	(10,12)	(10,14)
(12,6)	(12,8)	(12,10)	(12,12)	(12,14)
(14,6)	(14,8)	(14,10)	(14,12)	(14,14)

Laginen batezbestekoak

6	7	8	9	10
7	8	9	10	11
8	9	10	11	12
9	10	11	12	13
10	11	12	13	14

b) Itzulerarik gabe:

Lagin posibleak

(6,8)	(6,10)	(6,12)	(6,14)
(8,6)	(8,10)	(8,12)	(8,14)
(10,6)	(10,8)	(10,12)	(10,14)
(12,6)	(12,8)	(12,10)	(12,14)
(14,6)	(14,8)	(14,10)	(14,12)

Laginen batezbestekoak

7	8	9	10
7	9	10	11
8	9	11	12
9	10	11	13
10	11	12	13

Beraz, ondokoak dira batezbestekoaren lagin-banaketak:

a) Itzuleraz:

$\bar{x}$	6	7	8	9	10	11	12	13	14
f	1	2	3	4	5	4	3	2	1

$$E(\bar{x}) = \frac{250}{25} = 10 \qquad V(\bar{x}) = 4$$

b) Itzulerarik gabe:

$\bar{x}$	7	8	9	10	11	12	13
f	2	2	4	4	4	2	2

$$E(\bar{x}) = \frac{200}{20} = 10 \qquad Var(\bar{x}) = 3$$

Populazioak, $\mu = 10$ batezbestekoa eta $\sigma^2 = 8$ bariantza dauzkala kontutan hartzen badugu, hurrengo emaitzak egiaztatzen ditugu:

a) Populazio finituan, itzulerazko laginketa egiten dugunean, batezbestekoaren lagin-banaketak, $E(\bar{x}) = E(X) = \mu$ eta $Var(\bar{x}) = \sigma^2/n$ dauzka.

b) Populazio finituan, itzuleragabeko laginketa egiten dugunean, batezbestekoaren lagin-banaketak $E(\bar{x}) = E(X) = \mu$ eta $Var(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$ dauzka.

Egiaztapena:

a) Itzuleraz:

$E(\bar{x}) = 10 = \mu$

$Var(\bar{x}) = 4 = \frac{8}{2} = \frac{\sigma^2}{n}$

b) Itzulerarik gabe

$E(\bar{x}) = 10 = \mu$

$Var(\bar{x}) = 3 = \frac{8}{2} \cdot \frac{5-2}{5-1} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$

Batezbestekoaren ordez, aldagaia ere izango den laginaren balioen arteko T baturaren ezaugarriak kalkulatzea interesatzen bazaigu, erraz egingo dugu, zeren eta

$\bar{x} = \frac{x_1+\ldots+x_n}{n}$ berdintzak $T = x_1+\ldots+x_n = n\cdot\bar{x}$ dela halabehartzen duenez, $E(X) = n\mu$ eta $Var(T) = n^2\cdot\frac{\sigma^2}{n} = n\cdot\sigma^2$ izango bait dira.

Adibide gisa, lagin-banaketen zenbait ezaugarri ikusi ditugu, baina berauek bezain garrantzitsua da lagin-banaketaren forma funtzionala.

Bi egoera desberdin bereiztuko ditugu:

1) <u>Normalki banaturiko populaziotik</u> abiaturiko laginketa.

2) <u>Normalki ez banaturiko populaziotik</u> abiaturiko laginketa.

Batezbestekoaren lagin-banaketaren adibideari jarraituz, laginketa, normalki banaturiko populazioan egiten bada, orduan, $\bar{x}$-ren banaketa, μ batezbestekoa eta σ^2/n bariantza dituen banaketa normala dela frogatzen da matematikoki.

Laginketa, normalki ez banaturiko populazioan burutzen bada, limitearen teorema zentrala erabiliz, ondokoa baiezta dezakegu: Edozein forma funtzionaletako populazioa emanik, non μ batezbestekoa eta σ^2 bariantza finituak diren, populazio honen n tamainuko laginetatik abiatuz kalkulaturiko <u>$\bar{x}$-ren lagin-banaketa</u>, laginaren tamainua handia denean, μ batezbestekoa eta $\frac{\sigma^2}{n}$ bariantzako banaketa normalaren moduan egongo da banatuta gutxi gora behera.

Alegia, laginaren batezbestekoaren kasuan, lagin-banaketa, gutxi gora behera, banaketa normala izango deneko segurtasuna dago, baldin eta:

a) laginketa, normalki banaturiko populazioan burutzen bada, edo,

b) laginketa, normalki ez banaturiko populazioan edo forma funtzional ezezagunekoan burutzen bada, baina laginaren tamainua handia bada.

Puntu honetan sortzen den galdera logikoa hauxe da: "lagina handia dela", noiz esango dugu? Zabal-zabalik erabilitako arau enpirikoak, egoera praktiko gehienetan, 25 edo 25 baino handiagoa den laginaren tamainua nahikoa dela adierazten du, nahiz eta laginaren tamainua emendatuz baldin bagoaz, hurbilketa, jakina, hobetuz joan.

Orain adierazi ditugun emaitza hauek egiazkoak dira, laginketa itzuleraz burutzen baldin bada, edo, alegia, laginak populazio bukaezinetatik ateratzen baldin badira. Hau da, populazioa normalki banatzen dela esaterakoan suposatzen duguna.

Baina, laginketa populazio finituan eta itzulerarik gabe egiten denean, $\bar{x}$-ren lagin-banaketak μ batezbestekoa eta $\frac{\sigma^2}{n}\frac{N-n}{N-1}$ bariantza izango ditu. $\frac{N-n}{N-1}$ faktoreari "populazio finitugaitiko zuzenketa" deritzogu, laginaren tamainua, populazioarekin konparatuz txikia denean arbuia dezakegularik, zeren $\frac{N-n}{N-1} \simeq 1$ bait da.

"Populazio finitugaitiko zuzenketa" erabiliko da, baldin $\frac{n}{N} > 0,05$ bada.

Jakina, laginaren tamainua handia baldin bada, batezbestekoaren lagin-banaketa banaketa normala izango da gutxi gora behera.

6.adibidea:

Pertsonen seroaren Fe-ren edukinaren batezbestekoa eta desbidazio standarda, 120 eta 15 mikrogramo 100 ml-ko baldin

badira hurrenez hurren, zein da 50 pertsonatako laginak, 115 eta 125-en arteko batezbestekoa izateko probabilitatea?

Soluzioa:

Populazioa: Batezbestekoa $\mu = 120$
Desbidazio standarda $\sigma = 15$
Forma funtzional ezezaguna
Lagina: $n=50$

$\Rightarrow \bar{x}$ banaketa, $N(120, \frac{15}{50})$ gutxi gora behera.

Beraz,

$$p(115<\bar{x}<125) = p\left(\frac{115-120}{15/\sqrt{50}} < \frac{\bar{x}-120}{15/\sqrt{50}} < \frac{125-120}{15/\sqrt{50}}\right) =$$

$$= p(-2,36<Z<2,36) = 0,9818$$

7.adibidea:

Asignatura batetan matrikulaturiko 200 ikasleren batezbesteko kalifikazioa 5,7 da, bariantza 5,97 izanik. Populazio honetatik, 50 ikasle desberdin zoriz hartzen baldin baditugu, zein da lagin honen batezbesteko kalifikazioa 6 baino txikiagoa izateko probabilitatea?

Populazio finitua: $N=200$
Laginaren tamainua: $n=50$
Itzuleragabeko laginketa
$\frac{n}{N} = 0,25 > 0,05$

$\Rightarrow$ $E(\bar{x})=5'7$; $Var(\bar{x}) = \frac{\sigma^2(N-n)}{n(N-1)} = 0'09$
$\bar{x}$-ren banaketa, $N(5'7, 0'3)$ da gutxi gora behera.

Beraz,

$$p(\bar{x}<6) = p\left(\frac{\bar{x}-5'7}{0'3} < \frac{6-5'7}{0'3}\right) = p(Z<1) = 1-p(Z\geqslant 1) =$$

$$= 1-0,1587 = 0,8413$$

Oharra:

Maiz, laginketa burutzerakoan, populazioaren banaketa, gutxi gora behera normala dela susma daiteke, baina bere bariantza jakin gabe. Kasu honetan, emaitza interesgarriren bat lortu ahal izateko, bariantzaren balioa estimatu beharko dugu, eta hori,

$s^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$ laginaren kuasibariantzaren bidez lortzen da, aztertzen dugun lagina aski handia baldin bada.

Alegia, ondokoak bereiztu behar ditugu:

Populazioaren bariantza: $\sigma^2 = E\,[(X-E(X))^2]$

Laginaren bariantza: $s'^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n}$, $\{x_1,\ldots,x_n\}$ lagina izanik.

Laginaren kuasibariantza: $s^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$, $\{x_1,\ldots,x_n\}$ lagina izanik.

Eta baieztatzen duguna hauxe da:

$n \geqslant 20$ baldin bada, $s^2 \simeq \sigma^2$ (Are zehatzago, $E(s^2) = \sigma^2$)

8.adibidea:

Merkatuan dagoen ardo-marka batetako alkoholaren graduazio normalki banatuta dagoela ba dakigu, 12 batezbestekoa duelarik, baina desbidazio standarda ezezaguna da.

500 elementutako lagina hartzerakoan, hurrengo banaketa enpirikoa lortu baldin badugu, estima ezazu desbidazio standardaren balioa:

Graduak	Maiztasunak	x_i	z_i	$z_i f_i$	$z_i^2 f_i$
10,8-11,2	3	11	-2	-6	12
11,2-11,6	6	11,4	-1	-6	6
11,6-12	18	11,8	0	0	0
12 -12,4	16	12,2	1	16	16
12,4-12,8	6	12,6	2	12	24
12,8-13,2	1	13	3	3	9
	50			19	67

$$z_i = \frac{x_i - 11,8}{0,4}$$

$$\bar{z} = \frac{19}{50} = 0,38 = \frac{\bar{x}-11,8}{0,4} \Longrightarrow \bar{x} = 11,8+0,38\cdot 0,4 = 11,452$$

$$s_z'^2 = \frac{67}{50} - 0,38^2 = 1,1956 = \frac{s_x'^2}{0,4^2} \Longrightarrow s_x'^2 = 0,4^2 \cdot 1,1956 = 0,191296$$

$$s^2 = \frac{n}{n-1} s'^2 = \frac{50}{49} \cdot 0,191296 = 0,1952 \Longrightarrow s = \sqrt{0,1952} = 0,44$$

Beraz, normalki banaturiko populazioaren bariantza ezezaguna denean, laginaren tamainua $n>20$ izanik, $\bar{x} \approx N(\mu, \frac{s}{\sqrt{n}})$ dela ondorioztatzen da.

V.3.- BI LAGINEN BATEZBESTEKOEN ARTEKO KENDURAREN BANAKETA.

V.3.1.- Normalki banaturiko populazioan egindako laginketa.

9.adibidea:

Demagun ondoko bi pertsona-taldeak: Batak (I taldeak), adimen-atzerapenarekin lotuta dagoen afekzioa pairatu du, eta besteak (II taldeak), ordea, ez du afekzio hori pairatu.

Talde bakoitzean, adimen-koefizientea normalki banaturik dagoela suposatzen da, desbidazio standarda 20 izanik.

Har dezagun talde bakoitzeko 15 pertsonatako lagin bat. Bakoitzerako, batezbesteko adimen-koefizientea kalkulatzen da, $\bar{x}_1=92$ eta $\bar{x}_2=105$ lorturik.

Beren egiazko batezbesteko adimen-koefizienteen arteko desberdintasunik ez badago ($\mu_1=\mu_2$), zein da laginen batezbestekoen arteko desberdintasuna hain handia izateko probabilitatea?

Galdera honi erantzuteko, $\bar{x}_1-\bar{x}_2$ laginen batezbestekoen arteko kenduraren lagin-banaketa ezagutu behar dugu. Praktikan horrela egin ez arren, banaketa hori lortzeko ideia kontzeptuala ondokoa da:

I taldetik, 15 tamainuko lagin guztiak hartzen dira, lagin bakoitzaren batezbestekoak kalkulaturik. II taldean gauza bera egiten da. Orduan, laginen batezbestekoen bikote posible guztiak, I taldeko bat eta II taldeko beste bat hartuz eratzen dira, beren kendurak kalkulaturik. Horrela, $\bar{x}_1-\bar{x}_2$ delakoaren banaketa lortuko dugu. Prozedura luzeegi izango da, lorturiko banaketa, $\mu_1-\mu_2$ batezbesteko eta $\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$ bariantzako banaketa normalari jarraitzen zaiolarik.

Hau, $\bar{x}_1=N(\mu_1,\frac{\sigma_1}{\sqrt{n_1}})$ eta $\bar{x}_2=N(\mu_2,\frac{\sigma_2}{\sqrt{n_2}})$ izatetik ondorioztatzen da:

Alde batetik, $\bar{x}_1-\bar{x}_2$ estatistikoa ere banaketa normalari jarraituko zaio gutxi gora behera; eta beste aldetik,

$$E(\bar{x}_1-\bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1-\mu_2$$

$$Var(\bar{x}_1-\bar{x}_2) = Var(\bar{x}_1)+Var(\bar{x}_2) = \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

Beraz, $\bar{x}_1-\bar{x}_2$ estatistikoa, $N(\mu_1-\mu_2, \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})$ banaketari darraio.

Egiten ari garen adibiderako banaketa, 0 batezbesteko eta $\frac{20^2}{15} + \frac{20^2}{15} = 53,3$ bariantzako banaketa normala da. Hots, $\bar{x}_1-\bar{x}_2$ delakoa N(0,7'3) banaketari darraio.

Emandako laginerako, $\bar{x}_1-\bar{x}_2=92-105=-13$ da, eskaturiko galderaren erantzuna $p(\bar{x}_1-\bar{x}_2<-13)$ delarik.

Kasu honetan, $\bar{x}_1-\bar{x}_2$ aldagaia tipifikatuz, $\frac{(\bar{x}_1-\bar{x}_2)-0}{7,3}$, N(0,1) banaketari darraiola ikusten dugu, taulak kontsultatuz, ondokoa lortzen dugularik:

$$p(\bar{x}_1-\bar{x}_2<-13) = p(Z<\frac{-13}{7,3}) = p(Z<-1,78)=p(Z>1,78)=0,0375$$

V.3.2.- Normalki ez banaturiko populazioan egindako laginketa.

Sarritan, banaketa normalari jarraitzen ez zaizkion populazioetan edo beren banaketa ezagutzen ez dugunekoetan burutzen dugu laginketa.

Arazo honen soluzioa, lagin handiak hartzean datza, zeren kasu honetan, $n_1,n_2 \geqslant 25$ direnean, limitearen teorema zentrala aplikatuz, hauxe lortzen bait dugu:

$\bar{x}_1-\bar{x}_2 \approx N(\mu_1-\mu_2, \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})$, bariantzak ezagunak baldin badira, edo,

$\bar{x}_1-\bar{x}_2 \approx N(\mu_1-\mu_2, \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}})$, bariantzak ezagutzen ez baditugu.

10.adibidea:

Psikologo batek, labirintoan, A bitaminarik gabeko janeurrian dauden 37 arratoi aztertzen ditu, labirintoa kurritzeko batezbesteko aldia 4 minututakoa dela ikusirik, 0,5 mi-

nututako desbidazio standardarekin; geroago, A bitaminaz joria den janeurri peko 50 arratoi aztertzen ditu labirinto berberean, batezbestekoa 3,2 minututakoa dela, 0,4 minututako desbidazio standardarekin, neurtzen duelarik.

Labirintoa kurritzeko batezbesteko aldian, A bitaminak eraginik ez daukala suposatzen dugu.

A bitaminarik ez duen janeurri peko arratoiei dagozkien aldien eta A bitaminaz joria denari dagozkionen arteko kendura, lehenago lorturikoa baino txikiagoa izateko probabilitatea, zein da?

Kasu honetan, ez ditugu ezagutzen ez labirintoa kurritzeko aldiaren banaketa ez eta populazioen bariantzarik ere.

$\mu_1 = \mu_2$, $\bar{x}_1 = 4$, $s_1' = 0{,}5$, $\bar{x}_2 = 3{,}2$, $s_2' = 0{,}4$ direla ezagutzen dugu soilik.

$n_1 = 37$ eta $n_2 = 50$, biak 20 baino handiagoak direnez,

$$\bar{x}_1 - \bar{x}_2 \simeq N\left(\mu_1 - \mu_2 , \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$ dela badakigu.

$$\mu_1 - \mu_2 = 0$$

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{s_1'^2}{n_1 - 1} + \frac{s_2'^2}{n_2 - 1}} = \sqrt{\frac{0{,}5^2}{36} + \frac{0{,}4^2}{49}} =$$

$$= \sqrt{0{,}01} = 0{,}1$$

Hau da, $\bar{x}_1 - \bar{x}_2 \simeq N(0, 0'1)$

$$\text{Orduan, } p(\bar{x}_1-\bar{x}_2<0,8) = p\left(Z<\frac{0,8-0}{0,1}\right)= p(Z<8) =$$

$$= 1-p(Z\geqslant 8) = 1 - 0 = 1$$

V.4.- LAGINAREN PROPORTZIOAREN BANAKETA.

Demagun, 4 pertsonak (daltoniko batek eta hiru ez-daltonikok) osoturiko populazioa. Populaziotik ateratako 2 tamainuko laginetan, daltonikoen proportzioaren banaketa aztertuko dugu.

$E = \{D_1,\bar{D}_1,\bar{D}_2,\bar{D}_3\}$ modura adierazten dugu populazioa.

a) Itzuleraz:

Lagin posibleak

$(D_1,D_1)\ (D_1,\bar{D}_1)\ (D_1,\bar{D}_2)\ (D_1,\bar{D}_3)$

$(\bar{D}_1,D_1)\ (\bar{D}_1,\bar{D}_1)\ (\bar{D}_1,\bar{D}_2)\ (\bar{D}_1,\bar{D}_3)$

$(\bar{D}_2,D_1)\ (\bar{D}_2,\bar{D}_1)\ (\bar{D}_2,\bar{D}_2)\ (\bar{D}_2,\bar{D}_3)$

$(\bar{D}_3,D_1)\ (\bar{D}_3,\bar{D}_1)\ (\bar{D}_3,\bar{D}_2)\ (\bar{D}_3,\bar{D}_3)$

Laginaren daltonikoen proportzioa:

1	0,5	0,5	0,5
0,5	0	0	0
0,5	0	0	0
0,5	0	0	0

Proportzioaren lagin-banaketa:

$\hat{p}$	0	0,5	1
f_i	9	6	1

$$E(\hat{p}) = \frac{4}{16} = \frac{1}{4}$$

$$\text{Var}(\hat{p}) = \frac{2,5}{16} - \frac{1}{16} = \frac{3}{32}$$

b) Itzulerarik gabe:

Lagin posibleak

$(D_1,\bar{D}_1)\ (D_1,\bar{D}_2)\ (D_1,\bar{D}_3)$

$(\bar{D}_1,D_1)\ (\bar{D}_1,\bar{D}_2)\ (\bar{D}_1,\bar{D}_3)$

$(\bar{D}_2,D_1)\ (\bar{D}_2,\bar{D}_1)\ (\bar{D}_2,\bar{D}_3)$

$(\bar{D}_3,D_1)\ (\bar{D}_3,\bar{D}_1)\ (\bar{D}_3,\bar{D}_2)$

Laginaren daltonikoen proportzioa:

0,5	0,5	0,5
0,5	0	0
0,5	0	0
0,5	0	0

Proportzioaren lagin-banaketa:

$\hat{p}$	0	0,5
f_i	6	6

$$E(\hat{p}) = \frac{3}{12} = \frac{1}{4}$$

$$\text{Var}(\hat{p}) = \frac{1,5}{12} - \frac{1}{16} = \frac{6-3}{48} = \frac{1}{16}$$

Populazioak, $p=\frac{1}{4}$ daltonikoen proportzioa zeukala kontutan hartzen baldin badugu, ondoko emaitzak egiaztatzen ditugu:

a) Laginketa itzuleraduna denean:

$$E(\hat{p})=p \quad \text{eta} \quad Var(\hat{p}) = \frac{pq}{n}$$

$$E(\hat{p})= \frac{1}{4} \quad \text{eta} \quad Var(\hat{p}) = \frac{\frac{1}{4}\cdot\frac{3}{4}}{2} = \frac{3}{32}$$

direla eta.

b) Laginketa itzuleragabea denean:

$$E(\hat{p})=p \quad \text{eta} \quad Var(\hat{p}) = \frac{pq}{n}\,\frac{N-n}{N-1}$$

$$E(\hat{p}) = \frac{1}{4} \text{ eta } Var(\hat{p}) = \frac{\frac{1}{4}\cdot\frac{3}{4}}{2}\cdot\frac{4-2}{4-1} = \frac{3}{32}\cdot\frac{2}{3} = \frac{1}{16}$$

direla eta.

Laginaren tamainua handia denean, laginen proportzioen banaketa normala da gutxi gora behera, limitearen teorema zentrala dela kausa.

Hurbilketa normala baliagarria izateko, zenbatekoa izan behar du laginaren tamainuak? Zabal-zabalik erabilitako erizpidean, bai np bai nq direlakoek 5 baino handiagoak izan behar dute. Kasu horretan, $\hat{p}$ estatistikoa, $N(p,\sqrt{\frac{pq}{n}})$ legeari darraio gutxi gora behera.

11.adibidea:

Suposa dezagun, biztanlego baten 8%-a daltonikoa dela. Biztanlego honetatik, 150 pertsona zoriz aukeratzen baldin

badira, zein da lagin horren daltonikoen proportzioa 15%-a baino handiagoa edo berdina izateko probabilitatea?

$$\left.\begin{array}{l} np = 150\cdot 0,08 = 12 > 5 \\ nq = 150\cdot 0,92 = 138 > 5 \end{array}\right\} \Longrightarrow \hat{p}: N(0'08, \sqrt{\frac{0'08\cdot 0'92}{150}}) = N(0'08, 0'022)$$

Beraz,

$$p(\hat{p} \geqslant 0,15) = p(Z \geqslant \frac{0,15-0,08}{0,022}) = p(Z \geqslant 3,16) = 0$$

Alegia, benetan, daltonikoen 15%-a baino gehiago duen lagina aterako balitz, arrarotzat hartuko genuke gertaera.

V.5.- BI LAGINEN PROPORTZIOEN ARTEKO KENDURAREN BANAKETA.

Aldagai binomialetako bi populaziotatik, n_1 eta n_2 tamainuko zorizko lagin askeak ateratzen badira, non bi populazioetako ezaugarri interesgarriak dauzkaten behaketen proportzioak p_1 eta p_2 diren hurrenez hurren, $\hat{p}_1-\hat{p}_2$ laginen proportzioen arteko kenduraren banaketa, normala da gutxi gora behera, batezbestekoa $\mu = E(\hat{p}_1-\hat{p}_2) = E(\hat{p}_1)-E(\hat{p}_2) = p_1-p_2$, eta bariantza $\sigma^2 = Var(\hat{p}_1-\hat{p}_2) = Var(\hat{p}_1)+Var(\hat{p}_2) = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$ dira, n_1 eta n_2 handiak izanik.

12.adibidea:

Suposa dezagun, ospitale batetako tuberkulosia pairatzen duen gaixo-proportzioa 50%-a dela, eta beste ospitale batetakoa 33%-a.

Zein da, populazio bakoitzetik ateratako 100 tamainuko laginek, $\hat{p}_1-\hat{p}_2$ delakoaren balioa 0,3 baino handiagoa izateko probabilitatea?

$\hat{p}_1-\hat{p}_2$ delakoaren banaketa, normala da gutxi gora behera, batezbestekoa $\mu = 0,5-0,33 = 0,17$ eta bariantza

$\sigma^2 = \frac{0,5\cdot 0,5}{100} + \frac{0,33\cdot 0,67}{100} = 0,004711$ izanik.

Orduan:

$$p(\hat{p}_1-\hat{p}_2 > 0,3) = p(Z > \frac{0,3-0,17}{0,004711}) = p(Z > 1,86) = 0,0314$$

V.6.- χ^2, PEARSON-en JI KARRATU BANAKETA.

V.6.1.- Definizioa.

Bira N(0,1)-en arauera normalki banatuak eta askeak diren $Z_1,\ldots,Z_\nu$ ν zorizko aldagai. Aldagai hauen karratuen arteko batura, Pearson-en χ^2_ν delako banaketari darraion aldagai berria da. Alegia,

$$\chi^2_\nu = Z_1^2 +\ldots+ Z_\nu^2$$

Banaketa honek ondoko propietateak betetzen ditu:

1) $[0,\infty)$ tartean definiturik dago.

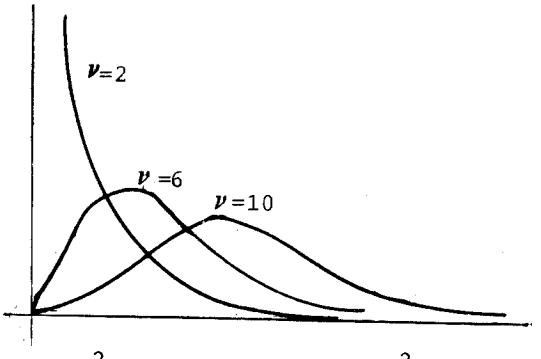
2) χ^2_ν banaketa, askatasun-graduen ν kopuruaren pean dago, non ν χ^2_ν -an parte hartzen duen N(0,1) zorizko aldagai-kopuruak emanik bait dator.

3) Banaketa hau jarraia da definizio-tarte osoan.

4) x= ν -2-rako, maximoa dauka, $\nu > 2$ suposaturik.

5) Ez da simetrikoa.

6) ν emendatzen den neurrian, beraren grafikoa normalarenera hurbilduz doa.



V.1.irudia

7) $E(\chi^2_\nu) = \nu$ eta $Var(\chi^2_\nu) = 2\nu$

V.6.2.- Baturaren teorema.

Bira $\chi^2_{\nu_1}, \ldots, \chi^2_{\nu_k}$, $\nu_1, \ldots, \nu_k$ askatasun-graduko "ji karratu" k zorizko aldagai aske. Beren batura, $\nu_1 + \ldots + \nu_k$ askatasun-graduko "ji karratu" zorizko aldagai berria izango da.

Hots, $\chi^2_{\nu_1} + \ldots + \chi^2_{\nu_k} = \chi^2_\nu$,

non $\nu = \nu_1 + \ldots + \nu_k$ den.

V.6.3.- Kuasibariantzaren banaketa.

σ^2 bariantzako populazio normala emanik, n tamainuko lagin bat ateratzen dugu, s^2 bere kuasibariantza kalkulaturik.

s^2 estatistikoa, laginketan sorturiko zorizko aldagaia da, beraren banaketa ondoko adierazpenaren bidez emana datorrelarik:

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2_{n-1}$$

V.6.4.- χ^2-aren beste aplikazio bat.

Suposa dezagun, populazio baten elementuak k klase exhaustibo eta baztertzailetan banatuta daudela, klase bakoitzaren barnekoa izateko probabilitateak $p_1,\ldots,p_k$ izanik hurrenez hurren.

Populazio honetatik, n tamainuko lagin bat ateratzen baldin badugu, klase bakoitzeko elementuetarako itxarotako kopuruak, $np_1,\ldots,np_k$ izango dira.

Azkenik, suposa dezagun, experientziaren bidez, n tamainuko laginean dauden klase bakoitzeko elementu-kopuruak behatu direla, $o_1,\ldots,o_k$ lorturik hurrenez hurren.

$\sum_{i=1}^{n} \frac{(o_i-np_i)^2}{np_i}$ adierazpena, χ^2 banaketari darraio, askatasun-graduen kopurua k-1 (populazioaren klase-kopuruaren eta baten arteko kendura) izanik. Klase desberdinen p_i probabilitateak ezagutzen ez direnean, eta c parametroren estimazioan oinarrituz kalkulatu behar ditugunean, k-c-1 dira χ^2-aren askataşun-graduak kasu honetan.

Metodo hau aplikatzeko, itxarotako maiztasunak ezin daitezke 5 baino txikiagoak izan; horrela gertatzen denean, klaseak bilduko ditugu, itxarotako maiztasunak 5 edo 5 baino handiagoak lortu ahal izateko.

V.7.- STUDENT-en "t" BANAKETA.

V.7.1.- Definizioa.

Lagin-batezbestekoaren banaketa aztertzerakoan, ondoko hauek ikusi ditugu:

- Populazioa, $N(\mu, \sigma)$ banaketari darraionean, $\bar{x}$ normalki banatuta dagoela, μ batezbestekoa eta $\sigma/\sqrt{n}$ desbidazio standarda izanik.

- n handia denean, $\bar{x}$ estatistikoa ere, $N(\mu, \frac{\sigma}{\sqrt{n}})$ banaketari darraiola gutxi gora behera, populazioaren banaketak axolarik ez duelarik.

- σ ezezaguna denean, bere balio hurbildua lor dezakegula ondoko adierazpenaren bidez hain zuzen ere:

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

Orduan, lagin-tamainua $n \geqslant 20$ baldin bada, $\bar{x}$ estatistikoa, $N(\mu, \frac{s}{\sqrt{n}})$ banaketari darraio.

Baina, σ ezezaguna eta $n < 20$ direnean, zer gertatzen da?

Kasu honetan, $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ kantitatea ez zaio banaketa normalari jarraitzen, n askatasun-graduko Student-en "t" delako banaketari baizik, non

$$t_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

den; alegia, N(0,1) banaketari darraion Z zorizko aldagaiaren, eta, beraren n askatasun-graduz zatitutako χ_n^2 zorizko aldagaiaren erro karratuaren arteko zatidura.

"t" banaketak ondoko propietateak betetzen ditu:

1) 0 batezbestekoa dauka.

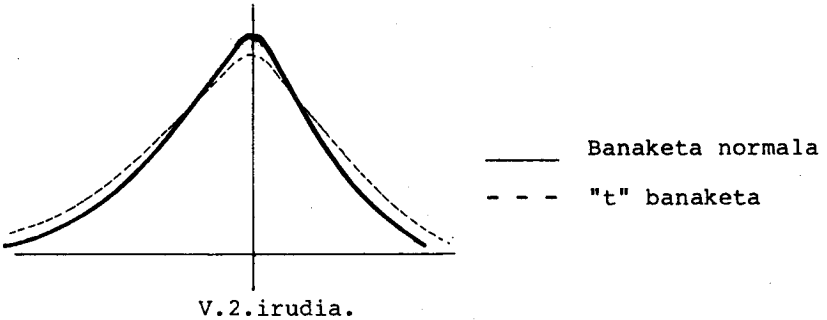
2) Orokorrean, bere bariantza 1 baino handiagoa da; baina, lagin-tamainua emendatuz doanean, 1-erantz jotzen du.

3) Batezbestekoarekiko simetrikoa da, kanpai-itxura daukalarik.

4) Balioak, $(-\infty,+\infty)$ tartean hartzen ditu.

5) Errealitatean, "t" banaketa banaketa-familia da, zeren askatasun-graduen n balio bakoitzerako banaketa desberdin bat bait daukagu.

6) Banaketa normalarekin konparatuz, "t" banaketa, erdian, normala baino txikiagoa da, beraren isatsak bestearenak baino altuagoak izanik. (V.2.irudia).



V.2.irudia.

σ ezezaguna eta $n<20$ direnean, $\bar{x}$ lagin-batezbestekoaren banaketa ezagutzea baimentzen digu banaketa honek.

Ikus dezagun, $\frac{\bar{x}-\mu}{s/\sqrt{n}} = t_{n-1}$ dela.

Izan ere, σ ezaguna denean, $\bar{x}:N(\mu, \frac{\sigma}{\sqrt{n}})$ dela ba dakigu; alegia, tipifikatuz,

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = z$$

Beste aldetik, $\frac{(n-1)s^2}{\sigma^2} = \chi^2_{n-1}$; hots, $\frac{\chi^2_{n-1}}{n-1} = \frac{s^2}{\sigma^2}$

Honetatik, $\sqrt{\frac{\chi^2_{n-1}}{n-1}} = \frac{s}{\sigma}$

Orduan, $\frac{z}{\sqrt{\chi^2_{n-1}/n-1}} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\frac{s}{\sigma}} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

eta definizioz, $\frac{z}{\sqrt{\chi^2_{n-1}/n-1}} = t_{n-1}$

Beraz, $\frac{\bar{x}-\mu}{s/\sqrt{n}} = t_{n-1}$

13.adibidea

Giza-populazio normalaren seroaren batezbesteko amilasa-edukina 90 unitate/100ml. dela badakigu. Itxura normaleko 17 pertsonatako laginaren seroaren amilasa-edukina neurtu zen. Laginak, 97unitate/100ml.-tako batezbestekoa eta 35unitate/100ml.-tako desbidazio standarda aurkeztu zituen. Zein da, laginaren batezbestekoa lortutakoa baino handiagoa izateko probabilitatea?

Alegia, $p(\bar{x}>97)$ kalkulatu behar dugu.

$\frac{\bar{x}-\mu}{s/\sqrt{n}} = t_{n-1}$ dela badakigu. Kasu honetan, $s'=35$;

$n=17$; $\mu=90$.

$ns'^2=(n-1)s^2$ denez, $\frac{s'^2}{n-1} = \frac{s^2}{n}$ ondorioztatzen da,

edo beste modu batetara, $\frac{s'}{\sqrt{n-1}} = \frac{s}{\sqrt{n}}$.

Beraz, $\frac{\bar{x}-90}{35/\sqrt{16}} = t_{16}$

Orduan, $p(\bar{x}>97) = p(t_{16}>\frac{97-90}{35/4}) = p(t_{16}>0,8) = 0,22$

V.7.2.- "t" banaketa eta batezbestekoen arteko kendura.

Populazioen bariantzak ezagutzèn ez direnean eta populazio bakoitzeko independenteki ateratako laginen batezbestekoen arteko kenduraren banaketa ezagutu nahi dugunean, populazioen bariantzak, ezezagunak izan arren, berdinak direla suposatuko dugu; hots, $\sigma_1^2 = \sigma_2^2$ (Suposamendu hau, hurrengo gai batetan ikusiko den erizpide egokiaren bidez frogatu beharko dugu).

Baldintza hau suposatuz, bi laginen kuasibariantzak gauza berberaren (bariantza komunaren) estimazioak direla interpreta daiteke. Orduan, bariantza komunaren estimazio mankomunatua aurki dezakegu, bi laginen kuasibariantzen batezbesteko astatua kalkulatuz. Lagin bakoitzaren kuasibariantza, beraren askatasun-graduez astatzen da ondokoa lorturik:

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$\bar{x}_1-\bar{x}_2$ delakoaren banaketak ondoko batezbestekoa eta bariantza izango ditu:

$$E(\bar{x}_1-\bar{x}_2) = E(\bar{x}_1)-E(\bar{x}_2) = \mu_1-\mu_2$$

$$Var(\bar{x}_1-\bar{x}_2) = Var(\bar{x}_1)+Var(\bar{x}_2) \simeq \frac{s_p^2}{n_1}+\frac{s_p^2}{n_2} \quad ,$$

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = t_{n_1+n_2-2}$$ adierazpena ondorioztaturik.

14.adibidea:

13.adibidearekin erlazionaturik, suposa dezagun, itxura normaleko pertsonez gain, 15 pertsona ospitaleraturengan ere amilasa-edukinaren determinazioak egin zirela. Pertsona ospitaleratuen talde honen seroaren batezbesteko amilasa-edukina eta desbidazio standarda 125 eta 40unitate/100ml. izan ziren hurrenez hurren.

Populazio ospitaleratuaren seroaren batezbesteko amilasa-edukina 100unitate/100ml.-takoa da, bariantza, pertsona normalen populazioarekiko aldatzen ez dela suposaturik.

Zein da laginen batezbestekoen arteko kendura lortutakoa baino txikiagoa izateko probabilitatea?

Lortutako kendura hauxe izan da: $\bar{x}_2-\bar{x}_1=125-97=28$.

Beraz, $p(\bar{x}_2-\bar{x}_1<28)$ kalkulatu behar dugu,

$$\frac{(\bar{x}_2-\bar{x}_1)-(\mu_2-\mu_1)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = t_{n_1+n_2-2} \text{ dela jakinez.}$$

$$\mu_2-\mu_1=100-90=10$$

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{n_1s_1'^2+n_2s_2'^2}{n_1+n_2-2} =$$

$$= \frac{17\cdot 35^2+15\cdot 40^2}{17+15-2} = \frac{44825}{30} = 1494,16$$

$$s_p = \sqrt{1494,16} = 38,65$$

$$\frac{(\bar{x}_2-\bar{x}_1)-(\mu_2-\mu_1)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(\bar{x}_2-\bar{x}_1)-10}{38,65\sqrt{\frac{1}{17}+\frac{1}{15}}} = \frac{(\bar{x}_2-\bar{x}_1)-10}{13,693} = t_{30}$$

$$\text{Beraz, } p(\bar{x}_2-\bar{x}_1<28)=p(t_{30}<\frac{28-10}{13,693})=p(t_{30}<1,31) =$$

$$= 1-p(t_{30}\geqslant 1,31) = 1 - 0,1 = 0,9$$

Bi populazioak normalak direla suposatu ahal izan arren, bien bariantzak berdinak direla suposatzeaz fidatzen ez garenean, deskribatu dugun "t" banaketaren erabilera ez dirudi egokia.

Bariantzak berdinak ez direnean, ondokoan datza arazoa:

$$t' = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

kantitatea ez zaiola n_1+n_2-2 askatasun-graduko banaketari jarraitzen.

Arazoa ebazteko era bat, askatasun-graduetarako ondoko balioa erabiltzean datza:

$$df' = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1}+\frac{(s_2^2/n_2)^2}{n_2}}$$

Normalitate-hipotesiak betetzen badira, "t" adierazpena, df' askatasun-graduko "t" banaketari darraio gutxi gora behera. (df' osoa ez bada, balio osorik hurbilena erabiltzen da).

15.adibidea:

Itxura normaleko 15 pertsonarengan eta 10 gaixorengan, osagarri serologikoaren (CH_{50}-aren) aktibitate osoa determinatu zen, hurrengo emaitzak lorturik:

Pertsonak	n	$\bar{x}$	s
Gaixoak	10	62,6	33,8
Normalak	20	47,2	10,1

Populazioen batezbestekoak, laginen batezbestekoetarako lorturiko balioen bidez hurbilduz, zein izango da $\bar{x}_1-\bar{x}_2$ kendura positiboa izateko probabilitatea?

$$\mu_1-\mu_2=62,6-47,2=15,4$$

$$df'=\frac{\left(\frac{(33,8)^2}{10}+\frac{(10,1)^2}{20}\right)^2}{\frac{\left(\frac{(33,8)^2}{10}\right)^2}{10}+\frac{\left(\frac{(10,1)^2}{20}\right)^2}{20}}=\frac{(114,244+5,1005)^2}{1305,1691+1,300755}=10,9\simeq 11$$

$$\frac{(\bar{x}_1-\bar{x}_2)-15,4}{\sqrt{\frac{(33,8)^2}{10}+\frac{(10,1)^2}{20}}}=\frac{(\bar{x}_1-\bar{x}_2)-15,4}{10,925}=t_{11}$$

$$p(\bar{x}_1-\bar{x}_2>0)=p\left(t_{11}>\frac{-15,4}{10,925}\right)=p(t_{11}>-1,41)=p(t_{11}<1,41)=$$

$$=1-p(t_{11}\geq 1,41)=1-0,1=0,9$$

V.8.- FISHER-SNEDECOR-en "F" BANAKETA.

V.8.1.- Definizioa.

ν_1 eta ν_2, hurrenez hurren, askatasun-graduko χ_1^2 eta χ_2^2 "ji karratu" aldagai emanik,

$$F_{\nu_1,\ \nu_2} = \frac{\chi_1^2/\ \nu_1}{\chi_2^2/\ \nu_2}$$

adierazpenari, ν_1 eta ν_2 askatasun-graduko Fisher-Snedecor-en "F" zorizko aldagaia deritzogu.

Alegia, aldez aurretik bakoitza beraren askatasun-graduez zatituriko bi "ji karratu" aldagai zatitzean lortzen da "F" banaketa.

Banaketa honek, hurrengo propietateak betetzen ditu:

1) $[0, \infty)$ tarteko balioak hartzen ditu.

2) Ez da simetrikoa.

3) ν_1 eta ν_2-ren balioen pean dago.

4) $p(F_{n_1,n_2} > A) = p(F_{n_2,n_1} < \frac{1}{A})$

V.8.2.- Kuasibariantzen arteko zatiduraren banaketa.

"F" banaketari darraion lagin-adierazpena ondokoa da:

σ_1^2 eta σ_2^2, hurrenez hurren, bariantzako bi populazio normal aske emanik, n_1 eta n_2 tamainuko laginak ateratzen 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-graduko "F" banaketari darraio.

Izan ere, $\frac{(n_1-1)s_1^2}{\sigma_1^2} = \chi^2_{n_1-1}$ eta $\frac{(n_2-1)s_2^2}{\sigma_2^2} = \chi^2_{n_2-1}$ dira.

Hemendik abiatuz, $\frac{\chi^2_{n_1-1}}{n_1-1} = \frac{s_1^2}{\sigma_1^2}$ eta $\frac{\chi^2_{n_2-1}}{n_2-1} = \frac{s_2^2}{\sigma_2^2}$

Bi ekuazioak zatituz, hauxe lortzen dugu:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \frac{\chi^2_{n_1-1}/n_1-1}{\chi^2_{n_2-1}/n_2-1} = F_{n_1-1,n_2-1}$$

Bariantzak berdinak diren kasuan, $\frac{s_1^2}{s_2^2} = F_{n_1-1,n_2-1}$ dela ondorioztatzen da.

16.adibidea:

14.adibideari dagokionez, zein izango da laginen bariantzen arteko zatidura 2,8 baino handiagoa izateko probabilitatea?

Hots, $p(s_1'^2/s_2'^2 > 2,8)$ kalkulatu behar dugu.

$\frac{s_1^2}{s_2^2} = F_{n_1-1,n_2-1}$ dela ba dakigu.

Beste aldetik:

$$(n_1-1)s_1^2 = n_1 s_1'^2 \Longrightarrow s_1^2 = \frac{n_1}{n_1-1} s_1'^2$$

$$(n_2-1)s_2^2 = n_2 s'^2_2 \Longrightarrow s_2^2 = \frac{n_2}{n_2-1}\, s'^2_2$$

Beraz:

$$\frac{s_1^2}{s_2^2} = \frac{\frac{n_1}{n_1-1}\, s'^2_1}{\frac{n_2}{n_2-1}\, s'^2_2} = \frac{n_1(n_2-1)s'^2_1}{(n_1-1)n_2 s'^2_2} = \frac{17\cdot 24\ s'^2_1}{25\cdot 16\ s'^2_2} = \frac{51 s'^2_1}{50 s'^2_2} = F_{16,24}$$

Orduan:

$$p\left(\frac{s'^2_1}{s'^2_2} > 2,8\right) = p\left(F_{16,24} > \frac{51}{50}\ 2,8\right) = p(F_{16,24} > 2,856) =$$

$$= 0,01$$

Estatistikoa	Populazioen baldintzak	Laginen baldintzak	Lagin-banaketa
$\bar{x}$	Normalki banatuta. σ ezaguna		$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = Z : N(0,1)$
$\bar{x}$	σ ezaguna	$n \geqslant 25$	$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$
$\bar{x}$	Normalki banatuta. σ ezezagu na	$n \geqslant 20$	$\frac{\bar{x} - \mu}{s/\sqrt{n}} \approx N(0,1)$
$\bar{x}$	Normalki banatuta. σ ezezagu na	$n < 20$	$\frac{\bar{x} - \mu}{s/\sqrt{n}} = t_{n-1}$
$\bar{x}_1 - \bar{x}_2$	Normalki banatuak eta askeak. σ_1, σ_2 eza gunak		$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = Z : N(0,1)$
$\bar{x}_1 - \bar{x}_2$	σ_1, σ_2 eza gunak. Popu lazio as- keak.	$n_1, n_2 \geqslant 25$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \approx N(0,1)$
$\bar{x}_1 - \bar{x}_2$	Normalki banatuak eta askeak. σ_1, σ_2 ez- ezagunak	$n_1, n_2 \geqslant 20$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \approx N(0,1)$
$\bar{x}_1 - \bar{x}_2$	Normalki banatuak eta askeak. σ_1, σ_2 ez- ezagunak baina $\sigma_1 = \sigma_2$	$n_1, n_2 < 20$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = t_{n_1+n_2-2}$

Estatistikoa	Populazioen baldintzak	Laginen baldintzak	Lagin-banaketa
$\bar{x}_1 - \bar{x}_2$	Normalki banatuak eta askeak. σ_1, σ_2 ezezagunak	$n_1, n_2 < 20$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = t_{df'}$ (*)
$\hat{p}$	p parametroko banaketa binomialari darraiona	$np > 5$ $nq > 5$	$\frac{\hat{p}-p}{\sqrt{pq/n}} \approx N(0,1)$
$\hat{p}_1 - \hat{p}_2$	p_1 eta p_2 parametroetako banaketa binomialei jarraitzen zaizkien populazio askeak	$n_1p_1 > 5$ $n_1q_1 > 5$ $n_2p_2 > 5$ $n_2q_2 > 5$	$\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} \approx N(0,1)$
s^2	Normalki banatuta		$\frac{(n-1)s^2}{\sigma^2} = \chi^2_{n-1}$
Behaturiko maiztasunak o_i		$np_i > 5$	$\sum_{i=1}^{k} \frac{(o_i - np_i)^2}{np_i} = \chi^2_{k-1-c}$
s_1^2 / s_2^2	Normalki banatuak eta askeak.		$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = F_{n_1-1, n_2-1}$
			(*) $df' = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1}+\frac{(s_2^2/n_2)^2}{n_2}}$

PROBLEMAK

1.- Aipatzen direnen artean, zeintzu dira proposaturiko populazioaren zorizko laginak?

Zorizko lagina ez denean, azal ezazu zergatia.

a) Biologoen Elkarteak, 5 azken urtetan Biologian Doktore-titulua lortu dutenek idatzitako lan teknikoen kopuruaz interesatzen da. 5 azken urtetan titulu hau eman zaien pertsoneei galdeketa bat bidaltzen die, kalkulua, jasotako erantzunetan oinarriturik.

b) Aldizkari batek produktu bati buruzko inkesta egiten du, posta-zerrendako barruti bakoitzeko harpidedun bati galdeketa bidaliz.

c) Txerto-mota bat ekoitzen duen laborategiak, ehungarren txerto bakoitza kalitate-kontrolera bidaltzen du probatan jartzeko.

d) Psikologoak, 30 baino handiagoa den taldeko 30 ikasle hartzen ditu, taldekide bakoitzari, aldez aurretik, 001, 002, ..., 485-en arteko zenbaki bat eman diolarik. Horretarako, zorizko zenbakien taulako 5 zenbakitako talde baten erdiko 3 digituak erabiltzen ditu, baina, 5 zenbakitako bi taldetatik bat, elkarren txandaka hartzen du, errepikatzen direnak eta 485 baino handiagoak direnak kendurik.

e) Aldizkari baten Bilboko 50 harpidedunetako lagina aukeratu nahi dugu. Bilboko telefono-zerrendatik zoriz aukeratzen dira zenbakiak, eta, arratsalde guztietan aukeraturiko zenbakiei telefonoz deitzen diegu, 50 erantzun lortu arte.

f) New York-eko ospitale ospetsu batetan minbiziaren aurkako ikerketa egiten duen sendagileak, birika-minbizia sendatzeko substantzia kimiko berria lortu du. Substantzia honen eraginkortasuna aztertzeko, birika-minbizia duen hirugarren gaixo bakoitzari, prest baldin badago, ematen dio, gaineratikoak, tratamendua konparatzeko kontroltzat utzirik.

2.- Pertsona heldu normalen azido urikoaren edukina, ehuneko 5,7 eta 1mg, hurrenez hurren, batezbesteko eta desbidazio standardeko banaketa normalari jarraitzen baldin bazaio gutxi gora behera, aurki itzazu 9 tamainuko laginak:

a) Ehuneko 6mg. baino batezbesteko handiagoa izateko probabilitatea.

b) Ehuneko 5 eta 6 mg.-ren arteko batezbestekoa izatekoa.

c) Ehuneko 5,2 mg. baino batezbesteko txikiagoa izatekoa.

3.- Ikerlariak, bi giza-populazioren gibeleko A bitaminaren maila normalki banaturik dagoela uste du. Bi populazioen bariantzak ondokoak direla suposatzen da:

1.populazioa: $\sigma_1^2 = 19.600$

2.populazioa: $\sigma_2^2 = 8.100$

Zein da 1.populazioaren 15 tamainuko zorizko laginak eta bigarrenaren 10 tamainukoak, $\bar{x}_1 - \bar{x}_2$ delakoaren balioa 50 baino handiagoa edo berdina emateko probabilitatea, populazioen batezbestekoen arteko diferentziarik ez badago?

4.- Gaixotasuna osatzeko erabilitako medikamendu standarda, 3 egunetan erabili deneko kasuen 75%-ean efektiboa izan dela ba dakigu. Gaixotasun berbera osatzeko medikamendu berriaren eraginkortasuna balioztatzeko, pairatzen zuten

150 pertsonari eman zitzaien. Hiru egunen buruan, 97 pertsona zendatu ziren. Medikamendu berria, lehenago erabilitako standarda bezain eraginkorra baldin bada, zein da pertsona sendatuen proportzioa hain txikia edo txikiagoa aurkitzeko probabilitatea?

5.- Ume atzeratuen populazioan, hiperaktiboak direnen proportzioa 0,4 dela ba dakigu. Populazio honetatik, 120 tamainuko zorizko lagin bat atera zen, 100 tamainuko beste lagin bat ume atzeratuen beste populazio batetatik aukeraturik. Bi populazioetan, ume hiperaktiboen proportzioa berdina baldin bada, zein da laginek 0,16 baino handiagoa edo berdina den f_1-f_2 kendura emateko probabilitatea?

6.- Bakterioen bizitza-aldia N(1.000h.,100h.) banaketari darraio. Bakterio hauen arteko lagina aukeratu nahi dugu, non laginaren batezbesteko bizitza-aldia 1000-tatik 50 ordu baino gehiagorik diferitzen ez bait da, 0,95 probabilitatea izanik.

Aurki ezazu laginaren tamainua.

Ebatzi ezazu, halaber, bakterioen bizitza-aldia darraion banaketa-mota ez dakigula suposaturik.

7.- Laborategiak disko moduko pastilak fabrikatzen ditu, X beren diametroa kontrolatu nahi duelarik. X kantitatea, μ=15mm. eta σ=10mm. parametroetako zorizko aldagaia da. Denbora-tarte batetan, 64 tamainuko laginak ateratzen dira, beren pastilen x_i diametroak neurturik.

Biz $\bar{x}$, emandako laginaren batezbesteko diametroa.

a) Kalkula itzazu $\bar{x}$-ren batezbestekoa eta bariantza.

b) X normalki banatzen baldin bada, nola banatzen da $\bar{x}$?

c) Kalkula ezazu $p(13<\bar{x}<16)$.

8.- Makinak, injekzioetarako orratzak ekoitzen ditu. Egindako kontrol-azterketak direla medio, 3%-a akasduna dela ba da kigu. Laborategian, 500 orratzetako kutxa bat aukeratzen da.

Zein da, akasdunen proportzioa 1%-a baino txikiagoa izateko probabilitatea?

Zein da, orratz akasdunak 25 baino gehiago izatekoa?

9.- Gaixotasun baten aurka erabilitako txertoak, A eta B laborategiek hornituak dira. A laborategiak ekoitzitakoek 2000h.-tako batezbesteko iraupena daukate, desbidazio standarda 400h.-takoa izanik. B laborategiak bidalitakoek 1800h.-tako batezbesteko iraupena daukate, desbidazio standarda 500h.-takoa izanik.

A laborategiko 200 txerto eta B-ko 150 txerto erosten dira.

Kalkula ezazu, A-ren laginaren batezbesteko iraupenak, B-arena 100h. baino gehiagotan ez gainditzeko probabilitatea.

Kalkula ezazu, halaber, 200h.baino gehiagotan gainditzekoa.

10.- Hegoen luzera, 1,67cm.-tako batezbesteko eta 0,032cm.-tako desbidazio standardeko banaketa normalari darraiola ba dakigu. Intsektu-populaziotik, 10 intsektutako zorizko lagina aukeratzen da.

Ondokoak eskatzen dira:

a) 10 intsektuen batezbesteko hego-luzera 1,65cm. baino laburragoa izateko probabilitatea.

b) 10 intsektuen hego-luzeren kuasibariantza, 0,00159cm. baino luzeagoa izatekoa.

11.- Antibiotikoen ekoizleak, 12 hartzidura-kultibo desberdinetan zorizko laginketa burutzen du, prestatzen ari den antibiotiko-erloaren batezbesteko ahalmena aztertu ahal izateko.

Irakurketak ondokoak izan ziren:

9 - 8,9 - 8,8 - 9,1 - 9 - 9,2 - 8,9 - 9,1 - 9 - 8,9 - 9,1 - 8,9

Orokorrean, antibiotiko hauen ahalmena, 9 batezbestekoa duen banaketa normalari jarraitzen zaiola jakinez, zein da edozein erloren batezbesteko ahalmena 9,1 baino handiagoa izateko probabilitatea?

12.- Bi analisi-eredu kimikoz lortutako emaitzak ikertzeko, hots, ea bi metodoen bidezko determinazio guztien batezbestekoak berdinak direnentz, bi lagin hartu ziren, lehen metodoaren bidez harturikoa 8 tamainukoa eta bigarrenaren bidezkoa 10 tamainukoa izanik.

Determinazioak, konposatu berberaren gainean burutzen direla suposatzen da.

$\mu_1 = \mu_2$, $\sigma_1 = 2$, $\sigma_2 = 1,7$ eta probabilitate-banaketak normalak eta askeak direla suposaturik, zein da $\bar{x}_1 - \bar{x}_2$ kendura, -2 eta 2-ren artean izateko probabilitatea?

Zein da $\bar{x}_1 - \bar{x}_2$ kendura, 3 baino handiagoa izatekoa?

13.- Ekoizleak, 5 termometroren zehaztasuna aztertu nahi du. Tenperatura, $N(0, \sigma)$ banaketari darraion zorizko erroreaz neurtzen dutela suposatzen du, σ ezezaguna izanik.

0°-tan, doi-doi, ondoko neurketak lortu ditu:

Termometroa	1	2	3	4	5
Seinalatzen duen tenperaturaren errorea	0,02°C	0,05°C	-0,01°C	-0,04°C	0,12°C

Tenperaturak beste 5 termometroz neurtzerakoan, zein da emandako taulan lortutakoa baino laginaren batezbesteko errore handiagoa lortzeko probabilitatea?

Eta, σ=0,05 dela ezagutuko balitz?

14.- Birusaren bizitza-aldia, $N(\mu, \sigma)$ banaketari jarraitzen zaio. 12 birusetako laginaren bizitza-aldiek ondoko emaitzak aurkeztu dituzte:

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{12} = 58h. \quad \text{eta} \quad S^2 = \sum_{i=1}^{12} (x_i - \bar{x})^2 = 99h.^2$$

a) Determina ezazu A zenbakia, non $p(\frac{S^2}{\sigma^2} < A) = 0,98$ den.

Ondoriozta ezazu σ_a balioa, non $p(\sigma > \sigma_a) = 0,98$ den.

b) Metodo analogoaren bidez, aurki ezazu σ_b balioa, non $p(\sigma < \sigma_b) = 0,9$ den.

15.- Koleoptero baten X hego-luzera, 60mm. batezbesteko eta 1,5mm. desbidazio standardeko banaketa normalari jarraitzen zaio. Koleoptero honen 9 aletako lagina aukeratzen da.

a) Laginaren 9 balioen $\bar{x}$ batezbestekorako kontrol-grafikoa ezarri nahi dugu.

i) Zein da $\bar{x}$-ren probabilitate-legea?

ii) Itxarotako $\bar{x}$-ren balioarekiko simetrikoak direnen artean, zein (a,b) tartetarako izango da $\bar{x}$-ren balioak aurkitzeko 95%-etako probabilitatea?

b) Laginaren s'^2 bariantza kontrolatzeko, $S^2 = \sum_{i=1}^{9}(x_i-\bar{x})^2$ delakoaren balioak aztertzen dira.

i) Zein da $\frac{S^2}{\sigma^2}$ -ren probabilitate-legea?

ii) Zein da S^2-rako itxarotako balioa? Eta bariantza?

iii) 95%-etako probabilitateaz, zein (0,a) tartetan egon behar dute S^2-ren balioek?

c) Azterturiko 9 aleen hego-luzerak hauxek izan dira:

60,4 - 60,8 - 59 - 61,6 - 62,6 - 58,3 - 59,2 - 60,2 - 60,5mm.

i) Kalkula itzazu lagin honen batezbestekoa eta bariantza.

ii) Lagin honen $\bar{x}$ eta S^2-ren balioak, 95%-eko mugen artean al daude?

16.- Ezintasun fisiko berbera pairatzen duten 9 gaixori, experimentu baten parte gisa, eginkizun bat burutzeko eskatu zitzaien. Eginkizuna burutzeko behar izan zuten batezbesteko denbora 7min.-takoa izan zen, desbidazio standarda 2min.-takoa izanik.

Orokorrean, eginkizun hori burutzeko beharrezkoa den denbora, 5min.-tako batezbestekoa duen banaketa normalari darraiola suposatzen dugu.

a) 9 gaixotako laginak harturik, zein da laginaren batez besteko denbora, emandako laginak erabilitakoa baino luzeagoa izateko probabilitatea?

b) Laginaren batezbesteko denbora, 5 baino luzeagoa eta A balioa baino laburragoa izateko probabilitatea 5%-takoa izatea nahi dugu. Kalkula ezazu A balioa.

17.- D bitaminaren eza duten 26 animalia, bi taldetan erdibitu ziren. 1.taldeak, D bitamina hornitzen duen janeurrian datzan tratamendua jaso zuen; 2.taldeak, osterantzean, ez zuen tratamendurik jaso. Aldi experimentalaren amaieran, seroko kaltzioaren determinazioak egin ziren, ondoko emaitzak lorturik:

Tratamendua jaso zuen taldea: $\bar{x}$ = 11,1mg./100ml. ;
s = 0,5mg./100ml.

Tratamendua jaso ez zuena: $\bar{x}$ = 7,8mg./100ml. ;
s = 0,75mg./100ml.

Tratamendua jaso zuten animalien kasuan, seroko kaltzio-edukina, $N(11, \sigma)$ banaketari, eta, tratamendurik gabekoenean, $N(8, \sigma)$ banaketari jarraitzen zaiola suposatzen dugu.

a) Tamainu berdineko laginetarako prozedura bera erabiliz, zein izango da tratamendua jaso duten animalien eta tratamendurik gabekoen lagin-batezbestekoen arteko kendura, emandako laginen artean lortutakoa baino txikiagoa izateko probabilitatea?

b) Zein da laginen kuasibariantzen arteko zatidura, 1 baino handiagoa izateko probabilitatea?

18.- Ospitale orokorrean sendaturiko 21 gaixotako laginaren egonaldiaren batezbesteko iraupena 7 egunetakoa izan zen, desbidazio standarda 2 egunetakoa izanik. Gaixotasun kronikoetarako ospitalean sendaturiko 25 gaixotako laginak 36 egunetako egonaldiaren batezbesteko iraupena aurkeztu zuen, 10 egunetako desbidazio standarda zuelarik.

Ospitale orokorreko egonaldiaren iraupena $N(5, \sigma_1)$ banaketari, eta gaixotasun kronikoetarakoa $N(35, \sigma_2)$ banaketari darraiola suposatzen dugu.

a) Ospitale orokorreko gaixoek osoturiko 21 tamainuko laginak eta gaixotasun kronikoetarako ospitalekoek osoturiko 25 tamainukoak hartzerakoan, zein da $\bar{x}_2-\bar{x}_1$ kendura, emandako laginen batezbestekoen arteko kendura baino handiagoa izateko probabilitatea?

b) Banaketak, N(5,2) eta N(35,10) direla suposatzen baldin badugu, zein izango da laginen kuasibariantzen arteko zatidura, emandako laginen artekoa baino txikiagoa izateko probabilitatea?

VI. GAIA: ESTIMAZIOA

VI.1.- Inferentzia estatistikoaren beharra.

VI.2.- Puntu-estimazioa.

VI.2.1.- Aurretiko adibidea.

VI.2.2.- Bariantza txikieneko estimatzaile zentratuak.

VI.3.- Konfidantza-tarteak.

VI.3.1.- Adibideari buruzko kontsiderazioak.

VI.3.2.- Konfidantza-tartearen kontzeptu oro korra.

VI.4.- Konfidantza-tarte arruntak.

VI.5.- Lagin-tamainuaren determinazioa.

ESTIMAZIOA

VI.1.- INFERENTZIA ESTATISTIKOAREN BEHARRA.

Populazioaren aleek aurkezten duten ezaugarria (aldakorra) aztertu nahi du edozein ikerketa estatistikok. Populazioaren ale bakoitzari eta guztiei buruzko informazioa hartzen denean, "zentsua" burutzen da. Hala ere, hau ez da beti posible edo egokia izaten. Batzutan, populazioa bukaezina delako; beste batzutan, probak errausleak direlako, esate baterako, materialen erresistentziarako neurketa; beste batzutan, populazioa elementu potentzialek osoturik dagoelako, adibidez, medikamenduaren pean jartzerakoan, ondorio sekundarioak aurkezten dituen ale-proportzioa kalkulatu nahi dugunean; gehienetan, behaketa bakoitzak "kostua" halabehartzen duelako.

Mota honetako kontsiderazioek populazioaren zati ("lagina" izenekoa) bati buruzko informazioa hartzera daramate ikertzailea. Laginaren elementu-kopuruari, lagin-tamainua deritzogu.

Berehala, Estatistikaren oinarrizko arazoa zein den igartzen da: Laginaren datuetatik abiatuz, populazioari buruzko ondorioak nola lor? Ondorio hauek zein fidagarritasun-maila daukate? Lagin-balioak, zein puntutaraino dira populazio-balioen adierazgarriak?

Beraz, Estatistikaren oinarrizko metodologia indukzioa da; baina, indukzioa edo inferentzia nola burutzen den finkatzeko, Matematikaren beste edozein atalenak bezain zorrotzak diren Probabilitate-kalkuluaren teoremak erabiltzen dira. Soilik aplikazio praktikora heltzerakoan, emaitzak "seguruak" ez direla kontutan hartu behar dugu, "konfidantza-maila" daukatela baizik.

Gai honetan aztertuko dugun arazo generikoa honela formula daiteke:

Demagun populazioa deskribatzen duen X zorizko aldagaiaren banaketa-funtzioa, parametro ezezagun (μ , σ , p,...) baten edo batzuren pean dagoela. Populaziotik, n tamainuko lagin bat aukeratzen dugu. Lorturiko informazioaren arauera, parametro ezezagunak "estimatu" behar ditugu. "Estimazioa", parametroari balio bat ematean baldin badatza, "puntu-estimazioa" burutzen dugu; populazioaren parametroaren balioa, tarte baten balioren bat dela estimatzen baldin badugu, "konfidantza-tartea" ematen dugularik.

VI.2.- PUNTU-ESTIMAZIOA.

VI.2.1.- Aurretiko adibidea:

Suposa dezagun, animalia ugaztunen espezie batetarako, kumaldiko X ume-kopurua ezagutu nahi dugula.

Poisson-en eredura jotzea aski arrazonagarria da. Baina, $P(\lambda)$ Poisson-en banaketa, ezezaguna den λ parametroaren menpean dago. Beraz, berari buruzko zenbait informazio lortu behar dugu. Horretarako, urtebetean zehar, 35 animalia behatzen ditugu, kumaldiko ume-kopurua zenbaturik. Horrela, ondoko taula eratzen dugu:

Kumaldiko ume-kopurua	0	1	2	3	4	5	6	7
Animali-kopurua	2	3	10	10	5	0	5	0

λ estimatzeko, nola erabil informazio hau?

Zenbait posibilitate dauzkagu:

1) $E(X) = \lambda$ ba dakigu, eta, lagin-batezbestekoa itxaropen teorikoaren balio "analogoa" denez, $\lambda = \bar{x}$ estima dezakegu.

Adibide honetan, $\bar{x}=2'94$; beraz, erizpide honen arauera, λ -ri 2'94 balioa dagokio.

2) X: $P(\lambda)$ baldin bada, $p(X=0)=e^{-\lambda}$ da; beraz, $\lambda=-\ln\{p(X=0)\}$.

Aurreko baldintza maiztasunen bidez interpretatuz, hauxe da λ -ri dagokion balioa:

$$\lambda=-\ln\{fr\ (X=0)\}=-\ln\frac{2}{35}=\ln\frac{35}{2}=2'86$$

3) $Var(X)=\lambda$ denez, ondoko lagin-balio analogoa kalkulatzen dugu:

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

Lorturiko laginerako $s^2=2'64$ denez, $\lambda=2'64$ estimatzen dugu.

Arrazonamendu-kopurua aski luza daiteke. Eta, bakoitzak λ-ren balio desberdinera garamatzanez, "hoberena" aukeratu behar dugu. Baina "hoberenaz" zer ulertzen dugu?

VI.2.2.- Bariantza txikieneko estimatzaile zentratuak.

Demagun, π populazioa deskribatzen duen X zorizko aldagaiaren $F(x,\theta)$ banaketa-funtzioa θ parametroaren pean dagoela. $T=T(x_1,\ldots,x_n)$ funtzio errealari, non $(x_1,\ldots,x_n)$ delakoak populazioaren zorizko lagin bakuna adierazten bait du, θ-ren estimatzailea deritzogu, θ-ren balio hurbilduak lortzeko erabiltzen dugularik.

Estimatzaileari, laginketako zorizko aldagaia denez, probabilitate-legea dagokio; lege hau, estimatzen nahi dugun parametroaren inguruan zenbat eta konzentratua izan, estimatzailea gero eta hobea delarik.

Estimatzailea eta estimazioa argi eta garbi bereiztu behar ditugu: Estimatzailea, probabilitate-banaketaren arauera, n tamainuko lagin bakoitzerako balio desberdinak hartzen dituen funtzio erreala da; estimazioa, bestalde, estimatzaileak lagin konkretuan hartzen duen balioa izanik.

Horrela, $\bar{x} = \frac{x_1+x_2}{2}$ bi tamainuko laginetarako estimatzailea da; $5 = \frac{6+4}{2}$, ordea, (6,4) laginari dagokion estimazioa izanik.

Edozein estimatzaile zorizko aldagaia denez, ea errepikapen-kopuru handian ematen den batezbestekoa, parametroaren berdina den galdetzea arrazonagarria da. Alboragabetasunaren nozioa garamatza honek: $T = T(x_1,\ldots,x_n)$ estimatzailea, θ-ren estimatzaile zentratua edo alboragabea da, berarentzat itxarotako balioa, hain zuzen ere, θ baldin bada; hots, $E(T) = \theta$ bada.

Halaber, bidezkoa da, estimatzaileak, θ parametroaren inguruko balio sakabanatuegirik eman ez dezala exijitzea: $Var(T) \leqslant Var(T')$ baldin bada, T estimatzailea, T' baino efizienteagoa izango da.

Hoberen-erizpidea formula dezakegu: $T = T(x_1,\ldots,x_n)$ estimatzailea θ-ren estimatzailerik hoberena da, baldin eta zentratua bada eta θ-ren edozein estimatzaile zentratuk baino bariantza txikiagoa badauka.

Bariantza txikieneko estimatzaile zentratua beti egon ez arren, hona hemen estimatzaile hoberenen zenbait adibide:

1.- Populazio dikotomikoa deskribatzen duen X:Bin(1,p) Bernouilliren zorizko aldagaiaren p parametrorako estimatzailerik hoberena,

$$\hat{p} = \frac{\text{laginaren arrakasta-kopurua}}{\text{laginaren tamainua}}$$ dugu.

$\hat{p}$-ren bariantza, $\frac{p(1-p)}{n}$ da, laginaren tamainua aski handia denean, $\hat{p}$ estimatzailea $N(p,\sqrt{\frac{p(1-p)}{n}})$ banaketari darraiolarik.

2.- Poisson-en banaketaren λ parametrorako estimatzailerik hoberena, $\hat{\lambda} = \bar{x}$ laginaren batezbestekoa da.

Beraren bariantza $\frac{\lambda}{n}$ da, eta, lagin-tamainu handirako, $N(\lambda, \sqrt{\frac{\lambda}{n}})$ banaketaren arauera banatzen da gutxi gora behera.

3.- Banaketa normalaren μ batezbesteko ezezagunerako estimatzailerik hoberena ere, $\hat{\mu} = \bar{x}$ lagin-batezbestekoa da.

Beraren bariantza σ^2/n da, dagokion banaketa $N(\mu, \frac{\sigma}{\sqrt{n}})$ izanik.

4.- Banaketa normalaren σ^2 bariantzarako estimatzailerik hoberena,

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i-\bar{x})^2$ lagin-kuasibariantza da.

Beraren bariantza $\frac{2\cdot\sigma^4}{n-1}$ da, $\frac{(n-1)s^2}{\sigma^2}$ delakoaren banaketa (n-1) askatasun-gradutako χ^2 banaketa izanik.

1.adibidea:

Esan dugun arauera, ugaztunen espezierako kumaldiko X ume-kopuruaren azterketan, λ-ren estimatzailerik hoberena lagin-batezbestekoa da.

Aurreko kasuan, $\bar{x}$=2'94 lortu dugu, datu teorikoak eta lagin-behaketak konparatzea interesa dakigukeelarik.

Kalkuluak errazteko, suposa dezagun λ=3 dela.

Kumaldiko ume-kopurua	Behaturiko maiztasunak	Maiztasun teorikoak
0	2	0,0498·35 = 1,7430
1	3	0,1494·35 = 5,2290
2	10	0,2240·35 = 7,8400
3	10	0,2240·35 = 7,8400
4	5	0,1680·35 = 5,8800
5	0	0,1008·35 = 3,5280
6	5	0,0504·35 = 1,7640
7	0	0,0216·35 = 0,7560

Taularen gaineko azterketak, datu teoriko eta experimentalen arteko komunztadura handirik ez dagoela adierazten digu. Beronen kausak desberdinak izan daitezke: Poisson-en eredua egoera honetarako baliagarria ez dela, nahiz lagina ez dela ondo aukeratua izan, nahiz zoria dela kausa,...

Geroago, atxekidura-maila edo doikuntz egokitasuna neurtuko dugu. Oraingoz, puntu-estimazioaren zehazgabetasuna azpimarratzea interesatzen zaigu: $\bar{x}$, λ-ren estimatzailerik "hoberena" da, baina, praktikan balio dutenak, λ-ren estimazioak dira, hots, laginaren pean dauden balio konkretuak. Bi ikertzailek, $\bar{x}$-en bidez estima dezakete λ, era berdinean aukera dezakete lagina,..., baina, lehenak, $\bar{x}$=2,94 lortzen du, eta, bigarrenak, $\bar{x}$=2,83. Zein da λ-ren estimaziorik hoberena? Nola deskriba, zoriaren eragina laginetan?

Estimatzaileen propietateetan oinarrituriko konfidantza-tarteen bidezko estimazio-metodoak argitzen dizkigu galdera hauek. Hala ere, estimatzaileen bariantzak n-rekin batera beheratzen direla kontutan hartu behar dugu; esate baterako, Poisson-en arazorako $\mathrm{Var}(\bar{x}) = \frac{\lambda}{n}$ da, eta, laginaren datu-kopuru handiagoa hartzerakoan, lortuko ditugun $\bar{x}$-ren balioak elkar hurbilagoak izango direla adierazten digu honek.

VI.3.- KONFIDANTZA-TARTEAK.

VI.3.1.- Adibideari buruzko kontsiderazioak.

Suposa dezagun, bulkada elektrikoa dela kausa, animali mota batek (arratoiek, adibidez) experimentatzen duen batezbesteko erreakzio-aldia estimatu nahi duela psikologoak. Gaiari buruzko ezagupenek, zorizko aldagai honek $N(\mu, 0'6sg.)$ banaketa betetzen duela suposatzea baimentzen diote. 36 experientzia egin eta gero, batezbesteko erreakzio-aldia $\bar{x}=0'45$ dela behatzen da. $\bar{x}$, μ-ren propietate oneko estimatzailea dela esan dugu; baina, termino probabilistikoen bidez, estimazio honen hurbilketa-maila ezagutzearen arazoa sortzen da berehala.

Beraz, $\bar{x} = \frac{x_1+\ldots+x_n}{n}$ lagin-batezbestekotik abiatuz, σ ezaguneko $N(\mu, \sigma)$ banaketaren batezbestekoa estimatzean datza arazoa:

Dakigunez, laginketan, $\bar{x}$, $N(\mu, \sigma/\sqrt{n})$ banaketa betetzen duen estatistikoa da. Banaketa normalaren taulatik, ondokoa ondorioztatzen dugu:

$$p(-1,96 < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < 1,96) = 0,95$$

Eta, hemendik, bi erlazio lortzen ditugu:

$$1)\ p(\mu - 1,96\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1,96\frac{\sigma}{\sqrt{n}}) = 0,95$$

$$2)\ p(\bar{x} - 1,96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1,96\frac{\sigma}{\sqrt{n}}) = 0,95$$

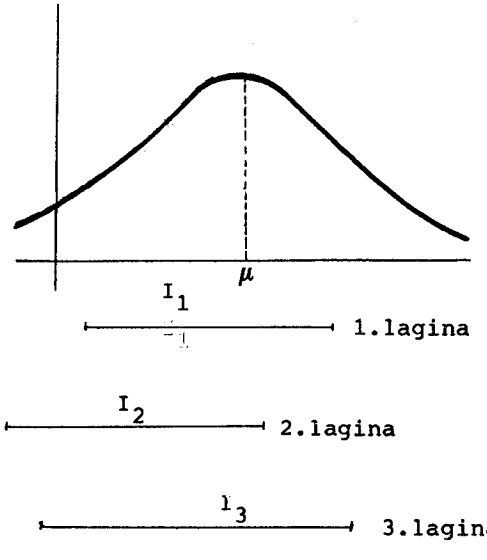
Bi erlazio hauen interpretazioak guztiz desberdinak dira. Lehendabizikoan, $\bar{x}$ zorizko aldagaia tarte fixuan egoteko probabilitatea 0,95 dela baieztatzen da. μ parametroa ezagutuko bagenu, baliagarria izango litzateke erlazioa. Baina, inferentzia estatistikoaren arazoan, μ ezezaguna da, hain zuzen ere.

Bigarrenak, $(\bar{x}-1{,}96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}+1{,}96\frac{\sigma}{\sqrt{n}})$ zorizko tarteak μ-ren balioa edukitzeko probabilitatea 0,95 dela azaltzen digu. Probabilitate honi, tartearen konfidantza-maila deritzogu.

Maiztasunen bidez azaldutako aurreko baiztapenak, hauxe adierazten digu: n heineko lagin-kopuru handian, 2) motako tarteen 95%-ak μ-ren balioa daukatela gutxi gora behera. Ondorioz, jokabide induktiboaren ondoko araua adieraz dezakegu: 0,95 balioa ("konfidantza-maila" izenekoa) finkaturik eta n heineko lagina aukeraturik, berari dagokion $\bar{x}$-ren balioa $\bar{x}_o$ izanik, $(\bar{x}_o-1{,}96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}_o+1{,}96\frac{\sigma}{\sqrt{n}})$ tartea, konfidantza-tartea da.

Tarte hau determinatuz gero, μ-ren balioa edukitzea ala ez gerta daiteke, eta, beraz, edukitzeko probabilitatea 0,95 dela esateak ez dauka zentzurik. Konfidantza-maila hau ez dagokio aurreko tarteari, 2) motako zorizko tarteari baizik.

VI.1.irudian, zenbait lagini dagozkien konfidantza-tarteak irudikatu ditugu, eta esaten duguna hauxe da: Tarte hauen kopuru handia irudikatzen badugu, kasuen arteko 95%-ak μ -ren balioa izango duela.



VI.1.irudia

Ikusten dugunez, konfidantza-mailak eta tartearen anplitudeak karakterizatzen dute estimazioaren zehaztasuna. Konfidantza-maila fixurako, zenbat eta konfidantza-tarte laburragoa, hainbat eta estimazio zehatzagoa. Edo, tarte-luzera berbererako, zenbat eta konfidantza-maila handiagoa, hurbilketa gero eta hobea izango da.

Gure kasuan, tartearen luzera, $L=2\cdot 1,96\frac{\sigma}{\sqrt{n}}$ da, konfidantza-maila emendatu nahi baldin badugu, luzera gehituko dela bistakoa delarik. Hala ere, laginaren tamainua egokiro emendatuz, helburu biak (konfidantza-tartearen luzeraren laburpena eta konfidantza-mailaren gehipena) lor daitezke.

Planteaturiko arazorako lortzen dugun 95%-eko konfidantza-tartea $(0'45-1'96\frac{0'6}{\sqrt{36}}\ ,\ 0'45+1,96\frac{0'6}{\sqrt{36}}) = (0'254,\ 0'646)$ da, luzera 0'392 izanik.

Laginaren tamainua n=100 izango balitz, konfidantza-maila berbererako tartearen luzera $L=2\cdot 1,96\frac{0'6}{\sqrt{100}} = 0'2352$ izango litzateke.

Hala ere, praktikan, laginaren tamainuak zenbait balio gainditzeak ez dauka interesik, zeren experimentua garestiegia bait da eta hurbilketa-maila determinatua gainditzeak zentzurik ez bait dauka.

VI.3.2.- Konfidantza-tartearen kontzeptu orokorra.

Ikuspuntu teorikotik, planteamendu orokorra oso erraza da:

θ parametroaren menpean dagoen $F(x,\theta)$ banaketa-funtzioko X zorizko aldagaia eta $(x_1,\ldots,x_n)$ zorizko lagin bakuna emanik, laginaren balioen $T_1(x_1,\ldots,x_n)$ eta $T_2(x_1,\ldots,x_n)$ bi funtzio erreal determinatu behar ditugu, non

$$p\{T_1(x_1,\ldots,x_n) < \theta < T_2(x_1,\ldots,x_n)\} = 1-\alpha$$

den, $0<\alpha<1$ zenbaki aurrefinkatua izanik.

Lagin konkretuaren balioez egindako $(T_1(x_1,\ldots,x_n), T_2(x_1,\ldots,x_n))$ tarteari, θ -rako konfidantza-tartea deritzogu, $(1-\alpha)$ zenbakia konfidantza-maila izanik. Praktikan, $1-\alpha=0,95$; $1-\alpha=0,99$; edo $1-\alpha=0,999$ balioak konfidantza-mailatzat hartzen dira, onartzeko prest gaudeneko ziurgabetasun-mailaren arauera.

Orokorrean, $T_1(x_1,\ldots,x_n)$ eta $T_2(x_1,\ldots,x_n)$ funtzioen determinazioa, θ parametroaren propietate oneko estimatzaileetatik abiatuz burutzen da, aurreko adibidean egin dugun antzera.

VI.4.- KONFIDANTZA-TARTE ARRUNTAK.

Konfidantza-tarteen mota desberdinak adieraziko ditugu soilik; beraien eraiketa-modua, aurreko adibidean erabilitakoaren antzekoa da, V.gaian emandako estatistikoen lagin-banaketetan oinarriturik.

n, laginaren tamainua, eta $1-\alpha$, konfidantza-maila direla suposatuko dugu.

a) Banaketa normalaren batezbestekorako konfidantza-tartea.

a_1.- σ^2 bariantza ezaguna baldin bada:

$$I_{\mu}^{1-\alpha} = (\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ , \ \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

a_2.- σ^2 bariantza ezezaguna eta laginaren tamainua handia $(n>20)$ baldin badira:

$$I_{\mu}^{1-\alpha} = (\bar{x}-z_{\alpha/2}\frac{s}{\sqrt{n}} \ , \ \bar{x}+z_{\alpha/2}\frac{s}{\sqrt{n}})$$

a_3.- σ^2 bariantza ezezaguna eta laginaren tamainua txikia ($n \leqslant 20$) baldin badira:

$$I_{\mu}^{1-\alpha} = (\bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} \ , \ \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}})$$

b) Banaketa normalaren bariantzarako konfidantza-tartea:

$$I_{\sigma^2}^{1-\alpha} = \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}} \ , \ \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;n-1}} \right)$$

2.adibidea:

Itxura normaleko 15 gizabanakoren laginaren seroaren amilasa-edukina neurtu zen. Laginak, 96 unitate/100ml. -tako batezbestekoa eta 35 unitate/100ml.-tako desbidazio standarda aurkeztu zituen. σ^2-rako 95%-eko konfidantza-tartea eraiki nahi dugu.

Laginak, $s'^2 = 35^2 = 1225$ eman zigun.

Askatasun-graduak, $n-1 = 15-1 = 14$ dira.

χ^2 banaketari dagokion taularen irakurketak, $\chi^2_{14;0,975} = 5,629$ eta $\chi^2_{14;0,025} = 26,119$ balioak ematen dizkigu.

Beraz, 95%-eko konfidantza-tartea hauxe izango da:

$$\left(\frac{ns'^2}{\chi^2_{\alpha/2;n-1}} \ , \ \frac{ns'^2}{\chi^2_{1-\alpha/2;n-1}} \right) = \left(\frac{15 \cdot 1225}{26,119} \ , \ \frac{15 \cdot 1225}{5,629} \right) =$$

$$= (656'61, 3046'72)$$

Halaber, σ-rako 95%-eko konfidantza-tartea ondokoa izango da:

$$\left(\sqrt{656'61} \ , \ \sqrt{3046'72} \right) = (25'62 \ , \ 55,20)$$

c) Bi banaketa normal askeren batezbestekoen arteko kendurarako konfidantza-tartea.

c_1.- σ_1^2 , σ_2^2 banaketa bien bariantzak ezagunak direnean:

$$I^{1-\alpha}_{\mu_1-\mu_2} = (\bar{x}_1-\bar{x}_2-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \ , \ \bar{x}_1-\bar{x}_2+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})$$

c_2.- Banaketen bariantzak ezezagunak eta laginen tamainuak handiak ($n_1, n_2 > 20$) direnean:

$$I^{1-\alpha}_{\mu_1-\mu_2} = (\bar{x}_1-\bar{x}_2-z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \ ,$$

$$\bar{x}_1-\bar{x}_2+z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}})$$

c_3.- Banaketen bariantzak ezezagunak baina berdinak direnean, laginen tamainuak txikiak izanik ($n_1, n_2 \leqslant 20$)

$$I^{1-\alpha}_{\mu_1-\mu_2} = (\bar{x}_1-\bar{x}_2 \pm t_{\alpha/2;n_1+n_2-2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}})$$

c_4.- Banaketen bariantzak ezezagunak eta laginen tamainuak txikiak ($n_1, n_2 \leqslant 20$) direnean:

$$I^{1-\alpha}_{\mu_1-\mu_2} = (\bar{x}_1-\bar{x}_2 \pm t_{\alpha/2;df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}})$$

3.adibidea:

Esnekiak laboratzeko zentralak, A eta B baserrietatik jasotzen du esnea egunero. Jasotako esnearen kalitatea aztertzeko, bi lagin aukeratzen dira zoriz, eta materia koipetsuaren edukina neurtzen da, ehunekotan adierazitako ondoko emaitzak lorturik:

A baserria: $n_A=33$; $\bar{x}_A=8,7\%$; $s_A^2=1,02\%$

B baserria: $n_B=27$; $\bar{x}_B=10,9\%$; $s_B^2=1,73\%$

Baserri bietako esnearen batezbesteko koipe-edukinen arteko kendurarako 95%-eko konfidantza-tartea eraikitzea eskatzen da.

Biz X_A, aztertzen dugun lehen populazioaren ezaugarria adierazten duen zorizko aldagaia (Kasu honetan, A baserriak emandako esnearen koipe-edukina).

Suposa dezagun, X_A μ_A batezbesteko eta σ_A^2 bariantza ezezagunetako banaketa normalari darraiola.

Eta, analogoki, biz X_B B baserriari dagokiona, μ_B batezbesteko eta σ_B^2 bariantza ezezagunetako banaketa normalekoa.

Zorizko aldagai hauen batezbestekoen arteko $\mu_A - \mu_B$ kendurarako konfidantza-tartea eraikitzea interesatzen zaigu.

Laginak handiak direnez, ondokoa izango da tartea:

$$(\bar{x}_A - \bar{x}_B \pm z_{\alpha/2} \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}) =$$

$$= (8'7 - 10'9 \pm 1'96 \sqrt{\frac{1'02}{33} + \frac{1'73}{27}}) =$$

$$= (-2'804 \ , \ -1'596)$$

Kontutan har dezagun, tarte honek 0 ez daukanez, bi baserrietako esnearen batezbesteko koipe-edukinen arteko diferentzia adierazgarria dela.

d) Bi banaketa normal askeren bariantzen arteko zatidurarako konfidantza-tartea.

$$I^{1-\alpha}_{\sigma_1^2/\sigma_2^2} = \left(\frac{s_1^2/s_2^2}{F_{\alpha/2;n_1-1,n_2-1}} \ , \ \frac{s_1^2/s_2^2}{F_{1-\alpha/2;n_1-1,n_2-1}}\right)$$

e) Banaketa binomialaren parametrorako konfidantza-tartea. (Lagin handiak, $n > 30$)

$$I^{1-\alpha}_{p} = \left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \ , \ \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

f) Bi banaketa binomial askeren proportzioen arteko kendurarako konfidantza-tartea. (Lagin handiak, $n_1, n_2 > 30$).

$$I^{1-\alpha}_{p_1-p_2} = (\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}})$$

4.adibidea:

Demagun, tuberkulosia pairatzen duten 100 pertsonatako talde bi. Farmako berrria, lehendabiziko taldeari ematen zaio. Bigarrenari, aldiz, ez (kontrol-taldea).

Lehendabiziko taldeko 75 pertsona eta bigarreneko 60 osatzen direla behatzen da. Aurki itzazu, osatzen diren gaixo tuberkulosoen proportzioen arteko kendurarako 95%-eko konfidantza-mugak.

Kasu honetan, hauxek dira datuak:

$\hat{p}_1=0'75$; $n_1=100$; $\hat{p}_2=0'6$; $n_2=100$; $z_{\alpha/2}=1'96$

Beraz, ondokoa da konfidantza-tartea:

$$(0'75-0'6 \pm 1'96 \sqrt{\frac{0'75\cdot 0'25}{100} + \frac{0'6\cdot 0'4}{100}}) =$$

$$= (0'15 \pm 0'13) = (0'02 \ , \ 0'28)$$

g) Batezbestekoen arteko kendurarako konfidantza-tartea. (Binakako datuak).

Suposa dezagun, tenperatura beheratzeko antitermikoaren eraginkortasuna egiaztatu nahi dugula. Informazio estatistikoa lortzeko era, gizabanako berberaren tenperatura, antitermikoa eman baino lehen eta eman eta gero neurtzean datza. Era hau, pertsona desberdinen talde bi kontsideratzean eta bati farmakoa ematean, bestea kontroltzat harturik, datzana baino hobea da, zeren faktore estrainio eta kontrolaezinak (talde biak osotzen dituzten gizabanakoen ezaugarri desberdinak) sartzen bait dira. Egoera honetan eta antzekoetan, binakako behaketak burutzea komenigarria da.

Farmakoa eman baino lehen, tenperatura neurtzen duen zorizko aldagaia, X, eta, farmakoa eman eta geroko tenperaturari dagokion zorizko aldagaia, Y deituak izanik, $\mu_1 - \mu_2$ delakorako konfidantza-tartea ematea interesatzen zaigu, $E(X) = \mu_1$ eta $E(Y) = \mu_2$ izanik.

Baldintza hauetan, X eta Y askeak ez direnez, aurrean azterturikoen desberdina den eredua eraiki behar dugu.

D=X-Y kendura neurtzen duen zorizko aldagaia, banaketa normalari darraiola suposa baldin badaiteke, $\mu_1 - \mu_2$ batezbestekoen arteko kendurarako konfidantza-tartea, laginaren tamainua handia ($n>20$) baldin bada,

$$(\bar{d} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{d} + z_{\alpha/2}\frac{s}{\sqrt{n}}) \text{ da, non } \bar{d}=\sum_{i=1}^{n}\frac{d_i}{n} \text{ eta}$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2 \text{ diren, } d_i=x_i-y_i \text{ izanik.}$$

Laginaren tamainua txikia ($n\leqslant 20$) bada, dagokion tartea

$$(\bar{d} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{d} + z_{\alpha/2}\frac{s}{\sqrt{n}}) \text{ da.}$$

5.adibidea:

Sukarra gutxitzeko antitermikoaren eraginkortasuna neurtu nahi dugu. Horretarako, gripea pairatzen duten 2 urteko 10 umeri, antitermikoa eman baino lehen eta eman eta geroko tenperaturak neurtu zitzaien, ondoko emaitzak lorturik:

Gaixoa	1	2	3	4	5	6	7	8	9	10
Lehenengo t.	39'3	39'7	39'9	40'0	38'9	39'7	39'3	39'6	39'9	40'0
Geroko tenp.	38'1	38'0	38'3	38'3	37,9	38'7	38'3	37'0	36'9	39'0

Tenperaturaren arteko batezbesteko kendurarako 90%-eko konfidantza-tartea aurkitzea eskatzen da.

Kontutan har dezagun, $d_i = x_i - y_i$ kenduren taula:

Gaixoa	1	2	3	4	5	6	7	8	9	10
Kendurak	1'2	1'7	1'6	1'7	1	1	1	2'6	3	1

Kalkulu errazek, $\bar{d} = 1'58$, $s = 0'71$ eta $t_{\alpha/2;n-1} = t_{0,05;9} = 1,833$ direla erakusten digute.

Beraz, hauxe izango da konfidantza-tartea:

$$(1'58 \pm 1'833 \frac{0'71}{\sqrt{10}}) = (1'17 , 1'99)$$

VI.5.- LAGIN-TAMAINUAREN DETERMINAZIOA.

Aztertu dugun lehendabiziko adibidean, ondoko hauek adierazi ditugu: parametroen estimazioaren zehaztasuna, konfidantza-maila eta estimaturiko tartearen luzeraren menpean dagoela; konfidantza-maila handirako lortzen den tartearen luzera ere handia dela, tartearen luzera laburrei dagozkien konfidantza-mailak oso handiak ez diren bitartean; beraz, aurrefinkaturiko lagin-tamainurako zehaztasunik ez dela bi zentzutan irabazten.

Orain, gure ikuspuntua, laginaren tamainua "a posteriori" finkatzea da, "a priori" finkaturiko zehaztasun-mailaren arauera. Praktikoki, honela egiten da: onartzeko prest gauden errore-marjinaren eta arazoaren ezaugarrien arauera, konfidantza-maila aukeratzen dugu, eta, orduan, n determinatzen dugu, tartearen luzera, estimazioan nahi dugun helburuari komeni zaiona izan dadin.

Kontsidera dezagun, adibidez, bariantza ezaguneko banaketa normalaren batezbestekorako konfidantza-tartea. $1-\alpha$ konfidantza-maila aukeraturik, dagokion tartea $(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$ da, beheko muga $\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, goiko muga $\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ eta luzera $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ izanik.

Populazioaren batezbestekoaren eta lagin-batezbestekoaren arteko kendura, balio absolutuan, δ-z adierazitako kantitatea baino txikiagoa izateko, n laginaren tamainua determinatzean datzan arazoa planteatzen diogu gure buruari. Hots,

$$|\bar{x} - \mu| < \delta \iff \bar{x} - \delta < \mu < \bar{x} + \delta \quad (*)$$

Baina, probabilitate-terminoz, $(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$ zorizko tarteak μ-ren balioa izateko probabilitatea $1 - \alpha$ dela ba dakigu; alegia,

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \ .$$

Hau dela eta, $\delta = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ da.

Beraz, $n = (z_{\alpha/2})^2 \frac{\sigma^2}{\delta^2}$ da.

Orokorrean $(z_{\alpha/2})^2 . \frac{\sigma^2}{\delta^2}$ kantitatea zenbaki osoa ez denez, $(z_{\alpha/2})^2 . \frac{\sigma^2}{\delta^2}$ delakoaren berdina edo $(z_{\alpha/2})^2 . \frac{\sigma^2}{\delta^2}$ baino handiagoa den lehendabiziko zenbaki osoa hartu behar dugu.

(*) baieztapena ez da "ziurra"; ondokoa adierazten duen zentzu probabilistikoa daukana baizik: lagin-batezbestekoaren eta teorikoaren arteko kendurak δ balio ez gainditzeko probabilitatea $1-\alpha$ dela, edo maiztasun-terminoz esanda, $n=(z_{\alpha/2})^2 \cdot \sigma^2 / \delta^2$ heineko lagin-kopuru handian zehar, aldien $100(1-\alpha)$%-n, $|\bar{x} - \mu| < \delta$ izango da, eta gaineratiko $100\,\alpha$%-n, $|\bar{x} - \mu| > \delta$

6.adibidea:

Melanogaster eulien arteko gurutzamenduan, 300-en arteko hego bestialetako 60 euli lortu dira. Ondokoa eskatzen da:

a) Aurki ezazu, gurutzamendu-kopuru handian lorturiko aleen arteko hego bestigialetako euli-proportziorako 95%-eko konfidantza-tartea.

b) Gurutzamendu-kopuru handian lorturiko aleen arteko hego bestigialetako euli-proportzioa eta laginarena, balio absolutuan, 0'01 baino gutxiago diferitzeko 95%-eko probabilitatearekin, zenbat gurutzamendu burutu behar da?

Soluzioa:

a) Aztertzen dugun ezaugarria dikotomikoa da, p = "ezaugarri honen azterketari elkarturiko aldagai binomialaren proportzioa" parametrorako konfidantza-tartea eraiki behar dugularik.

Laginaren tamainua, n=300, aski handia denez, hurbilketa normala erabil daiteke; p-rako $1-\alpha=0'95$-eko konfidantza-tartea ondoko adierazpenak emana dago:

$$(\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$$

$\hat{p} = \frac{60}{300} = 0'2$; $n=300$; $z_{\alpha/2} = 1'96$ direnez, hauxe da tartea:

$$(0'2 - 1'96 \sqrt{\frac{0'2 \cdot 0'8}{300}} , 0'2 + 1'96 \sqrt{\frac{0'2 \cdot 0'8}{300}}) =$$
$$= (0'15474 , 0'2453)$$

b) Orain, konfidantza-maila fixurako, proportzioaren estimazio puntuala eta proportzio teorikoa, balio absolutuan, 0'01 baino gutxiago diferitzeko, laginaren tamainuaren determinazioaren arazoa planteatzen zaigu, hots, $|\hat{p} - p| < 0'01 \iff \hat{p} - 0'01 < p < \hat{p} + 0'01$ izatekoa.

Baina, 95%-eko konfidantzaz,

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

dela ba dakigu.

Beraz, $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0'01$

$$n = (z_{\alpha/2})^2 \cdot \frac{\hat{p}(1-\hat{p})}{0'01^2}$$

Kontutan har dezagun, n, $\hat{p}$ estimatzailearen menpean dagoela, berau n-ren menpean egonik. n-ren balioa ezin dela determinatu halabehartzen du honek.

Gurpil zoro hau apurtzeko era, $\hat{p}$-ri n_o=300 tamainuko aurretiko laginerako lorturiko p_o balioa ematean datza, zeren lagin handietarako lorturiko p-ren balioen "erregulartasuna" suposatzea aski arrazonagarria bait da.

Horrela, ondoko lagin-tamainua lortuko dugu:

$$n = 1'96^2 \cdot \frac{0'2 \cdot 0'8}{0'01^2} = 6146'56 \cong 6147$$

Alegia, 6147 gurutzamendu burutu behar ditugu.

PROBLEMAK

1.- Biztanlegoaren laginaren odoleango sedimentazio-indizeak, on doko taulan adierazten den arauera banatzen dira:

Indizeak	Maiztasunak
15 - 17	4
18 - 20	8
21 - 23	12
24 - 26	10
27 - 29	6

a) Kalkula itzazu laginaren batezbestekoa eta desbidazio standarda.

b) 95%-eko konfidantza-mailaz, estima itzazu populazioaren batezbestekoa eta desbidazio standadaren balioak.

2.- Hirikoa eta baserrikoa, bi mailatan sailkaturiko familien populazioa kontsideratzen dugu. Bira p_1 eta p_2, ezaugarria daukan maila bakoitzeko famili proportzioak. Proportzio hauek estimatzeko, maila bakoitzeko n_1 eta n_2 familiatan, laginketa itzuleraduna burutzen da, ezaugarri hori aurkez ten duen lagin bakoitzeko famili kopurua X_1 eta X_2-z adie razirik.

a) Zeintzu dira X_1 eta X_2-ren probabilitate-legeak?

b) $p_1=p_2=p$ dela suposatzen da, $\hat{p}(a)= a\frac{X_1}{n_1} + (1-a)\frac{X_2}{n_2}$ $(0 \leqslant a \leqslant 1)$ adierazpenak definituriko p-ren estimatzailea kontsidera tzen delarik.
Edozein a-tarako, egiazta ezazu $\hat{p}(a)$ zentratua dela.
Kalkula ezazu bere bariantza.
Determina ezazu, bariantza minimoa egiten duen a-ren ba lioa.

c) Orain, p_1 eta p_2 desberdinak izan daitezkeela suposatzen da.
$n_1 = 200$; $X_1 = 152$; $n_2 = 100$; $X_2 = 62$ emaitzen arauera, determina itzazu p_1 eta p_2-rako 99%-ko konfidantza-tarteak.

3.- Orain dela 10 urte egindako herrialde batetako pertsona heldu en altuerari buruzko zentsuan, batezbesteko altuera 1'70m. zela ikusi zen. Aztertu nahi dugun zorizko aldagaia banaketa normalari darraiola suposatzen dugu.

a) 101 pertsonatako lagina aukeratzen badugu, bertatik 1'80m.-tako batezbestekoa eta 0,1m.-tako desbidazio standarda lortu rik, eraiki itzazu oraingo populazioaren batezbestekorako 95%-eko eta 99%-ko konfidantza-tarteak.

b) Laginaren batezbestekoaren eta benetakoaren arteko kendura, edozein norantzatan, 1cm. baino laburragoa izateko probabilitatea 0,98 izan dadin, zenbatekoa izan behar du hartzen dugun laginaren tamainuak?

4.- 30 arraultzeren hazkunde-aldiaren batezbesteko iraupena, 37 egunetakoa izan da, bariantza 9 egunetakoa izanik. Erradiazioaren pean egon ziren 40 arraultzetako laginaren batezbesteko iraupena 38 egunetakoa izan zen, bariantza 15 egunetakoa izanik. 95%-eko konfidantza-mailaz, estima itzazu lehen dabiziko populazioaren desbidazio standarda eta bigarrenaren batezbestekoa.

5.- 10 partzela-bikote daude. Bikote bakoitzeko batek (1. partzelak) ongarri fosforikoa dauka, eta besteak (2. partzelak) ez, gaineratiko baldintzak berberak izanik. Laborearen etekina adierazten du ondoko taulak:

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

Etekinaren banaketa normala dela eta ongarriak etekinaren desbidazio standardean eraginik ez daukala onartzen da.

Kalkula itzazu ongarririk gabeko partzelen etekinaren bariantzarako 99%-ko konfidantza-tartea, eta batezbestekoen arteko kendurarako maila berbereko tartea.

Emaitza hauetatik, ongarri fosforikoaren eraginkortasuna onduriozta al daiteke?

6.- 1 desbidazio standardeko banaketa normalari darraion X zorizko aldagaiaren lagina, μ batezbestekoa estimatzeko erabiltzen da.

a) Determina ezazu, $\bar{x}$ lagin-batezbestekorako 95%-eko konfidantza-tartea.

b) Kalkula ezazu $p(\bar{x} - 2{,}58 \leqslant \mu \leqslant \bar{x} + 2{,}58)$. Suposa ezazu $n = 4$ dela.

c) Determina ezazu, $p(\bar{x} - 0{,}1 \leqslant \mu \leqslant \bar{x} + 0{,}1) = 0{,}99$ izateko, behar den behaketa-kopurua.

7.- Aurki ezazu, populazio normalaren batezbestekorako tartearen konfidantza-maila 90%-etik 99,9%-ra emendatzeko, behar den lagin-tamainuaren gehikuntza.

8.- Hurbil ezazu, $n_1 = 31$ tamainuko laginaren bariantza, $n_2 = 25$ tamainuko laginarena bi bider gutxienez izateko probabilitatea. Suposa ezazu bi laginak populazio normal berberetik eta era independentez aukeratzen direla.

9.- Ondoren azaltzen diren datuak, arrago desberdinetatik atera tako burdina urtuaren bi egoeratako tenperaturen arteko kendurak (°C-tan) dira:

Arragoa	Tenperaturen arteko kendurak								
1	50	90	100	90	100	80			
2	120	110	120	110	120	130	130	110	130

Aurki ezazu, batezbestekoen arteko kendurarako 95%-eko konfidantza-tartea.

10.- Azterketa egiterakoan, tentsio pean dauden umeen kalifikazioak, tentsiorik gabekoenak baino aldakorragoak izango direla kontsideratzen du psikologoak; zeren ume argiak, beldur izango diren ume atzeratuak baino hobeago moldatuko bait dira. Tentsiorik gabeko 11 umetako kontrol-taldeak 8'6 -tako desbidazio standarda daukan bitartean, tentsio pean dauden 16 umek 15-tako desbidazio standarda daukatela aurkitzen du.

Eraiki ezazu, bariantzen arteko zatidurarako 95%-eko konfidantza-tartea eta atera itzazu ondorioak.

11.- A(normal) eta B(adrenalizatua), bi arratoi-motatan egindako experimentuan, ondoko odol-biskositatea aurkitu zen:

Arratoiak	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

Determina itzazu, bi populazioen batezbestekoen arteko kendurarako eta bi bariantzen arteko zatidurarako 95%-eko konfidantza-mugak. Interpreta itzazu emaitzak.

12.- Tuberkulosiaren kausaz hildako 19 pertsonatako laginerako heriotza-adinak, 50 urtetako batezbestekoa eta 6 urtetako desbidazio standarda aurkezten ditu.

a) Normalitatearen hipotesia suposatuz, aurki ezazu μ batez besteko heriotza-adinerako 95%-eko kondidantza-tartea.

b) Estimaturiko batezbesteko adina eta benetakoa, 2 urte baino gehiagotan diferitzeko konfidantza-maila 95%-ekoa izateko, zenbatekoa izango da hartu beharko genukeen lagin-tamainua?

13.- Poisson-en banaketaren parametrorako konfidantza-tartea. Limitearen teorema zentralaren aplikazioa.

Biz X, λ parametroko Poisson-en banaketari darrion zorizko aldagai diskretua. Zorizko lagin bakuna hartzen badugu, limitearen teorema zentralaren ondorioz, $\bar{x}$ lagin-batez bestekoa banaketa normalari jarraitzen zaio asintotikoki. Hain zuzen ere, $N(\lambda, \sqrt{\lambda/n})$ banaketari, zeren $E(\bar{x}) = \lambda$ eta $Var(\bar{x}) = \lambda/n$ bait dira.

Adibidez, $1-\alpha = 0'95$ konfidantza-maila finkaturik, eta n aski handia ($n>30$) izanik,

$$p(-1'96 < \frac{\bar{x}-\lambda}{\sqrt{\lambda/n}} < 1'96) \approx 0'95$$

dela ba dakigu.

Kontutan har dezagun, λ ezin dela zuzenki bakandu. Horretarako, $1'96 = \frac{\bar{x}-\lambda}{\sqrt{\lambda/n}}$ eginez, $\lambda^2-(2\bar{x}+\frac{3'84}{n})\lambda+\bar{x}^2 = 0$ ekuazioa lortzen da.

Ekuazioa ebatziz:

$$\lambda = \bar{x} + \frac{1'92}{n} \pm \sqrt{\frac{3'84\bar{x}}{n} + \frac{3'69}{n^2}}$$

Horrela, tartearen goiko muga, $\bar{x} + \frac{1'92}{n} + \sqrt{\frac{3'84\bar{x}}{n} + \frac{3'69}{n^2}}$ da; baina, n^{-1} ordenako gaiak arburiatuz, $\bar{x} + 1'96\sqrt{\frac{\bar{x}}{n}}$ lortzen da.

Beheko mugarako kontsiderazio berberak eginez, ondoko 95%-eko konfidantza-tartea lortzen dugu:

$$(\bar{x} - 1'96 \sqrt{\frac{\bar{x}}{n}} \ , \ \bar{x} + 1'96 \sqrt{\frac{\bar{x}}{n}} \ )$$

Zein izango da $(1 - \alpha)$-ko konfidantza-tarte hurbildua?

14.- Gramo bat lurreko mikroorganismo-kopurua, λ parametroko Poisson-en banaketari darraiola suposatzen da. 20 tamainuko laginean, 4.330 mikroorganismo aurkitu dira. Kalkula ezazu, λ-rako 90%-eko konfidantza-tartea.

15.- Hoberen-erizpide berria. Egiantz handieneko estimatzailea.

Demagun 4 bola dauzkan kutxontzia. Zuriak eta beltzak direla ba dakigu, baina ez dugu proportzioa ezagutzen. p, bola zurien proportzioa baldin bada, p = 0, $\frac{1}{4}$, $\frac{2}{4}$, $\frac{3}{4}$ ala 1 izan daiteke.

Suposa dezagun, bi ateraldi itzuleradun burutzen ditugula, bola zuri bat eta beste beltz bat lorturik. Emaitza honetatik abiatuz, kutxontziaren konposaketa estimatu nahi dugu.

Egiantz handieneko metodoaren ideia zuzentzailea, kontutan harturiko gertaerari probabilitaterik handiena ematen diona dela oinarrizko populazioa onartzean datza. Hona hemen kasu posibleak eta dagozkien probabilitateak:

Kutxontziko bola zurien proportzioa (ezezaguna)	0	1/4	2/4	3/4	1
"Itzuleraz, bola zuri bat eta beste beltz bat" gertaeraren probabilit.	0	6/16	8/16	6/16	0

Gure kasuan, probabilitaterik handiena 8/16 da; beraz; egiantz handieneko metodoaren ideia zuzentzailea onartuz, kutxontziaren konposaketa bi bola zuri eta bi beltzetakoa dela estimatzen dugu; alegia, p = 1/2.

Orokorrean, biz $f(x,\theta)$ θ parametro ezezagunaren menpean dagoen populazioaren dentsitate-funtzioa. Bertatik, $(x_1,..,x_n)$ zorizko lagin bakuna aukeratzen dugu, lagin hori lortzeko probabilitatea handiena egiten duen θ-ren balioaz gure buruari galdetzen diogularik.

Lagina zorizkoa bada, bere dentsitate-funtzioa ondoko adierazpenak emanik dator:

$$L(\theta) = f(x_1,\theta)f(x_2,\theta)...f(x_n,\theta)$$

Hau da egiantz-funtzioa izenekoa eta θ parametroaren pean dagoena.

$L(\theta)$ egiantz-funtzioa maximoa egiten duen θ parametroaren estimazioa bilatzen dugu. Baina, $L(\theta)$ positiboa denez, $L(\theta)$-ren maximoak eta $\log L(\theta)$-renak batera datoz.

Beraz, $\frac{\partial \log L(\theta)}{\partial \theta} = 0$ ekuazioaren $\hat{\theta}$ edozein soluzioak $L(\theta)$ maximizatzen du, θ-ren egiantz handieneko estimatzailea deitua izanik.

Adibidez, $f(x,p) = p^x(1-p)^{1-x}$ $x=0, 1$ dentsitate-funtzioa daukan aurrekoarena bezalako Bernouilliren populaziorako p-ren egiantz handieneko estimatzailea,

$$\hat{p} = \frac{\text{arrakasta-kopurua}}{n} = \frac{\text{laginaren 1-kopurua}}{\text{laginaren tamainua}} \text{ da.}$$

Izan ere, hauxe da egiantz funtzioa:

$$L(p) = f(x_1,p)...f(x_n,p) = p^{x_1+..+x_n}(1-p)^{n-(x_1+...+x_n)}$$

$$\log L(p) = (x_1+...+x_n)\log(p) + \left[n-(x_1+...+x_n)\right]\log(1-p)$$

$$\frac{\partial \log L(p)}{\partial p} = \frac{x_1+...+x_n}{p} - \frac{n-(x_1+...+x_n)}{1-p} = 0$$

$$\hat{p} = \frac{x_1+...+x_n}{n} = \frac{\text{laginaren 1-kopurua}}{n}$$

Orain, onar dezagun animali koloniaren bizitza-aldia, $f(x,\beta) = \beta e^{-\beta x}$ dentsitate-funtzioko zorizko aldagai jarraia dela. Koloniatik zoriz aukeraturiko 10 animaliren heriotza-uneak behatu dira, ondoko aldiak (ordutan) lorturik:

5'1 - 5'3 - 4'8 - 6 - 5 - 5'5 - 6'1 - 4'9 - 5 - 5'3

Zein da β parametroaren egiantz handieneko estimazioa?

Lehenik, froga ezazu, n tamainuko laginaren egiantz-funtzioa ondokoa dela:

$$L(\beta) = \beta^n e^{-\beta(x_1+\ldots+x_n)}$$

Halaber, egiazta ezazu, $\frac{\partial \log L(\beta)}{\partial \beta} = 0$ ekuazioaren soluzioak, $\hat{\beta} = \frac{n}{x_1+\ldots+x_n}$ balioa duela.

Ondoriozta ezazu, β-ren egiantz handieneko estimazioa $\hat{\beta} = \frac{10}{53} = 0'188$ h. dela.

16.- (0, φ) tarteko banaketa uniformearen φ parametroaren estimazioa.

Suposa dezagun, X zorizko aldagai jarraiaren dentsitate-funtzioa $f(x) = 1/\varphi$ dela $(0, \varphi)$ tartean, φ parametroa ezezaguna izanik.

$(x_1,\ldots,x_n)$ n tamainuko laginaren bidez, φ estimatzeko arazoa planteatzen da.

Zein estimatzaile erabili behar duguneko galdera sortzen da.

Egiantz handieneko estimatzailea erabiltzeko erizpideari jarraituz, estimatzaile hori $U_1 = \max(x_1,\ldots,x_n)$ dela egiaztatzen da. (Frogapena ez da nabaria).

Azter ditzagun estimatzaile honen propietateak:

U_1-en banaketa-funtzioa hauxe da:

$$F_{U_1}(x) = p(U_1 \leqslant x) = p(\max(x_1,\ldots,x_n) \leqslant x) =$$

$$= p\{(x_1 \leqslant x) \cap \ldots \cap (x_n \leqslant x)\} = p(x_1 \leqslant x)\ldots p(x_n \leqslant x)$$

zeren $x_1,\ldots,x_n$ askeak bait dira.

Aurreko aldagai bakoitzak X-en banaketa-funtzioa berbera daukanez, hauxe ondorioztatzen da:

$$F_{U_1}(x) = \left(\frac{x}{\varphi}\right)^n \quad \text{baldin} \quad 0<x<\varphi$$

Beraz, ondokoa da dentsitate-funtzioa:

$$f_{U_1}(x) = \frac{n}{\varphi}\left(\frac{x}{\varphi}\right)^{n-1} \quad \text{baldin} \quad 0<x<\varphi$$

Hemendik:

$$E(U_1) = \int_0^{\varphi} x\,\frac{n}{\varphi}\left(\frac{x}{\varphi}\right)^{n-1} dx = \frac{n}{n+1}\varphi$$

$$\mathrm{Var}(U_1) = \int_0^{\varphi} x^2\,\frac{n}{\varphi}\left(\frac{x}{\varphi}\right)^{n-1} dx - \left(\frac{n}{n+1}\varphi\right)^2 = \frac{n}{(n+2)(n+1)^2}\varphi^2$$

Orokorrean, egiantz handieneko ideiari jarraituz eginda ko hoberen-erizpidea, eta zentratua eta bariantza txikienekoa dena, batera ez datozela ikusten da berehala, zeren eta, $E(U_1) = \frac{n}{n+1}\varphi$ denez, U_1 alboratua bait da.

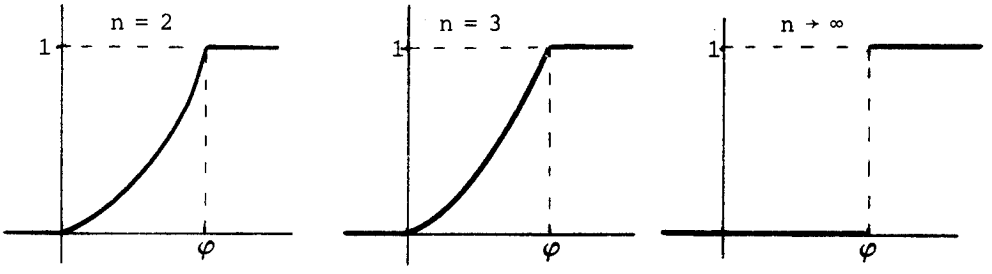
Hala ere, kontutan har dezagun, lagin-tamainua handia denean, $E(U_1) = \frac{n}{n+1}\varphi \underset{n\to\infty}{\longrightarrow} \varphi$ dela. Beraz, φ-ren estimazio onak lortzen dira.

φ-rako konfidantza-tartea eraiki nahi dugunean planteatzen da arazoa. U_1 asintotikoki normala izango balitz, limitearen teorema zentrala aplikatuz lortuko litzateke tartea, aurreko kasuan egin dugun antzera. Baina hau ez da daukagun egoera, zeren

$$F_{U_1}(x) \xrightarrow[n\to\infty]{} \begin{cases} 1 \text{ baldin } x = \varphi \\ 0 \text{ baldin } 0 < x < \varphi \end{cases}$$

betetzen bait da.

Ondoko irudiek ematen digute ideia grafikoa:



(n lagin-tamainuaren arauerako maximoaren banaketa-funtzioak)

Oharra: Maximoaren banaketa funtzioen limitea diskretua da!

$U_1 = \max(x_1,\ldots,x_n)$ estimatzailean oinarrituz, φ-rako konfidantza-tartea eman liteke, baina eztabaidan ez sartzea zuhurra dirudi.

Kontsidera dezagun, ordea, asintotikoki normala den banaketako $\bar{x} = \frac{x_1+\ldots+x_n}{n}$ estimatzailea (limitearen teorema zentrala).

$(0,\varphi)$ tarteko banaketa uniformerako, $E(X) = \frac{\varphi}{2}$ eta $Var(X) = \frac{\varphi^2}{12}$ direnez, $E(\bar{x}) = \frac{\varphi}{2}$ eta $Var(\bar{x}) = \frac{\varphi^2}{12n}$ direla frogatzen da errazki.

(Kontutan har dezagun, $\bar{x}$, U_1 baino estimatzaile txarragoa dela, alboratuagoa eta bariantza handiagokoa bait da).

Adibidez, φ-rako 90%-eko konfidantza-tartea lortu nahi badugu, aski handia den lagina hartuz eta hurbilketa normalaren bidez, hauxe lortzen dugu:

$$p(-1'645 < \frac{\bar{x} - \varphi/2}{\varphi/\sqrt{12n}} < 1'645) \approx 0'90$$

Eta, $\frac{\bar{x} - \varphi/2}{\varphi/\sqrt{12n}} = \pm 1'645$ ekuazioak ebatziz, ondoko 90%-eko konfidantza-tartea lortzen da:

$$\left(\frac{\bar{x}}{\frac{1}{2} + \frac{1'645}{\sqrt{12n}}} < \varphi < \frac{\bar{x}}{\frac{1}{2} - \frac{1'645}{\sqrt{12n}}}\right)$$

X-en 20 baliotako lagina aterako bagenu eta lorturiko lagin-batezbestekoa $\bar{x} = 3'2$ izango balitz, φ-rako 90%-eko konfidantza-tartea (5'28 , 8'12) izango litzateke.

17.- Parametroaren estimazio desberdinak, nola konbina?

Suposa dezagun, biztanlegoaren entzima-maila 45 unitatetako bariantza daukan banaketa normalari darraiola, batez besteko entzima-maila ezezaguna izanik.

Estimatzeko, independenteki lan egiten duten bi ikertzailek 10 pertsonatako lagin bana ateratzen dute biztanlegotik, $\bar{x}_1 = 22$ eta $\bar{x}_2 = 20$ lagin-batezbestekoak lorturik, hurrenez hurren.

μ-rako 95%-eko konfidantza-tartea lortu nahi da. Biek, ondoko tarteak lortuko dituzte:

$$(22 \mp 1'96\sqrt{\frac{45}{10}}) = (17'884 , 26'116)$$

$$(20 \mp 1'96\sqrt{\frac{45}{10}}) = (15'884 , 24'116)$$

Ikertzaile bien lanaz baliatuz, nork bere aldetik egiten duena baino batezbestekoaren estimazio hobeagorik lorote daiteke?

Beharbada, $\bar{x}_1$ eta $\bar{x}_2$-ren arteko batezbesteko balioa emangeniezaioke μ-ri; edo, orokorragoan, balio biei p_1 eta p_2

pisuak ($p_1 + p_2 = 1$) emanez, $p_1\bar{x}_1 + p_2\bar{x}_2$ har genezake μ-ren baliotzat.

Honek, ondoko arazo generikora garamatza:

Bira $T_1,\ldots T_n$, μ parametroaren estimatzaile zentratuak, $T = p_1T_1+\ldots+p_nT_n$ ($p_1+\ldots+p_n = 1$, $p_i \geqslant 0$) estimatzailea kontsideratzen dugularik.

Estimatzaile hau, aurrekoen konbinazio lineala da, μ-ren estimazio astatuak ematen dizkigularik.

Kontutan har dezagun, T ere zentratua dela:

$$E(T) = p_1E(T_1)+\ldots+p_nE(T_n) = (p_1+\ldots+p_n)\mu = \mu$$

edozein $p_1,\ldots,p_n$ pisuen baliotarako.

Beraz, T estimatzaileak, $T_1,\ldots,T_n$-k baino "aldakortasun" txikiagoa ematen baldin badu, T-ren erabilera bidezkoa izango da. (Kontutan har dezagun, T estimatzaileak kasu partikulartzat dauzkala $T_1,\ldots, T_n$, zeren $p_1 = 1$, $p_2=0,..,p_n=0$ baldin badira, $T = T_1$ bait da, eta abar).

Termino zehatzagoetan esandako galdera hauxe da: T-ren bariantza, zeintzu $p_1,\ldots p_n$ pisutarako da minimoa?

$Var(T_1) = Var(T_2) =\ldots= Var(T_n) = \sigma^2$ den kasuan ebatziko dugu arazoa.

Kasu honetan, $Var(T)= p_1^2\ Var(T_1)+\ldots+p_n^2Var(T_n)=(p_1^2+..+p_n^2)\sigma^2$

Azkenik, $\sum_{i=1}^{n} p_i= 1$, $p_i \geqslant 0$ $i=1,\ldots,n$ lotura-baldintzen menpeko $\sum_{i=1}^{n} p_i^2$ funtzioaren minimoa bilatu behar dugu.

Minimoa, $p_1=\ldots=p_n= \frac{1}{n}$ direnean lortzen dela adierazten digu Lagrange-ren biderkatzaileen metodoak.

Beraz, $E(T) = \mu$ eta $Var(T) = \frac{\sigma^2}{n}$ betetzen dituen $T = \frac{1}{n} T_1 + \ldots + \frac{1}{n} T_n$ estimatzailea erabiliko dugu.

Egiten ari garen adibiderako, ondokoa daukagu:

$$Var(\bar{x}_1) = \frac{45}{10} \qquad Var(\bar{x}_2) = \frac{45}{10}$$

Beraz, $T = \frac{\bar{x}_1 + \bar{x}_2}{2}$ eta $Var(T) = \frac{45}{20}$

Honez gainera, T estimatzaileak $N(\mu, \sqrt{\frac{45}{20}})$ banaketa betetzen duenez, T-an oinarritutako 95%-eko konfidantza-tartea hauxe izango da:

$$(21 \pm 1'96\sqrt{45/20}) = (18'060 \ , \ 23'940)$$

Ikusten denez, aurrekoak baino tarte hobeagoa da.

18.- Mikroskopioak, zelularen zatien neurketak burutzeko gauza den koadrikuladun okulare mikrometriko moldatua dauka. Jaukitzen diren zorizko erroreak, $\mu = 0$ batezbesteko eta desbidazio standard ezezaguneko banaketa normalari jarraitzen zaizkio.

n tamainuko laginetarako, $s'^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$ (x_i: i.neurketaren errorea) estatistikoa kontsideratzen dugu.

a) Froga ezazu populazioaren σ^2 bariantzaren egiantz handieneko estimatzailea dela. Zentratua al da?

b) Behaturiko ondoko erroreetatik abiatuz, kalkula ezazu σ^2-rako 95%-eko konfidantza-tartea:

0'3, 0'8, -0'6, 1'2, -0'7, -1'8, 0'2, 0, 2, -1'1

VII. GAIA: HIPOTESI-KONTRASTEA.

VII.1.- Sarrera eta adibideak.

VII.1.1.- Kontsiderazio orokorrak.

VII.1.2.- Bariantza ezaguneko banaketa normalaren batezbestekorako kontrasteari buruzko kontsiderazio intuikorrak.

VII.1.3.- Kontrastearen ahalmena.

VII.1.4.- II motako errorearen bornea garantizatzeko lagin-tamainuari buruzko kontsiderazioak.

VII.1.5.- Konfidantza-tarte eta hipotesi-testen arteko erlazioa.

VII.2.- Planteamendu orokorra.

VII.2.1.- Oinarrizko definizioak.

VII.2.2.- Populazioaren parametroari buruzko hipotesia kontrastatzeko lau urratsak.

VII.3.- Ohizko zenbait test.

VII.3.1.- Desbidazio standard ezaguneko banaketa normalaren batezbesteko testa.

VII.3.2.- Desbidazio standard ezezaguneko banaketa normalaren batezbeskorako testa. ($n<20$).

VII.3.3.- Populazio normalaren desbidazio standarderako testa.

VII.3.4.- σ_X^2 eta σ_Y^2 bariantza ezagunetako bi banaketa normal askeren batez bestekoen arteko berdintasunerako testa.

VII.3.5.- Bariantza ezezagunak berdinak direnean, bi banaketa normal askeren batezbestekoen arteko berdintasunerako testa. ($n_X < 20$, $n_Y < 20$).

VII.3.6.- Bariantza ezezagunetako bi banaketa normal askeren batezbestekoen arteko berdintasunerako testa. ($n_X < 20$, $n_Y < 20$).

VII.3.7.- Behaketak binaka emanik daudenean, banaketa normal biren batezbestekoen arteko berdintasunerako testa.

VII.3.8.- Banaketa normalari darrazkion bi populazio askeren desbidazio standarden arteko berdintasunerako testa.

VII.3.9.- Banaketa binomialaren parametrorako testa. ($n > 30$).

VII.3.10.- Bi banaketa binomial askeren parametroen arteko berdintasunerako testa. ($n_1 > 30$, $n_2 > 30$).

HIPOTESI-KONTRASTEA

VII.1.- SARRERA ETA ADIBIDEAK

VII.1.1.- Kontsiderazio orokorrak.

Ikerketa-metodo emankorrenetariko bat, "lan-hipotesiarena" da. Ikertzailea, alor berberean harturiko experientzia eta bere intuizioa direla kausa, fenomeno naturalei buruzko zenbait hipotesi formulatzen ausartzen da, berauek, hasierako hipotesia baieztatzeko ala errefusatzeko behar den beharretako informazioa lortzea baimentzen dion experientzi plangintzaren diseinuan gidatzen dutelarik.

Askotan, populazioaren parametro batetarako edo batzutarako balioak finkatzean datza hipotesia. Adibidez, biokimikariak, bere ezagupenetan eta intuizioan oinarriturik, odolplasmaren albumina-edukina, droga-dosi egokia aplikatuz, 20%-a gutxituko dela uste du. Beste ikertzaile batek, Ipar Europako emakumeen ziklo menstrualaren batezbesteko iraupena, herri tropikaletakoena baino luzeagoa dela suposatzen du. Eta abar...

Kasu hauetan guztietan, datu experimentalen arauera, ea hasierako hipotesia onartzen ala errefusatzen dugun, termino probabilistikoz ondorioztatzea baimentzen digun experientziatik harturiko zenbakizko informazioa aztertzeko metodo zientifiko egokiak ematen dizkigu Estatistika Induktiboak.

Gai honetan azalduko ditugun metodo edo erabideek, populazioen normalitatea eskatzen dute ea beti; eta, beren aplikaziorako, laginen batezbestekoak eta desbidazio standardak kalkulatu behar dira.

Kontrastearen arazoaren formulazio orokorra azaldu baino lehen, kasu konkretu eta ohizkoa aztertuko dugu zehatzmehatz. Geroago era abstraktuan definituko ditugun oinarrizko

kontzeptuak, azterketa honen barnekoak dira.

VII.1.2.- Bariantza ezaguneko banaketa normalaren batezbestekorako kontrasteari buruzko kontsiderazio intuikorrak.

σ desbidazio standard ezaguneko $N(\mu, \sigma)$ banaketari darraion populazioa emanik, suposa dezagun, $H_o: \mu = \mu_o$ hipotesia, $H_a: \mu \neq \mu_o$ aukeraren aurka kontrastatu nahi dugula, n tamainuko $(x_1,\ldots,x_n)$ laginaren bidez.

Intuikorki, honela arrazona dezakegu: $H_o: \mu = \mu_o$ hipotesia egiazkoa izango balitz, $\bar{x}$ lagin-batezbestekoa $N(\mu_o, \sigma/\sqrt{n})$ banaketari jarraituko litzaioke, edo, beste modu batetara esanda,

$$\frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} : N(0,1) \text{ izango litzateke.}$$

Hau dela eta, "gehienetan" $\bar{x}$-en balioak μ_o-rekiko "hurbil" (gehiegitzaz zein gutxiegitzaz) egongo direla eta "gutxitan" μ_o-tik asko diferituko direla espero izango dugu.

Beraz, ondoko jokabide-arau induktiboa hartuko dugu:

"Experimentalki, n tamainuko lagina aukeraturik, laginaren batezbestekoak μ_o-rekiko balio "hurbila" hartzen badu, hipotesia onartzen dugu. Lorturiko balioa, ordea, μ_o-tik "urrun" badago, $\mu = \mu_o$ deneko hipotesia errefusatuko dugu".

Aurreko araua onarturik, jakina, bi errore-mota jauki ditzakegu:

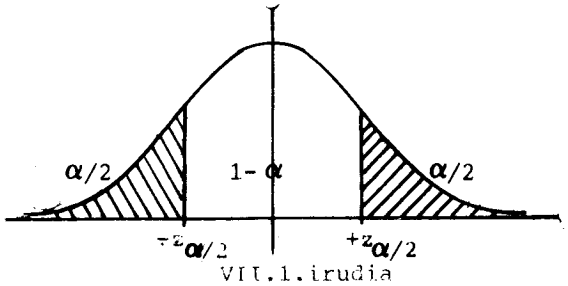
- I motako errorea: $\mu = \mu_o$ egiazkoa izanik, lorturiko $\bar{x}$-en balio experimentala μ_o-tik "urrun" dagoela eta, $\mu = \mu_o$ hipotesia errefusatzen dugu. Alegia, egiazkoa den hipotesia errefusatzen dugu.

- II motako errorea: $\mu = \mu_o$ faltsua izanik, $\bar{x}$-en balioa μ_o-tik "hurbil" dagoela eta, $\mu = \mu_o$ hipotesia onartzen dugu. Hots, hipotesi faltsua onartzen dugu.

Berehala, bi errore-moten magnitudearen idazkera (posible bada) eta "hurbil" eta "urrun" adierazpenen zehaztasuna eskatzen dira.

(Idazkerari buruzko oharra: $p(A/\mu = \mu_o)$ ikurraz, $\mu = \mu_o$ deneko hipotesiaren menpeko A gertaeraren probabilitatea adierazten dugu).

Suposa dezagun, I motako errorearen arriskua α ($\alpha>0$) dela aurrefinkatzen dugula. Errore hau, aldien 100α %-n jauki dezakegula adierazten du honek.



N(0,1) banaketaren ntsitate-funtzioa

VII.1.irudia

Biz $\mu = \mu_o$ hipotesiaren pean, N(0,1) banaketari darraion $Z = \dfrac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}$.

Determina ditzagun $z_{\alpha/2}$ eta $-z_{\alpha/2}$ balioak non

$$1-\alpha = p(-z_{\alpha/2} < Z < z_{\alpha/2} / \mu = \mu_o) = p\left(-z_{\alpha/2} < \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) =$$

$$= p\left(\mu_o - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu_o + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \text{ den.}$$

Hurrenez hurren, onarpen-eskualdea eta eskualde kritikoa deritzegun ondoko eskualdeak kontsideratzen ditugu:

$$S_o = \left(\mu_o - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \ \mu_o + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$S_1 = \bar{S}_o = \left(-\infty, \ \mu_o - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \cup \left[\mu_o + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \ +\infty \right)$$

Ondoko erara birplantea dezakegu arazoa:

"Experimentalki, n tamainuko lagina aukeraturik, $\bar{x} \in S_o$ (hots, μ_o-tik, $\frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ baino gutxiago diferitzen du) gertatzen bada, $\mu = \mu_o$ hipotesia onartzen dugu. $\bar{x} \in S_1$ (eskualde kritikoa) baldin bada, $\mu = \mu_o$ dela errefusatzen dugu.

Aurreko erregelaren arauera, kontutan har dezagun, egiazkoa den hipotesia errefusatzeko probabilitatea α dela; hots, $p(\bar{x} \in S_1 / \mu = \mu_o) = \alpha$. Horrela, I motako errorea jaukitzeko probabilitatea zeharo bornaturik dago.

Hala ere, $z_{\alpha/2}$ zenbakien ordez, beste u_1 eta u_2 zenbaki batzu jarriz, non

$$1 - \alpha = p(u_1 < z < u_2 / \mu = \mu_o) = p\left(u_1 < \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} < u_2\right) =$$

$$= p\left(\mu_o + u_1 \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu_o + u_2 \frac{\sigma}{\sqrt{n}} \right) \text{ den,}$$

S'_o beste onarpen-eskualde bat eta dagokion $S'_1 = \bar{S}'_o$ eskualde kritikoa determina genitzake.

Orain:

$$S'_o = \left(\mu_o + u_1 \frac{\sigma}{\sqrt{n}} , \ \mu_o + u_2 \frac{\sigma}{\sqrt{n}} \right) \text{ (onarpen-eskualde berria)}$$

$$S_1' = (-\infty, \mu_o + u_1 \frac{\sigma}{\sqrt{n}}] \cup [\mu_o + u_2 \frac{\sigma}{\sqrt{n}}, +\infty)$$ (eskualde kritiko berria)

eta berriro $p(\bar{x} \in S_1' / \mu = \mu_o) = \alpha$ lor genezake.

Onarpen-eskualde eta eskualde kritikoaren aukera zein puntutaraino den hautazkoa galde diezaiokegu gure buruari.

Oraingoz berari buruzko ezer esan ez duguneko II motako errorearen arriskuaren azterketak ematen digu planteaturiko galdera erantzuteko arrasto egokia.

VII.1.3.- Kontrastearen ahalmena.

S_o eta S_1 (onarpen-eskualdea eta eskualde kritikoa, hurrenez hurren) emanik, honela planteatzen da II motako errorearen probabilitatea:

p(II motako errorea) = p(faltsua den hipotesiaren onarpena) = $p(\bar{x} \in S_o / \mu \neq \mu_o)$

Probabilitate honen kalkulua, μ-ren menpekoa denez, I motako errorearen probabilitatearena baino zailagoa da.

Kontsidera dezagun probabilitate osagarria:

p(faltsua den hipotesia errefusatzea) =

$$= p(\bar{x} \in S_1 / \mu \neq \mu_o) = 1 - p(\bar{x} \in S_o / \mu \neq \mu_o) = P(\mu)$$

μ-ren edozein baliotarako definituriko $P(\mu)$ funtzioari, ahalmen-funtzioa deritzogu.

Ahalmena ez da balio fixurik, μ-ren funtzioa baizik.

I motako errorea jaukitzeko probabilitatea, α edo α baino txikiagoa deneko test posibleen (edo S_o, S_1 eskualdebikoteen) artean, ahalmena maximizatzen duen edo, hau da, II motako errorea jaukitzeko probabilitatea minimizatzen duen testa aukeratzea desiragarria da.

Termino formaletan esanda:

$p(\bar{x} \in S_1 / \mu = \mu_o)$ eta $p(\bar{x} \in S_1^* / \mu = \mu_o) \leqslant \alpha$ betetzen duten (S_o, S_1), (S_o^*, S_1^*) bi test emanik, (S_o, S_1) testa (S_o^*, S_1^*) baino <u>ahaltsuagoa</u> da, baldin

$$\beta(\mu) = p(\bar{x} \in S_1 / \mu \neq \mu_o) > p(\bar{x} \in S_1^* / \mu \neq \mu_o) =$$

$$= \beta^*(\mu) \quad \forall \mu \neq \mu_o$$

(S_o, S_1) testak beste edozein (S_o^*, S_1^*)-rekiko erlazio hau betetzen badu, <u>testik uniformeki ahaltsuena</u> da.

Beraz, uniformeki ahaltsuena den testa aukeratzea desiragarria izango litzateke; baina, kasu honetan, ez dago.

Ikus dezagun zergatia:

Demagun $S_o = (a_1, a_2)$ eta $S_1 = (-\infty, a_1] \cup [a_2, +\infty)$ motako (S_o, S_1) testa, $p(\bar{x} \in S_1 / \mu = \mu_o) = \alpha$ (*) beterik.

Gogora dezagun eraiketa-modua:

$\mu = \mu_o$ hipotesiaren pean, $\dfrac{\bar{x} - \mu_o}{\sigma / \sqrt{n}}$ kantitatea $N(0,1)$ banaketari darraio.

u_1 eta u_2 balioak aukera ditzakegu, non

$$p(u_1 < Z < u_2 / \mu = \mu_o) = p(u_1 < \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}} < u_2) =$$

$$= p(\mu_o + u_1\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu_o + u_2\frac{\sigma}{\sqrt{n}}) = 1-\alpha \text{ den.}$$

Ondoko testak, (*) baldintza betetzen du:

$$S_o = (\mu_o + u_1\frac{\sigma}{\sqrt{n}} \quad , \quad \mu_o + u_2\frac{\sigma}{\sqrt{n}})$$

$$S_1 = (-\infty, \mu_o + u_1\frac{\sigma}{\sqrt{n}}] \cup [\mu_o + u_2\frac{\sigma}{\sqrt{n}}, +\infty)$$

u_1 eta u_2 aukeratzeko hautazkotasunak, test-familia bukaezina ematen digu.

Ahalmen-funtzioak ondokoa balio du:

$$p(\bar{x} \in S_1 / \mu) = 1-p(\bar{x} \in S_o / \mu) = 1-p(\mu_o + u_1\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu_o + u_2\frac{\sigma}{\sqrt{n}} / \mu) =$$

$$= 1-p(\frac{\mu_o + u_1\frac{\sigma}{\sqrt{n}} - \mu}{\sigma/\sqrt{n}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{\mu_o + u_2\frac{\sigma}{\sqrt{n}} - \mu}{\sigma/\sqrt{n}} / \mu) =$$

$$= 1- \phi(u_2 - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) + \phi(u_1 - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) =$$

$$= P(\mu)$$

non ϕ funtzioa, N(0,1) banaketaren banaketa-funtzioa den.

u_1 eta u_2 balio aurrefinkatuetarako ahalmen-kurbak eraiki daitezke, banaketa normalaren taulen bidez.

Kurba guztiek, ondoko bi propietate komunak dauzkate:

i) Ahalmenaren definizio berberaz, 0 eta 1-en artean bornaturik daude.

ii) (μ_o, α) koordenatuak dituen puntutik pasatzen dira.

Izan ere, $P(\mu_o) = 1 - \phi(u_2) + \phi(u_1) =$

$= 1 - (1-\alpha) = \alpha$.

Azter ditzagun 3 kurba partikular:

a) $u_1 = -\infty$-rako, $P(\mu) = 1 - \phi\left(u_2 - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)\right)$ da.

Kasu honetan, $P(\mu)$ gorakorra da eta $\lim_{\mu \to -\infty} P(\mu) = 0$

eta $\lim_{\mu \to +\infty} P(\mu) = 1$ dira.

Kurba honi, C_1 deritzogu.

b) $u_2 = +\infty$-rako, $P(\mu) = \phi\left(u_1 - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)\right)$ da.

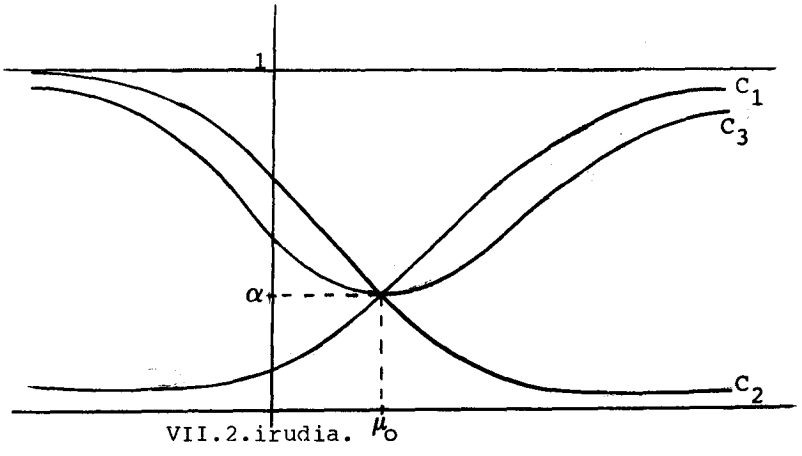
C_2-z adierazitako kurba hau beherakorra da 1-etik 0-ra.

c) $u_1 = -u_2 = -z_{\alpha/2}$ balioetarako,

$$P(\mu) = 1 - \phi\left(z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)\right) + \phi\left(-z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)\right) \text{ da.}$$

C_3 kurba honek, $\mu = \mu_o$ puntuan, minimoa dauka; $\mu = \mu_o$ zuzenarekiko simetrikoa da; eta ondokoa betetzen du:

$$\lim_{\mu \to -\infty} P(\mu) = \lim_{\mu \to +\infty} P(\mu) = 1$$



VII.2.irudia.

Ahalmenak konparatuz, μ_o-ren eskuinean, C_1 gaineratikoen gainean dagoenez, $\mu = \mu_o$ hipotesia $\mu = \mu_1$, $\mu_1 > \mu_o$ izanik, hipotesiaren aurka kontrastatzeko testik hoberena da. Era berean, $\mu_1 < \mu_o$ denean, C_2 kurbak ematen du testik hoberena. Halaber, kontutan harturiko test-familiaren artean, testik uniformeki hoberenik ez dagoela ikusten da.

Ohar ezazu oso test-kopuru txikia kontsideratu dugula; kopuru handia kontsideratuz, testik uniformeki ahaltsuena aurkitzea gerta liteke. Dena dela, adibide honetan, test posible guztiak kontsideratzen baditugu, testik uniformeki ahaltsuenik ez dagoela froga daiteke.

$$\left.\begin{aligned} S_o &= \left(\mu_o - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \mu_o + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\ S_1 &= \left(-\infty,\ \mu_o - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \cup \left[\mu_o + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ +\infty\right) \end{aligned}\right\} \quad (1)$$

(1) testari dagokion C_3 ahalmen-kurba aztertuz, μ_o-ren hurbil dauden μ-ren balioetarako $P(\mu)$ txikia dela eta μ_o-tik urrun daudenetarako, $P(\mu)$-k 1-erantz jotzen duela ikusten da. Beste hitz batzutan, μ benetako batezbestekoa, μ_o-ren "oso" desberdina denean, $H_o: \mu = \mu_o$ hipotesia errefusatzeko probabilitatea handia da. $H_o: \mu = \mu_o$ hipotesia, $H_1: \mu \neq \mu_o$ hipotesi alternatiboaren aurka kontrastatzeko aurreko testaren aukeraren arrazoia hau da, nahiz eta ahalmenik handienekoa ez izan.

Azkenik, kontutan har dezagun, μ_o-tik hurbil dauden μ-ren balioetarako, $P(\mu)$, α-ren ordena berberekoa dela, eta, beraz, $1-P(\mu)$ delakoaren berdina den II motako errorea jaukitzeko probabilitatea, $(1-\alpha)$-ren ordena berberekoa dela.

Lagin-tamainua, hots, informazio-kantitatea aldatu gabe, I eta II motako erroreen arriskuak ezin direla batera gutxitu adierazten du honek.

VII.1.4.- <u>II motako errorearen bornea garantizatzeko lagin-tamainuari buruzko kontsiderazioak</u>.

Aztertzen ari garen kontraste-arazoan, (1)-ez adierazitako testa aukeratzeko arrazoiak emanak ditugu. Test honetarako ahalmen-kurbak ondoko adierazpena ere daukala ikusi dugu:

$$P(\mu) = 1 - \phi\left(z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)\right) + \phi\left(-z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)\right)$$

n lagin-tamainua aldatzen badugu, n-rekin batera $P(\mu)$ ere aldatzen dela behatzen dugu, $P(\mu)$-ren ordez, $P(\mu,n)$, hots, bi aldagairen funtzioan idaztea egokiagoa izanik.

μ-ren balioak fixu dirauelarik, $P(\mu,n)$ n-rekin batera emendatzen dela ikusten dugu, $\lim_{n\to\infty} P(\mu,n) = 1 - \phi(-\infty) + \phi(-\infty) = 1$ izanik.

n lagin-tamainua egokiro emendatuz (I motako errorea jaukitzeko α probabilitatea mantenduz), ahalmenak baliorik hoberenak hartzen dituela adierazten du honek, edo bestela bate tara esanda, II motako errorearen posibilitatea gero eta txikiagoa dela.

Maila teorikoan, aurreko baieztapenak betegarriak dira; baina, praktikan, ondoko arazoa planteatzen da:

$P(\mu,n)$ testaren ahalmenak, zein lagin-tamainutatik gainditzen du aurrefinkaturiko balioa?

II motako errorearen probabilitateak aurrefinkaturiko bornea gaindi ez dezan, zenbat behaketa egin behar da?

Matematikoki: β txikia ($0 < \beta < 1$) finkaturik, n-ren zein baliotatik betetzen da $P(\mu,n) \geqslant 1-\beta$?

Edo, era baliokidean:

$$\phi\left(z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu-\mu_o)\right) - \phi\left(-z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu-\mu_o)\right) \leqslant \beta \quad (2$$

$$\frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}-\frac{\sqrt{n}}{\sigma}(\mu-\mu_o)}^{z_{\alpha/2}-\frac{\sqrt{n}}{\sigma}(\mu-\mu_o)} e^{-t^2/2}\,dt$$

adierazitako (2) desberdintza ebazteko zaila da, eta ebazpen hurbildua lor daiteke soilik.

Demagun, $\phi(z_\beta) = 1-\beta$ ($\Longleftrightarrow \phi(-z_\beta) = \beta$) betetzen duen z_β abzisako puntua dugula, eta, suposa dezagun, $\phi(-z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o))$, β-rekin konparatuz, txikia dela (hau egia da, $|\mu - \mu_o| \geqslant A > 0$ denean).

$\mu > \mu_o$ baldin bada:

$$\phi(z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) \leqslant \beta = \phi(-z_\beta).$$

Hemendik:

$$z_{\alpha/2} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_o) \simeq -z_\beta$$

Eta beraz:

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_o)^2}$$

$\mu < \mu_o$ baldin bada, honela idazten da (2) adierazpena:

$$1 - \phi(-z_{\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) - 1 + \phi(z_{\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) \leqslant \beta$$

Beraz:

$$\phi(z_{\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) - \phi(-z_{\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) \leqslant \beta = \phi(-z_\beta)$$

Hemendik:

$$\phi(z_{\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu - \mu_o)) \lesssim \phi(-z_\beta) \Longrightarrow z_{\alpha/2} + \frac{\sqrt{n}}{\sigma}(\mu - \mu_o) \simeq -z_\beta$$

$$\Longrightarrow n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{(\mu - \mu_o)^2}$$

Hortaz, edozein kasutan, $n \gtrsim \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu - \mu_o)^2}$ dela ikusten dugu.

Espero dugunez, lagin-tamainua μ-ren menpekoa da, $\mu \longrightarrow \mu_o$-rantz doan neurrian emendatzen delarik.

Ondorioz, $P(\mu, n) \geqslant 1-\beta \quad \forall \mu \neq \mu_o$ betetzen duen n-ren baliorik ez dago.

Baina, $\mu_1 \neq \mu_o$ aukeratzen badugu, orduan, $n \gtrsim \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_o)^2}$ hartuz, $P(\mu, n) \geqslant 1-\beta$ izango da, $|\mu - \mu_o| \geqslant |\mu_1 - \mu_o|$ betetzen duen edozein μ-ren baliotarako, hots, μ_o-tik $D = |\mu_1 - \mu_o|$ baino distantzia luzeagora dagoen edozein μ-ren baliotarako.

Laburbilduz, praktikan, σ ezaguneko $N(\mu, \sigma)$ banaketa normalaren μ batezbestekoa μ_o-ren berdina deneko hipotesia kontrastatzeko, ondokoak aurrefinkatzen dira:

i) I motako errorearen α arriskua

ii) II motako errerearen β arriskua

iii) $\mu_1 \neq \mu_o$ balioa

Orduan, $n \gtrsim \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_o)^2}$ lagin-tamainua hartzen da, ondoko testa eraikirik:

$$S_o = \left(\mu_o - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \; , \; \mu_o + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$S_1 = (-\infty\ ,\ \mu_o - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \cup [\mu_o + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ +\infty)$$

Adierazitako tamainuko zorizko lagina aukeratzen da populaziotik, $\bar{x} = \frac{x_1+\ldots+x_n}{n}$ laginaren estatigrafoaren balioa kalkulaturik. Balio hau, S_o eskualdearen barnekoa bada, $\mu = \mu_o$ hipotesia onartzen da; aldiz, S_1-en barnekoa bada, hipotesia errefusatzen delarik.

Horrela eginez, μ_o-tik μ_1-ek baino gehiago diferitzen duen edozein μ-ren baliotarako, I motako errorearen arriskuak ez duela α gainditzen eta II motakoarenak ez duela β gainditzen garantizatzen da.

α , β eta μ_1 aukeratzeko, zein erizpideri jarraitu behar zaio?

Aukera hartzera eramaten gaituen arrazoi matematikorik ez dago. Kontsideratzen den problemaren menpekoak diren erizpide guztiz extraestatistikoetan oinarritzen da aukera. Ondorioetan, segurtasun handia nahi badugu, α eta β oso txikiak eta μ_o-tik oso hurbil dauden μ_1-en balioak aukeratu beharko ditugu. Azkenik, ikertzailea da erabakitzen duena.

VII.1.5.- Konfidantza-tarte eta hipotesi-testen arteko erlazioa.

Konfidantza-tarteen teknika eta hipotesi-testen metodoa, inferentzia estatistikoaren arazoa ebazteko edo populaziotik ateratako laginak emandako datuetan oinarrituriko populazioari buruzko ondorioak ateratzeko bi era dira.

Populazioaren parametroa estimatzeko, bi metodoak dira aplikagarriak, elkar osotzen dezaketelarik.

Gure ezagupenak urriak badira, informazio handiagoa ematen digu konfidantza-tarteak eraikitzeak. Erabakia hartu nahi badugu, kontrastearen metodoa erabiltzea egokiagoa izan daiteke. Esate baterako, experimentatzen ari den medikamenduaren arrakastaren proportzioa estimatzea, eta, arrakastaren proportzioa balio determinatua baino txikiagoa den ala ez den arauera, X medikamendua merkaturik kentzea ala ez erabakitzea ez dira gauza bera.

Inferentzia estatistikoaren arazoa, populazioaren parametro baten (edo batzuren) estimazioan ez datzanean, testen metodoa erabil daiteke; konfidantza-tarteena, osterantzean, ez. Hau da geroago aztertuko ditugun zenbait normalitate-test eta doikuntz egokitasunaren kasua.

Aurreko adibidea erabiliko dugu zenbait analogia azaltzeko:

Gogora dezagun, α adierazgarritasun-mailako konfidantza-tartea, ondoko adierazpenak emanik zetorrela:

$$I_{\mu}^{1-\alpha} = (\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} , \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$$

I motako errorearen α arriskuko testa hauxe da:

$$S_o = (\mu_o - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} , \mu_o + \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$$

$$S_1 = (-\infty , \mu_o - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}] \cup [\mu_o + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} , +\infty)$$

$\mu = \mu_o$ izanik, $\bar{x} \leqslant \mu_o - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ edo

$\bar{x} \geqslant \mu_o + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ denean, I motako errorea jauki-

tzen dugu; hots, $I_{\mu}^{1-\alpha}$ konfidantza tarteak μ_o balioa ez daukanean.

Adierazgarritasun-maila eta I motako errorearen magnitudea batera datoz.

$\bar{x} \in S_o$ denean, hots, $I_{\mu}^{1-\alpha}$ konfidantza-tarteak μ_o balioa daukanean, $\mu = \mu_o$ dela onartzen dugu.

$\mu \neq \mu_o$ izanik eta $\bar{x} \in S_o \Longleftrightarrow \mu_o - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \bar{x} < \mu_o + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \Longleftrightarrow \bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu_o < \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ denean, II motako errorea jaukitzen dugu; hots, konfidantza-tarteak μ_o balioa (erroreduna) daukanean.

Konfidantza-tarteetan, estimaturiko tartearen luzera laburtu nahian, lagin-tamainua emendatu behar genuen. Hemen, lagin-tamainu handiagoak, testaren ahalmenaren gehikuntza ondorioztatzen du.

VII.2.- PLANTEAMENDU OROKORRA.

VII.2.1.- Oinarrizko definizioak.

Fisherrek hasitako eta Neyman-Pearsonek garaturiko hipotesi estatistikoen test-teoria, inferentzia estatistikoa bideratzeko beste metodologia bat da.

Hona hemen, teoria honen bidez zehaztu eta ebatzi nahi diren problemen zenbait adibide konkretu:

1) Patologi ikertzaileak, giltzurrun-afekzioa pairatzen duten pertsonen odol-seroaren batezbesteko albumina-portzentaia ea 70%-a edo 70%-a baino handiagoa den jakin nahi du.

2) Substantzia irradiaktiboaren atomoen batezbesteko bizitza-aldia "b" dela ea onar daitekeen erabaki nahi du ikertzaileak.

3) Zoologi ikertzaileak, zenbait ingurune-baldintza aldatuz, ea animali koloniaren ugalketa gehitzen den erabaki behar du.

4) Zenbait experimentu eta datutatik abiatuz, gizabanakoak tuberkulosia pairatzen duen erabaki behar du medikuak.

Eskematikoki, estatistikatzaileak honela heltzen die problema hauei:

H_o hipotesia, (adibidez, 4.probleman, gizabanakoa gaixorik dago) eta H_a alternatiba (gizabanakoa osasuntsua da) kontsideratzen ditu. Hipotesi hauen artean, zein den egiazkoa erabaki behar du.

Experimentu-multzo bat burutzen da, H_o egiazkoa bada, S_1 gertaera jazotzeko probabilitate oso txikia (adibidez, $p(S_1/H_o)=0,05$) dagoela jakinik.

Orduan, experimentu-multzo konkretua burutzerakoan, S_1 gertaera jazotzen bada, H_o errefusatzen dugu, ondorioz, H_a alternatiba onarturik. Justifikazioa hauxe da: Ondoko bi kasue tariko bat ondorioztatu behar dugu: hipotesia egiazkoa da eta 0,05-eko probabilitatea daukan gertaera jazo da; ala, hipotesia faltsua da eta errefusatu behar dugu. Lehena ez onartzea zilegi dirudienez, bigarrena onartzen da, H_o hipotesia errefusaturik.

Neyman-Pearsonen metodoaren ondoko adibideak egiaztatzen duenez, testa eraikitzeko orain arte egindako era intuikorra osotu beharra dago:

Demagun dato bat daukagula eta 1 ateratzeko p probabilitatea $p \geqslant \frac{1}{6}$ dela.

H_o: $p = \frac{1}{6}$ hipotesia kontrastatzeko, "datoa 50 aldiz jaurtikitzerakoan, 1 zenbakia inoiz ez lortzea" gertaera kontsideratzen dugu. H_o hipotesiaren menpeko gertaera honen probabilitatea

$$p(S_1/H_o) = (\frac{5}{6})^{50} \simeq 0,0001 = \alpha \text{ da.}$$

Experientzia burutzen da, 1 zenbakia lortzen ez delarik. Orduan, $p = \frac{1}{6}$ hipotesia errefusatzen dugu; baina, problemaren baldintzak direla kausa, $p > \frac{1}{6}$ hipotesia onartu behar dugu eta honek α baino txikiagoa den probabilitatea ematen dio S_1 gertaerari.

Beraz, H_o soilik kontsideratzearen interesa ikusi ordez, H_o eta H_a, hau da, biak kontutan hartzekoaréna ikusten da. Hau da Neyman-Pearsonek sorturiko ahalmen-nozioaren jatorria.

Atal honetan, populazioaren parametro bati edo batzuri dagozkien hipotesiak aztertuko ditugu soilik. Beste hipotesi-kontraste orokorragoak, geroago aztertuko dira.

Demagun, θ parametroaren menpeko $F(x,\theta)$ banaketa-funtzioko X zorizko aldagaiak deskribaturiko populazioa. Suposa dezagun, θ -ren balio posibleek osoturiko Ω multzoari "espazio parametrikoa" deritzogula. (Esate baterako, $X:N(\mu,1)$ baldin bada, μ-ren espazio parametrikoa $\Omega = \mathbb{R}$ da; $X:Bin(1,p)$ baldin bada, p-ren espazio parametrikoa $\Omega = [0,1]$ da; eta abar,...).

θ parametroari buruzko hipotesia egitea, Ω espazio parametrikoa, W eta $\Omega - W$ bi multzo baztertzailetan sailka-

tzea da; geroago, laginaren emaitzaren bidez, ea $\theta \in W$ (H_o hipotesi nulua) ala $\theta \in \Omega - W$ (H_a hipotesi alternatiboa) den erabakitzeko.

Ω espazio parametrikoa, tartea izaten da; halaber, W eta $\Omega - W$ ere tarteak dira.

W puntu bakar bat bada, H_o hipotesi nulua bakuna da eta kontrastea, aldebikoa. Kasu honetan, $H_o: \theta = \theta_o$ hipotesia, $H_a: \theta \neq \theta_o$ hipotesiaren aurka kontrastatzen da.

Osterantzean, hipotesi nulua konposatua da, eta kontrastea, aldebatekoa. Ondokoak dauzkagu:

$H_o: \theta \geq \theta_o$, $H_a: \theta < \theta_o$ hipotesiaren aurka (ezkerretiko aldebateko kontrastea).

$H_o: \theta \leq \theta_o$, $H_a: \theta > \theta_o$ hipotesiaren aurka (eskuinetiko aldebateko kontrastea).

Balio errealak hartzen dituen laginaren $T(x_1,\ldots,x_n)$ estatistikoa (lagin-batezbestekoa, lagin-bariantza, lagin-proportzioa, eta abar,...) aukeraturik, $H_o: \theta \in W$ hipotesia, $H_a: \theta \in \Omega - W$ hipotesiaren aurka kontrastatzeko testa honela definitzen dugu:

$T(x_1,\ldots,x_n)$-ren balio posibleak, S_1 (eskualde kritikoa) eta S_o (onarpen-eskualdea) bi eskualde osagarri eta baztertzailetan sailkatzen ditugu.

Laginari dagokion balioa lortzean, $T(x_1,\ldots,x_n) \in S_1$ bada, $H_o: \theta \in W$ hipotesia errefusatzen dugu; $T(x_1,\ldots,x_n) \in S_o$ bada, onartzen dugularik. Hots, testa, ondoko jokabide-arau induktiboaren baliokidea da:

$T(x_1,\ldots,x_n) \in S_o$ bada, $\theta \in W$ da, hots, H_o onartzen dugu.

$T(x_1,\ldots,x_n) \in S_1$ bada, $\theta \in \Omega - W$ da, hots, H_o errefusatzen dugu.

Bi errore-mota halabehartzen ditu honek:

- I motako errorea: hipotesia egiazkoa denean, berau errefusatzean datza, eta "testaren adierazgarritasun-maila" izeneko ondoko adierazpenak neurtzen du errore honi dagokion probabilitatea:

$$\sup_{\theta \in W} [p(T(x_1,\ldots,x_n) \in S_1 / \theta)]$$

- II motako errorea: faltsua denean, hipotesia onartzean datza, errorea jaukitzeko probabilitatea ondoko adierazpenak bornatua izanik:

$$\sup_{\theta \in \Omega - W} [p(T(x_1,\ldots,x_n) \in S_o / \theta)]$$

$P(\theta)$ testaren ahalmena, parametroaren balioa $\theta \in \Omega - W$ eskualdean egotearen menpeko H_o hipotesia errefusatzeko probabilitatetzat definitzen da.

Hau da:

$$P(\theta) = p[T(x_1,\ldots,x_n) \in S_1 / \theta \in \Omega - W]$$

Kontutan har dezagun,

$p[T(x_1,\ldots,x_n) \in S_o / \theta] = 1 - P(\theta)$ dela.

Hau da,

$$\sup_{\theta \in \Omega - W} [p(T(x_1,\ldots,x_n) \in S_o / \theta)] = 1 - \inf_{\theta \in \Omega - W} [P(\theta)]$$

Beraz, II motako errorea jaukitzeko probabilitatea txikitu nahi izatea, testaren ahalmena emendatzearen baliokidea da.

α adierazgarritasun-mailaren bidez, testaren hedadura (edo, I motako errorearen probabilitatea) bornatzean, baldintza hau betetzen duten testen artean, ahalmenik handiena daukana aukeratzean datza Neyman-Pearsonen ideia.

Adierazgarritasun-mailaren aukera, I motako errorearen arriskuak daukan garrantziaren menpekoa da. Normalean, $\alpha = 0{,}005$, $\alpha = 0{,}01$ edo $\alpha = 0{,}05$ balioak hartzen dira, emaitza oso adierazgarria, adierazgarria edo ia adierazgarria izanik hurrenez hurren.

(S_o, S_1) eta (S_o^*, S_1^*), α adierazgarritasun-mailari dagozkion H_o hipotesia kontrastatzeko bi test baldin badira eta

$$p(S_1/\theta) \leqslant p(S_1^*/\theta) \text{ baldin } \theta \in W$$

$$p(S_1/\theta) > p(S_1^*/\theta) \text{ baldin } \theta \in \Omega - W$$

betetzen badituzte, bigarrena baino ahaltsuagoa da lehena.

(S_o, S_1) testak, beste edozein testekiko aurreko erlazioak betetzen baditu, bera da testik uniformeki ahaltsuena.

VII.2.2.- Populazioaren parametroari buruzko hipotesia kontrastatzeko lau urratsak.

Populazioaren parametroari buruzko hipotesia kontrastatu nahi dugunean eman behar dugun lehen urratsa, I motako errorearen magnitudea finkatzea da. Esan dugunez, finkapen honek izaera extraestatistikoa du.

Bigarrenik, beraren banaketa ezagutzen duguneko lagin-estatistikoa aukeratu behar dugu, beronek probabilitateak kalkulatzea baimendu behar digularik. Estatistiko hau, kontraste-motaren menpean dago, orain arte aztertutako adierazpenetariko bat izanik.

Hirugarrenik, II motako errorearen magnitudea bornatu behar dugu, bornapena garantizatzeko behar den lagin-tamainua determinaturik. Aztertu dugun adibidean egiaztatu dugunez, hirugarren urrats honek oztopoak aurkezten ditu. Praktikan, test eta adierazgarritasun-maila bakoitzeko kurba-sorten ("eragiketa-kurba karakteristikoak" izenekoen) taula daukagu, zeinak I eta II motako erroreen borneak garantizatzeko behar den lagin-tamainua zuzenki "irakurtzea" baimentzen bait digu.

Azkenik, aurreko baldintzak aseguratzen dituzten onarpen-eskualdea eta eskualde kritikoa determinatu behar ditugu. Orokorrean, tarte edo tarte-bilduraren itxura daukate eskualde biek.

Aplikazio praktikoetan erabateko garrantzia eduki arren, ez gara hirugarren urratsaz arduratuko. Azaldutako adibidean egin dugunez, beraren forma analitikoa eta esangura agerian jartzea interesatzen zaizkigu. Eragiketa-kurba karakteristikoen sorten taulak izatea dela kausa, beraren burutzapenak ez du eragozpenik aurkezten.

Ondoren aztertuko ditugun zenbait test-mota, era eskematikora eta justifikatzen duen argudiorik gabe azalduko ditugu. Hala ere, beren garapen argumentala, lehenago aipaturiko adibidean burutu dugunaren antzekoa da.

VII.3.- OHIZKO ZENBAIT TEST.

VII.3.1.- Desbidazio standard ezaguneko banaketa normalaren batezbestekorako testa.

a) Aldebiko kontrastea: $H_o: \mu = \mu_o$

$H_a: \mu \neq \mu_o$

Proba-estatistikoa: $\dfrac{(\bar{x} - \mu_o)}{\sigma / \sqrt{n}}$

Testa: Onarpen eskualdea: $S_o=(-z_{\alpha/2}\ ,\ z_{\alpha/2})$

Eskualde kritikoa:

$$S_1=(-\infty,-z_{\alpha/2}]\ \cup\ [z_{\alpha/2},+\infty)$$

(Test guztiak, α adierazgarritasun-mailakoak direla suposatzen da).

b) Ezkerretiko aldebateko kontrastea: $H_o:\ \mu \geqslant \mu_o$

$$H_a:\ \mu < \mu_o$$

Proba-estatistikoa: $\dfrac{(\bar{x}-\mu_o)}{\sigma/\sqrt{n}}$

Testa: Onarpen-eskualdea: $S_o=(-z_\alpha\ ,+\infty)$

Eskualde kritikoa: $S_1=(-\infty,-z_\alpha\ ]$

c) Eskuinetiko aldebateko kontrastea: $H_o:\ \mu \leqslant \mu_o$

$$H_a:\ \mu > \mu_o$$

Proba-estatistikoa: $\dfrac{(\bar{x}-\mu_o)}{\sigma/\sqrt{n}}$

Testa: Onarpen-eskualdea: $S_o=(-\infty,z_\alpha\ )$

Eskualde kritikoa: $S_1=[\,z_\alpha\ ,+\infty)$

<u>1.adibidea</u>:

Osasunerako Zuzendaritza Orokorrean, pastilen B_{12} bitaminaren batezbesteko edukina 300 µg edo 300 µg baino handiagoa deneko hipotesia, 300 µg baino txikiagoa denekoaren aurka kontrastatu nahi dute.

Pastilak fabrikatzen dituen laborategiak, $\alpha=0,01$ oso txikia den I motako errorea (ekoizlearen arriskua) finkatzea lortzen du.

Beste aldetik, Inspekzioak ondoko erabide estatistikoa aplikatu nahi du: B_{12} bitaminaren edukinen populazioaren batezbestekoa 280 µg-takoa ($\mu_1=280$ µg) baldin bada, hipotesia onartzeko probabilitatea 0,05 izatea ($\beta=0,05$).

Produktua dosifikatzean eta pastilak egitean datzan metodoaren bidez buruturiko oso experientzia hedatuak, $\sigma=30$ µg dela azaldu digu.

Diseina ezazu, H_o: $\mu \geqslant 300$ µg hipotesia, H_a: $\mu < 300$ µg alternatibaren aurka kontrastatzeko erabide estatistikoa, $\alpha=0,01$ eta $\beta=0,05$ direla garantizaturik.

Egitez, lagin-tamainua determinatu behar dugu. Kasu honetan, hauxe da:

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_o)^2} = \frac{(2'326 + 1'645)^2 30^2}{(280-300)^2} = 35,46$$

Beraz, n=36 hartzen dugu.

Demagun, n=36 tamainuko zorizko lagina aukeratu dugula, $\bar{x}=290$ µg lorturik. Behaketa honetarako proba-estatistikoak ondokoa balio du:

$$\frac{(\bar{x} - \mu_o)}{\sigma / \sqrt{n}} = \frac{(290-300)}{30/\sqrt{36}} = -2$$

Onarpen-eskualdea, $S_o=(-z_\alpha, +\infty)=(-z_{0,01}, +\infty)=$

$=(-2'326, +\infty)$ da.

$-2 \in S_o$ denez, $\mu \geqslant 300$ µg hipotesia onartzen dugu.

Laginaren batezbestekoa $\bar{x}$=290 µg, μ_o=300 µg balio hipotetikoa baino txikiagoa izatearen zergatia, zoria izan daiteke, eta ez da μ<300 µg izatea.

VII.3.2.- Desbidazio standard ezezaguneko banaketa normalaren batezbestekorako testa. ($n \leq 20$).

a) Aldebiko kontrastea: H_o: $\mu = \mu_o$

H_a: $\mu \neq \mu_o$

Proba-estatistikoa: $\dfrac{(\bar{x} - \mu_o)}{s/\sqrt{n}}$

Testa: Onarpen-eskualdea: $S_o = (-t_{\alpha/2;n-1} , t_{\alpha/2;n-1})$

Eskualde kritikoa:

$S_1 = (-\infty, -t_{\alpha/2;n-1}] \cup [t_{\alpha/2;n-1}, +\infty)$

b) Ezkerretiko aldebateko kontrastea: H_o: $\mu \geqslant \mu_o$

H_a: $\mu < \mu_o$

Proba-estatistikoa: $\dfrac{(\bar{x} - \mu_o)}{s/\sqrt{n}}$

Testa: Onarpen-eskualdea: $S_o = (-t_{\alpha;n-1} , +\infty)$

Eskualde kritikoa: $S_1 = (-\infty , -t_{\alpha;n-1}]$

c) Eskuinetiko aldebateko kontrastea: H_o: $\mu \leqslant \mu_o$

H_a: $\mu > \mu_o$

Proba-estatistikoa: $\dfrac{(\bar{x} - \mu_o)}{s/\sqrt{n}}$

Testa: Onarpen-eskualdea: $S_o=(-\infty , t_{\alpha;n-1})$

Eskualde kritikoa: $S_1=[t_{\alpha;n-1} , +\infty)$

2.adibidea:

Itxura normaleko 15 pertsonatako laginean, seroaren amilasa-determinazioak egin ziren. Laginak, 96 unitate/100ml. -tako batezbestekoa eta 35 unitate/100ml.-tako desbidazio standarda aurkeztu zituen.

Banaketa normalari darraion populazioaren bariantza ezezaguna da.

Populazioaren batezbestekoa 120 unitate/100ml.-ren desberdina den ala ez ea ondoriozta daitekeen jakin nahi dugu, adierazgarritasun-maila $\alpha=0,05$ izanik.

Kasu honetan, $H_o: \mu=120$, $H_a: \mu\neq120$ alternatibaren aurka kontrastatu behar dugu (Aldebiko kontrastea).

H_o hipotesiaren menpean, $\frac{\bar{x} - \mu_o}{s/\sqrt{n}}$ proba-estatistikoa t_{n-1} banaketari darraio.

Adibide honetan, $-t_{\alpha/2;n-1} = -t_{0'025;14} = -2'1448$ eta $t_{\alpha/2;n-1} = t_{0'025;14} = 2'1448$ dira.

Beraz, $S_o=(-2'1448 , 2'1448)$ eta

$$S_1=(-\infty , -2'1448] \cup [2'1448, +\infty).$$

Laginerako estatigrafoak hartzen duen balioa hauxe da:

$$\frac{(\bar{x} - \mu_o)}{s/\sqrt{n}} = \frac{(\bar{x} - \mu_o)}{s'/\sqrt{n-1}} = \frac{(96 - 120)}{35/\sqrt{14}} = -2'5657$$

$-2'5657 \in S_1$ denez, $\alpha = 0'05$ adierazgarritasun-mailarako, $H_o: \mu = 120$ hipotesia errefusatzen dugu.

VII.3.3.- Populazio normalaren desbidazio standarderako testa.

a) Aldebiko kontrastea: $H_o: \sigma = \sigma_o$

$Ha: \sigma \neq \sigma_o$

Proba-estatistikoa: $\frac{(n-1)s^2}{\sigma_o^2}$

Testa: Onarpen-eskualdea:

$$S_o = (\chi^2_{1-\alpha/2;n-1} , \chi^2_{\alpha/2;n-1})$$

Eskualde kritikoa:

$$S_1 = [0, \chi^2_{1-\alpha/2;n-1}] \cup [\chi^2_{\alpha/2;n-1} , +\infty)$$

b) Ezkerretiko aldebateko kontrastea: $H_o: \sigma \geqslant \sigma_o$

$H_a: \sigma < \sigma_o$

Proba-estatistikoa: $\frac{(n-1)s^2}{\sigma_o^2}$

Testa: Onarpen-eskualdea: $S_o = (\chi^2_{1-\alpha;n-1} , +\infty)$

Eskualde kritikoa: $S_1 = [0 , \chi^2_{1-\alpha;n-1}]$

c) Eskuinetiko aldebateko kontrastea: $H_o: \sigma \leqslant \sigma_o$

$H_a: \sigma > \sigma_o$

Proba-estatistikoa: $\frac{(n-1)s^2}{\sigma_o^2}$

Testa: Onarpen-eskualdea: $S_o=[0\ ,\ \chi^2_{\alpha;n-1})$

Eskualde kritikoa: $S_1=[\ \chi^2_{\alpha;n-1}\ ,\ +\infty)$

3.adibidea:

Urin gastrikoan, drogaren disoluzio-aldien(segundotan) desbidazio standarda 1'6s. edo 1'6s. baino laburragoa dela suposatzen dugu. Hipotesi hau kontrastatzeko, ondoko erabide estatistikoa diseinatzen dugu:

Hipotesia egiazkoa bada, 0,95 da onartua izateko probabilitatea ($1-\alpha=0,95$).

Suposa dezagun, urin gastrikoango drogaren 90 disoluzio-alditako lagina aukeratu dugula, $s^2=2,8s^2$. lorturik.

Kasu honetan, eskuinetiko kontrastea, $H_o: \sigma \leqslant 1,6$, $H_a: \sigma > 1,6$ hipotesiaren aurka kontrastatzean datza.

Eta, $\alpha=0,05$ adierazgarritasun-mailarako onarpen-eskualdea $S_o=[0,\ \chi^2_{0,05;89}) = [0,112'022)$ da.

$\frac{(n-1)s^2}{\sigma_o^2} = \frac{89\ 2'8}{1'6} = 97,34 \in S_o$ denez, hipotesia onartzen dugu, aurrefinkaturiko errorea garantizaturik.

VII.3.4.- σ_X^2 eta σ_Y^2 bariantza ezagunetako bi banaketa normal askeren batezbestekoen arteko berdintasunerako testa.

a) Aldebiko kontrastea: $H_o:\ \mu_X=\mu_Y$

$H_a:\ \mu_X \neq \mu_Y$

Proba-estatistikoa: $\dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Testa: Onarpen-eskualdea: $S_o = (-z_{\alpha/2} , z_{\alpha/2})$

Eskualde kritikoa:

$$S_1 = (-\infty , -z_{\alpha/2}] \cup [z_{\alpha/2} , +\infty)$$

b) Ezkerretiko aldebateko kontrastea: H_o: $\mu_X \geqslant \mu_Y$

H_a: $\mu_X < \mu_Y$

Proba-estatistikoa: $\dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Testa: Onarpen-eskualdea: $S_o = (-z_{\alpha} , +\infty)$

Eskualde kritikoa: $S_1 = (-\infty , -z_{\alpha}]$

c) Eskuinetiko aldebateko kontrastea: H_o: $\mu_X \leqslant \mu_Y$

H_a: $\mu_X > \mu_Y$

Proba-estatistikoa: $\dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$

Testa: Onarpen-eskualde: $S_o = (-\infty , z_{\alpha})$

Eskualde kritikoa: $S_1 = [z_{\alpha} , +\infty)$

4.adibidea:

Adimen-atzeratuen tratamendurako ospitalean, mongolismoa pairatzen duten 12 gizabanakotako laginerako seroko azido urikoaren batezbesteko kontzentrazioa $\bar{x}_1$=4,5mg./100ml. izan zen.

Ospitale orokorrean, adin eta sexu berbereko 15 gizabanako normaletako laginaren batezbestekoa, $\bar{x}_2$=3'4mg./100ml. izan zen.

Kontzentrazio-populazio biak, 1 bariantzako banaketa normalari jarraitzen zaizkiola suposatuz, mongolismoa pairatzen dutenen eta gizabanako normalen seroko azido urikoaren batezbesteko mailen arteko desberdintasuna dagoela adierazten al digute datu hauek?

H_o: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 \neq 0$ hipotesiaren aurkako kontrastea burutzen dugu, α =0,05 izanik.

Onarpen-eskualdea $S_o = (-z_{\alpha/2} , z_{\alpha/2}) = (-1'96 , 1'96)$ da.

Halaber, proba-estatistikoaren balioa ondokoa da:

$$\frac{4'5 - 3'1}{\sqrt{\frac{1}{12} + \frac{1}{15}}} = 2,82 \notin S_o$$

Beraz, batezbesteko mailak berdinak ez direla ondorioztatzen dugu.

VII.3.5.- Bariantza ezezagunak berdinak direnean, bi banaketa normal askeren batezbestekoen arteko berdintasunerako testa. ($n_X < 20$, $n_Y < 20$)

a) Aldebiko kontrastea: H_o: $\mu_X = \mu_Y$

H_a: $\mu_X \neq \mu_Y$

Proba-estatistikoa:

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \sqrt{\frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X + n_Y - 2}}}$$

Testa: Onarpen-eskualdea:

$$S_o = (-t_{\alpha/2;n_X+n_Y-2} \ , \ t_{\alpha/2;n_X+n_Y-2})$$

Eskualde kritikoa:

$$S_1 = (-\infty, -t_{\alpha/2;n_X+n_Y-2}] \cup [t_{\alpha/2;n_X+n_Y-2}, +\infty)$$

b) Ezkerretiko aldebateko kontrastea: $H_o: \ \mu_X \geq \mu_Y$

$H_a: \ \mu_X < \mu_Y$

Proba-estatistikoa:

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \sqrt{\frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X + n_Y - 2}}}$$

Testa: Onarpen-eskualdea: $S_o = (-t_{\alpha;n_X+n_Y-2} \ , \ +\infty)$

Eskualde kritikoa: $S_1 = (-\infty \ , -t_{\alpha;n_X+n_Y-2}]$

c) Eskuinetiko aldebateko kontrastea: $H_o: \ \mu_X \leq \mu_Y$

$H_a: \ \mu_X > \mu_Y$

Proba-estatistikoa:

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \sqrt{\frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X + n_Y - 2}}}$$

Testa: Onarpen-eskualdea: $S_o=(-\infty , t_{\alpha;n_X+n_Y-2})$

Eskualde kritikoa: $S_1=[t_{\alpha;n_X+n_Y-2} , +\infty)$

VII.3.6.- Bariantza ezezagunetako bi banaketa normal askeren batezbestekoen arteko berdintasunerako testa. ($n_X < 20$, $n_Y < 20$)

a) Aldebiko kontrastea: H_o: $\mu_X = \mu_Y$

H_a: $\mu_X \neq \mu_Y$

Proba-estatistikoa: $$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Testa: Onarpen-eskualdea: $S_o=(-t_{\alpha/2;df'} , t_{\alpha/2;df'})$

Eskualde kritikoa:

$S_1=(-\infty,-t_{\alpha/2;df'}] \cup [t_{\alpha/2;df'},+\infty)$

b) Ezkerretiko aldebateko kontrastea: H_o: $\mu_X \geq \mu_Y$

H_a: $\mu_X < \mu_Y$

Proba-estatistikoa: $$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Testa: Onarpen-eskualdea: $S_o=(-t_{\alpha;df'} , +\infty)$

Eskualde kritikoa: $S_1=(-\infty , -t_{\alpha;df'}]$

c) Eskuinetiko aldebateko kontrastea: $H_o: \mu_X \leqslant \mu_Y$

$H_a: \mu_X > \mu_Y$

Proba-estatistikoa: $$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Testa: Onarpen-eskualdea: $S_o=(-\infty , t_{\alpha;df'})$

Eskualde kritikoa: $S_1=[t_{\alpha;df'} , +\infty)$

VII.3.7.- Behaketak binaka emanik daudenean, banaketa normal biren batezbestekoen arteko berdintasunerako testa.

Problema askotan, populazioaren elementuen gainean, bi faktoreren eraginak konparatzeko, komenigarria da bi faktoreak elementu bakoitzean aplikatzea, beren eraginen arteko diferentzia behaturik.

Adibidez, A eta B bi hipnotikoren eragina konparatu nahi badugu, egokia da gizabanako bakoitzari bi hipnotikoak ematea, zeren pertsona-lagin bat A hipnotikoaz, eta pertsona desberdinek osoturiko beste lagin bat, B hipnotikoaz tratatzen badira, lotorduen arteko desberdintasunaren zergatia ez bait datza hipnotikoen eraginean soilik, baizik eta gizabanakoen baldintza biologikoak (konstituzio genetikoa, elikadura, ohitura fisiologikoak, eta abar) ere izan bait daitezke kausa.

Beraz, bi faktoreak aplikatzen dira elementu bakoitzean, A faktoreak sorturiko zorizko aldagaia X-en bidez, eta B-k sorturikoa Y-ren bidez adierazten ditugularik.

μ_X eta μ_Y konparatzeko (hots, $\mu_X \geqslant \mu_Y$, $\mu_X \leqslant \mu_Y$ ala $\mu_X = \mu_Y$), D=X-Y kendurak kalkulatzen dira, $\mu_D \geqslant 0$, $\mu_D \leqslant 0$ ala $\mu_D = 0$ hipotesiak kontrastaturik, zeren X eta Y, $N(\mu_X, \sigma_X)$ eta $N(\mu_Y, \sigma_Y)$ banaketei jarraitzen zaizkien zorizko aldagaiak badira, D=X-Y zorizko aldagai berria $N(\mu_D = \mu_X - \mu_Y, \sigma_D)$ banaketari bait darraio, σ_D ezezaguna izanik. Beraz, D-ren batezbestekorako kontrastea, VII.3.2 atalaren kasu partikularra da, $\mu_o = 0$ izanik.

5.adibidea:

A eta B bi hipnotikoren eraginen arteko desberdintasuna egiaztatu nahi dugu. Ezagupen farmakologikoen arauera, A hipnotikoak B-k baino lotordu gehiago edo kopuru berbera sorterazten duela, batez beste, uste dugu.

$\mu_A \geqslant \mu_B$ hipotesia kontrastatzeko, ondoko erabide estatistikoa aplikatu nahi dugu: Hipotesia egiazkoa bada, 0'95 ($\alpha = 0'05$) probabilitatearekin onartua izatea.

Problema honen izaera dela kausa eta hipnotikoen eraginak bereizteko, 7 elementutako laginaren pertsona bakoitzari aplikatzen zaizkio bi hipnotikoak, sorterazitako lotorduen arteko kendurak kalkulaturik:

Pertsonak	A hipnotikoak sorterazitako X lotorduak	B hipnotikoak soterazitako Y lotorduak	D=X-Y	D^2
1	9	8	1	1
2	10	11	-1	1
3	12	9	3	9
4	11	10	1	1
5	8	9	-1	1
6	14	12	2	4
7	10	12	-2	4
Denetara			3	21

$$\bar{D} = \frac{3}{7} = 0'4286$$

$$s_D = \sqrt{\frac{1}{6}\left(21 - \frac{1}{7}\,3^2\right)} = 1'8127$$

Proposaturiko kontrastea hauxe da: H_o: $\mu_D \geqslant 0$

H_a: $\mu_D < 0$

VII.3.2 ataleko metodoak aplikatzen ditugu, σ_D ezagutzen ez dugu eta.

Proba-estatistikoak ondokoa balio du:

$$\frac{(\bar{D} - 0)}{s_D/\sqrt{7}} = \frac{0'4286}{1'8127/2'6458} = 0'6256$$

Onarpen-eskualdea: $S_o = (-t_{0,05;6}\ ,\ +\infty) = (-1'943\ ,\ +\infty)$

$0'6256 \in S_o$ denez, batez beste, A hipnotikoak B-k baino lotordu gehiago sorterazten duela onartzen dugu.

VII.3.8.- <u>Banaketa normalari darrazkion bi populazio askeren desbidazio standarden arteko berdintasunerako testa</u>.

a) Aldebiko kontrastea: H_o: $\sigma_X = \sigma_Y$

H_a: $\sigma_X \neq \sigma_Y$

Proba-estatistikoa: $\frac{s_X^2}{s_Y^2}$ (H_o hipotesiaren menpean, $F_{n_1-1;n_2-1}$ banaketari darraiona)

Testa: Onarpen-eskualdea: $S_o = (F_{1-\alpha/2;n_1-1,n_2-1}\ ,\ F_{\alpha/2;n_1-1,n_2-1})$

Eskualde kritikoa:

$$S_1 = [\ 0\ ,\ F_{1-\alpha/2;n_1-1,n_2-1}\ ] \cup [\ F_{\alpha/2;n_1-1,n_2-1},\ +\infty)$$

b) Eskuinetiko aldebateko kontrastea: H_o: $\sigma_X \leqslant \sigma_Y$

H_a: $\sigma_X > \sigma_Y$

Proba-estatistikoa: $\dfrac{s_X^2}{s_Y^2}$

Testa: Onarpen-eskualdea: $S_o = [0\ ,\ F_{\alpha;n_1-1,n_2-1})$

Eskualde kritikoa: $S_1 = [F_{\alpha;n_1-1,n_2-1}\ ,\ +\infty)$

Kontraste-mota honetan, adierazitako aldebateko kontrastea bakarra da. Beraz, oso kontu handiz ibili behar dugu idazkera hartzerakoan. σ_Y baino txikiagoa edo berdina den egiaztatu nahi dugun desbidazio standarda adierazi behar du σ_X delakoak.

6.adibidea:

Aurreko adibidean aipaturiko bi hipnotikoek, ea alda kortasun berbereko eraginak (lotorduak) sorterazten dituzten egiaztatu nahi dugu. Horretarako, hipotesia egiazkoa denean, bera errefusatzeko probabilitatea 0,05 (α=0,05) deneko prozedura diseinatu nahi dugu.

Suposa dezagun, 61 pertsona aukeratzen direla zoriz, eta A eta B hipnotikoak aplikatzen zaizkiola bakoitzari, sorterazitako lotorduen kuasibariantzak $s_A^2 = 3,28h^2$ eta $s_B^2 = 4,5h^2$ izanik.

H_o: $\sigma_A = \sigma_B$ hipotesia, H_a: $\sigma_A \neq \sigma_B$ hipotesiaren aurka kontrastatu behar dugu.

Kasu honetan, $S_o=(F_{0,975;60,60} , F_{0,025;60,60}) =$

$=(\frac{1}{1,6668} , 1'6668) = (0'5999 , 1'6668)$ da.

$\frac{s_A^2}{s_B^2} = \frac{3'28}{4'50} = 0,7289 \in S_o$ denez, bi hipnotikoek aldakortasun berbereko eraginak sorterazten dituztela onartuko dugu.

VII.3.9.- Banaketa binomialaren parametrorako testa. ($n>30$).

a) Aldebiko kontrastea: $H_o: p = p_o$

$H_a: p \neq p_o$

Proba-estatistikoa: $\frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$

Testa: Onarpen-eskualdea: $S_o=(-z_{\alpha/2} , z_{\alpha/2})$

Eskualde kritikoa:

$S_1=(-\infty , -z_{\alpha/2}] \cup [z_{\alpha/2} , +\infty)$

b) Ezkerretiko aldebateko kontrastea: $H_o: p \geqslant p_o$

$H_a: p < p_o$

Proba-estatistikoa: $\frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$

Testa: Onarpen-eskualdea: $S_o=(-z_{\alpha} , +\infty)$

Eskualde kritikoa: $S_1=(-\infty , -z_{\alpha}]$

c) Eskuinetiko aldebateko kontrastea: $H_o: p \leqslant p_o$

$H_a: p > p_o$

Proba-estatistikoa: $$\frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

Testa: Onarpen-eskualdea: $S_o = (-\infty , z_\alpha)$

Eskualde kritikoa: $S_1 = [z_\alpha , +\infty)$

7.adibidea:

4 haur baino gehiago dituen famili proportzioa ez dugu ezagutzen. Proportzio hau determinatzeko, 800 familiatako lagina aukeratzen dugu, bertan behaturiko proportzioa 0,18 izanik.

$H_o: p = 0,2$ hipotesi nulua, $H_a: p \neq 0,2$ alternatibaren aurka formulatzen dugu, $\alpha = 0,05$ adierazgarritasun-mailaz kontrastatu nahi dugularik.

Kasu honetan, $S_o = (-z_{\alpha/2} , z_{\alpha/2}) = (-1'96 , 1'96)$ da.

Lagin honetarako proba-estatistikoaren balioa hauxe da:

$$\frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{0'18-0'2}{\sqrt{\frac{0'2 \cdot 0'8}{800}}} = -1,41 \in S_o$$

Hau dela eta, $p = 0,2$ hipotesia onartzen dugu.

VII.3.10.- Bi banaketa binomial askeren parametroen arteko berdintasunerako testa. ($n_1 > 30$, $n_2 > 30$)

a) Aldebiko kontrastea: $H_o: p_1 = p_2$

$H_a: p_1 \neq p_2$

Proba-estatistikoa: $$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Testa: Onarpen-eskualdea: $S_o = (-z_{\alpha/2} , z_{\alpha/2})$

Eskualde kritikoa:

$$S_1 = (-\infty , -z_{\alpha/2}] \cup [z_{\alpha/2} , +\infty)$$

b) Ezkerretiko aldebateko kontrastea: $H_o: p_1 \geq p_2$

$H_a: p_1 < p_2$

Proba-estatistikoa: $$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Testa: Onarpen-eskualdea: $S_o = (-z_{\alpha} , +\infty)$

Eskualde kritikoa: $S_1 = (-\infty , -z_{\alpha}]$

c) Eskuinetiko aldebateko kontrastea: $H_o: p_1 \leq p_2$

$H_a: p_1 > p_2$

Proba-estatistikoa: $$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Testa: Onarpen-eskualdea: $S_o = (-\infty, z_{\alpha})$

Eskualde kritikoa: $S_1 = [z_{\alpha} , +\infty)$

8.adibidea:

Izurria dela kausa hil den hegazti gazteen proportzioa, hegazti helduena baino txikiagoa dela susmatzen du baserritarrak. Arazoa argitzeko, 200 hegazti gaztetako lagina isolatzen du, hauetatik, hilabetearen buruan, 58 hil direlarik; halaber, 150 hegazti heldu isolatzen ditu, aldi berean, 36 hiltzen direlarik.

5%-eko adierazgarritasun-mailaz, zein ondorio atera dezake baserritarrak?

H_o: $p_1 \leqslant p_2$ eta H_a: $p_1 > p_2$ hipotesiak konparatu nahi ditugu.

5%-eko adierazgarritasun-mailaz, hipotesi nuluaren onarpen-eskualdea $S_o=(-\infty , 1'64)$ da. Halaber,

$\hat{p}_1 = \frac{58}{200} = 0'29$; $\hat{p}_2 = \frac{36}{150} = 0'24$; $n_1=200$ eta $n_2=150$ dira.

Beraz, proba-estatistikoaren balioa $1'055 \in S_o$ denez, baserritarraren susmoa egiazkoa da.

OHARRA.- Bi azken test-moten izaera, aurrekoenaren desberdin samarra dela kontutan hartu behar dugu. Izan ere, H_o: $p = p_o$ hipotesia, H_a: $p \neq p_o$ alternatibaren aurka kontrastatzerakoan erabiltzen dugun

$$\frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

estatistikoaren banaketa ez da zehazki normala, gutxi gora behera baizik. (Lagin-tamainua emen datuz doan neurrian, hurbilketa gero eta hobea da).

Ondoko erreflexiora garamatza honek: Populazioaren zenbait parametrori buruzko hipotesiak kontrastatzeko lehenago azterturiko testek, populazioaren normalitatea halabehartzen

zuten beti; baina, populazioa normala izango ez balitz, parametro hauek kontrastatzeko testak diseinatu ahal izango genituzke? Proba-estatistikoak banaketa asintotikoki normala daukanean, hots, limitearen teorema zentrala betetzen denean, baiezkoa da erantzuna. Kasu honetan, onarpen-eskualdea eta eskualde kritikoa hurbilduak izango dira soilik, n aski handia hartuz, hurbilketa ona lorturik.

PROBLEMAK

1.- 20 eta 25 urteren arteko pertsonen taupada-kopuruak, minutuko 72-tako batezbestekoa eta 9-tako desbidazio standarda dauzka. 100 pertsonatako laginaren batezbestekoa, $\bar{x} = 64$ baldin bada, adierazgarritzat jo al daiteke lagin hau?

2.- Materia koipetsuaren edukina kontrastatu nahi duguneko gazta-erloak dauzkagu. Gazta-mota baten materia koipetsuaren edukina, $m_o = 30\%$ batezbestekoa daukan banaketa normalari darraio, eta gaineratikoena, $m > m_o$ betetzen duen m batezbestekoa daukanari, kasu bien desbidazio standarda 16% izanik.

H_o: $m = m_o \leqslant 30\%$ eta H_a: $m > 30\%$ hipotesien artean aukeratzeko, 16 gaztatako lagina aukeratzen da.

Ondokoa eskatzen da:

a) Adierazgarritasun-maila 0,02 dela jakinik, kalkula ezazu eskualde kritikoa.

b) Determina ezazu II motako errorearen arriskua.

c) Lorturiko laginaren batezbestekoa $\bar{x} = 40\%$ baldin bada, zein hipotesi aukeratuko dugu?

3.- Bonbila akasdunen kopurua kontrolatu nahi dugu. Biz p, bonbila akasdunen proportzioa. H_o: $p = p_o = 0,05$ eta H_a: $p < p_o$ hipotesien artean aukeratzeko, n = 144 tamainuko zorizko lagina hartzen da.

Ondokoa eskatzen da:

a) Adierazgarritasun-maila 0,05 dela jakinik, kalkula ezazu

eskualde kritikoa.

b) Aurki ezazu kontrastearen ahalmena, p-ren funtzioan.

4.- Pentsu-ekoizleak, beraren produktuak behien esnearen ekoizpena gehitzen duela baiezten du. Baserritarrak, 2.800l.-tako behi bakoitzeko batezbesteko ekoizpena lortzen zuen, 100 l.-tako desbidazio standardarekin. Pentsu berria erostea erabaki baino lehen, 81 behiri ematen zaie urtebetean zehar, behiko batezbesteko ekoizpena 3.000 l. esne izanik.

α= 0,01 mailaz, froga ezazu pentsu berria hobea deneko hipotesia.

5.- Ondoko taulak, 10 gaixok somniferoa hartuz irabazitako lotordu-kopurua adierazten du:

Gaixoa	1	2	3	4	5	6	7	8	9	10
Irabazitako lotordu-kop.	1,2	-1,3	0,7	0,2	3,4	0,8	3,1	1,8	2	3,1

Somniferoak lotordu-kopurua emendatzen duela egiaztatzen ote dute datu hauek?

6.- Eredu genetikoaren arauera, elkar baztertzaileak diren populazioko A, B eta C hiru karakterek ondoko probabilitateak aurkezten dituzte:

$p(A) = p^2$; $p(B) = pq$; $p(C) = q$

p eta q, p+q = 1 baldintza betetzen duten parametroak izanik.

120 aletako laginean, 32 alek A karakterea, 26 alek B karakterea eta 62 alek C karakterea aurkezten zutela behatu zen.

Ondokoak eskatzen dira:

a) Egiantz handieneko metodoaz, estima ezazu p parametroa.

b) α= 0,01 adierazgarritasun-mailaz, froga ezazu, datu hauek eta eredu genetikoak elkar komunztatzen dutela.

7.- Artikulu berberak ekoizten dituzten bi makinatatik lagin bana aukeratu zen, akasdunen portzentaiak berdinak zirenentz egiaztatzeko. Lehenetik, 600 unitate, eta bigarrenetik 900 hartu ziren, 45 eta 18 artikulu akasdun lorturik hurrenez hurren.

Emaitza hau, eta makina bien akasdunen proportzioak berdinak direneko hipotesia bateragarriak ote dira? Har ezazu α = 0,05

8.- Erreuma pairatzen duten gaixotako taldeari, antirreumatikoa eman zitzaion, gaixoen hasierako egoerari 4 balioa emanik. Medikamendua hartuz gero, ondoko emaitzak lortu ziren:

3 - 5 - 4,5 - 7 - 6 - 6,5 - 4 - 5,5 - 7 - 7

Batez beste, gaixoek onera egin zutela onar al daiteke?

α = 0,05 izanik, egizu testa, ondoko kasuetan:

a) σ = 2 dela eta normalitatea suposaturik.

b) σ ezezaguna dela eta normalitatea suposaturik.

9.- Suposa dezagun, urteroko herrialde batetan jausitako X euri-kantitatea, 30 batezbesteko eta 9 bariantzako banaketa normalari darraiola. Azken urteotan, zenbait klima-aldaketak urteroko euri-jaustean eragin dutela ematen du. Urtero jausitako euria gutxitu egin deneko hipotesia formulatzen dugu. Azken 9 urteotan, urteroko euri-jaustea neurtzen dugu, ondokoa lorturik:

24,1 - 33,7 - 27,4 - 31,1 - 30,9 - 35,2 - 28,4 - 32,1 - 26,2

a) 1%-eko adierazgarritasun-mailaz, zein izango da onartuko dugun hipotesia?

b) Lor ezazu β, μ-ren funtzioan.

β = 0,05 izateko, zein izango litzateke μ-ren balioa?

10.- Orri-mota baten luzera, σ=2 desbidazio standardeko banaketa normalari darraio, batezbestekoa, 35 mm. ala 40 mm.-takoa izan daitekeelarik. μ=35 hipotesia kontrastatzerakoan, I motako errorearen probabilitatea 0,05 eta II motakoarena 0,01 izateko, determina itzazu laginaren tamainua eta onarpen-eskualdea.

11.- Ilarrez buruturiko experimentuetan, 315 biribil eta horiak, 108 biribil eta berdeak, 101 luze eta horiak eta 32 luze eta berdeak lortu zituen Mendelek. Heredentziaren teoria mendeliarraren arauera, emaitzek, 9:3:3:1 proportzioetan egon beharko lukete hurrenez hurren.

1%-eko adierazgarritasun-mailaz, teoria honi buruzko zalantzarik ote dago?

12.- a) Espezie baten ar-portzentaia 0,2 deneko hipotesi nulua, 0,3 deneko hipotesi alternatiboaren aurka formulatu dugu. 100 tamainuko lagina aukeratzen dugu, laginean, 25 ar baino gehiago baldin badaude, hipotesia errefusatzen delarik.

Aurki itzazu I eta II motako erroreen arriskuak.

b) 30 eta 40 tamainuko lagin bi aukeratzen dira, 6 eta 10 ar aurkiturik hurrenez hurren. 1%-eko adierazgarritasun-mailaz, portzentaien arteko kendura adierazgarria ez dela onar al dezakegu?

13.- Urin gastriko astinduago drogaren disoluzio-aldien (segundotan) desbidazio standarda, 1,60 sg. edo 1,60 sg. baino labu rragoa dela suposatzen dugu. 90 disoluzio-alditako lagina aukeratu dugu, $s^2 = 2,8\ sg^2$ kalkulaturik.

$\alpha = 0,05$ adierazgarritasun-mailaz, kontrasta ezazu hipotesia.

14.- A eta B bi hipnotikok, aldakortasun berbereko eraginak (lotorduak) egiten dituztenentz egiaztatu nahi dugu. Suposa dezagun, hipnotikoak 7 pertsonatako bi talderi eman zaizkiela, ondoko emaitzak lorturik:

A hipnotikoaren bidezko lotorduak	B hipnotikoaren bidezko lotorduak
9	8
10	11
12	9
11	10
8	9
14	12
10	12

$\alpha = 0,01$ adierazgarritasun-mailaz, kontrasta ezazu hipotesia.

15.- Era intuikorrean egindako testik hoberenen determinazioa.

A eta B bi pertsonak, txanpon bat jaurtikitzen dute. 100 jokaldiren buruan, aurpegia aukeratu zuen A jokalariak, 62 aldiz irabazi du. Emaitza honen ondoren txanpona amainaturik dagoela baieztatzen du B jokalariak, aurpegia lortzeko probabilitatea, gutxienez, 2/3 izanik. A-k, txanpona zuzena dela esaten dirau. Nork dauka arrazoia? Aukera ezazu 0,05-eko adierazgarritasun-maila.

Biz p, txanpona jaurtikitzerakoan, aurpegia lortzeko probabilitatea.

H_o: p = 1/2 (A-k dauka arrazoia) eta H_a: p = 2/3 (B-k dauka arrazoia) hipotesien artean aukeratu behar dugu.

100 jokalditan lorturiko X aurpegi-kopurua, H_o egiazkoa bada, p = 1/2 eta n = 100 parametroetako banaketa binomialari darraio, eta, egiazkoa H_a bada, p = 2/3 eta n = 100 parametroetakoari. Banaketa binomial hauek, normalaren bidez hurbil ditzakegu. Hurbilketaren emaitzak, ondokora garamatza:

H_o: $X \equiv N(50,5)$

H_a: $X \equiv N(66'67,4'71)$

Intuikorki, $p(X \in W/H_o) = 0'05$ betetzen duen edozein W tarte, 0,05 adierazgarritasun-maileko eskualde kritikoa izango da. Baina, erraz ikusten denez, tarte hau zenbat eta ezkerragoan egon, testa gero eta ahaltsuagoa da, zeren eta, H_a egiazkoa denean, behaturiko X balioa bere barruan egoteko probabilitatea handiagoa bait da. Tarte hau, W = [a,+∞) motakoa izango da.

$p(X \geqslant a/H_o) = 0,05$ betetzen duen a balioa kalkulatu behar dugu.

H_o hipotesiaren pean, $X \equiv N(50,5)$ denez, ondokoa daukagu:

$$p\left(\frac{X-50}{5} \geqslant 1'645\right) = 0,05 \Longleftrightarrow p(X \geqslant 50 + 5 \cdot 1,645) = 0,05.$$

Beraz, a = 50 + 8'225 = 58'225 da, eskualde kritikorik hoberena W = [58'225,+∞) izanik.

Ondoriozta ezazu, 0,05-eko adierazgarritasun-mailaz, arrazoia B jokalariak daukala.

Halaber, egiazta ezazu testaren ahalmena handia dela: $P(p_1) = 0'945$.

16.- Egiantz-arrazoiaren testa.

Testen eraiketa-mota hau, egiantz handieneko estimatzaileenaren antzekoa da, Neyman-ek eriden zuelarik.

Demagun, $\varphi_1,\ldots,\varphi_k$ k parametrok definituriko $f(x, \varphi_1,\ldots, \varphi_k)$ funtzioa.

Suposa dezagun, n heineko $(x_1,\ldots,x_n)$ laginaren bidez, ondoko kontrastea burutu nahi dugula:

$$H_o : (\varphi_1,\ldots,\varphi_k) \in W$$

$$H_a : (\varphi_1,\ldots,\varphi_k) \notin W$$

non W, $(\varphi_1,\ldots,\varphi_k)$ parametroen posibleen Ω multzoaren azpimultzoa den.

Biz, $L(x_1,\ldots,x_n,\varphi_1,\ldots,\varphi_k)= f(x_1,\varphi_1,..,\varphi_k)\ldots f(x_n,\varphi_1,..\varphi_k)$ n tamainuko laginerako egiantz-funtzioa.

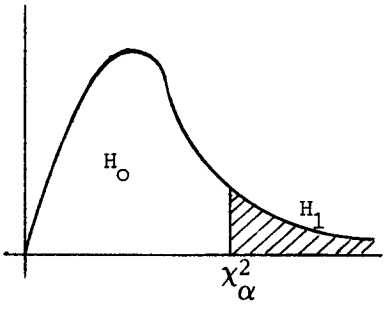
$(x_1,\ldots,x_n)$ lagina finkaturik, kontsidera ditzagun parametroak W-n daudeneko hipotesiaren (H_o hipotesiaren) pean, $L(x_1,\ldots,x_n,\varphi_1,\ldots,\varphi_k)$ funtzioak hartzen duen $L(\hat{W})$ balio maximoaren eta Ω-n hartzen duen $L(\hat{\Omega})$ balio maximoaren arteko zatidurari $\gamma = L(\hat{W})/L(\hat{\Omega})$ egiantz-arrazoia deritzogu.

$(x_1,\ldots,x_n)$ laginaren pean dagoen zorizko aldagaia den γ zatidurak, $0<\gamma\leqslant 1$ betetzen du. H_o egiazkoa bada, $L(\hat{W})$ eta $L(\hat{\Omega})$ hurbil egongo dira, eta, beraz, γ zatidurak 1-erantz jotzen du. Osterantzean, egiazkoa H_a bada, γ-k 0-rantz jotzen du. Orduan, γ-ren banaketa ezagutuz, α adierazgarritasun-mailarako, $k\in[0,1]$ konstantea aukeratu behar dugu, non $p(\gamma\leqslant k)= \alpha$ den.

Kontrasterako lortzen dugun testa hauxe da: $\gamma>k$ baldin bada, H_o onartzen dugu, eta, $\gamma\leqslant k$ baldin bada, H_o errefusatzen dugu, H_a onarturik.

Baina, askotan, ez dugu γ -ren banaketa ezagutzen. Hala ere, lagin handietarako ondoko teorema daukagu:

Wilks-en teorema: W-ren dimentsioa, r, eta Ω -rena k baldin badira, orduan, H_o hipotesia egiazkoa dela suposatu rik, $\chi^2 = -2\log\gamma$ estatigrafoa k-r askatasun-gradutako χ^2 banaketari jarraitzen zaio.



Praktikan, taulak ematen duen χ^2_α balioa bilatzen dugu.

$\chi^2 \leqslant \chi^2_\alpha$ baldin bada, H_o onartzen dugu; $\chi^2 > \chi^2_\alpha$ bada, H_a onartzen delarik.

Demagun $C_1,\ldots,C_k$ k zelulatan sailkaturiko populazio multinomiala. Alea, C_i zelularen barnekoa izateko probabilitatea p_i-z adierazten dugu. $\{\pi_1,\ldots,\pi_k\}$ zenbaki positiboak baldin badira, $\pi_1+\ldots+\pi_k=1$ izanik, zelula desberdinen p_i probabilitateak, $p_1=\pi_1,\ldots,p_k=\pi_k$ direneko hipotesia kontrastatu nahi dugu, n behaketak osoturiko zorizko lagin bakunaren bidez. Haien artean, C_1 klaseko $n_1,\ldots,C_k$ klaseko n_k daudela suposatzen baldin badugu, laginaren egiantz-funtzioa

$$L = p_1^{n_1} \ldots p_k^{n_k} \quad \text{izango da.}$$

Kasu honetan, Ω espazio parametrikoaren dimentsioa k-1 da, zeren $p_1,\ldots,p_k$, $p_1+\ldots+p_k=1$ baldintza betetzen duten zenbaki positiboak bait dira (hots, k-1 zenbaki "aske" daude soilik, zeren $p_1,\ldots,p_{k-1}$ ezaguturik, $p_k = 1-p_1-\ldots-p_{k-1}$ bait da); eta W-ren dimentsioa 0 da, $(\pi_1,\ldots,\pi_k)$ puntu bilakatzen bait da.

Berehala, $L(\hat{W}) = \pi_1^{n_1} \ldots \pi_k^{n_k}$ dela ondorioztatzen da. $L(\hat{\Omega})$ lortzea zailagoa da. Horretarako, kontutan har dezagun L-ren maximoak, logL-k dituen berberak direla.

Beraz, $p_1+\ldots+p_k = 1$ baldintzaren menpeko $\log L = n_1 \log p_1+\ldots+n_k \log p_k$ funtzioaren maximoak aurkitzean datza problema.

Lagrange-ren biderkatzaileen teknika erabiliz, $p_1 = \frac{n_1}{n}, \ldots, p_k = \frac{n_k}{n}$ direnean, logL-ren maximoa lortzen dela ondorioztatzen dugu,

$L(\hat{\Omega}) = \frac{1}{n^n} n_1^{n_1} \ldots n_k^{n_k}$ izanik.

Orduan, $\gamma = \frac{L(\hat{W})}{L(\hat{\Omega})} = n^n \left(\frac{\pi_1}{n_1}\right)^{n_1} \ldots \left(\frac{\pi_k}{n_k}\right)^{n_k}$

Eta, n handia bada, $-2\log\gamma$, k-1 askatasun-gradutako χ^2 banaketari darraio.

Adibidea: Suposa dezagun, 100 ilarretako lagina aukeratzen dela zoriz, lore zurikoak(Z) 20, horikoak (H) 54 eta berdekoak (B) 26 izanik.

Lore zuriak, horiak eta berdeak dauzkaten ilar-proportzioak $\frac{1}{4}$, $\frac{1}{2}$ eta $\frac{1}{4}$ direla hurrenez hurren baieztatzen al dute datu hauek?

Kasu honetan, Z, H eta B hiru zelulatan sailkaturik dago populazioa.

$p_1 = p(Z)$, $p_2 = p(H)$ eta $p_3 = p(B)$ izanik, H_o: $p_1 = \frac{1}{4}$, $p_2 = \frac{1}{2}$, $p_3 = \frac{1}{4}$ hipotesia kontrastatu nahi dugu.

Egiantz-arrazoiak ondokoa balio du:

$$\gamma = 100^{100} \left(\frac{1/4}{20}\right)^{20} \left(\frac{1/2}{54}\right)^{54} \left(\frac{1/4}{26}\right)^{26}$$

Hemendik:

$-2 \log \gamma = -2(100\log100 - 20\log80 - 54\log108 - 26\log104) = M$

$-2 \log \gamma$, χ^2_2 banaketari darraionez, $\alpha = 0,05$ adierazgarritasun-maila finkaturik,

$\chi^2_{0,05;2}$ = 5,991 lortzen da taula irakurtzean.

Lorturiko M balioa, 5,991 baino txikiagoa baldin bada, hipotesia onartzen dugu; osterantzean, errefusaturik gera tzen da.

17.- Ultzera gastrikoa diagnostikatzeko probari buruzko erreflexioak.

Ultzera gastrikoaren diagnostikoan maiz erabilitako pro ba, histamina administratu eta 15 minuturen buruan, urdaile ko azido askeren edukin osoa determinatzean datza. Proba hau, diagnostiko-erabidetzat baloratzeko, ondoko erara bildu ziren datuak:

Ultzera gastrikoaren sintomak aurkeztu zituzten 19 eta 30 urteren arteko 450 pertsonatako taldeari administratu zi tzaion histamina-dosi kinakorra, 15 minuturen buruan, azidotasun gastriko aske osoa determinatu zelarik. Gaixo haue tariko 200-engan, ikerketa erradiologikoaz eta geroko bila kaera klinikoaz baieztatu zen diagnostikoa. 250 kasutan, baieztatu gabe geratu zen diagnostikoa.

Aurkitutako azidotasun osoaren kantitateen (datu bakoi tza, urin gastrikoaren 4 laginetako batezbestekoan oinarri turik) arauerako bi talde hauen banaketa azaltzen da ondoko taulan:

Azidotasun osoa	Ez baieztaturiko diagnostikoa	Diagnostiko baieztatua
0 - 9	3	0
10 - 19	8	0
20 - 29	9	2
30 - 39	19	4
40 - 49	15	9
50 - 59	28	15
60 - 69	28	21
70 - 79	34	31
80 - 89	36	38
90 - 99	38	30
100 - 109	13	24
110 - 119	7	15
120 - 129	9	6
130 - 139	3	5
	250	200

a) Aipa itzazu, laginak aukeratzeko experientzian, aintzat hartu diren faktoreak.

b) Histaminaren probaren diagnostikorako interesaren balora penean, aukeratuak izan diren modura erabil al daitezke laginak?
Gizabanako normaletako konparazioarekiko, zein da azter keta honen konparazio-motaren abantaila?

c) Proba hau, ultzera gastrikoaren diagnostiko-eratzat go- menda al liteke?

d) Histamina administratu eta 15 minuturen buruan, 40 eta 100 unitateren arteko ultzera gastriko osoa daukan giza banakoaren aurrean, ultzera gastrikoaren diagnostikoa egitera jo al dezakegu?

18.- "Eupagurus prideauxi" delakoaren luzeraren azterketa.

Hamarratz ermitariren mota bati ("Eupagurus prideauxi" delakoari) buruzko azterketan, 2000 ale harrapatu ziren, bi talde nagusitan sailkatuak izanik:

a) Ur ez-sakonetako espezimenak (35m. baino gutxiagoko sakonerakoak).

b) Ur sakonetako espezimenak (35m. baino gehiagoko sakonerakoak).

Animali talde bakoitza, eme eta arretan sailkatu zen, gutxi gora behera 500 aletako 4 talde desberdin emanik. Erdiko lerroan zehar, hamarratzen oskolaren luzera neurtu zen. Lau taldetakoen oskolen luzeraren (mm.-tan eta 0,1mm.-z hurbilketarekin) maiztasun-banaketa aurkezten du ondoko taulak.

Oskolaren luzera	Ur sakonetako arrak	Ur sakonetako emeak	Ur ez-sakonetako arrak	Ur ez-sakonetako emeak
4,1	0	3	2	0
4,6	3	3	7	8
5,1	10	12	5	13
5,6	23	20	22	42
6,1	23	24	29	73
6,6	40	39	56	115
7,1	49	59	97	119
7,6	37	57	107	108
8,1	51	77	98	57
8,6	49	63	36	16
9,1	51	58	13	3
9,6	52	52	2	0
10,1	27	43	0	0
10,6	33	28	0	0
11,1	18	9	0	0
11,6	15	3	0	0
12,1	2	0	0	0
Denetara	483	550	474	554

i) Kalkula itzazu, lau maiztasun-banaketen batezbestekoa eta kuasibariantza, lau zifra hamartarretako hurbilketaz.

ii) Arren (bai ur sakonetakoen bai ez-sakonetakoen) oskolaren luzeren populazioek bariantza berbera daukatela suposatzen bada, determina ezazu bariantza komunaren estimatzailerik hoberena.
Egizu gauza bera emeetarako.

iii) 5%-eko adierazgarritasun-mailaz, egiazta ezazu bariantzen arteko berdintasunaren aurreko hipotesia.

iv) ii) atalean lorturiko estimazioa erabiliz, kalkula ezazu ondoko taldeen oskolen batezbesteko luzeren arteko kendurarako 95%-eko konfidantza-tarteak:

a) Ur sakonetako eta ur ez-sakonetako arrak.

b) Ur ez-sakonetako emeak eta arrak.

v) Bariantzen arteko berdintasuna suposatzen ez bada, nola aldatzen da a) atalean emandako konfidantza-tartea?

vi) 0,5%-eko adierazgarritasun-mailaz, kontrasta itzazu arren batezbesteko luzera, emeena baino luzeagoa deneko hipotesia eta populazio bien bariantzak berdinak direnekoa.

VIII. GAIA: JI KARRATU BANAKETAREN APLIKAZIOAK.

VIII.1.- Doikuntz egokitasunaren probak.

VIII.2.- Bi faktoreren arteko menpekotasun ala independentzi proba.

VIII.3.- 2 × 2 kontingentzi taulak. Yates-en zuzenketa.

VIII.4.- Homogenotasun-probak.

JI KARRATU BANAKETAREN APLIKAZIOAK

Estimazio eta hipotesi-probei buruzko gaietan, populazioaren bariantzarako konfidantza-tartearen eraiketan eta berari buruzko hipotesi-probetan, ji karratu banaketa aipatzen da laburki.

Aplikazio estatistikoetan erabilienetarikoa den banaketa honek beste erabilpen asko ditu, gai honetan, arruntenak diren ondokoak azaltzen direlarik.

Behaturiko maiztasun-banaketak, maiztasun-banaketa teorikoarekin elkar komunztatzen duenentz zehazten duteneko "doikuntz egokitasunaren" arazoak; bi aldagai askeak direnentz erabakitzen direneko arazoak; eta, azkenik, zenbait lagin aske populazio berberetik datozenentz kontrastatzen saiatzen gareneko arazoak.

VIII.1.- DOIKUNTZ EGOKITASUNAREN PROBAK.

Ondoko kasuetan erabiltzen dira: Populazioak banaketa-funtzio determinatua daukan edo gertaera posible guztiek probabilitate konkretuak dauzkaten hipotesia egiten dugunean, eta populaziotik ateratako laginaren bidez, ea hipotesia arrazonagarria den egiaztatu nahi dugunean.

Laginean lorturiko balioen banaketa, banaketa teorikoari zein puntutaraino egokitzen zaioneko estimazioan datzan prozedura erabiltzen dugu.

Planteamendu orokorrera heldu baino lehen, doikuntz adibide erraza aztertuko dugu:

1.adibidea:

1950.ean egindako azterketen arauera, 30 urtetako per

tsonen artean, 20%-ak goi-mailako ikasketak egin zituela, 50%-ak bigarren mailako ikasketen zertifikatua zeukala, baina goi-mailakorik ez, eta 30%-ak bigarren mailako eskolara heldu ez zela egiaztatu zen.

Oraingo populazioaren banaketa, 1950.ekoaren antzekoa al da?

Galdera honi erantzun ahal izateko, 30 urtetakoetatik zoriz aukeraturiko 1000 pertsonaren artean, azterketa burutzen da ondokoa behaturik:

250, goi-mailako eskolatan graduatuak dira
520, bigarren mailako eskolatakoak
230, bigarren mailako eskolara ez dira heldu

Ondoko hipotesiak formula daitezke:

H_o: 1950.eko banaketaren eta gaurkoaren arteko aldaketarik ez dago.

H_1: Gaurko populazioak (30 urtetako pertsonak) eta 1950.ekoak (1950.ean 30 urte zeuzkaten pertsonak) banaketa desberdinak dauzkate.

(Kontutan har dezagun, H_o, populazioaren banaketari buruzko hipotesia dela, eta ez parametroari buruzkoa, orain arte aztertu dugun bezala).

0,01-eko adierazgarritasun-mailako proba burutuko dugu:

Laginean behaturiko maiztasunak, O-z, eta itxarotakoak E-z adierazten ditugu. H_o egiazkoa izango balitz, hots, 1950.etik honantza banaketa aldatuko ez balitz, ondokoa itxarongo genuke:

Laginaren aleen 20%-ak, hots, 200 pertsonak goi-mailako ikasketak egin dituztela; laginaren aleen 50%-a bigarren mailako eskolako graduatua dela eta 30%-a ez dela bigarren mailako graduatua.

Ondoko erara aurkez daitezke datuak:

	O	E
Goi-mailako eskolatan graduatuak	250	200
Bigarren mailako eskolatan graduatuak	520	500
Bigarren mailako eskolatan ez graduatuak	230	300
	1000	1000

H_o hipotesia onartzeko ala errefusatzeko, zein erizpide erabili behar dugu?

Datuetatik abiatuz, nola erabakitzen dugu bi banaketen arteko desberdintasuna adierazgarria denentz?

$\sum(O-E) = \sum O - \sum E = 1000 - 1000 = 0$ denez, banaketen arteko desberdintasuna aztertzeko beste erizpide-mota bat ezarri behar dugu.

Horretarako, bi banaketen artean dagoen desberdintasuna adieraziko digun ondoko estatistikoa definitzen dugu:

$$\chi^2_p = \sum \frac{(O-E)^2}{E}$$ (p azpiindizea Pearson-i dagokio)

χ^2_p "oso handia" bada, desberdintasuna adierazgarria dela adierazten du; beraz, 1950.eko eta gaurko banaketak batera datozeneko hipotesia errefusatu behar dugu.

Hipotesia onargarria izan dadin, χ^2_p-ren balio maximoa kalkulatzeko, kasu bakoitzean determinatu behar ditugun askatasun-graduetako ji karratu banaketaz baliatzen gara.

Gure datuetarako χ^2_p-ren balio maximoa kalkulatuko dugu:

	O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
G-M.E.Gr.	250	200	50	2500	12,5
B-M.E.Gr.	520	500	20	400	0,8
B-M.E. ez Gr.	230	300	-70	4900	16,3
	1000	1000	0		29,6

Beraz, $\chi^2_p = 29,6$ da.

Horrela definituriko χ^2_p, χ^2_k banaketari darraionez, puntu kritikoa deritzogun $\chi^2_{\alpha;k}$ bilatzen dugu.

$\chi^2_p > \chi^2_{\alpha;k}$ baldin bada, 1950.eko eta gaurko banaketak berdinak direneko hipotesia errefusatzen dugu, desberdintasuna adierazgarria da eta (χ^2_p-ren balioa oso handia izan da).

χ^2_k-ren askatasun-graduak aukeratzeko, O edo E zutabeetako bi behaketa emanik, $\sum O = \sum E = n$ (laginaren tamainua) baldintza ezarriz, 3.behaketa ondorioztatzen dela kontutan hartu behar dugu.

Beraz, k = askatasun-graduak = 3 - 1 = 2 da.

$\chi^2_{2;0,01} = 9,21$ puntu kritikoa bilatzen dugu taulan.

$\chi^2_p = 29,6 > 9,21$ denez, H_o hipotesia errefusatzen dugu, hezkuntzaren arauerako 30 urtetako pertsonen gaurko banaketaren eta 1950.ekoaren arteko desberdintasun adierazgarria dagoela ondorioztaturik.

Komenigarria da azpimarratzea, behin-betiko erabaki orokorrak ez direla χ^2 testak. Aurrefinkaturiko arriskuz, zenbait hipotesi onartzeko ala errefusatzeko erizpide arrazonagarriak ditugu. Alegia, χ^2 testik abiatuz, desberdintasun adierazgarririk ez dagoela (edo, doikuntza egokia dela) baiezten dugunean, populazioak, suposatu dugun banaketa betetzen

duela matematikoki frogatu dugunik ez da adierazten, χ^2 testaren erizpidea erabiliz, ez dugula desberdintasunik ikusi baizik.

Honek ez du halabehartzen, beste erizpide baten bidez desberdintasuna aurki dezakegunik; are gehiago, χ^2 test berberaz, baina beste lagin desberdin bat erabiliz, desberdintasunak ager daitezke.

PLANTEAMENDU OROKORRA.

Demagun $C_1,\ldots, C_k$ k klase baztertzaile eta exhaustibotan sailkaturiko populazioa, klase hauen probabilitateak, $\sum_{i=1}^{k} p_i=1$ betetzen duten $p_1,\ldots,p_k$ direneko hipotesia kontrastatu nahi dugularik.

n elementu askek osoturiko lagina aukeratzen dugu zoriz, klase bakoitzerako itxarotako maiztasunak edo maiztasun teorikoak ondokoak izanik:

$e_1=np_1,\ldots, e_k=np_k$, $n=e_1+\ldots+e_k$ beterik.

Laginean behaturiko maiztasunak, $o_1,\ldots, o_k$-ren bidez adierazten baditugu, $o_1+\ldots+o_k=n$ izanik,

$$\chi^2_p = \sum_{i=1}^{k} \frac{(o_i-np_i)^2}{np_i} \quad (1)$$

adierazpena, (k-1) askatasun-gradutako χ^2 banaketari darraio.

(Askatasun-graduen kopurua, behaturiko maiztasun askeen kopurutzat jotzen da. Kasu honetan, k-1 maiztasun emanik $\sum_{i=1}^{k} o_i=n$ denez, beste maiztasuna determinaturik dago; beraz, k-1 dira askatasun-graduak).

Hau kontutan harturik, ea $np_1,\ldots,np_k$ maiztasun teorikoak onartzen ala errefusatzen diren kontrastatzeko, aski izango da k klaseen $o_1,\ldots,o_k$ maiztasunak zenbaturiko n ateraldi askek osoturiko lagina lortzea.

Gero, proba-estatistikoa den (1) adierazpenaren balioa kalkulatzen da, $\chi^2_{\alpha;k-1}$ puntu kritikoarekin konparaturik:

$\sum_{i=1}^{k} \frac{(o_i-np_i)^2}{np_i} \leqslant \chi^2_{\alpha;k-1}$ baldin bada, $C_1,\ldots,C_k$ k klaseen probabilitateak $p_1,\ldots,p_k$ direneko hipotesia onartzen da.

$\sum_{i=1}^{k} \frac{(o_i-np_i)^2}{np_i} > \chi^2_{\alpha;k-1}$ bada, hipotesi hau errefusatzen da, adierazgarritasun-maila α izanik.

Praktikan, χ^2-ren testa aplikatu ahal izateko, klase guztietarako itxarotako maiztasunek 5 baino handiagoak izan behar dute. Baldintza hau betetzen ez denean, ondoz-ondoko klaseak batzen dira, edozein i-tarako $np_i > 5$ lortu arte.

Honez gainera, askatasun-graduen determinazioa azpimarratzea interesatzen zaigu. Itxarotako maiztasunak edo maiztasun teorikoak lortzeko, c parametro estimatu behar baldin badira, parametro bakoitzaren estimazioak, o_i-en arteko erlazio berri bat ($\sum o_i=n$ erlazioaz gainera) ezartzen du, eta, beraz, estimaturiko parametro bakoitzeko unitate batez gutxitu egiten dira χ^2 banaketaren askatasun-graduak.

Azken batez, askatasun-graduen kopurua K-c-1 da, non k batu eta geroko klase-kopurua eta c estimaturiko parametro-kopurua diren.

2.adibidea:

Kultibo-tanta bat hedatzen denean, plaka koadrikulatuaren erretikulu bakoitzeko bakterio-kopurua, Poisson-en ba-

naketari darraion zorizko aldagaia deneko hipotesia kontrastatu nahi dugu (α=0,05 mailaz). Horretarako, kultiboaren disoluzio-tanta bat, 400 erretikulu txikitako plakan hedatu zen, mikroskopioaren laguntzaz, erretikulu bakoitzeko bakterio-kopuruak zenbatu zirelarik.

Lorturiko o_i emaitzak, ondoko taulan adieraztzen dira. Itxarotako maiztasunak, determinatu behar dugun λ parametroko Poisson-en banaketari jarraitzen zaizkio; o_i emaitzetatik abiatuz, lagin-batezbestekoaren bidez estimatuko dugu λ , hots,

$$\hat{\lambda} = \bar{x} = \frac{\sum_{i=1}^{k} x_i o_i}{n} = 4,68$$

	x bakterio-kopurua daukan erretikulu-kopuruak			
x bakterio-kopurua	Behaturikoak		Itxarotakoak	
0	0	20	3,71	21,08
1	20		17,37	
2	43		40,65	
3	53		63,41	
4	86		74,19	
5	70		69,44	
6	54		54,16	
7	37		36,21	
8	18		21,18	
9	10		11,02	
10	5	9	5,16	8,66
11	2		2,19	
12	2		0,86	
13	0		0,31	
14	0		0,10	
15	0		0,03	
16	0		0,01	
	400		400,00	

$\forall i\ x_i > 5$ izateko, taulan azaltzen den moduan bildu ditugu klaseak, 10 klase izanik. Beraz, askatasun-graduen kopurua, k-c-1=10-1-1=8 da.

Ondoko modu errezagoaz ere kalkula dezakegu proba-estatistikoaren balioa:

$$\chi^2_p = \sum_{i=1}^{k} \frac{(o_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{o_i^2}{np_i} - 2\sum_{i=1}^{k} o_i + \sum_{i=1}^{k} np_i =$$

$$= \sum_{i=1}^{k} \frac{o_i^2}{np_i} - n$$

Gure kasuan, kalkulu hauek burutuz, hauxe lortzen dugu:

$$\chi^2_p = 404,39 - 400 = 4,39$$

α=0,05-erako puntu kritikoa, $\chi^2_{0,05;8} = 15,507$ da; beraz, $4,39 < 15,507$ denez, erretikulu bakterio-kopurua, λ=4,68 parametroko Poisson-en banaketari darraion zorizko aldagaia dela onartzen dugu.

VIII.2.- BI FAKTOREREN ARTEKO MENPEKOTASUN- ALA INDEPENDENTZI PROBA.

Beharbada naharoena den χ^2 banaketaren beste erabilpen bat, bi faktore (klase edo mailatan sailkatuak), elementu-multzo berberari aplikatzen zaizkionean, askeak direneko hipotesi nulua kontrastatzea da; edo, beste modu batetara esanda, ea bi aldagai erlazionatuta dauden ala askeak diren aztertzeko erabiliko da χ^2 banaketa.

Adibidez, hautagai errepublikarrei botoa ematen dieten gizonezkoen portzentaia emakumezkoena baino handiagoa da, ala ez ote dago sexuaren eta partidu politikoen arteko erlaziorik? Familiaren diru-sarrera eta hiriko biztanleen bizilekua askeak al dira?

Hauetako arazo bakoitzean, baliagarria izango da χ^2 proba.

Arazo-mota hau ebazteko metodo estatistikoa, ondokoan datza: populazioaren elementuen gainean, laginketa burutzean, bi faktoreen maila gurutzatuak agertzen direneko maiztasunak zenbaturik; eta, maiztasun hauetan oinarriturik, faktoreen independentzia ea onartzen ala errefusatzen den erabakitzean.

"Kontingentzi taulak" izeneko tauletan edo koadroetan adierazten dira maiztasunak.

Adibidez:

		ADINA				
		5-10	10-15	15-20	20-25	...
ALTUERA	75-100					
	100-110					
	110-120					
	120-130					
	⋮					

Maila bien arteko ebakidura bakoitzari gelaska deritzogu.

Ondoko adibide errazaren bidez azalduko dugu metodo hau:

3.adibidea:

Ea adinaren eta tuberkulinaren aurkako erreakzioaren arteko menpekotasunik dagoen kontrastatu nahi dugu. Horretarako, experimentua burutu zen, adin desberdinetako 1969 pertsonari xiringatu zitzaielarik. Emaitzak, ondokoak izan ziren:

		B_1 Erreakzio positiboa	B_2 Erreakzio negatiboa	Baturak
Adinak	A_1 17-tik 19-ra	285	983	$1268 = f_{1.}$
	A_2 20-tik 24-ra	172	401	$573 = f_{2.}$
	A_3 25 baino gehiago	59	69	$128 = f_{3.}$
Baturak		$c_{.1}=516$	$c_{.2}=1453$	$1969 = N$

Behaturiko maiztasun hauek o_{ij}-z adieraziko ditugu.

Ondoko hipotesi nulua formulatzen dugu:

H_o:"Adina eta erreakzioa faktore askeak dira".

H_o egiazkoa bada, hau da, bi faktoreak askeak badira, $p(A_i \cap B_j)=p(A_i)p(B_j)$ da, eta, beraz, ondoko erara kalkula daitezke itxarotako maiztasunak:

$$e_{ij} = Np(A_i \cap B_j) = Np(A_i)p(B_j)$$

non N laginaren tamainua den; kasu honetan, 1969 da.

$p(A_i)$ eta $p(B_j)$ ezagutzen ez ditugunez, laginaren datuetatik estimatu beharko ditugu; hots:

$$p(A_i) \simeq \frac{f_{i.}}{N} \quad \text{eta} \quad p(B_j) \simeq \frac{c_{.j}}{N}$$

Beraz,

$$e_{ij} = N \frac{f_{i.}}{N} \frac{c_{.j}}{N} = \frac{f_{i.} c_{.j}}{N}$$

Kalkula ditzagun, adibide honetarako itxarotako maiztasunak:

$$e_{11} = \frac{1268 \cdot 516}{1969} = 332,29 \qquad e_{12} = \frac{1268 \cdot 1453}{1969} = 935,71$$

$$e_{21} = \frac{573 \cdot 516}{1969} = 150,16 \qquad e_{22} = \frac{573 \cdot 1453}{1969} = 422,84$$

$$e_{31} = \frac{128 \cdot 516}{1969} = 33,54 \qquad e_{32} = \frac{128 \cdot 1453}{1969} = 94,46$$

Emaitza hauek, parentesi artean, datu-taulan idazten dira, bakoitzari dagokion gelaskan. Gelaska guztietarako (independentziaren hipotesiaren menpean) itxarotako maiztasunak kalkulatuz gero, χ^2 banaketari darraion ondoko estatistikoa eraikitzen da:

$$\chi^2_p = \sum_i \sum_j \frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

Zenbatekoa da askatasun-graduen kopurua?

Askatasun-graduen kopurua, errenkada eta zutabeen batura osoak zehatzak direneko murrizketaren menpean jarri baino lehen, libreki bete daitezkeen gelaska-kopuruak emanik dator.

Adibidez:

	B_1	B_2	
A_1	235	*	1268
A_2	172	*	573
A_3	*	*	128
	516	1453	1969

* ikurra daukaten gelaskei dagozkien maiztasunak, total marginalak begiratuz ezagut ditzakegula ikusten da. Beraz, askatasun-graduen kopurua 2 da.

Orokorrean, n×m taularen askatasun-graduen kopurua, (n-1).(m-1) da.

Adibidera itzulirik, taularen datuetarako χ_p^2-k hartzen duen balioa kalkulatuko dugu:

$$\chi_p^2 = \sum_{i=1}^{3}\sum_{j=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{3}\sum_{j=1}^{2}\frac{o_{ij}^2}{e_{ij}} - N =$$

$$= 2008,61 - 1969 = 39,61$$

χ_p^2-ren balioa "oso handia" bada, behaturiko maiztasunak eta itxarotakoak (faktoreak askeak direla suposaturik) oso desberdinak direla pentsatzea zilegi dirudi; beraz, faktoreak menpekoak direla ondorioztatuko da.

χ_p^2-ren zein baliotaraino onartuko dugun independentzia jakiteko, $\chi_{\alpha;2}^2$ (non α adierazgarritasun-maila den) puntu kritikoa kalkulatzen dugu.

α=0,05 aukeratzen badugu, $\chi_{0,05;2}^2$=5,991 da.

39,61$>$5,991 denez, independentziaren hipotesia errefusatzen dugu 5%-eko adierazgarritasun-mailaz. Beraz, tuberkulinaren aurkako erreakzioa adinaren menpekoa da.

Orokorrean, k eta l, hurrenez hurren sailkaturiko ($A_1,\ldots,A_k$ eta $B_1,\ldots,B_l$) bi faktoreren arteko independentzia kontrastatu nahi dugunean, ondoko erara burutuko dugu χ^2 proba:

1) Azterturiko populaziotik, N tamainuko lagina aukeratuko dugu.

2) Bi faktoreen klase gurutzatu desberdinetarako behaturiko o_{ij} maiztasunak kalkulatuko ditugu, k×l kontingentzi taularen barruan dagozkien gelaskatan idatzirik. Taula honetan, errenkaden total marginalak ($f_{i.}$), zutabeenak ($c_{.j}$) eta N totala ere agertu beharrekoak dira.

B faktorea

A faktorea	B_1	...	B_j	...	B_l	Baturak
A_1	o_{11}	...	o_{1j}	...	o_{11}	$f_{1.}$
⋮	⋮		⋮		⋮	⋮
A_i	o_{i1}	...	o_{ij}	...	o_{il}	$f_{i.}$
⋮	⋮		⋮		⋮	⋮
A_k	o_{k1}	...	o_{kj}	...	o_{kl}	$f_{k.}$
Baturak	$c_{.1}$	...	$c_{.j}$	...	$c_{.l}$	N

3) Faktoreak askeak direla suposatzen dugunean, itxarotako e_{ij} maiztasunak kalkulatzen ditugu ondoko formularen bidez:

$$e_{ij} = \frac{f_{i.} c_{.j}}{N} \qquad \begin{matrix} i=1,\ldots,k \\ j=1,\ldots,l \end{matrix}$$

$f_{i.}$, A_i-ren total marginala eta $c_{.j}$, B_j-ren total marginala izanik.

e_{ij} maiztasun bakoitza, dagokion taularen gelaskan idatz daiteke parentesiren artean.

Bestalde, garrantzitsua da azpimarratzea, ezen, $e_{ij} < 5$ deneko kasuetarako, $e_{ij} > 5$ $\forall i$ $\forall j$ lortu arte, klaseak bildu behar direla. (Ikus ezazu erlazioaren 5.problema).

4) Taularen datuetarako proba-estatistikoak hartzen duen balioa kalkulatzen da:

$$\chi^2_p = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{o_{ij}^2}{e_{ij}} - N$$

Gogora dezagun, estatistiko hau $(k-1)\cdot(l-1)$ askatasun-gradutako ji karratu banaketari darraiola.

5) α adierazgarritasun-maila zehazten da eta $\chi^2_{\alpha;D}$ puntu kritikoa bilatzen da, $D=(k-1)\cdot(l-1)$ izanik.

6) Azkenik, ondoko erara arrazonatzen dugu:

$\chi^2_p > \chi^2_{\alpha;D}$ baldin bada, independientziaren hipotesia errefusatzen dugu, hau da, A eta B faktoreak menpekoak direla kontsideratuko dugu.

$\chi^2_p < \chi^2_{\alpha;D}$ baldin bada, independientziaren hipotesia onartzen dugu, hau da, A eta B faktoreak askeak direla kontsideratuko dugu.

VIII.3.- 2×2 KONTINGENTZI TAULAK. YATES-EN ZUZENKETA.

Batzutan, aztertzen ditugun bi faktoreak, bi klase edo bi mailatan sailka daitezke soilik. Datu gurutzatuak sailkatzerakoan, bi errenkada eta bi zutabetako kontingentzi taula lortzen da. Orokorrean, mota honetako taula, "2×2 taula" izenaz ezagutzen da, kasu partikular honetan, askatasun-graduen kopurua 1 delarik.

4.adibidea:

Ondoko 2×2 taulan aurkezten diren A eta B faktoreak askeak al dira?

B faktorea

A faktorea	B_1	B_2	
A_1	40	105	145
A_2	60	95	155
	100	200	300

$\chi^2_p = \sum_{i,j} \frac{(o_{ij}-e_{ij})^2}{e_{ij}}$ kalkulatzen dugu:

o_{ij}	e_{ij}	$(o_{ij}-e_{ij})$	$(o_{ij}-e_{ij})^2$	$(o_{ij}-e_{ij})^2/e_{ij}$
40	48,3	-8,3	68,89	1,426
105	96,7	8,3	68,89	0,712
60	51,7	8,3	68,89	1,332
95	103,3	-8,3	68,89	0,642
				$4,112 = \chi^2_p$

$\chi^2_{0,05;1} = 3,84$ da. Beraz, $4,112 > 3,84$ denez, A eta B faktoreak menpekoak direla ondorioztatzen da.

Arazo-mota honetan, banaketa diskretua (zenbaki osoa den maila bakoitzaren maiztasuna), χ^2 banaketa jarraiaren bidez hurbiltzen dugula kontutan hartu behar dugu. Beraz, jaukitzen dugun errore txikia zuzendu beharrekoa da.

Askatasun-graduen kopurua $D>1$ deneko kasuan, zuzenketa ez da beharrezkoa.

$D=1$ denean, $\chi^2_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(o_{ij}-e_{ij})^2}{e_{ij}}$ (1) estatistikoaren banaketa χ^2_1 delakoari darraio gutxi gora behera. Hau dela eta, ondoko proba-estatistiko berria erabiliko dugu:

$$\chi^2_Y = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|o_{ij}-e_{ij}| - \frac{1}{2})^2}{e_{ij}} \qquad (2)$$

(2) estatistikoa, orain arte erabilitako (1) estatistikoaren Yates-en zuzenketa da.

Yates-ek erideniko (2) estatistiko berri hau, χ^2_1 banaketarako hurbilketa hobea da. Zuzenketa honen funtzionamendua aztertzeko, 4.adibidea berrartuko dugu, adibide honen datuetarako ondokoa daukagularik:

o_{ij}	e_{ij}	$(o_{ij}-e_{ij})$	$\lvert o_{ij}-e_{ij}\rvert-0,5$	$(\lvert o_{ij}-e_{ij}\rvert-0,5)^2/e_{ij}$
40	48,3	-8,3	7,8	1,26
105	96,7	8,3	7,8	0,63
60	51,7	8,3	7,8	1,18
95	103,3	-8,3	7,8	0,57
				$3,64=\chi^2_Y$

Balio kritikoak, $\chi^2_{0,05;1} = 3,84$ izaten dirau.

$3,64 < 3,84$ denez, independentziaren hipotesia onartzen dugu; hots, A eta B faktore askeak direla ondorioztatzen dugu. Datu berberek, beste modu batetara azterturik, emaitza guztiz desberdinak eman ditzaketela ikusten da.

VIII.4.- HOMOGENOTASUN-PROBAK.

Lehenago azterturiko adibideetan zehar errepikatzen den ezaugarria, hauxe da: gelaska bakoitzerako behaturiko elementu-kopurua, lagina atera eta gero kalkulatzen dela; eta, beraz, total marginalak, ikertzaileak erabil ezin ditzakeen zorizko kantitateak (laginaren menpekoak direnak) direla. Kasu hauetan, aukeraturiko lagina, populazio bakar batetatik ateratako lagin bakartzat jotzen dugu.

Hala ere, batzutan, total marginalak (errenkada eta zutabeenak) kontrolatzea interesatzen zaio ikertzaileari. Kasu honetan, total hauen multzoa finkatuko du, horretarako, dagozkien populazioetatik lagin aske desberdinak aukeraturik.

Lehen prozedurak, independentziarako χ^2 probara garamatza. Bigarrenak, aldiz, homogenotasunerako χ^2 probara. Egoera biek, laginketa-prozedurak desberdinak izatez gain, hipotesi nulu desberdinetara garamatzate.

Gogora dezagun, independientzi probatako hipotesi nulua "A eta B faktoreak askeak dira" dela. Homogenotasun-probako hipotesi nuluak, aldiz, "laginak populazio berberetik ateratzen direla" baiezten du.

Kontzeptuari eta laginketa-prozedurari dagokienez, nabariak dira desberdintasunak. Hala ere, ondoko adibidean ikusten denez, proba biak berdinak dira kalkulu matematikoen aldetik.

5.adibidea:

Droga-erabiltzaileak zirela deklaratu zuten batxilergoko ikasleen arteko droga-erabilpenaren maila aztertzerakoan, ikertzaileak, erabiltzaile-talde honetatik, lehenengo ikasturteko 150 ikasle, bigarreneko 135 ikasle, hirugarreneko 125 ikasle eta azken ikasturteko 100 ikasletako lagina aukeratu zuen.

Ikasle bakoitzak, drogen ondoko erabilpen-mailak adierazten zituen galdeketa bete zuen: experimentala, kasuala eta moderatua izatetik bortitza izatera. Emaitzak ondokoak izan ziren:

Burutzen duten ikasturtearen arauera sailkaturiko 510 ikasleren drogen erabilpen-mailak:

		Experimentala	Kasuala	Moderaturik bortitzera	Baturak
Ikasturtea	1	57	50	43	150
	2	57	58	20	135
	3	56	45	24	125
	Azkena	45	22	33	100
	Baturak	215	175	120	510

Laginak aukeratzeko erak, hots, populazio bakoitzaren kopuru zehaztua aukeratzeak, taulako errenkaden totalak determinatzen ditu. Datu hauek eta lau populazioak drogen erabilpen-mailarekiko homogenoak direneko hipotesia bateragarriak al dira? Hau da, lau populazioek drogen erabilpen-mailarekiko banaketa berbera al dute?

Proba-estatistikoa, $\chi^2_p = \sum_{i,j} \frac{(o_{ij}-e_{ij})^2}{e_{ij}}$ da, eta, bere balioa kalkulatzeko, taulako gelaska bakoitzerako itxarotako maiztasunak kalkulatu behar ditugu.

Izan ere, populazioak homogenoak baldin badira, edo era baliokidean, laginak, drogen erabilpenarekiko populazio berberetik aukeratzen badira, drogak maila experimentalean soilik erabili dituztenek osoturiko populazioko proportzioaren estimaziorik hoberena, $\frac{215}{510}$ = 0,4216 da (zeren populazio berberaren barruan, lagina zenbat eta handiagoa izan, informazioa hainbat eta hobea izango bait da).

Lau populazioak homogenoak baldin balira, probabilitate hau, populazio bakoitzari indibidualki aplikatuko balitzaion interpretatzen da. Adibidez, hipotesi nuluaren menpean, lehenengo ikasturtea osotzen duten 150 ikasleetatik $\frac{215}{510} \cdot 150 = 63,24$ ikasle, erabiltzaile experimentalak lirateke eela aurkitzea espero genuke.

Berriro, $e_{ij} = \frac{f_{i.} c_{.j}}{N}$ dela ikusten da, non $f_{i.}$=i. errenkadari dagokion total marginala, $c_{.j}$=j.zutabearen total marginala eta N=laginen elementu-kopuru osoa diren.

Itxarotako maiztasun hauek kalkulaturik, parentesi artean idazten dira taulan, eta ondoko proba-estatistikoa kalkulatzen da:

$$\chi^2_p = \sum_{i,j} \frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \frac{(57-63,24)^2}{63,24} + \ldots + \frac{(33-23,52)^2}{23,52} = 19,4$$

Askatasun-graduen kopurua, D = 3×2 = 6 da; beraz, α =0,005 adierazgarritasun-mailarako puntu kritikoa $\chi^2_{0,005;6}$ = 0,676 da.

19,4>0,676 denez, hipotesi nulua errefusatzen da. Ondorioz, populazioak, drogen erabilpen-mailarekiko homogenoak direla baztertzen da.

PROBLEMAK

1.- Zoriz aukeraturiko 200 haziri aplikaturiko hiru tratamendu kimiko probatzean, hoziketa-probak burutuz, ondoko 3x2 kontingentzi taulan adierazitako emaitzak lortu ziren.

Tratamendu kimikoa	Hozi zirenak	Hozi ez zirenak	Denetara
A	190	10	200
B	170	30	200
C	180	20	200
Denetara	540	60	600

Kontrasta ezazu, tratamendu kimikoaren eta hoziketaren arteko independentzia, 5%-eko adierazgarritasun-mailaz.

2.- Urin gastrikoaren azidotasun-maila desberdinen (aklorhidria, hipoklorhidria, normal eta hiperklorhidria) eta gaixotasun-motaren (ultzera gastrikoa ala minbizia) arteko menpekotasunik dagoen aztertu nahi dugu. Lorturiko emaitzak, ondoko taulan adierazten dira:

	Aklorhidria	Hipoklorhidria	Normal	Hiperklorhidria
Ultzera gastrikoa	3	7	35	9
Minbizia	22	2	6	0

Kontrasta ezazu, azidotasun gastrikoaren eta gaixotasun-motaren arteko independentzia, 1%-eko adierazgarritasun-mailaz.

3.- 2×2 kontingentzi taula. χ^2_p-ren kalkulua, itxarotako maiztasunak estimatu gabe.

2×2 kontingentzi taularen itxura, ondokoa da:

B faktorea	A faktorea A_1	A_2	Baturak
B_1	o_{11}	o_{12}	$f_{1.}$
B_2	o_{21}	o_{22}	$f_{2.}$
Baturak	$c_{.1}$	$c_{.2}$	$n = f_{1.}+f_{2.} = c_{.1}+c_{.2}$

χ^2_1 banaketarantz jotzen duen estatistikoa,

$$\chi^2_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(o_{ij}-e_{ij})^2}{e_{ij}} \qquad (1)$$

da; eta χ^2_1-etarako hurbilketarik hoberena, Yates-en zuzenketaren bidezko ondoko estatistikoa dugu:

$$\chi^2_Y = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|o_{ij}-e_{ij}|-0,5)^2}{e_{ij}} \qquad (2)$$

(1) eta (2) estatistiko bietan, χ^2_p-ren balioa kalkulatzeko, e_{ij} itxarotako maiztasunak estimatu behar ditugula ikusten da. Baina, transformazio egokien bidez, (1) eta (2) adierazpenek ondoko erak onartzen dituzte:

$$\chi^2_p = \frac{n(o_{11}o_{22}-o_{12}o_{21})^2}{f_{1.}f_{2.}c_{.1}c_{.2}} \qquad (3)$$

eta

$$\chi^2_Y = \frac{n(|o_{11}o_{22}-o_{12}o_{21}| - \frac{n}{2})^2}{f_{1.}f_{2.}c_{.1}c_{.2}} \qquad (4)$$

(4) estatistikoa, (3) estatistikorako Yates-en zuzenketa da, berau $|o_{11}o_{22}-o_{12}o_{21}| > \frac{n}{2}$ denean erabiliko dugularik.

Beraz, metodo berriaren bidez, ikus dezagun nola ebazten den ondoko arazoa:

108 pertsonatatik, tratamendu profilaktikoaren menpean egon eta gero, 20 gaixotu egin ziren, kontrasterako plazeboak hartu zituzten 237 pertsonatatik 74 gaixotu zirelarik. Medikamendu errealaren (tratamendu profilaktikoaren) eta itxurazkoaren (kontrasterako plazeboen) arteko aktibitate-desberdintasuna adierazgarria al da? (Har ezazu $\alpha = 0,05$).

4.- Genetika-ikertzaileak, gaixotasunaren heredagarritasuna aztertzen du. 180 haurretako laginean ondoko datuak biltzen ditu:

		Gaixotasuna, haurrak pairatzen al du?	
		Bai	Ez
Amak pairatzen al du?	Bai	10	70
	Ez	8	72

0,01-eko adierazgarritasun-mailaz, gaixotasuna amak pairatu badu, haurrak pairatzea probableagoa al da?

5.- Pearson-en χ^2 banaketa erabiliz, egiazta ezazu, hemorragia anormalak dauzkaten emakume-proportzioak, haurdunaldian zehar droga hartu zutenentzat eta hartu ez zutenentzat, berdinak direneko hipotesia.

	Droga hartu zutenak	Droga hartu ez zutenak	Denetara
Hemorragia normalak	14	6	20
Hemorragia anormalak	347	334	681
Denetara	361	340	701

6.- Ondoko taulak, ibai batetatik ateratako 100 ur-laginetan aurkitutako organismo-kopuruak 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 hauek Poisson-en banaketatik atera zireneko hipotesi nulua ($\alpha = 0,05$).

7.- 70 gaixori 5 proba egitean datzan experimentua burutu da. Probak positiboa emateari, "arrakasta" deritzogu, lorturiko arrakasta-banaketa ondokoa izanik:

Arrakasta-kopurua	Gaixo-kopurua
0	3
1	8
2	14
3	18
4	15
5	12

Datu hauek banaketa binomiala betetzen duteneko hipotesia onar al daiteke, $\alpha = 0,05$ izanik?

8.- Kuiren arteko gurutzamenduaren 64 ondorengoren artean, 34 gorri, 10 beltz eta 20 zuri lortu ziren. Genetikak emandako ereduaren arauera, zenbaki hauek 9:3:4 proportzioan agertu behar dute.

$\alpha = 0,05$ izanik, azter ezazu, ea datu hauek eta ereduak elkar komunztatzen duten.

9.- 250 gaixorengan jasotako azido urikoaren determinazioen banaketa adierazten digu ondoko taulak:

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
$\geqslant 10$	5

Froga ezazu, $\mu = 5,74$ eta $\sigma = 2,01$ parametroetako banaketa normalarekiko datu hauen doikuntzaren egokitasuna ($\alpha = 0,01$).

10.- Zenbait lagin aske erabiliz, ea populazio binomialaren parametroak aurrefinkaturiko balioa duen kontrastatzeko Pearson-en χ^2-ren aplikazioa. (Laginetarako homogenotasun-testa)

Demagun, $p_1, p_2, \ldots, p_k$ probabilitateko $C_1, C_2, \ldots, C_k$ k klase edo gelaska baztertzaile eta exhaustibok osoturiko populazioa, $p_1 + \ldots + p_k = 1$ izanik.

C_1-en o_1 elementu,..., C_k-ren o_k elementu dituen n tamainuko lagina aukeratzen da zoriz.

Orduan, $\sum_{i=1}^{k} \frac{(o_i - np_i)^2}{np_i}$ (1) adierazpena, k-1 askatasun-gradutako χ^2 banaketari darraio.

k=2 deneko kasu partikularrean, populazioa binomiala izan go da, p_1 eta p_2 probabilitateko C_1 eta C_2 klase bi izanik, hurrenez hurren.

n ateraldi asketako laginean, C_1-en x elementu eta C_2-ren n-x elementu aurkitzen badira, (1) adierazpena honela agertzen zaigu:

$$\frac{(x-np)^2}{np} + \frac{\{(n-x)-n(1-p)\}^2}{n(1-p)} = \frac{(x-np)^2}{np(1-p)} \qquad (2)$$

eta (2) adierazpen hau, 1 askatasun-gradutako banaketari jarraituko zaio.

Adierazpen hau, $n_1, n_2, \ldots, n_k$ tamainuko k lagin askeren datuen bidez eraikitzen badugu,

$$\frac{(x_1-n_1p)^2}{n_1p(1-p)} + \frac{(x_2-n_2p)^2}{n_2p(1-p)} + \ldots + \frac{(x_k-n_kp)^2}{n_kp(1-p)} = \frac{1}{p(1-p)} \sum_{i=1}^{k} \frac{(x_i-n_ip)^2}{n_i} \qquad (3)$$

daukagu.

1 askatasun-gradutako ji karratu k aldagairen arteko batura denez, (3) adierazpena k askatasun-gradutako χ^2 banaketari darraio; hots:

$$\frac{1}{p(1-p)} \sum_{i=1}^{k} \frac{(x_i-n_ip)^2}{n_i} = \chi^2_k \qquad (4)$$

k laginen tamainua berbera deneko kasuan, hots, $n_1=\ldots=n_k=n$ denean, (4) adierazpena era honetan agertzen da:

$$\frac{1}{np(1-p)} \sum_{i=1}^{k} (x_i-np)^2 = \chi^2_k \qquad (5)$$

Izan ere, (4) adierazpena da, banaketa binomialaren parametroak p balioa dueneko hipotesia kontrastatzeko erabiliko dugun estatistikoa.

Kontrastea beti bezala burutzen da, proba-estatistikoa eta $\chi^2_{\alpha;k}$ puntu kritikoa konparatuz,

$$\frac{1}{p(1-p)} \sum_{i=1}^{k} \frac{(x_i - n_i p)^2}{n_i} \leqslant \chi^2_{\alpha;k}$$

baldin bada, binomialaren parametroa p deneko hipotesia onartzen dugu; hau da, lagin aske guztiak, p parametroko populazio binomialetik datoz.

Desberdintza kontrakoa bada, hipotesia errefusatzen dugu.

p ezezaguna bada, lagin guztien datuen bidez estimatu beharko dugula azpimarratzea garrantzitsua da, eta, beraz, χ^2-ren askatasun-graduak unitate batez gutxituko direla, hots, (4) estatigrafoa χ^2_{k-1} banaketari jarraituko zaiola.

p-ren estimazioa hauxe izango da: $\hat{p} = \dfrac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$

Ezagupen hauek aplikatuz, azter dezagun ondoko arazoa:

Ale horiko ilarren eta ale berdeko ilarren arteko gurutzamendua aztertu zuen Mendelek. Hibridoen lehen benaunaldia, ale horiko (karaktere meneratzailea) landarek osoturik dago. Bigarren belaunaldiko karaktereen disjuntzioa aztertzeko, ale hauetariko 10 lagin erein ziren, lorturiko emaitzak ondoko taulan adierazten direlarik.

Teoriak aurrefinkaturiko $p = \frac{3}{4}$ balioa, bigarren belaunaldiko ale horiko ilarra izateko probabilitatetzat onar al dezakegu? ($\alpha = 0{,}05$)

Laginak	Horiak(x_i)	Berdeak(n_i-x_i)	Denetara(n_i)	$(4x_i-3n_i)^2/3n_i$
1	25	11	36	0,5926
2	32	7	39	1,0342
3	14	5	19	0,0175
4	70	27	97	0,4158
5	24	13	37	2,0270
6	20	6	26	0,0512
7	32	13	45	0,3630
8	44	9	53	1,8176
9	50	14	64	0,3333
10	44	18	62	0,5376
Denetara	355	123	478	7,224

$\chi^2_{0,05;10}$= 18,31 eta 7,224 < 18,31 denez, p = $\frac{3}{4}$ denetako hipotesia onartzen da; hots, Mendelen legeak experientziak egiaztatzen ditu.

Aurreko taula, zenbait azterketa estatistiko burutzeko gai da. Adibidez, 1 askatasun-gradutako χ^2-ren balioak diren azken zutabeko balioak, $\chi^2_{0,05;1}$= 3,841 delakoarekin konparatzen badira, lagin bakoitzerako p = 3/4 hipotesia onarturik dagoela ikusten da.

10 laginak, p = $\frac{3}{4}$ parametroko populazio binomial berberetik datozela adierazten digu honek.

Homogenotasuna egiaztatu eta gero, n = $\sum n_i$ = 478 tamainuko lagin bakar batetan bil daitezke 10 laginak, x = $\sum x_i$ = 355 izanik. Lagin orokor honetarako estatigrafoaren balioa, $(4x-3n)^2/3n$ = 0,1367 da, $\chi^2_{0,05;1}$ = 3,841 baino txikiagoa izanik. Beraz, p = 3/4 hipotesia, lagin guztiak lagin bakar batetan biltzetik ondorioztatzen deneko lagin orokorrean ere onartzen da.

k lagin askeri dagozkien χ^2_1 banaketen arteko baturatik ondorioztaturiko χ^2_k banaketa erabiliz, p parametroak aurrefinkaturiko balioa (3/4, adibide honetan) dueneko hipotesia

onartzen denean, "k laginen homogenotasuna" kontrastatu de la esaten da, lagin guztiak p = 3/4 parametroko populazio binomialetik etorriko bailiren jo dezakegu. Edo beste hitz batzutan esanda, p = 3/4 balioak konstante iraun du laginez lagin.

11.- Poisson-en legearen arauera banaturiko populaziorako homogenotasun-testa, χ^2 banaketaren bidez.

Bakterio-kultiboaren disoluziorako teknikak, bakterioen zorizko banaketa ondorioztatzen badu, eta berauek plakaren gainean independenteki garatzen badira, plaka bakoitzaren gainean zenbaturiko bakterio-kopurua Poisson-en banaketari darraiola ba dakigu.

Baldintza berberetan diluituriko bakterio-kultiboa dau katen 15 plaka erein ziren, plakako ondoko bakterio-kopuruak lorturik:

193 - 168 - 161 - 153 - 183 - 152 - 171 - 156 - 159 - 140 - 151 - 152 - 133 - 164 - 157

Lorturiko kopuruak eta plaka guztietarako konstantea den m parametroko Poisson-en banaketari darrazkioneko hipotesiak elkar komunztatzen dutenentz egiaztatu nahi dugu.

Poisson-en populazioa homogenoa (m=kte) deneko hipotesi hau, ondoko proba-estatistikoa erabiliz kontrasta daitekе:

$$\chi^2_p = \sum_{i=1}^{k} \frac{(x_i - \bar{x})^2}{\bar{x}} \quad (1), \quad \bar{x} = \frac{1}{k} \sum_{i=1}^{k} x_i \quad \text{izanik.}$$

Hipotesia egiazkoa bada (lagin guztiak, m=kte parametroko Poisson-en populaziotik badatoz), (1) adierazpena k-1 askatasun-gradutako χ^2 banaketari darraio.

Gure kasuan, x_i = lagin(plaka) bakoitzeko bakterio-kopurua, eta k = lagin(plaka)-kopurua dira.

Beraz,

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = 159,53$$

$$\chi^2_p = \sum_{i=1}^{15} \frac{(x_i - \bar{x})^2}{\bar{x}} = 20,2953$$

Balio hau, "aski handia" bada, m konstantea ez dela ondorioztatuko dugu, diluzio-teknika egokia ez izanik.

$\alpha = 0,05$ adierazgarritasun-mailako puntu kritikoa bilatzen dugu, hau da, $\chi^2_{0,05;14} = 23,685$.

$20,2953 < 23,685$ denez, plaken gainean zenbaturiko bakterio-kopurua, m = 159,53 parametroko Poisson-en banaketari darraiola ondorioztatzen dugu.

12.- Jatorri desberdinetako 30 igelatako 8 talde, G estrofantinato (inglukosiko kardiakoa) 0,30 g./g.-z xiringatuak ziren, hurrengo egunean talde bakoitzeko hildako igelen ondoko kopuruak zenbaturik:

15 - 12 - 12 - 18 - 14 - 22 - 17 - 13

$\alpha = 0,01$ adierazgarritasun-mailaz, talde desberdinetako igelak G estrofantinato adierazitako dosiaz xiringatzean, probabilitate berberaz hiltzen direla onar al dezakegu?

13.- Plaka koadrikulatuaren gainean, 8 odol-tanta hedatu ziren, mikroskopioaren laguntzaz, tanta bakoitzeko globulu gorrien kopurua zenbaturik:

1870 - 1994 - 1872 - 1880 - 1869 - 1820 - 1830 - 1836

8 odol-tantak (bolumen berberekoak) Poisson-en populazio berberetik datozela onar al dezakegu? Har ezazu $\alpha = 0,025$

14.- Bost eskualde desberdinetan jaiotzen diren haur deformeak ez direla uniformeki banatzen susmatzen da; hots, zenbait industria kaltegarri direla kausa, haur deformeen portzentaia, eskualde batzutan beste batzutan baino handiagoa dela susmatzen da. Susmoa egiaztatzeko ala baztertzeko, ondoko datuak hartu ziren kontutan:

Eskualdea	Haur deformeen kopurua	Jaiotze-kopuru osoa
1	62	1740
2	38	1498
3	25	942
4	38	1312
5	50	1326
	213	6818

Kontrasta ezazu, eskualdearen eta haur deformeen proportzioen arteko independentzia ($\alpha = 0,05$).

IX. GAIA: ERREGRESIOA ETA KORRELAZIOA.

IX.1.- Sarrera.

IX.2.- Erregresio-lerroa.

IX.3.- Aurresana eta minimo karratuen metodoa.

IX.4.- Pearson-en korrelazio linealaren koefizientea.

IX.5.- Estimazio-errore standarda.

IX.6.- Erregresio eta korrelazio-arazoetarako konfidantza-tarteak.

ERREGRESIOA ETA KORRELAZIOA

IX.1.- SARRERA.

Bi magnitude neurgarri emanik, batzutan, beraien arteko erlazioa lortzea interesatzen zaigu, existitzen bada behintzat. X eta Y-z adierazitako aldagaien artean, zentzu matematikoan, erlazio funtzioanala dagoenean, egoera ideala daukagu. Hots, $Y = f(X)$ denean. ($Y = X^2$, $Y = \sin X$, eta abar).

Mota honetako erlazio funtzionala, non X (denbora, adibidez) aldagai askearen balio bakoitzari, $Y = K \cdot X$ erlazioaren bidez esate baterako (K, abiadura izango da), Y (espazioa, adibidez) menpeko aldagaiaren balio bakar bat dagokion, maiz agertzen da Fisikan.

Biologian ez da hain ugaria; izan ere, kontsidera dezagun ondoko adibidea:

Demagun ξ experimentua, non X giro-tenperatura, eta Y, ugaztun batek ordu batetan, beraren pisuaren Kg.-ko kontsumituriko oxigeno-kantitatea diren. X tenperaturaren balio berberarako, Kg./h.-tan adierazitako O_2-ren kontsumoaren balioak, experientziaz experientzia era aurrikustezinean aldatzen direla behazten da. Baina, X-ekiko Y-ren menpekotasuna dagoela "susma" dezakegu.

Era berean, animaliaren $mm.^3$ odoleko μ gr.-tan adierazitako drogaren X kontzentrazioaren baliorako (experimentatzaileak finkaturikoa), experientziaz experientzia, mm.Hg-tan neurturiko Y odol-presioaren aldaketa aurrikustezinak daude.

Arratoi-kumaldietako X kume-kopuruarekiko kumeen Y pisua; "Oecanthus fultari" izeneko kilkirretan, X tenperaturarekiko Y batezbesteko kirrinka-kopurua; X tenperaturarekiko,

"Empoasca fabae" delako patataren matxinsaltoaren enbrioialdiaren Y iraupena (egunetan) ere, mota honetako adibideak dira.

Hots, egoera biologiko ugaritan, X aldagai deterministiko askearekiko (tenperatura, drogaren kontzentrazioa, eta abar), non beraren balioak experimentatzaileak eman (edo kontrola edo eraen) ditzakeen, Y zorizko aldagai menpekoaren (O_2-ren kontsumoa, odol-presioa, eta abar) menpekotasuna agertzen da. Menpekotasun honi, menpekotasun estokastiko edo zorizkoa deritzogu.

IX.2.- ERREGRESIO-LERROA.

Erlazio edo menpekotasun honen azterketa, erlazio funtzionalaren azterketa bihurtzen da ondoko erara:

X (deterministiko, aske eta kontrolagarria) eta Y (zorizkoa eta X-en menpekoa), menpekotasun estokastikoan dauden bi aldagai emanik, aldagai askearen $X = x_1$ balio batetarako, berari dagokion Y-ren balio-multzoak (aurrikustezina dena), $Y/X=x_1$ eran adieraziko dugun zorizko aldagai berriaren heina osotzen duela hartu behar dugu kontutan. Era berean, $X= x_2$-ri, $Y/X=x_2$ beste zorizko aldagai baten heina den balio-multzo desberdina dagokiola, eta abar.

Orduan, x_1-i $E(Y/X=x_1)$ ($Y/X=x_1$ delakoaren itxaropen matematikoa), x_2-ri $E(Y/X=x_2)$, eta, orokorrean, x-i $E(Y/X=x)$ elkartzen dion funtzioa eraikitzen dugu.

$f(x) = E(Y/X=x)$ funtzioa, X gaineko Y-ren erregresio-funtzioa deitzen da, bere irudikapena X gaineko Y-ren erregresio-lerroa izanik.

IX.3.- AURRESANA ETA MINIMO KARRATUEN METODOA.

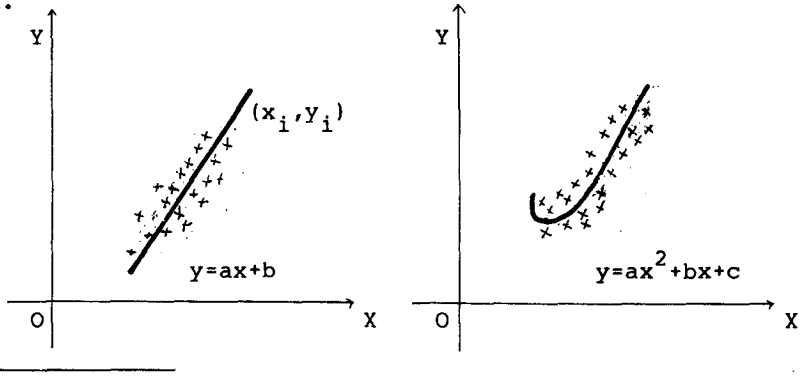
Orokorrean, ezin dugu erregresio-funtzioaren adierazpide matematikoa lortu, zeren eta $Y/X=x$ zorizko aldagai bakoitzaren probabilitate-banaketa, eta, beraz, $E(Y/X=x)$ bere itxa-

ropena lortzea baimentzen digun sistema ez bait dago beti.

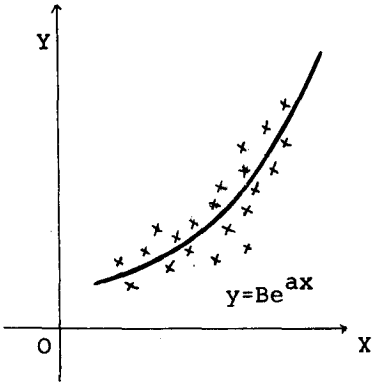
Hala ere, lehen azaldutako motako erlazio estokastikoa agertzen deneko arazo biologiko bakoitzaren E(Y/X=x)-en ezagumendua interesgarria izango litzateke, zeren eta, modu honetan, experimentatzaileak berak emandako X aldagai askearen x balio bakoitzerako (esan dugunez, "a priori" aurrikustezina dena) Y/X=x zorizko aldagaiaren jokabidea "aurresan" bait lezake.

E(Y/X=x) kalkulatzeko metodorik egon ez arren, Y/X=x delakoari buruzko aurresanak egitea interesatzen zaigunez, berau lortzea baimentzen diguten metodo alternatiboak bilatuko ditugu.

Metodorik emankorrena hauxe da: X aldagai askeari x_1 balioa ematean eta Y aldagaiaren y_1 balioa lortzean,..., X-i x_n balioa ematean eta Y-ren y_n balioa lortzean dautzaten experimentuaren n errepikapen burutzen dira. Geroago, XOY planoan irudikatzen dira $(x_1,y_1),\ldots,(x_n,y_n)$ puntuak, "puntuhodeia" deituriko puntu-multzoa lorturik. Irudikapen hau ikusirik(*), X eta Y-ren arteko erlazio estokastikoa egokiro deskribatzen duen lerroa, $(x_1,y_1),\ldots,(x_n,y_n)$ n datu-bikoteek zirriborraturikoa, ea lerro zuzena, parabolikoa, exponentziala, eta abar, den erabakitzen da, berau $y = ax + b$ funtzio lineal, $y = ax^2 + bx + c$ funtzio koadratiko, $y = B \cdot e^{ax}$ funtzio exponentzial eta abarren irudikapena izanik hurrenez hurren.

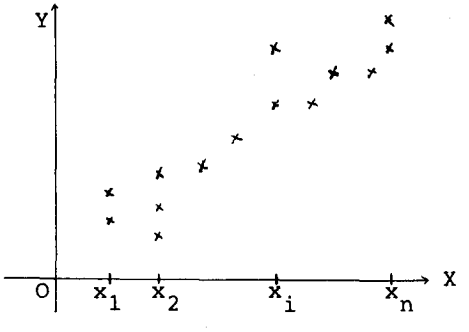


(*) eta arazo biologikoa sakonki ezaguturik.



Erregresio-lerroa adierazten duen eredu teorikoa aukeratu eta gero, ereduaren parametroen estimazioaren arazoa agertzen zaigu, hots, posible guztien artean, zein zuzen, edo zein parabola, eta abar, den egokiena bi magnitude biologikoen arteko menpekotasuna "matematikoki" adierazteko, eta, eman nahi diogun X=x balioa funtzioan (lineal, koadratiko, exponentzial, eta abarrean) ordezkatuz, Y-ren balioa aurresan ahal izateko.

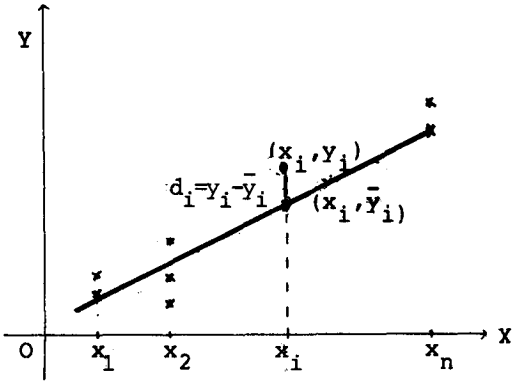
Azkenik, "puntu-hodeiari doituriko lerroa" izenaz ere ezagutzen den erregresio-lerroa lortu eta gero, "doikuntza" "egokia" denentz erabakitzeko metodoak eman beharko ditugu, hots, doikuntz egokitasuna aztertzekoak, aurreko atalean egin dugun antzera.



Irudikoaren antzeko "puntu-hodeia" emanik, erregresio-lerroa adierazteko eredu egokia, y=ax+b zuzena dela erabakitzen dugu.

a eta b parametroak estimatzeko, eta beraz, bilatzen dugun funtzioa lortzeko, ondokoan datzan Gauss-en minimo karratuen metodoa delakoari jarraituko gatzazkio:

x_i bakoitzerako, berari elkarturiko y_i guztiak (ohar dezagun, irudian, $X=x_2$-ri, adibidez, $Y=y_2'$, $Y=y_2''$, $Y=y_2'''$ Y-ren hiru balio elkartu zaizkiola) eta, gainera, $\bar{y}_i=ax_i + b$ kontsideratzen dira.



$d_i = y_i - \bar{y}_i$ izanik, $G = \sum_{i=1}^{n} d_i^2$

funtzioa kontsideratzen badugu, a eta b parametroen balioek, zeintzuetarako G funtzioak balio posiblerik txikiena hartzen duen, $y = ax + b$ ekuazioan ordezkatuz emango digute, (x_1,y_1), $\ldots,(x_n,y_n)$ puntuei adierazitako zentzuan hoberen "doitzen" zaien zuzenaren ekuazioa, hots, "bertikalki" (hau da, OY ardatzaren paraleloki) neurturiko zuzenaren eta puntuen arteko distantzien karratuen arteko batura minimoa egiten dueneko zentzuan.

Bilatzen ditugun a eta b kalkulatzea baimentzen digun ekuazio-sistema, deribatu partzialak anulatuz lortzen dugu:

$$\phi \begin{cases} \dfrac{\partial G}{\partial a} = 0 \\ \dfrac{\partial G}{\partial b} = 0 \end{cases} \Longleftrightarrow \begin{cases} n\cdot b + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

Aurreko ekuazio-sistema ebatziz, hauxe da erdiesten dugun a-ren balioa:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$$

"laginaren erregresio-koefizientea" deitua izanik.

Beste aldetik, ϕ sistemaren lehen ekuaziotik, ondokoa lortzen da:

$$b = \frac{1}{n} \sum_{i=1}^{n} y_i - a \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{y} - a \cdot \bar{x}$$

Eta, $y = a \cdot x + b$ ekuazioa erabiliz, lorturiko b-ren ordezkapenaren bidez,

$$y = a \cdot x + \bar{y} - a \cdot \bar{x}$$

ondorioztatzen da; hau da,

$$y - \bar{y} = a(x - \bar{x})$$

berau, X gaineko Y-ren erregresio-zuzenaren ekuazioa izanik.

1.adibidea:

Sei disoluziotatik abiatuz lorturiko datuak (X: kolorimetroaren neurketak; Y: erretxina-kontzentrazioak, mgr./100ml.-tan) adierazten dira ondoko taulan:

X	8	50	81	102	140	181
Y	0,12	0,71	1,09	1,38	1,95	2,50

Datu hauei dagokien X gaineko Y-ren erregresio-zuzena erdiesteko, ondoko koadro edo taula eraikitzen dugu:

x_i	y_i	x_i^2	$x_i y_i$
8	0,12	64	0,96
50	0,71	2500	35,50
81	1,09	6561	88,29
102	1,38	10404	140,76
140	1,95	19600	273,00
181	2,50	32761	452,50
562	7,75	71890	991,01

Hemendik:

$$\bar{x} = \frac{562}{6} = 93,667 \qquad \bar{y} = \frac{7,75}{6} = 1,292$$

$$a = \frac{991,01 - \frac{1}{6}\,562 \cdot 7,75}{71890 - \frac{1}{6}\,562^2} = 0,0137715$$

Beraz, hauxe da zuzenaren ekuazioa:

$$y - 1,292 = 0,0137715(x - 93,667)$$

edo, zenbait kalkulu burutuz:

$$y = 0,0137715 \cdot x + 0,0017362$$

Kontutan har dezagun, planteaturiko egoeran, X neurketa kolorimetrikoa aldagai "kontrolagarritzat" jotzen dugula. Baina, hemen, datu-bikoteak batera hartu ditugunez, Y erretxina-kontzentrazioaren gaineko X datu kolorimetrikoaren erregresio-zuzena bila eta lor genezakeela azpimarratu behar dugu, orain, Y aldagaia kontrolagarritzat hartuz.

Honela, balio kolorimetrikotzat 90 duen disoluzioaren (aurreko seien artean ez dagoena) erretxina-kontzentrazioa

$$y = 0,0137715 \cdot 90 + 0,0017362 = 1,2411762$$

dela aurresan dezakegun antzera, emandako erretxina-kontzentrazioari dagokion balio kolorimetrikoaren aurresana lor genezake (Y gaineko X-en erregresio-zuzenaren aldez aurreko -eta egindakoaren antzeko- kalkuluaren bidez).

Ondoko adibidean, arazo hau argitzen saiatuko gara:

2.adibidea:

X ordu pasa eta gero, kultiboan agertzen den bolumen unitateko Y bakterio-kopurua, ondoko taulak emanik dator:

X ordu-kopurua	0	1	2	3	4	5
Bolumen unitateko Y bakterio-kopurua	12	19	23	34	56	62

Extrapolazioa egin nahi badugu, hots, esate baterako, 6 orduren buruan, zenbat bakterio espero dugun galdetu nahi badiogu gure buruari, eta puntu-hodeia zuzenaren bidez doitzea erabaki badugu, lehenago deskribaturiko erara egingo dugu:

$\bar{x} = 2,5$; $\bar{y} = 34,43$; $a = \frac{31}{2,9}$ lortuko ditugu, eta, beraz, $y - \bar{y} = a(x - \bar{x})$ hauxe izango da:

$$y = 10,68 \cdot x + 7,68$$

Hau dela eta,

$y_6 = 10,68 \cdot 6 + 7,68 \simeq 72$ bakterio espero izango ditugu.

Aitzitik, beste zentzutan extrapolatu nahi badugu, eta 100 bakterio izan arte zenbat orduk igaro behar duen estimatu nahi badugu, Y gaineko X-en erregresio-zuzena ($x = a'y + b'$ zuzena, non OX ardatzaren paraleloki neurturiko puntuen eta zuzenaren arteko distantzien karratuen arteko batura txikiena den) kalkulatuko dugu.

Aurreko egoeran egin dugun antzera, bere adierazpena ondokoa izango da:

$$x - \bar{x} = a'(y - \bar{y})$$

non $a' = \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} y_i \right)^2}$ den.

Ditugun datuetarako, $x = 0{,}088 \cdot y - 0{,}55$ dela ondorioztatzen da.

Beraz, $y = 100$ bakteriotarako, $x = 0{,}088 \cdot 100 - 0{,}55 \simeq 8{,}25$ lortzen da, hots, 8 ordu eta 15 minutu.

IX.4.- PEARSON-en KORRELAZIO LINEALAREN KOEFIZIENTEA.

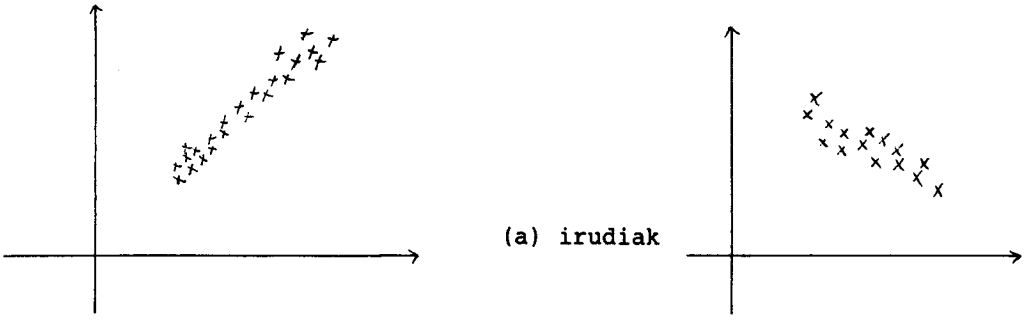
Aztertu dugun erregresioaren kontzeptuari estuki lotuta dagoen kontzeptua, korrelazioarena da.

Lehen baieztaturikoaren ordez, orain, ez dugu suposatuko aldagai askea eta aurrekoaren menpekoa den zorizko aldagaia bereiztu behar ditugunik. Aitzitik, X eta Y bi zorizko aldagai emanik, (X,Y) zorizko aldagai bidimentsionalaren $(x_1,y_1),\ldots,(x_n,y_n)$ laginaren bidez, bien arteko menpekotasun linealaren neurria aurkeztu nahi dugu.

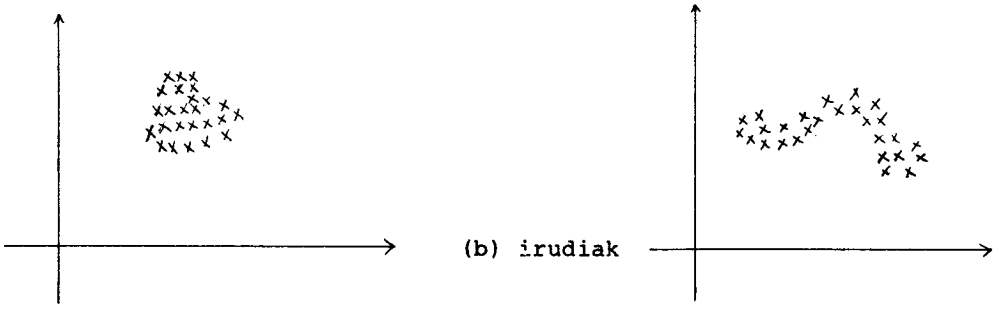
Beste era batetara esanda, $(x_1,y_1),\ldots,(x_n,y_n)$ puntuak irudikatzen badira, orokorrean lerro zuzenaren gainean kokatua izango ez den puntu-hodeia lortuko dugu, zeren eta X eta Y-ren arteko menpekotasuna ez bait da $Y = a \cdot X + b$ (edo $X = a' \cdot Y + b'$) motako funtzionala, estatistikoa baino, gaiaren hasieran adierazi dugun legez.

Hala ere, mendel estuan ((a) irudiak) edo itxura gabeko eran ((b) irudiak) edo lerro ez-zuzena gogorarazten duen mendelean ((b) irudiak) kokatuak izateak, menpekotasun linealaren mailaketa desberdina eta beraren linealitate-maila isladatzen

duen menpekotasunaren zenbait neurri-mota sartzearen premia adierazten du.



(a) irudiak



(b) irudiak

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1} x_i\right)^2}\sqrt{\sum_{i=1} y_i^2 - \frac{1}{n}\left(\sum_{i=1} y_i\right)^2}}$$

definituriko Pearson-en korrelazio linealaren koefizienteak, X eta Y zorizko aldagaien arteko menpekotasun linealaren (eta estokastikoaren) maila estimatzea baimentzen digu, $(x_1,\ldots, x_n)$ eta $(y_1,\ldots,y_n)$ laginetatik abiatuz.

$-1 \leqslant r \leqslant +1$ dela frogatzen da.

$|r| = 1$ baldin bada, menpekotasun lineala absolutua da, eta ez da zorizkoa, funtzionala baino.

$r > 0$ baldin bada, menpekotasuna gorakorra da (X-en balio handiagoetarako, Y-ren balio handiagoak lortuko dira zorizko fluktuazioak izan ezik).

$r < 0$ baldin bada, osterantzean, menpekotasuna beherakorra da (X-en balio handiagoetarako, Y-ren balio txikiagoak lortuko dira fluktuazioak izan ezik).

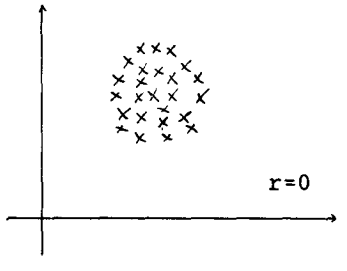
Orokorrean:

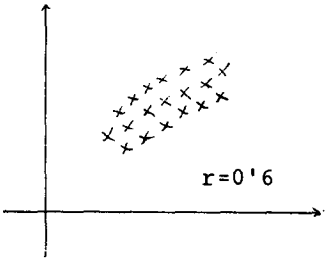
$|r| \geqslant 0,8$ betetzen duten "r"-ren balioek, menpekotasun lineal handia adierazten dute.

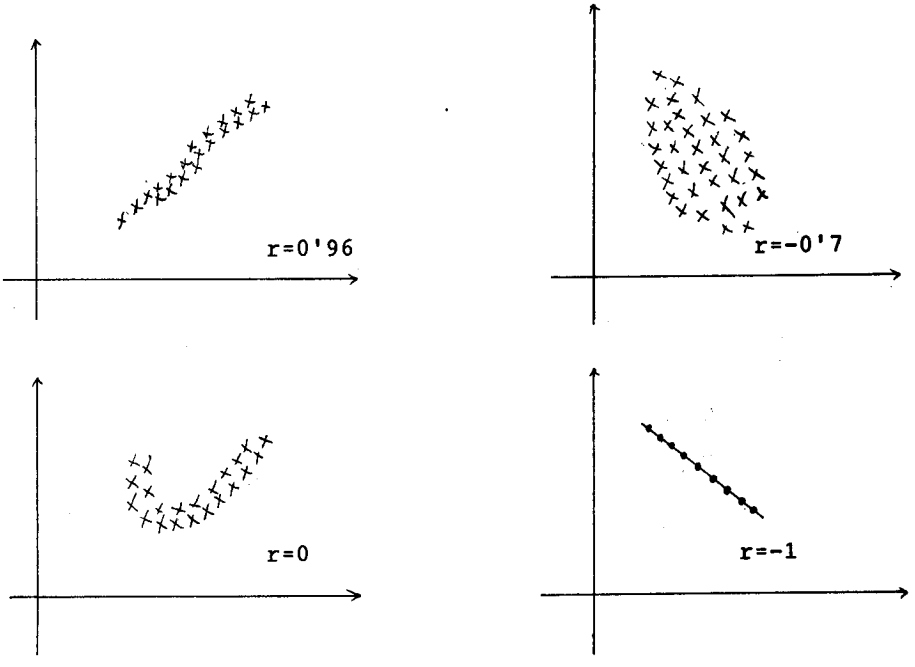
$0,5 \leqslant |r| < 0,8$ betetzen duten "r"-ren balioek, menpekotasun lineala onargarria dela adierazten dute.

$|r| < 0,5$ betetzen duten "r"-ren balioek, menpekotasun lineala deneko interpretazioa egokia ez dela adierazten dute, edota, menpekotasunik ez dagoela.

Ondoko irudiek argitzen dizkigute aurreko ideiak:







3.adibidea:

Kalkula dezagun, 1.adibidean aurkeztutako X eta Y laginei dagokien korrelazio-koefizientea, non X-ek kolorimetroaren neurketak eta Y-k erretxina-kontzentrazioak deskribatzen dituzten.

x_i, y_i, x_i^2, x_iy_i zutabeetako taulari, y_i^2 bostgarren zutabea erantsi behar diogu:

y_i^2
0,0144
0,5041
1,1881
1,9044
3,8025
6,25
13,6635

Eta beraz:

$$r = \frac{991{,}1 - \frac{1}{6}\,562 \cdot 7{,}75}{\sqrt{71890 - \frac{1}{6}\cdot 562^2}\ \sqrt{13{,}6635 - \frac{1}{6}\cdot 7{,}75^2}} = 0{,}997$$

"r"-ren balio honek, menpekotasun lineala ia-ia osoa dela adierazten du.

IX.5.- ESTIMAZIO-ERRORE STANDARDA.

Hodeiaren puntuekiko erregresio-zuzenaren doikuntzaren egokitasuna neurtzeko, hots, $y = a \cdot x + b$ (a eta b, laginetik abiatuz kalkulatuak izan dira azterturiko formulen bidez) ekuazioaren bitartez emandako x balioari dagokion y-ren balioa aurresatean jaukitzen den erroreari buruzko informazioa lortzeko, "estimazio-errore standarda" delakoa sartzen da. Honela definitzen da:

$$S_{ey} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - ax_i - b)^2}{n-2}}$$

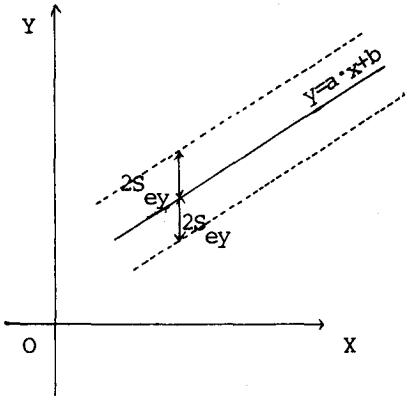
edo, kalkulu errezaren(*) bidez:

$$S_{ey} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a \sum_{i=1}^{n} x_i y_i - b \sum_{i=1}^{n} y_i}{n-2}}$$

Erregresio-arazo askotan gertatzen den bezala, Y/X=x zorizko aldagai guztiak normalak eta bariantza berberekoak (X-en edozein baliotarako) baldin badira, Y-aX-b zorizko aldagaia ere normala da, batezbestekoa 0 eta desbidazio standarda $\sigma \approx S_{ey}$ izanik.

(*) Begira ezazu hurrengo orrialdea.

Baldintza hauetaz gainera, n laginaren tamainua 25 edo 25 baino handiagoa baldin bada, $(-2S_{ey}\ ,\ 2S_{ey})$ tarteak desbidazioen 95%-a dauka gutxi gora behera; hots, Y = a X + b erregresio-zuzenaren bi zuzen paralelo, bata behetik eta bestea goitik, ordenatuen ardatzaren paraleloki neurturiko $2S_{ey}$ distantzian irudikatzen baditugu, 95%-eko konfidantza-mendela determinaturik dago, hau da, hodeiaren puntuen 95%-a mendelaren barruan dago.



4.adibidea:

1.adibideko doikuntzari dagokion estimazio-errore standarda ondokoa da:

$$S_{ey} = \sqrt{\frac{13,6635 - 0,0017362\cdot 7,75 - 0,0137715\cdot 991,01}{4}} = 0,0242435$$

Hau da, kolorimetroaren X neurketen bidez Y erretxina-kontzentrazioa estimatzean jaukitako errorea,

$$2S_{ey} = 2\cdot 0,0242435 = 0,048487$$

baino txikiagoa da, 0,95 probabilitatea duelarik.

(*) Izan ere:

$$\sum_{i=1}^{n}(y_i-ax_i-b)^2 = \sum_{i=1}^{n}(y_i-ax_i-b)(y_i-ax_i-b) = \sum_{i=1}^{n} y_i(y_i-ax_i-b) -$$

$$-a\sum_{i=1}^{n} x_i(y_i-ax_i-b) - b\sum_{i=1}^{n}(y_i-ax_i-b) =$$

$$= \sum_{i=1}^{n} y_i(y_i-ax_i-b)$$

zeren eta $\sum_{i=1}^{n} x_i(y_i-ax_i-b)=0$ eta $\sum_{i=1}^{n}(y_i-ax_i-b)=0$ bait dira, lehenago adierazitako a eta b kalkulatzeko ϕ ekuazio-sistemaren arauera.

Lagin txikien kasuan, doikuntzaren egokitasuna aztertzeko, ezin dugu S_{ey} estimazio-errore standarda erabili, azterketa zehatzagoa egin behar bait dugu, eta Y/X=x zorizko aldagaien normalitatearen eta bariantzen arteko berdintasunaren hipotesiek baliagarriak izaten dirautela suposatu behar dugu.

IX.6.- ERREGRESIO ETA KORRELAZIO-ARAZOETARAKO KONFIDANTZA-TARTEAK.

Nahiz eta $y = a.x + b$ erregresio-zuzena kalkulatu, berau, $(x_1, y_1),\ldots,(x_n, y_n)$ datuak soilik erabiliz lorturikoa dela azpimarratu behar dugu. Beraz, ea zuzen hau, populazio osoaren deskribapen linealari dagokion $y = \alpha x + \beta$ zuzenarekiko nahiko ondo doitzen den (hots, ea doikuntza egokia den) jakin behar dugu. Helburu hau lortzeko, zenbait konfidantza-tarte desberdin kalkulatuko ditugu:

1) Populazioaren α erregresio-koefizienterako konfidantza-tartea:

$(1-\gamma)$-ko konfidantza-mailarako, α-ren konfidantza-tartea hauxe da:

$$\left(a - t_{\gamma/2;n-2} \frac{S_{ey}}{\sqrt{\sum (x_i-\bar{x})^2}},\ a + t_{\gamma/2;n-2} \frac{S_{ey}}{\sqrt{\sum (x_i-\bar{x})^2}}\right)$$

2) $y = \alpha x_o + \beta$ ordenaturako konfidantza-tartea:

Populazio osoari doituriko $y = \alpha x + \beta$ zuzenaren ordenaturako ($x=x_o$ balioari dagokiona) $(1-\gamma)$-ko konfidantza-tartea ondokoa da:

$$\left(ax_o + b - t_{\gamma/2;n-2}\, S_{ey} \sqrt{\frac{1}{n} + \frac{(x_o-\bar{x})^2}{\sum (x_i-\bar{x})^2}}\right.,$$

$$ax_o + b + t_{\gamma/2;n-2}\, S_{ey} \sqrt{\frac{1}{n} + \frac{(x_o-\bar{x})^2}{\sum(x_i-\bar{x})^2}}\ )$$

3) Jatorriko β ordenaturako konfidantza-tartea:

$(1-\gamma)$ mailako β-rako konfidantza-tartea hauxe da:

$$(b - t_{\gamma/2;n-2}\, S_{ey} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}\ ,$$

$$b + t_{\gamma/2;n-2}\, S_{ey} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}\ )$$

4) Populazioaren ρ korrelazio-koefizienterako konfidantza-tartea:

Erregresioari buruz ikusi ditugun egoeretan gertatzen den antzera, lehenago sarturiko Pearson-en "r" korrelazio-koefizientea, laginaren korrelazio-koefizientea da, zeinaren bidez X eta Y bi zorizko aldagairen arteko menpekotasuna neurtzen duen populazioaren ρ korrelazio-koefizientea estimatu nahi bait dugu.

Beraz, "r" kalkulatu eta gero, lorturiko hurbilketa estatistikoki neurtzen duen ρ-rako konfidantza-tartea eraikitzea komeni zaigu.

Tarte honen adierazpena korapilatsua (*) da.

Alde batetik, ondokoa kalkulatu behar dugu:

$$(\frac{1}{2}\ln\frac{1+r}{1-r} - z_{\gamma/2}\sqrt{\frac{1}{n-3}}\ ,\quad \frac{1}{2}\ln\frac{1+r}{1-r} + z_{\gamma/2}\sqrt{\frac{1}{n-3}}\ )$$

(*) Begira ezazu Viedma J.: Bioestadística.

non "r", laginaren korrelazio-koefizientea, eta n, laginaren tamainua diren, $z_{\gamma/2}$ N(0,1) banaketaren taulak emanik datorrelarik.

Beste aldetik, $\frac{1}{2} \ln \frac{1+\rho}{1-\rho}$ delakorako $(1-\gamma)$-ko konfidantza-mugak diren tarte honen muturretatik, ρ-rako konfidantza-mugak lortzen dira, ondoko taularen alderantzizko irakurketaz:

$$Z = \ln \frac{1+r}{1-r}$$

r	Z	r	Z	r	Z	r	Z
.00	.000						
.01	.010	.26	.266	.51	.563	.76	.996
.02	.020	.27	.277	.52	.576	.77	1.020
.03	.030	.28	.288	.53	.590	.78	1.045
.04	.040	.29	.299	.54	.604	.79	1.071
.05	.050	.30	.310	.55	.618	.80	1.099
.06	.060	.31	.321	.56	.633	.81	1.127
.07	070	.32	.332	.57	.648	.82	1.157
.08	.080	.33	.343	.58	.663	.83	1.188
.09	.090	.34	.354	.59	.678	.84	1.221
.10	.100	.35	.365	.60	.693	.85	1.256
.11	.110	.36	.377	.61	.709	.86	1.293
.12	.121	.37	.388	.62	.725	.87	1.333
.13	.131	.38	.400	.63	.741	.88	1.376
.14	.141	.39	.412	.64	.758	.89	1.422
.15	.151	.40	.424	.65	.775	.90	1.472
.16	.161	.41	.436	.66	.793	.91	1.528
.17	.172	.42	.448	.67	.811	.92	1.589
.18	.182	.43	.460	.68	.829	.93	1.658
.19	.192	.44	.472	.69	.848	.94	1.738
.20	.203	.45	.485	.70	.867	.95	1.832
.21	.213	.46	.497	.71	.887	.96	1.946
.22	.224	.47	.510	.72	.908	.97	2.092
.23	.234	.48	.523	.73	.929	.98	2.298
.24	.245	.49	.536	.74	.950	.99	2.647
.25	.255	.50	.549	.75	.973		

5.adibidea:

39 neurketa-bikotetako laginaren korrelazio-koefizientea 0,68 da.

ρ-rako 95%-eko konfidantza-tartea lortzeko, ondokoa eraikitzen da:

$$\left(\frac{1}{2}\ln\frac{1+0,68}{1-0,68} - z_{0,025}\sqrt{\frac{1}{36}}\ ,\ \frac{1}{2}\ln\frac{1+0,68}{1-0,68} + z_{0,025}\sqrt{\frac{1}{36}}\right) = (0,82911 - 1,96\,\frac{1}{6}\ ,\ 0,82911 + 1,96\,\frac{1}{6}) =$$

$$= (0,50244\ ,\ 1,15578)$$

Eta, aurreko taularen alderantzizko irakurketaren bidez (orokorrean, interpolatu beharra izango dugu), ρ-rako 95%-eko konfidantza-tartea aurkitzen dugu:

$$(0,465\ ,\ 0,819)$$

PROBLEMAK

1.- Funtzio exponentzialen doikuntza.

Fenomeno biologiko askotan, hazkundean batez ere, balioak sistematikoki ematen zaizkion X zorizko aldagaia Y zorizko aldagaiarekin lotzen duen erregresio-lerroa,

$$Y = B \cdot e^{C \cdot X} \qquad (1)$$

exponentziala da.

Problema honetan, B eta C parametroak nola estima eta doikuntzaren egokitasuna nola juzka azalduko dugu.

a) Parametroen estimazioa:

(1)-en bi ataletan logaritmoak hartuz, ondokoa daukagu:

$$\ln Y = \ln B + C \cdot X$$

eta $\ln Y = Y'$, $\ln B = \beta$ eta $C = \alpha$ ordezkatuz, zuzenaren ekuazioa den ondoko adierazpena lortzen dugu:

$$Y' = \beta + \alpha X \qquad (2)$$

non α eta β parametroak, minimo karratuen metodoaren bidez estimatzen diren.

Estimazio hauek, a eta b-z adieraziko ditugu:

$$a = \frac{\sum x_i y'_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} \left(\sum x_i\right)^2} \quad (3)$$

$$b = \bar{Y}' - a\,\bar{X} \quad (4)$$

(3) eta (4) adierazpenen bidez, a eta b estimatuz gero, $B = e^{\beta}$, e^b-ren bidez, eta $C = \alpha$, a-ren bidez estimaturik daude.

Beraz, erregresio-lerroa hauxe izango da:

$$Y = e^b . e^{a.X}$$

b) Doikuntzaren egokitasuna

α, β eta $Y'_o = \alpha x_o + \beta$ -tarako konfidantza-tarteak lor daitezke. Orduan, β_1 eta β_2, β-ren mugak eta α_1, α_2 α-renak baldin badira, B-ren mugak e^{β_1} eta e^{β_2}, eta C-renak α-ren berberekoak izango dira; Y'_o-ren mugak Y'_{o1} eta Y'_{o2} badira, Y_o-renak $e^{Y'_{o1}}$ eta $e^{Y'_{o2}}$ izanik.

Problema.

Denboran zehar, oilasko-enbrioien pisuaren emendioa, $Y = B \cdot e^{C \cdot X}$ menpekotasun estokastikoaren lege exponentzialari darraiola ezagutzen da experimentalki. Enbrioiaren jaiotzetiko 6.egunetik 16.erako pisuak ondoko taulan azaltzen dira:

Enbrioiaren adina (egunetan)	Enbrioiaren pisua (gr.-tan)
X	Y
6	0,029
7	0,052
8	0,079
9	0,125
10	0,181
11	0,261
12	0,425
13	0,738
14	1,130
15	1,882
16	2,812
121	

a) Kalkula itzazu erregresio-lerroa eta estimazio-errore standarda.

b) Azter ezazu doikuntzaren egokitasuna.

2.- Funtzio potentzialen doikuntza.

Puntu-hodeiari $Y = B \cdot X^{\alpha}$ kurba exponentziala doitzeko, aurreko problemaren antzera egingo dugu.

Logaritmoak hartzen dira, ondokoa lorturik:

$$\log Y = \log B + \alpha \log X$$

Eta $\log Y = Y'$, $\log B = \beta$, $\log X = X'$ ordezkatuz, zuzenaren ekuazioa den ondoko adierazpena daukagu:

$$Y' = \alpha X' + \beta$$

Beraz, α eta β parametroak estimatzeko, beti bezala egingo dugu, a eta b-ren formulak ezagunak izanik.

a eta b estimatuz gero, B = 10^{β} , 10^{b}-ren bidez, eta α, a-ren bidez estimaturik daude.

Beraz, erregresio-lerroa Y = $10^{b} \cdot X^{a}$ da.

Problema.

Tenperaturak konstante dirauelarik, 6 presio desberdinetan jarritako gas-masak beteriko bolumenak azaltzen dira ondoko taulan:

Presioa (Kgf/cm^2-tan) X	Bolumena (l.-tan) Y
0,5	1,62
1	1
1,5	0,76
2	0,62
2,5	0,52
3	0,46

Doi iezaiozu Y = B·X^{α} motako kurba, 6 neurri-bikote experimentaletako laginari.

Kalkula itzazu estimazio-errore standarda, eta B eta α-rako konfidantza-tarteak.

3.- Ondoko taula emanik eta X zorizko aldagaia askea izanik, zein da erregresio-lerroaren ekuazioa?. X = 72 denean, zein da Y-rako itxarotako balioa?

X	Y
70	127
66	112
75	146
71	131
67	116
62	97
411	729

4.- 1969-1973.urteetan zehar, 1000 gizabanakoren natalitatea adierazten du ondoko taulak:

Urteak	1969	1970	1971	1972	1973
Natalitatea	56	59	64	65	73

a) Kalkula ezazu korrelazio linealaren koefizientea eta interpreta ezazu beraren balioa.

b) Zenbat jaiotze espero izango dugu 1978.urterako?

5.- Denboran zehar, preparatu hormonalaren aktibitatea determinatzen da, ondoko emaitzak lorturik:

Denbora (hilabetetan)	1	2	3	4	5
Aktibitatea (%-tan)	90	75	42	30	21

Kalkula itzazu:

a) Denborarekiko aktibitatearen erregresio-zuzena.

b) Erregresio-zuzenaren arauera, 90%-eko aktibitateari dagokion denbora (hilabetetan).

c) Aktibitatea, noiz izango da nulua?

6.- Bost pertsonaren altuerak eta pisuak ondokoak dira:

Altuerak (hazbetetan)	64	68	70	72	74
Pisuak (libratan)	160	170	180	190	195

a) 69 hazbetetako altuera duen pertsonarentzat, zein izango da itxarotako pisua?

b) 185 libratako pisua den pertsonarentzat, zein izango da itxarotako altuera?

7.- Eguneko kaloria-kopuruari buruzko eta asteko pisu-galerari buruzko ondoko datuak biltzen ditu medikuak:

Kaloria-kopuruak	2000	1200	1500	500	800
Pisu-galerak	1	3	2	5	4

a) Aurki ezazu eguneko kaloria-kopuruaren eta asteko pisu-galeraren arteko korrelazioa.

b) Eguneko 700 kaloriatako janeurria egiten duen pertsonarentzat, zein pisu-galera espero izango dugu?

8.- Austriako Worther lakuko Y mgr./l. oxigenoa neurtu da X metrotako sakoneran, ondoko datuak lorturik:

X	15	20	30	40	50	60	70
Y	6,5	5,6	5,4	6	4,6	1,4	0,1

a) Doi iezaiezu erregresio-zuzena lorturiko datuei. (Eragiketak, hamarren batez egin itzazu).

b) Azter ezazu aldagai bien arteko korrelazioa.

c) 75 eta 80 metroren arteko sakonerako, zein oxigeno-kantitate aurresan daiteke?

9.- Glukogeno-kantitate fixuko bost lagin hartu dira. X glukogenasa-kantitatea (mmol/l.-tan) aplikatu zaie, kasu bakoitzaren Y erreakzio-abiadura neurturik mol./ml.-tan.

Ondoko taula daukagu:

X	1	2	3	0,2	0,5
Y	18	35	60	8	10

a) Datu hauetatik, erreakzio-abiadura glukogenasa-kantitatearekin batera emendatzen dela ondorioztatzen

ote da? Justifika ezazu zehatz-mehatz erantzuna.

b) Lagin bati, 5 mmol./l.-tako glukogenasa-kantitatea aplikatuko bagenio, zein izango litzateke itxarotako erreakzio-abiadura? Zein zehaztasun-maila izango du?

10.- Giltzurrun-funtzioari buruzko experimentuan, txakur bati, iodo irradiaktiboa (I^{131}) zuen konposatua xiringatu zitzaion, gernuaren Y iodo-kontzentrazioa neurturik, jarraian adierazten diren X denbora-tartetan (minututan):

X	Y
0,5	21,5
3	18
3,5	14
4	11,82
5	7,87
6	6,36
7	5,07
8	3,72
9	3,33
10	2,86
11	2,46
12	2,34
13	2,03
14	1,98
15	1,69

Estima ezazu dagokion erregresio-kurba exponentziala.

11.- Ingurugiroko tenperatura desberdinetan neurturik, txoriek kontsumituriko oxigeno-kantitateak dira ondoko datuak:

O_2-ren kontsumoa (ml./gr-h.-tan)	5,2	4,7	4,5	3,6	3,4	3,1	2,7	1,8
Tenperatura (°C-tan)	-18	-15	-10	-5	0	5	10	19

Tenperatura, X aldagai asketzat harturik, ondokoa eskatzen da:

a) Estima ezazu erregresio-zuzena. (Estima itzazu a eta b).

b) Kalkula ezazu estimazio-errore standarda.

c) Lor itzazu α , β eta $Y_o = \alpha X_o + \beta$ parametroetarako konfidantza-mugak. X_o-ri balio batzu ematen baldin bazaizkio, $X_o = \bar{X}$, $X_o = \bar{X} - 5$, $X_o = \bar{X} - 10$, $X_o = X - 15$, $X_o = \bar{X} + 5$, $X_o = \bar{X} + 10$, $X_o = \bar{X} + 15$ esate baterako, eta balio bakoitzerako $Y_o = \alpha X_o + \beta$ -ren konfidantza-mugak lortzen badira, beheko mugak eta goiko mugak lotuz, bi kurba (bata, behekoa, eta bestea, goikoa) lortzen dira, zeintzuek erregresio-zuzenaren konfidantza-mendela mugatzen duten. Lor ezazu konfidantza-mendela, $\gamma = 0,05$ izanik.

12.- Presio sistolikoaren (mm.Hg-tan) eta pertsona helduen adinaren arteko menpekotasuna aztertzeko, 12 pertsonatako lagina behatu zen, ondoko emaitzak lorturik:

Adina (urtetan)	Odolaren presio sistolikoa (mm.Hg-tan)
X	Y
30	107
30	109
40	118
40	120
40	124
50	134
50	136
60	145
60	148
60	150
70	158
70	162

Doi ezazu erregresio-lerroa eta kalkula ezazu estimazio-errore standarda.

13.- Zoriz aukeratu diren 10 ikasleren bidez, Unibertsitate-ikasturteko bost asignaturatan saiaturiko irakaskuntz metodoaren eraginkortasuna kontrastatu nahi dugu. Ikasle hauen "X" azken kalifikazioen emaitzak hartu dira. Adimen-testa ere aplikatu zaie, beraren balioak Y izanik.

Azterketa honen helburua, lorturiko kalifikazioen eta adimen-indizeen arteko erlazioa lortzea da.

Ondoko taularen datuetatik abiatuz, aurki itzazu:

a) X gaineko Y-ren erregresio-zuzena.

b) α eta β -rako, 95%-eko konfidantza-tarteak.

Adimen-indizea	Kalifikazioak
Y	X
120	48
100	40
95	35
105	45
80	20
90	24
110	40
100	34
85	18
90	32
975	336

14.- "Neurospora crassa" onddoa, 30°C-tako agar-giroan, oxigenoko 5%-a duen gas geldoz beteriko hodietan hazi zen.

mm./h.-tan emandako ondoko hazkunde-abiadurak, non bakoitza 5 edo 6 probaren batezbestekoa den, gas bakoitzaren pisu molekularraren funtzioarekin erlaziona daitezke.

Gasa	He	Ne	N_2	Ar	Kr	Xe
X P.M.	4	20,2	28,2	39,9	83,8	131,3
Y mm./h.	3,51	3,14	3,03	2,71	2,27	1,86

Doi ezazu kurba potentziala eta kalkula ezazu estimazio-errore standarda.

X. GAIA: BARIANTZ ANALISIA.

X.1.- Sarrera.

X.2.- Aldakuntz faktore bakar batekiko bariantz analisia.

X.2.1.- Batezbestekoen kontrastea. Bariantz analisia.

X.2.2.- Bariantz analisiaren kalkuluetarako lan-koadroa.

X.3.- Bi aldakuntz faktore askerekiko bariantz analisia.

X.3.1.- Zutabe- eta errenkada-populazioen batezbestekoen arteko berdintasunerako kontrastea.

X.3.2.- Lan-koadroa.

X.4.- Hiru aldakuntz faktore askerekiko bariantz analisia.

BARIANTZ ANALISIA

X.1.- SARRERA

Azterturiko faktorearen balioa, berarekin erlazionaturiko beste faktore baten balioaren bidez aurresateko formula aurkitzea interesatuko balitzaio ikertzaileari, berak erabiliko lituz keen teknikak azaldu ditugu aurreko gaian. Gai honetan, desberdin samarra den egoera aztertuko dugu, non, interesa, berarekin erlazionaturiko faktoreen bidez faktore baten balioa aurresatean ez datzan, baizik eta magnitude baten gaineko faktore bat edo batzuren eraginak konparatzean.

VII. gaian, Student-en t banaketaren bidez, bariantza berdineko bi populazio normal askeren batezbestekoen arteko berdintasuna kontrastatzeko metodo estatistikoa garatu zen. Gai honetan, "bariantz analisia" deituriko teknika oso ahaltsua garatuz orokortuko dugu metodo hura. Snedecor-en F banaketa erabiliz, bariantza berdineko zenbait populazio normal askeren batezbestekoen arteko berdintasuna batera kontrastatzea baimentzen digu teknika honek.

X.2.- ALDAKUNTZ FAKTORE BAKAR BATEKIKO BARIANTZ ANALISIA.

Askotan, maila desberdinetan sailkaturiko faktoreak, magnitude neurgarriaren batezbesteko balioetan eragina duenentz jakitea interesatzen zaigu.

Ikerketa-mota hau sorterazi zuen adibide historikoaren bidez azalduko dugu metodoa.

Suposa dezagun, ongarri-mota aldatzerakoan, zenbait galsororen batezbesteko etekinaren aldaketa posible aztertu nahi dugula. Kasurik sinpleena, atal honetan aztertuko duguna, hauxe da: Ongarria ezik, gariaren etekinean eragin dezaketen beste faktoreek (hazi-mota, lurraren kalitatea, hezetasuna, eta abar) lur

zatiz lurzati konstante dirautela suposatzen da, hau da, aldatzen den faktore bakarra ongarri-mota dugu.

Experimentuan, $A_1,\ldots,A_k$, k ongarri-mota erabiltzen badira, eta N lurzati homogeno aukeratzen badira, haietariko n_1 A_1 ongarriaz, n_2 A_2 ongarriaz,..., n_k A_k-z ereinik, ondoko taulan azaltzen dira lorturiko etekinak:

		A faktorearen mailak (ongarri-motak)					
		A_1	A_2	...	A_j	...	A_k
Lurzati-zenbakia	1	x_{11}	x_{12}	...	x_{ij}	...	x_{1k}
	2	x_{21}	x_{22}	...	x_{2j}	...	x_{2k}
	⋮	⋮	⋮		⋮		⋮
	i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ik}
	⋮	⋮	⋮		⋮		⋮
		$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_j j}$	...	$x_{n_k k}$
Baturak		x_1	x_2	...	x_j	...	x_k
Batezbestekoak		$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_j$	...	$\bar{x}_k$

⟶ Zutabeetako laginen tamainuak, $n_1,\ldots, n_k$ dira, $\sum_{j=1}^{k} n_j = N$ izanik. Beraz, zutabeak luzera berberekoak izateko zergatirik ez dago. Hau ikusten da aurreko koadroan, non lehen zutabeak n_1 elementu dauzkan, bigarrenak n_2 eta abar... Honek, ordea, ez du inolako garrantzirik.

⟶ x_{ij} elementuak, A_j ongarriaz ereiniko n_j lurzatien arteko i.lurzatiaren etekina adierazten du.

⟶ Zutabeen azpian, beraien baturak eta batezbestekoak agertzen dira.

Problema horrela planteaturik, normalak, askeak eta bariantza berdinekoak direla suposatuko duguneko k populazio imajina ditzakegu.

Bira, P_1, A_1 ongarriaz ereiniko lurzati guztien etekin-populazioa

P_2, A_2 ongarriaz ereiniko lurzati guztien etekin-populazioa

⋮

P_k, A_k ongarriaz ereiniko lurzati guztien etekin-populazioa eta bira $\mu_1, \ldots, \mu_k$ populazio hauen batezbesteko ezezagunak.

Planteatzen dugun arazoa hauxe da:

H_o: $\mu_1 = \mu_2 = \ldots = \mu_k$ kontrastatzea, α adierazgarritasun-mailaz.

Kontraste hau burutzeko, k populazioen bariantza komuneko bi estimazio kalkulatzen dira:

1) Klasebarneko bariantza, S_I^2

Biz, j.zutabeko laginaren kuasibariantza den

$$s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1}$$

$$j = 1, 2, \ldots, k$$

Orduan, klasebarneko bariantza honela definitzen da:

$$S_I^2 = \frac{(n_1-1)s_1^2 + \ldots + (n_k-1)s_k^2}{(n_1-1) + \ldots + (n_k-1)} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2}{N - k} \quad (1)$$

Hau da, S_I^2 klasebarneko bariantza, koadroaren zutabeetako k laginen kuasibariantzen batezbesteko astatua da.

k laginen kuasibariantzen batezbesteko astatuaren bidezko

$\sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2$-ren estimazio hau, era naturalaz burutzen da.

- S_I^2-k (N-k) askatasun-gradu ditu.

- S_I^2, σ^2-ren estimatzaile alboragabea da beti; hots, $E(S_I^2) = \sigma^2$

2) Klasearteko bariantza, S_{II}^2

Koadroari dagokionez, bere elementu guztien batezbestekoa

$$\bar{x} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{N} \qquad (2)$$

da eta klasearteko bariantza honela definitzen dugu:

$$S_{II}^2 = \frac{\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2}{k-1} \qquad (3)$$

- S_{II}^2-k (k-1) askatasun-gradu ditu.

- Populazioen batezbestekoak berdinak direnean, S_{II}^2, σ^2-ren estimatzaile alboragabea da.

- Eta (3) adierazpena behaturik, $\bar{x}_j$-ren kuasibariantza dela ikusten da, bakoitzari n_j maiztasuna, hots, dagokion laginaren tamainua elkarturik.

X.2.1.- Batezbestekoen kontrastea. Bariantz analisia.

H_o egiazkoa bada, S_I^2 eta S_{II}^2, σ^2-ren estimatzaile alboragabeak dira. Beraz, S_{II}^2 / S_I^2 zatidura, k-1 eta N-k askatasun-gradutako Snedecor-en F banaketari darraio. Hau da,

$$\frac{S_{II}^2}{S_I^2} = \frac{\frac{1}{k-1} \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2}{\frac{1}{N-k} \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2} = F_{k-1,N-k} \qquad (4)$$

Beraz, H_o hipotesia kontrastatzen duen testa burutzeko, honela egingo dugu:

- S^2_{II}/S^2_I proba-estatistikoa kalkulatzen da.
- Tauletan, $F_{\alpha;k-1,N-k}$ puntu kritikoa bilatzen da, non α, aukeraturiko adierazgarritasun-maila den.
- Bi balioak konparatzen dira:
  - $\frac{S^2_{II}}{S^2_I} \leqslant F_{\alpha;k-1,N-k}$ baldin bada, $H_o: \mu_1=\ldots=\mu_k$ hipotesia onartzen da.
  - $\frac{S^2_{II}}{S^2_I} > F_{\alpha;k-1,N-k}$ baldin bada, H_o errefusatzen da, α adierazgarritasun-mailaz.

Baina, testaren burutzapen praktikorako, $\frac{S^2_{II}}{S^2_I}$ proba-estatistikoaren balioaren kalkulua, ondokoa kontutan hartuz erraz daiteke:

a) $$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}x_{ij}^2 - \sum_{j=1}^{k}x_j\bar{x}_j$$

b) $$\sum_{j=1}^{k}n_j(\bar{x}_j-\bar{\bar{x}})^2 = \sum_{j=1}^{k}x_j\bar{x}_j - x\bar{\bar{x}}$$

non x_j, koadroaren j.zutabeko elementuen batura eta x, koadroaren x_{ij} guztien batura diren.

Orduan,

$$\frac{S^2_{II}}{S^2_I} = \frac{\frac{1}{k-1}\left(\sum_{j=1}^{k}x_j\bar{x}_j - x\bar{\bar{x}}\right)}{\frac{1}{N-k}\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j}x_{ij}^2 - \sum_{j=1}^{k}x_j\bar{x}_j\right)}$$

eta formula honen bidez, testa burutzeko beharrezkoa den proba-estatistikoaren balioa arinago kalkulatzen da, balio hau kalkulatzeko ondoko lan-koadroa erabiltzen delarik:

X.2.2.- Bariantz analisiaren kalkuluetarako lan-koadroa:

Bariantz analisirako urrats guztiak zehaztu eta gero, lan-koadroa ondoko erara eraikitzen da:

Populazioak	A_1	A_2	...	A_j	...	A_k
	x_{11}	x_{12}	...	x_{1j}	...	x_{1k}
	x_{21}	x_{22}	...	x_{2j}	...	x_{2k}
	⋮	⋮		⋮		⋮
	$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_j j}$	...	$x_{n_k k}$
Baturak	x_1	x_2	...	x_j	...	x_k
Batezbestekoak	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_j$	...	$\bar{x}_k$
Biderkadurak	$x_1\bar{x}_1$	$x_2\bar{x}_2$	...	$x_j\bar{x}_j$	...	$x_k\bar{x}_k$

Koadro honetan, bere aldetik kalkulatzen den $\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}^2$ batura ezik, proba-estatistikoaren kalkulurako beharrezkoak diren kantitateak agertzen dira.

1. adibidea:

Galsoroen batezbesteko etekinarekiko ongarri-motaren eragin posiblea aztertzeko, 16 lurzati homogeno aukeratu ziren, eta launaka banatu ziren zoriz, talde bakoitzekoak ongarri-mota berberaz ereinik. Lorturiko emaitzak ondokoak ziren, kinta-le/Ha.-tan azaldurik:

A faktorea (ongarri-mota)

	Ongarria	A	B	C	D	
	Lurzatia 1	218	263	274	207	
	Lurzatia 2	198	201	248	180	
	Lurzatia 3	215	219	243	318	
	Lurzatia 4	174	226	325	183	
(1)	Baturak	805	909	1090	888	$x=3692; \bar{x}=230,75$
(2)	Batezbestekoak	201,25	227,25	272,50	222	$x\bar{x} = 851929$
(3)	Biderkadurak	162006,25	206570,25	297025	197136	$\sum_j x_j\bar{x}_j = 862737,5$

$\longrightarrow$ Erabil ezazu $\alpha = 0,05$

$\longrightarrow$ Bariantz analisiaren kalkuluetarako beharrezkoak diren errenkadez osotzen da taula. Erantsitako errenkadak, (1), (2) eta (3) dira.

$\longrightarrow \sum_j \sum_i x_{ij}^2 = 882952$ ere kalkulatzen da.

Eta, orain, "testa" burutzen dugu:

$$\frac{S_{II}^2}{S_I^2} = \frac{\frac{1}{k-1}\left(\sum_{j=1}^{k} x_j \bar{x}_j - x\bar{x}\right)}{\frac{1}{N-k}\left(\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^{k} x_j \bar{x}_j\right)} = \frac{\frac{1}{3}(862737,5 - 851929)}{\frac{1}{16-4}(882952 - 862737,5)} =$$

$$= 2,1388$$

$$F_{\alpha;k-1,N-k} = F_{0,05;3,12} = 3,4903$$

$2,1383 < 3,4903$ denez, $H_o: \mu_1 = \ldots = \mu_4$ hipotesia onartzen da, hots, lau ongarri-motekiko lau populazioen batezbesteko etekinen arteko berdintasuna onartzen da. Beraz, ez dago batezbesteko etekinen arteko desberdintasun adierazgarririk, lau ongarri-motetariko edozein erabiltzeak axolarik ez izanik.

X.3.- BI ALDAKUNTZ FAKTORE ASKEREKIKO BARIANTZ ANALISIA.

Atal honetan, magnitude baten batezbesteko balioaren gaineko bi faktore askeren eragin posiblearen aldibereko azterketa burutuko dugu. Bi faktoreak askeak direla esatean, beraien arteko elkarrekintzarik ez dagoela adierazi nahi dugu. Galsoroen batezbesteko etekinen adibidera itzulirik, (A) ongarri-mota eta (B) hazi-mota faktoreak kontutan hartzen baditugu, eta askeak baldin badira, ongarri-motaren aldaketek eragin berbera dute (eraginik baldin badute) hazi-mota desberdinen batezbesteko etekinetan. Gertatzen ez dena ondokoa da: faktoreen maila desberdinek, magnitudearen batezbesteko balioetarako ekarpenean elkar berrindartzen ala ahultzen dutela, hots, ez dago elkarrekintzarik.

Bi aldakuntz faktore aske daudenean, bariantz analisiaren azterketan, magnitudearen batezbesteko balioan eragin dezaketen beste faktoreek konstante dirautela suposatzen da.

Beste atalean egin dugun antzera, adibidearen bidez azaldu duko dugu metodoa. Demagun, ongarri-mota eta hazi-mota aldatzean, galsoroen batezbesteko etekinaren aldaketa posiblea aztertu nahi dugula.

Experimentuan, $A_1,\ldots, A_k$, k ongarri-mota eta $B_1,\ldots, B_n$, n hazi-mota erabiltzen badira, lorturiko emaitzak ondoko koadroan azaltzen dira:

		A faktorea					
Populazioak		A_1	...	A_j	...	A_k	Batezbestekoak
B faktorea	B_1	x_{11}	...	x_{1j}	...	x_{1k}	f_1
	B_2	x_{21}	...	x_{2j}	...	x_{2k}	f_2
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	B_n	x_{n1}	...	x_{nj}	...	x_{nk}	f_n
Batezbestekoak		c_1	...	c_j	...	c_k	$\bar{x}$

⟶ Honela, zutabe guztietako laginen tamainuak n eta errenkada guztietakoenak k direla ikusten da, A_j zutabeko elementuen batezbestekoa c_j-z eta B_i errenkadako elementuen bateztestekoa f_i-z adierazten ditugularik.

⟶ x_{ij} elementuak, B_i hazi-motaz ereiniko eta A_j-z ongarriztaturiko lurzatiaren etekina adierazten du.

Arazoa honela planteaturik, P_{A_1}, $P_{A_2},\ldots, P_{A_k}$, k zutabe-populazio eta $P_{B_1},\ldots, P_{B_n}$, n errenkada-populazio imajina ditzakegu eta n+k populazioak (errenkadetakoak eta zutabetakoak), normalak, askeak eta bariantza berdinekoak direla suposatuko dugu.

Bira $\mu_{A_1},\ldots, \mu_{A_k}$ eta $\mu_{B_1},\ldots, \mu_{B_n}$, $P_{A_1},\ldots, P_{A_k}$ eta

$P_{B_1}, \ldots, P_{B_n}$ populazioen batezbesteko ezezagunak.

Honela planteatzen da arazoa:

1) H_1: $\mu_{A_1} = \ldots = \mu_{A_k}$ (zutabe-populazioen batezbestekoen arteko berdintasuna) hipotesia kontrastatzea.

2) H_2: $\mu_{B_1} = \ldots = \mu_{B_n}$ (errenkada-populazioen batezbestekoen arteko berdintasuna) hipotesia kontrastatzea.

Kontraste biak batera burutzen dira.

Kontrasteak burutzeko, komuna dela suposatzen dugun n+k populazioen bariantzaren hiru estimazio kalkulatzen dira:

1) Errenkaden arteko bariantza, S_f^2:

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

- H_2 betetzen denean, S_f^2, σ^2-ren estimatzaile alboragabea da, (n-1) askatasun-gradu dituelarik.

2) Zutabeen arteko bariantza, S_c^2:

$$S_c^2 = \frac{n\sum_{j=1}^{k}(c_j-\bar{x})^2}{k-1}$$

- H_1 betetzen denean, S_c^2, σ^2-ren estimatzaile alboragabea da, (k-1) askatasun-gradu dituelarik.

3) Erroreen bariantza edo hondar-bariantza, S_e^2:

$$S_e^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k} d_{ij}^2}{(n-1)\cdot(k-1)} \; ; \; d_{ij} = x_{ij}-f_i-c_j+\bar{x} \text{ izanik.}$$

- S_e^2, beti da σ^2-ren estimatzaile alboragabea, (n-1)·(k-1) askatasun-gradu dituelarik.

X.3.1.- Zutabe- eta errenkada-populazioen batezbestekoen arteko berdintasunerako kontrastea.

S_f^2 , S_c^2 eta S_e^2-ren propietateek, ondoko adierazpenak eraikitzea baimentzen digute:

$$\frac{S_f^2}{S_e^2} = \frac{\frac{k}{n-1}\sum_{i=1}^{n}(f_i-\bar{x})^2}{\frac{1}{(n-1)(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}d_{ij}^2} = F_{n-1,(n-1)(k-1)}, H_2 \text{ egiazkoa denean.}$$

$$\frac{S_c^2}{S_e^2} = \frac{\frac{n}{k-1}\sum_{j=1}^{k}(c_j-\bar{x})^2}{\frac{1}{(n-1)(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}d_{ij}^2} = F_{k-1,(n-1)(k-1)}, H_1 \text{ egiazkoa denean.}$$

Beraz, H_1 eta H_2 (biak) hipotesiak kontrastatzen dituen "testa" burutzeko, honela egiten da:

- $\frac{S_f^2}{S_e^2}$ eta $\frac{S_c^2}{S_e^2}$ proba-estatistikoak kalkulatzen dira.

- $F_{\alpha;n-1,(n-1)(k-1)}$ eta $F_{\alpha;k-1,(n-1)(k-1)}$ puntu kritikoak bilatzen dira taulan, α, aukeraturiko adierazgarritasun-maila izanik.

- Ondoko konparaketak egiten dira:

  - $\frac{S_f^2}{S_e^2} \leqslant F_{\alpha;n-1,(n-1)(k-1)}$ baldin bada, $H_2: \mu_{B_1} = \ldots = \mu_{B_n}$ hipotesia onartzen da,

  - $\frac{S_c^2}{S_e^2} \leqslant F_{\alpha;k-1,(n-1)(k-1)}$ baldin bada, $H_1: \mu_{A_1} = \ldots = \mu_{A_k}$ hipotesia onartzen da,

  α adierazgarritasun-maila dutelarik.

Baina, testaren burutzapenerako, $\frac{S_f^2}{S_e^2}$ eta $\frac{S_c^2}{S_e^2}$ proba-estatistikoen balioen kalkulua, ondoko adierazpenen laguntzaz sistematizatzen da:

a) $k \cdot \sum_{i=1}^{n} (f_i - \bar{x})^2 = \sum_{i=1}^{n} F_i f_i - x\bar{x}$, non F_i, i.errenkadako elementuen batura den.

b) $n \sum_{j=1}^{k} (c_j - \bar{x})^2 = \sum_{j=1}^{k} C_j c_j - x\bar{x}$, non C_j , j.zutabeko elementuen batura den.

c) $\sum_{i=1}^{n} \sum_{j=1}^{k} d_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \sum_{i=1}^{n} F_i f_i - \sum_{j=1}^{k} C_j c_j + x\bar{x}$

Beraz,

$$\frac{S_f^2}{S_e^2} = \frac{\frac{1}{n-1} \left(\sum_{i=1}^{n} F_i f_i - x\bar{x} \right)}{\frac{1}{(n-1)(k-1)} \left(\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \sum_{i=1}^{n} F_i f_i - \sum_{j=1}^{k} C_j c_j + x\bar{x} \right)}$$

$$\frac{S_c^2}{S_e^2} = \frac{\frac{1}{k-1} \left(\sum_{j=1}^{k} C_j c_j - x\bar{x} \right)}{\frac{1}{(n-1)(k-1)} \left(\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \sum_{i=1}^{n} F_i f_i - \sum_{j=1}^{k} C_j c_j + x\bar{x} \right)}$$

Azken bi adierazpenak dira proba-estatistikoak praktikan kalkulatzeko erabiltzen direnak, eta beraietan agertzen diren kantitate guztiak (bere aldetik kalkulatzen den $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2$ batura ezik) erraz lortzen dira, datuak, ondoko erara eraikitako lan-koadroan jarriz:

X.3.2.- Lan-koadroa

Populazioak	A_1	...	A_j	...	A_k	Baturak	Batezbestekoak	Biderkadurak
B_1	x_{11}	...	x_{1j}	...	x_{1k}	F_1	f_1	$F_1 f_1$
⋮	⋮		⋮		⋮	⋮	⋮	⋮
B_i	x_{i1}		x_{ij}		x_{ik}	F_i	f_i	$F_i f_i$
⋮	⋮		⋮		⋮	⋮	⋮	⋮
B_n	x_{n1}		x_{nj}		x_{nk}	F_n	f_n	$F_n f_n$
Baturak	C_1		C_j		C_k	x		$\sum_{i=1}^{n} F_i f_i$
Batezbeste.	c_1		c_j		c_k			
Biderkadur.	$C_1 c_1$		$C_j c_j$		$C_k c_k$	$\sum_{j=1}^{k} C_j c_j$	$\bar{x} = \frac{x}{nk}$	$x\bar{x}$

$\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2$ batura bere aldetik kalkulatuz gero, proba-estatistikoen kalkulurako behar diren kantitateak dauzkagu.

Jarraian, 1.adibidearen antzekoa den adibidea azalduko dugu:

2. adibidea:

16 lurzati homogeno dauzkagularik, lurzatien batezbesteko etekinarekiko hazi-motaren eta ongarri-motaren eragin posiblea aztertu nahi dugu batera.

Honetarako, lau taldetan sailkatzen dira lurzatiak, talde bakoitzekoetan hazi-mota eta ongarri-mota berbera erabiltzen direlarik.

Era experimentalean lorturiko emaitzak (Kintale/Ha-tan) ondoko taulan azaltzen dira:

Populazioak		Ongarri-mota A_1	A_2	A_3	A_4	Baturak	Batezbestekoak $\bar{x}$	Biderkadurak
Hazi-mota	B_1	310	353	366	299	1328	332	440896
	B_2	284	293	335	264	1176	294	345744
	B_3	307	306	339	311	1263	315,75	398792,25
	B_4	267	308	312	266	1153	288,25	332352,25
Baturak		1168	1260	1352	1140	$x=4920$	$\sum_{i=1}^{4} F_i f_i = 1517784,5$	
Batezbestek.		292	315	338	285	$\bar{x}=\frac{4920}{16}=307,5$	$\bar{x}\bar{x}=1512900$	
Biderkadurak		341056	396900	456976	324900	$\sum_{j=1}^{4} C_j c_j = 1519832$		

non, taula, errenkadaka eta zutabekako baturen, batezbestekoen, biderkaduren eta beraien kopuru osoen kalkuluek osoturik dagoen.

$\longrightarrow \sum_{i=1}^{4}\sum_{j=1}^{4} x_{ij}^2 = 1526272$ kalkulatzen da, eta "testa" burutzen hasten gara:

$$\longrightarrow \frac{s_f^2}{s_e^2} = \frac{\frac{1}{3}(1517784,5 - 1512900)}{\frac{1}{3\cdot 3}(1526272 - 1517784,5 - 1519832 + 1512900)} = 9,42$$

$$\longrightarrow \frac{s_c^2}{s_e^2} = \frac{\frac{1}{3}(1519832 - 1512900)}{\frac{1}{3\cdot 3}(1526272 - 1517784,5 - 1519832 + 1512900)} = 13,369$$

$$\longrightarrow F_{\alpha;(n-1),(n-1)(k-1)} = F_{\alpha;(k-1),(n-1)(k-1)} = F_{0,05;3,9} = 3,8626$$

$9,42 > 3,8626$ denez, errenkada-populazioen batezbestekoen arteko berdintasuna errefusatzen da.

Eta $13,369 > 3,8626$ denez, zutabe-populazioen batezbestekoen arteko berdintasuna ere errefusatzen da.

Beraz, batezbesteko etekinak hazi-motaren eta ongarri-motaren menpean daude.

X.4.- HIRU ALDAKUNTZ FAKTORE ASKEREKIKO BARIANTZ ANALISIA.

Teknika experimentalak, bariantz analisiaz baliatzen dira erabaki-fasean. Aurreko ataletan azaldutako ereduetariko bakoitzak, faktore bat edo bi daudeneko analisi faktorialaren ere du matematikoa eratzen du.

Faktoreen maila-kopurua txikia ez denean, analisi faktorialak eskatzen duen experimentu-kopurua handia izan daiteke, neketsua eta burutezina gerta dakigukeelarik. Adibidez, 10 maila dauzkaten A, B eta C, hiru faktoreren eraginaren azterketak, 1.000 experimenturen burutzapena (A, B-rekin: 100; eta AB, C-rekin: 1.000) eskatzen du, saiakuntza, bi faktorerekiko analisi faktorialaren ikuspuntutik egiten bada.

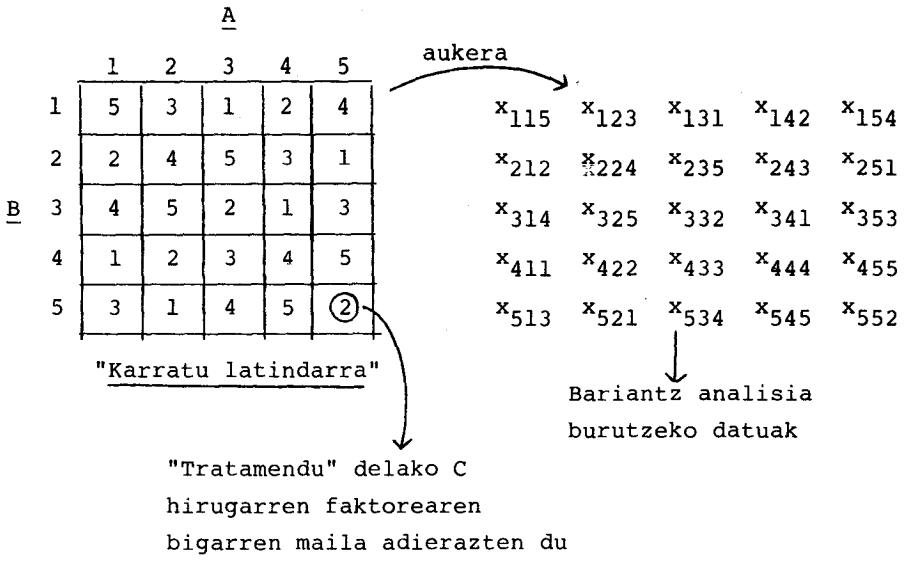
Eragozpen hau errazteko, diseinu experimentalak asmatu dira, zeintzuek, eraginkortasun estatistiko txikiagoa eduki arren, nahi dugun informazioa ematen diguten. Teknika hauen adibide gisa, azterketa trifaktorialean erabilgarria den "karratu latindarren" diseinua aurkeztuko dugu.

Aplikagarria ahal izateko, ondoko baldintza bete behar dira:

- Hiru faktoreetan, maila-kopuru berbera aztertzen da, kopuru hau, N-z adierazten dugularik.

- Normalitate, independentzia eta bariantzen arteko berdintasunaren hipotesiak bete behar dira.

Diseinu osoan N^3 x_{ijk} behaketa kontsideratuko liratekeen bitartean, karratu latindarren diseinuan, N^3-en artean aukeraturiko N^2 experimentutatik datozen N^2 behaketek soilik hartzen dute parte. N^2 (i,j,k) hirukoteen aukera, N dimentsioko karratu latindarren bidez burutzen da. Aukera-mota honen adibide gisa, N=5 dimentsioko karratu latindarra azaltzen dugu, beraren bidez N^2 behaketak aukeraturik:



Karratu latindarren aldez aurreko eraiketa zorizkoa da, zorizko zenbakien taularen bidez eraiki daitekeelarik. Gai honetan, ordea, ez dugu karratuaren hasierako lorpenik azalduko; zoriz emango da kasu bakoitzean.

Karratu latindar honen egiturak, faktore baten maila bakoitza beste faktoreen maila bakoitzarekin behin eta soilik behin elkarturik dagoela aseguratzen du.

Kasu honetan, kontutan hartzen ditugun hipotesiak hauxek dira:

H_1: $\mu_{A_1} = \ldots = \mu_{A_N}$ (ez dago zutabe efekturik)

H_2: $\mu_{B_1} = \ldots = \mu_{B_N}$ (ez dago errenkada efekturik)

H_3: $\mu_{C_1} = \ldots = \mu_{C_N}$ (ez dago tratamendu efekturik)

A eta B faktoreen eraginak konparatzeko, bi faktoretarako analisia burutzen da, non egin behar dugun gauza bakarra, hiru faktoretarako idazkera moldatzea den.

Tratamendua, hots, C faktorea konparatzeko ondokoa daukagu:

$$S_T^2 = \frac{N}{N-1} \left(\sum_t \bar{x}^2_{..t} - N\bar{x}^2 \right)$$

eta

$$S_e^2 = \frac{1}{(N-1)(N-2)} \sum_i \sum_j \sum_t (x_{ijt} - \bar{x}_{i..} - \bar{x}_{.j.} - \bar{x}_{..t} + 2\bar{x})^2$$

$$\frac{S_T^2}{S_e^2} = F_{N-1,(N-1)(N-2)} \text{ da.}$$

Beraz, $\frac{S_T^2}{S_e^2} \leqslant F_{\alpha;N-1,(N-1)(N-2)}$ baldin bada, H_3 onartzen da, desberdintza osagarria, hau da, kontrakoa betetzen denean, H_3 errefusatzen dugularik. Hots, tratamenduaren edo C faktorearen eragina errefusatzen ala onartzen da.

Arazoa, A eta B faktoreetarako birplanteatzen badugu, ondokoa daukagu:

$$S_f^2 = \frac{N}{N-1} \left(\sum_i \bar{x}^2_{i..} - N\bar{x}^2 \right)$$

$$S_c^2 = \frac{N}{N-1} \left(\sum_j \bar{x}^2_{.j.} - N\bar{x}^2 \right)$$

(arestian azterturiko formula berberak dira, baina hiru faktoreren kasurako moldatuak)

Ondoko zatidurak eraikitzen dira:

$$\frac{S_f^2}{S_e^2} = F_{N-1,(N-1)(N-2)} \longrightarrow \text{B faktorerako } (H_2 \text{ hipotesia onartzen ala errefusatzen da})$$

$$\frac{S_c^2}{S_e^2} = F_{N-1,(N-1)(N-2)} \longrightarrow \text{A faktorerako } (H_1 \text{ hipotesia onartzen ala errefusatzen da})$$

Eta honen arauera, H_1 eta H_2 hipotesiak onartzen ala errefusatzen dira.

3. adibidea:

Demagun, A, B eta C hiru farmako lasaigarriren arteko konparaketa aztertzen dugula. "Gizabanako faktoreaz" gain, beste bigarren faktore bat hartu behar dugu kontutan: farmakoa administratzen deneko aldia. Beraz, hiru faktorek hartzen dute parte: gizabanakoa, aldia eta farmakoa.

Karratu latindarraren diseinua, oso sistema zorrotza da, behaketa-kopuru txikienaren bidez hiru faktore koordinatzeko.

Adibidea, hiru farmakoak hiru gizabanakoen gainean eta hiru aldietan probatzean datza (zeren hiru faktoreetan, maila-kopuru berbera aztertu behar bait dugu). Farmako bakoitzak, behin eta soilik behin hartzen du parte aldi bakoitzean eta gizabanako bakoitzarengan.

Esate baterako, ondoko karratu latindarra eratzen dugu zoriz:

Gizabanakoa \ Aldia	1	2	3
1	B (2)	C (3)	A (1)
2	A (1)	B (2)	C (3)
3	C (3)	A (1)	B (2)

$\Longrightarrow$

Ondoko behaketak hartzen dira kontutan:

x_{112} , x_{123} , x_{131}

x_{211} , x_{222} , x_{233}

x_{313} , x_{321} , x_{332}

zeintzuek honako balio hauek dituzten:

Gizabanakoa \ Aldia	1	2	3
1	519 (2)	637 (3)	548 (1)
2	619 (1)	776 (2)	818 (3)
3	701 (3)	641 (1)	678 (2)

Honela, 818 balioa, hirugarren aldian, C farmakoa eman zaioneko bigarren gizabanakoaren behaketa da.

Kalkuluak burutuz, ondokoa daukagu:

B faktorerako, S_f^2 behar dugu:

$$S_f^2 = \frac{N}{N-1}\left(\sum_i \bar{x}_{i..}^2 - N.\bar{x}^2\right)$$

$$\bar{x} = \frac{1}{N^2}\sum_i\sum_j\sum_t x_{ijt} = \frac{5937}{9} = 659,666$$

$$\bar{x}_{1..} = \frac{1}{N}\sum_j\sum_t x_{1jt} = \frac{1}{3}(519+637+548) = \frac{1704}{3} = 568 \longrightarrow \bar{x}_{1..}^2 = 322624$$

$$\bar{x}_{2..} = \frac{2213}{3} = 737,666 \longrightarrow \bar{x}_{2..}^2 = 544152,111$$

$$\bar{x}_{3..} = \frac{2020}{3} = 673,333 \longrightarrow \bar{x}_{3..}^2 = 453377,777$$

$$\left.\begin{array}{l}\sum_i \bar{x}_{i..}^2 = 1320153,888 \\ N.\bar{x}^2 = 1305480,333\end{array}\right\} \Longrightarrow S_f^2 = 22010,332$$

A faktorerako, S_c^2 behar dugu:

$$S_c^2 = \frac{N}{N-1}\left(\sum_j \bar{x}_{.j.}^2 - N\,\bar{x}^2\right)$$

$\bar{x} = 659,666 \qquad N\cdot\bar{x}^2 = 1305480,333$

$$\bar{x}_{.1.} = \frac{1}{N}\sum_i\sum_t x_{ilt} = 613 \longrightarrow \bar{x}_{.1.}^2 = 375769$$

$$\bar{x}_{.2.} = 684,666 \longrightarrow \bar{x}_{.2.}^2 = 468768,444$$

$$\bar{x}_{.3.} = 681,333 \longrightarrow \bar{x}_{.3.}^2 = 464215,111$$

$$\sum_j \bar{x}_{.j.}^2 = 1308752,555$$

$$S_c^2 = 4908,333$$

C faktorerako, S_T^2 behar dugu:

$$S_T^2 = \frac{N}{N-1}\left(\sum_t \bar{x}_{..t}^2 - N\,\bar{x}^2\right)$$

$$N\cdot\bar{x}^2 = 1305480,333$$

$$\bar{x}_{..1} = \frac{1}{N}\sum_i\sum_j x_{ij1} = \frac{1}{N}(x_{131}+x_{211}+x_{321}) = \frac{1}{3}(548+619+641) =$$

$$= 602,666$$

$$\bar{x}_{..1}^2 = 363207,111$$

$$\bar{x}_{..2} = \frac{1}{3}(519+776+678) = 657,666 \longrightarrow \bar{x}_{..2}^2 = 432525,444$$

$$\bar{x}_{..3} = \frac{1}{3}(637+818+701) = 718,666 \longrightarrow \bar{x}_{..3}^2 = 516481,777$$

$$\sum_t \bar{x}_{..t}^2 = 1312214,333$$

$$S_T^2 = 10100,998$$

Orain, S_e^2 kalkulatzen dugu:

$$S_e^2 = \frac{1}{(N-1)(N-2)}\sum_i\sum_j\sum_t (x_{ijt}-\bar{x}_{i..}-\bar{x}_{.j.}-\bar{x}_{..t}+2\bar{x})^2$$

$$(x_{112}-\bar{x}_{1..}-\bar{x}_{.1.}-\bar{x}_{..2}+2\bar{x})^2 = 0,1111$$

$$(x_{123}-\bar{x}_{1..}-\bar{x}_{.2.}-\bar{x}_{..3}+2\bar{x})^2 = 225$$

$$(x_{131}-\bar{x}_{1..}-\bar{x}_{.3.}-\bar{x}_{..1}+2\bar{x})^2 = 235,1111$$

$$(x_{211}-\bar{x}_{2..}-\bar{x}_{.1.}-\bar{x}_{..1}+2\bar{x})^2 = 225$$

$$(x_{222}-\bar{x}_{2..}-\bar{x}_{.2.}-\bar{x}_{..2}+2\bar{x})^2 = 235,1111$$

$$(x_{233}-\bar{x}_{2..}-\bar{x}_{.3.}-\bar{x}_{..3}+2\bar{x})^2 = 0,1111$$

$$(x_{313}-\bar{x}_{3..}-\bar{x}_{.1.}-\bar{x}_{..3}+2\bar{x})^2 = 235,1111$$

$$(x_{321}-\bar{x}_{3..}-\bar{x}_{.2.}-\bar{x}_{..1}+2\bar{x})^2 = 0,1111$$

$$(x_{332}-\bar{x}_{3..}-\bar{x}_{.3.}-\bar{x}_{..2}+2\bar{x})^2 = 225$$

$$2\bar{x} = 1319,32$$

$$S_e^2 = \frac{1}{2\cdot 1} \quad 1380,666 = 690,333$$

Beraz, proba-estatistikoak kalkula ditzakegu:

$$\frac{S_f^2}{S_e^2} = \frac{22010,332}{690,33} = 31,884$$

$$\frac{S_c^2}{S_e^2} = \frac{4908,333}{690,33} = 7,11$$

$$\frac{S_T^2}{S_e^2} = \frac{10100,998}{690,33} = 14,632$$

Hirurak, $F_{N-1,(N-1)(N-2)}$ banaketari darrazkionez, puntu kritikoa $F_{\alpha;N-1,(N-1)(N-2)} = F_{0,05;2,2} = 19$ da.

Orduan,

$31,884 > 19$ denez, H_2 hipotesia errefusatzen dugu.

$7,11 < 19$ denez, H_1 hipotesia onartzen dugu.

$14,632 < 19$ denez, H_3 hipotesia onartzen dugu.

Beraz, aldien arteko eta farmakoen arteko desberdintasuna ez da adierazgarria; gizabanakoen artekoa, osterantzean, bai.

PROBLEMAK

1.- Hiru analgesikoren arteko desberdintasun adierazgarririk dagoen aztertu nahi dugu. Horretarako, buruko mina pairatzen duten pertsonarengandik, hiru lagin aukeratu dira eta analgesiko bakoitzaren dosi egokia eman zaie, eragin lasaigarriaren iraupen-aldiak(ordutan) neurturik. Lorturiko emaitzak ondoko taulan azaltzen dira:

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		

5%-eko adierazgarritasun-mailaz, kontrasta ezazu, minaren batezbesteko inhibizio-aldien arteko berdintasunaren hipotesia.

2.- Landare baten printzipio aktiboaren edukina determinatzeko, analisi kimikoaren lau metodo desberdin aplikatu dira. Hiru baldintza desberdinetan biltegiraturiko landareetan burutzen dira analisiak: 1) landare freskoetan; 2) urtebetean biltegiraturiko landareetan; 3) urtebetean biltegiraturiko landare tratatuetan.

Datu experimentalak ondoko koadroan azaltzen dira:

Biltegirapen-baldintzak	Analisi-metodoa			
	A	B	C	D
1	1,35	1,13	1,06	0,98
2	1,4	1,23	1,26	1,22
3	1,49	1,46	1,4	1,35

Zutabeetako populazioen batezbestekoak berdinak al dira?
Errenkadetako populazioen batezbestekoak berdinak al dira?
Erabil ezazu $\alpha = 0,05$

3.- Itsas-artropodoen lau espezietarako aminoazido askeen edukina (µmol/g.-tan, pisu sikua) determinatu zen. Datuak ondokoak dira:

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		403,7	379,4
			381,3

Kontrasta ezazu, lau artropodo-espezietarako aminoazido askeen batezbesteko edukina berdina deneko hipotesia. Har ezazu $\alpha = 0,05$.

4.- Farmako-laborategian, kapsulak, hautsezko antibiotikoz betetzeko, makina automatikoa erabiltzen da, zeina 500 mg.-tako kapsulak betetzeko prestatu zen. Kapsulen pisuetan behaturiko irregulartasunak, ondoko bi faktoreek sorterazitakoak izan daitezkeela pentsatzen da:

1) Makinaren albeoloen arteko desberdintasunak

2) Hautsaren ezaugarrien (hezetasuna, dentsitatea, e.a.) aldakuntzak.

5 une desberdinetan (1-etik 5-era zenbaturikoak), 4 kapsula (albeolo bakoitzeko bat) aukeratu dira, lorturiko pisuak ondoko taulan adierazten direlarik:

	Albeoloak			
Uneak	1	2	3	4
1	525	528	536	537
2	526	537	532	529
3	539	526	532	540
4	519	528	536	532
5	518	530	533	531

α= 0,05 adierazgarritasun-mailaz, kontrasta itzazu erren kada-populazioen batezbestekoen arteko berdintasuna eta zutabe-populazioenen artekoa.

5.- Erretzen dituen hiru marka desberdinetako zigarroen batezbesteko nikotina-edukinaz oso kezkaturik dagoen erretzaileak, zoriz aukeraturiko zigarroen azterketa burutzen du. mg.-tan adierazitako nikotina-edukinak ondokoak izan ziren:

Marka	Nikotina-edukinak(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-edukina banaketa normalari darraiola, eta hiru populazioen bariantza berbera dela suposatzen baldin badugu, zein ondorio atera daiteke emaitza hauetatik?

6.- Arraza desberdinetako hiru animali motaren erraietan dagoen batezbesteko droga-edukina neurtzeko, experimentua burutu nahi dugu. Arraza bakoitzeko 10 animaliatako hiru zorizko lagin hartzen ditugu eta droga eman ondoren hil egiten ditugu.

Lorturiko erraietako droga-edukinak ondokoak dira:

	Arrazak		
	A	B	C
Droga-edukinak	69	71	71
	72	75	74
	71	72	72
	72	72	70
	72	74	75
	73	76	74
	71	73	73
	68	70	68
	69	75	73
	75	76	74

Hiru populazioen droga-edukina, bariantza berbereko banaketa normalari darraizkiola suposatzen dugu.

Hiru populazioetako erraietako batezbesteko droga-edukinak berdinak direla baieztatzen al dute emaitza hauek?

7.- Termometro klinikoen ekoizpenaz arduratzen den lantegiak, marka desberdinetako lau makina erabiltzen ditu, non makina bakoitza lantegiko lau langiletariko batek erabilia den.

Eguneroko langile eta makinako ekoitzitako unitate-kopuruak (ehunekotan) ondokoak dira:

Makina / Langilea	A	B	C	D
1	14	9	7	8
2	12	11	10	9
3	16	8	8	11
4	14	8	6	10

Egin ezazu dagokion bariantz analisia, suposaketak eta ondorioak zehazturik.

8.- Hiru konplexu bitaminiko konparatu nahi ditugu. Horretarako, urtebeteko hirukitako lau talde bildu ziren. Hirukien arteko haur bakoitzak, bi urte kontsekutibotan zehar, hiru konplexu bitaminikoetariko bat hartu zuen zoriz. Erantzunaldagaia hazkundea dela suposatuz, kg.-tako hamarrenetan adierazitako datuak ondokoak ziren:

		konplexu bitaminikoa		
		A	B	C
Familia	1	68	70	72
	2	74	71	69
	3	66	67	73
	4	70	68	68

Egin ezazu dagokion bariantz analisia.

9.- Suposa dezagun, edozein kinadak (entzumenezkoak, usaimenezkoak, ukimenezkoak), aldaketak sorterazten dituela iluntasunera moldaturik dauden begien sentikortasunean. Baiezpen hau frogatzeko, ondoko experimentua burutu zen: Pertsona bati, bere begiak iluntasunera moldatzeko, 15 minututako denbora eman zitzaion, eta orduan, testa egiten zuen bitartean, soinu gogorrak, soinu ahulak, presio handiak (eskuaren ganbilaren gaineko pisuak), presio txikiak edo ezerezak eragindakoak ziren bere begiak. Testaren erantzun zuzenen kopurua,

iluntasunera moldaturiko begien sentikortasunaren neurritzat hartu zen.

5 pertsonak, "tratamendutzat" harturiko eta arestian adierazitako bost kinadak, eta bost egunek (eguneko proba bakar bat izanik) osoturiko 5 x 5-eko karratu latindarra erabili zen.

Magnitude neurgarria, kontrastearen aurreko sentikortasuna izan zen.

Erabili zen karratu latindarra (egunak, parentesi artean) eta neurketak ondokoak izan ziren:

	Kinadak				
Pertsonak	Soinu gogorra	Soinu ahula	Presio handia	Presio txikia	Ezer ez
A	21 (2)	22 (3)	20 (4)	22 (5)	22 (1)
B	22 (4)	16 (1)	23 (5)	19 (2)	23 (3)
C	14 (1)	14 (5)	23 (2)	24 (3)	20 (3)
D	29 (5)	24 (4)	24 (3)	24 (1)	28 (2)
E	16 (3)	15 (2)	14 (1)	15 (4)	13 (5)

5%-eko mailaz, kontrasta ezazu, entzumenezko eta ukimenezko eragingarriek sorterazitako kontrastearen aurreko sentikortasunaren gaineko eragina eta kinada-ezakoa berberak direneko hipotesia.

10.- Paratinoideseko aterakinaren proba biologikoan, ahalmen ezaguna duten gertakin standardeko bi dosi (0,125 eta 0,205 cm^3/kg) eta ahalmen ezezaguna duten gertakin berriko bi do

si berdin konparatu ziren, horretarako, txakurren seroko kaltzio-mailaren gaineko eraginak neurtu zirelarik.

Lau tratamendu hauei, S_1, S_2, N_1 eta N_2 deritzegu. Tratamendu bakoitza, txakur bakoitzari (lau txakur) behin eman zion lau denbora-tartetan zehar.

Erabili zen karratu latindarra eta behaturiko seroko kaltzio-mailak (mg.-ko ehunekotan) ondokoak ziren:

	Dosia xiringatu zeneko eguna			
Txakurrak	1	2	3	4
1	14,2 (S_1)	14,1 (N_1)	15 (S_2)	14,4 (N_2)
2	13 (N_1)	13,4 (S_1)	13,8 (N_2)	14 (S_2)
3	15,8 (N_2)	16 (S_2)	15 (N_1)	15,4 (S_1)
4	15,2 (S_2)	16,2 (N_2)	15 (S_1)	15,3 (N_1)

α = 0,05 izanik, egiazta ezazu, S_1, N_1, S_2 eta N_2 tratamenduen arteko desberdintasunik ez dagoeneko hipotesi nulua.

11.- Aurreko problemaren datuetarako egin ezazu hiru faktoreretako bariantz analisi osoa.

12.- Hiru lur-mota desberdinetarako batezbesteko etekinen arteko berdintasunaren hipotesia eta bost arto-motatarako batezbesteko etekinen artekoarena kontrastatu nahi ditugu batera.

Lur-mota bakoitzeko 5 lurzati aukeratzen dira zoriz, bakoitza arto-mota bakar batez ereinik. Artoa biltzean, lor

turiko batezbesteko etekinak (Kg/area-tan) ondokoak dira:

		arto-mota				
		A	B	C	D	E
lur-mota	1	280	300	310	270	330
	2	250	240	260	270	240
	3	310	304	290	280	300

Erabil ezazu $\alpha = 0,05$

1. taula

B(n,p) banaketa binomiala

$$p(\xi = k) = \binom{n}{k} p^k q^{n-k}$$

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	0,9026	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	0,3219	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	0,0312
	1		0,0480	0,2036	0,3280	0,3915	0,4096	0,3955	0,3602	0,3292	0,3124	0,2592	0,2059	0,1657	0,1562
	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	0,0004	0,0022	0,0064	0,0146	0,0284	0,0412	0,0488	0,0768	0,1128	0,1470	0,1562
	5		0,0000	0,0000	0,0000	0,0001	0,0003	0,0010	0,0024	0,0041	0,0053	0,0102	0,0185	0,0283	0,0312
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	0,2437	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	0,0139	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	0,0603	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	0,1918	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	0,0312
	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	0,0312
	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	9		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	0,0636	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	0,0495	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

2. taula

$P(\lambda)$ Poisson-en banaketa

$$P(\xi=k)=e^{-\lambda}\frac{\lambda^k}{k!}$$

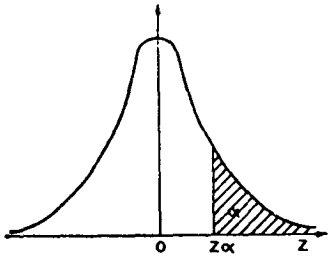
λ \ k	0	1	2	3	4	5	6	7	8	9	10	11	12
0,1	0,9048	0,0905	0,0045	0,0002	0,0000								
0,2	0,8187	0,1637	0,0164	0,0011	0,0001	0,0000							
0,3	0,7408	0,2222	0,0333	0,0033	0,0002	0,0000							
0,4	0,6703	0,2681	0,0536	0,0072	0,0007	0,0001	0,0000						
0,5	0,6065	0,3033	0,0758	0,0126	0,0016	0,0002	0,0000						
0,6	0,5488	0,3293	0,0988	0,0198	0,0030	0,0004	0,0000	0,0000					
0,7	0,4966	0,3476	0,1217	0,0284	0,0050	0,0007	0,0001	0,0000					
0,8	0,4493	0,3595	0,1438	0,0383	0,0077	0,0012	0,0002	0,0000					
0,9	0,4066	0,3659	0,1647	0,0494	0,0111	0,0020	0,0003	0,0001	0,0000				
1,0	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0005						
1,1	0,3329	0,3662	0,2014	0,0738	0,0203	0,0045	0,0008	0,0001	0,0000				
1,2	0,3012	0,3614	0,2169	0,0867	0,0260	0,0062	0,0012	0,0002	0,0000				
1,3	0,2725	0,3543	0,2303	0,0998	0,0324	0,0084	0,0018	0,0003	0,0001	0,0000			
1,4	0,2466	0,3452	0,2417	0,1128	0,0395	0,0111	0,0026	0,0005	0,0001	0,0000			
1,5	0,2231	0,3347	0,2510	0,1255	0,0471	0,0141	0,0035	0,0008	0,0001	0,0000			
1,6	0,2019	0,3230	0,2584	0,1378	0,0551	0,0176	0,0047	0,0011	0,0002	0,0000			
1,7	0,1827	0,3106	0,2640	0,1496	0,0636	0,0216	0,0061	0,0015	0,0003	0,0001	0,0000		
1,8	0,1653	0,2975	0,2678	0,1607	0,0723	0,0260	0,0078	0,0020	0,0005	0,0001	0,0000		
1,9	0,1496	0,2842	0,2700	0,1710	0,0812	0,0309	0,0098	0,0027	0,0006	0,0001	0,0000		
2,0	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120	0,0034	0,0009	0,0002	0,0000		
2,2	0,1108	0,2438	0,2681	0,1966	0,1082	0,0476	0,0174	0,0055	0,0015	0,0004	0,0001	0,0000	
2,4	0,0907	0,2177	0,2613	0,2090	0,1254	0,0602	0,0241	0,0083	0,0025	0,0007	0,0002	0,0000	
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	0,0000
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	0,0000
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	0,0176	0,0348	0,0148	0,056	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	0,0264
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	0,0157	0,0337	0,0607	0,0911	0,1171	0,1318	0,1381	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
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	0,0193	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

3. taula

N(0,1) banaketa normala

$$\int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \alpha$$

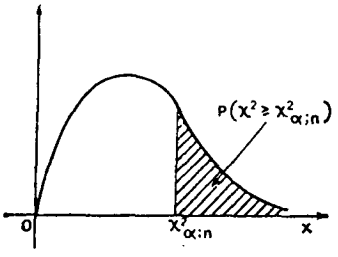


z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	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	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
2,7	0,00256	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264
2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139

z_α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	0,00135	$0,0^3968$	$0,0^3687$	$0,0^3483$	$0,0^3337$	$0,0^3233$	$0,0^3159$	$0,0^3108$	$0,0^4723$	$0,0^4481$
4	$0,0^4317$	$0,0^4207$	$0,0^4133$	$0,0^5854$	$0,0^5541$	$0,0^5340$	$0,0^5211$	$0,0^5130$	$0,0^6793$	$0,0^8479$
5	$0,0^6287$	$0,0^6170$	$0,0^7996$	$0,0^7579$	$0,0^7333$	$0,0^7190$	$0,0^7107$	$0,0^8599$	$0,0^8332$	$0,0^8182$
6	$0,0^9987$	$0,0^9530$	$0,0^9282$	$0,0^9149$	$0,0^{10}777$	$0,0^{10}402$	$0,0^{10}206$	$0,0^{10}104$	$0,0^{11}523$	$0,0^{11}260$

4. taula

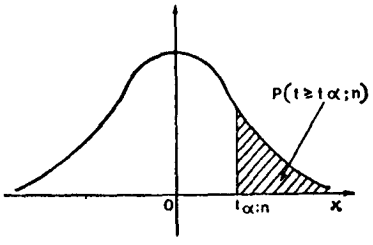
χ^2 banaketa



n \ α	0,995	0,99	0,98	0,975	0,95	0,90	0,10	0,05	0,025	0,02	0,01
1	$0,0^4393$	$0,0^3157$	$0,0^3628$	$0,0_3982$	0,00393	0,0158	2,706	3,841	5,024	5,412	6,635
2	0,0100	0,0201	0,0404	0,0506	0,103	0,211	4,605	5,991	7,378	7,824	9,210
3	0,0717	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	1,646	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	7,042	19,812	22,362	24,736	25,472	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	0,312	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	11,688	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	12,879	14,125	14,573	16,151	18,114	36,741	40,113	43,194	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

5.taula

Student-en "t" banaketa



n \ α	0,40	0,30	0,20	0,10	0,050	0,025	0,010	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
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	5,893	6,859
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	5,405
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	3,611	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	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,192	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	3,690
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	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,259	0,851	1,303	1,648	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,256	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

6. taula

Fisher-Snedecor-en F banaketa

$\alpha = 0{,}10$

n_2 \ n_1	1	2	3	4	5	6	7	8	9
1	39,864	49,500	53,593	55,833	57,241	58,204	58,906	59,439	59,858
2	8,5263		9,1618	9,2434	9,2926	9,3255	9,3491	9,3668	9,3805
3	5,5383	5,4624	5,3908	5,3427	5,3092	5,2847	5,2662	5,2517	5,2400
4	4,5448	4,3246	4,1908	4,1073	4,0506	4,0098	3,9790	3,9549	3,9357
5	0,0604	3,7797	3,6195	3,5202	3,4530	3,4045	3,3679	3,3393	3,3163
6	3,7760	3,4633	3,2888	3,1808	3,1075	3,0546	3,0145	2,9830	2,9577
7	3,5894	3,2574	3,0741	2,9605	2,8833	2,8274	2,7849	2,7516	2,7247
8	3,4579	3,1131	2,9238	2,8064	2,7265	2,6683	2,6241	2,5893	2,5612
9	3,3603	3,0065	2,8129	2,6927	2,6106	2,5509	2,5053	2,4694	2,4403
10	3,2850	2,9245	2,7277	2,6053	2,5216	2,4606	2,4140	2,3772	2,3473
11	3,2252	2,8595	2,6602	2,5362	2,4512	2,3891	2,3416	2,3040	2,2735
12	3,1765	2,8068	2,6055	2,4801	2,3940	2,3310	2,2828	2,2446	2,2135
13	3,1362	2,7632	2,5603	2,4337	2,3467	2,2830	2,2341	2,1953	2,1638
14	3,1022	2,7265	2,5222	2,3947	2,3069	2,2426	2,1931	2,1539	2,1220
15	3,0732	2,6952	2,4898	2,3614	2,2730	2,2081	2,1582	2,1185	2,0862
16	3,0481	2,6682	2,4618	2,3327	2,2438	2,1783	2,1280	2,0880	2,0553
17	3,0262	2,6446	2,4374	2,3077	2,2183	2,1524	2,1017	2,0613	2,0284
18	3,0070	2,6239	2,4160	2,5858	2,1958	2,1296	2,0785	2,0379	2,0047
19	2,9899	2,6056	2,3970	2,2663	2,1760	2,1094	2,0580	2,0171	1,9836
20	2,9747	2,5893	2,3801	2,2489	2,1582	2,0913	2,0397	1,9985	1,9649
21	2,9609	2,5746	2,3649	2,2333	2,1423	2,0751	2,0232	1,9819	1,9480
22	2,9486	2,5613	2,3512	2,2193	2,1279	2,0605	2,0084	1,9668	1,9327
23	2,9374	2,5493	2,3387	2,2065	2,1149	2,0472	1,9949	1,9531	1,9189
24	2,9271	2,5383	2,3274	2,1949	2,1030	2,0351	1,9826	1,9407	1,9063
25	2,9177	2,5283	2,3170	2,1843	2,0922	2,0241	1,9714	1,9292	1,8947
26	2,9091	2,5191	2,3075	2,1745	2,0822	2,0139	1,9610	1,9188	1,8841
27	2,9012	2,5106	2,2987	2,1655	2,0730	2,0045	1,9515	1,9091	1,8743
28	2,8939	2,5028	2,2906	2,1571	2,0645	1,9959	1,9427	1,9001	1,8652
29	2,8871	2,4955	2,2831	2,1494	2,0566	1,9878	1,9345	1,8918	1,8560
30	2,8807	2,4887	2,2761	2,1422	2,0492	1,9803	1,9269	1,8841	1,8498
40	2,8354	2,4404	2,2261	2,0909	1,9968	1,9269	1,8725	1,8289	1,7929
60	2,7914	2,3932	2,1774	2,0410	1,9457	1,8747	1,8194	1,7748	1,7380
120	2,7478	2,3473	2,1300	1,9923	1,8959	1,8238	1,7675	1,7220	1,6843
∞	2,7055	2,3026	2,0838	1,9449	1,8473	1,7741	1,7167	1,6702	1,6315

6. taula (jar.)

Fisher-Snedecor-en F banaketa

$\alpha = 0{,}10$

10	12	15	20	24	30	40	60	120	∞
60,195	60,705	61,220	61,740	62,002	62,265	62,529	62,794	63,061	63,328
9,3916	9,4081	9,4247	9,4413	9,4496	9,4579	9,4663	9,4746	9,4829	9,4913
5,2304	5,2156	5,2003	5,1845	5,1764	5,1681	5,1597	5,1512	5,1425	5,1337
3,9199	3,8955	3,8689	3,8443	3,8310	3,8174	3,8036	3,7896	3,7753	3,7607
3,2974	3,2682	3,2380	3,2067	3,1905	3,1741	3,1573	3,1402	3,1228	3,1050
2,9369	2,9047	2,8712	2,8363	2,8183	2,8000	2,7812	2,7620	2,7423	2,7222
2,7025	2,6681	2,6322	2,5947	2,5753	2,5555	2,5351	2,5142	2,4928	2,4708
2,5380	2,5020	2,4642	2,4246	2,4041	2,3830	2,3614	2,3391	2,3162	2,2926
2,4163	2,3789	2,3396	2,2983	2,2768	2,2547	2,2320	2,2085	2,1843	2,1592
2,3226	2,2841	2,2435	2,2007	2,1784	2,1554	2,1317	2,1072	2,0818	2,0554
2,2482	2,2087	2,1671	2,1230	2,1000	2,0762	2,0516	2,0261	1,9997	1,9721
2,1878	2,1474	2,1049	2,0597	2,0360	2,0115	1,9861	1,9597	1,9323	1,9036
2,1376	2,0966	2,0532	2,0070	1,9827	1,9576	1,9315	1,9043	1,8759	1,8462
2,0954	2,0537	2,0095	1,9625	1,9377	1,9119	1,8852	1,8572	1,8280	1,7973
2,0593	2,0171	1,9722	1,9243	1,8990	1,8728	1,8454	1,8168	1,7867	1,7551
2,0281	1,9854	1,9399	1,8913	1,8656	1,8388	1,8108	1,7816	1,7507	1,7182
2,0009	1,9577	1,9117	1,8624	1,8362	1,8090	1,7805	1,7506	1,7191	1,6856
1,9770	1,9333	1,8868	1,8368	1,8103	1,7827	1,7537	1,7232	1,6910	1,6567
1,9557	1,9117	1,8647	1,8142	1,7667	1,7382	1,7083	1,6988	1,6659	1,6308
1,9367	1,8924	1,8449	1,7938	1,7873	1,7592	1,7298	1,6768	1,6433	1,6074
1,9197	1,8750	1,8272	1,7756	1,7481	1,7193	1,6890	1,6569	1,6228	1,5862
1,9043	1,8593	1,8111	1,7590	1,7312	1,7021	1,6714	1,6389	1,6042	1,5668
1,8903	1,8450	1,7964	1,7439	1,7159	1,6864	1,6554	1,6224	1,5871	1,5490
1,8775	1,8319	1,7831	1,7302	1,7019	1,6721	1,6407	1,6073	1,5715	1,5327
1,8658	1,8200	1,7708	1,7175	1,6890	1,6589	1,6272	1,5934	1,5570	1,5176
1,8550	1,8090	1,7596	1,7059	1,6771	1,6468	1,6147	1,5805	1,5437	1,5036
1,8451	1,7989	1,7492	1,6951	1,6662	1,6356	1,6032	1,5686	1,5313	1,4906
1,8359	1,7895	1,7395	1,6852	1,6560	1,6252	1,5925	1,5575	1,5198	1,4784
1,8274	1,7808	1,7306	1,6759	1,6465	1,6155	1,5825	1,5472	1,5090	1,4670
1,8195	1,7727	1,7223	1,6673	1,6377	1,6065	1,5732	1,5376	1,4989	1,4564
1,7627	1,7146	1,6624	1,6052	1,5741	1,5411	1,5056	1,4672	1,4248	1,3769
1,7070	1,6574	1,6034	1,5435	1,5107	1,4755	1,4373	1,3952	1,3476	1,2915
1,6524	1,6012	1,5450	1,4821	1,4472	1,4094	1,3676	1,3203	1,2646	1,1926
1,5987	1,5458	1,4871	1,4206	1,3832	1,3419	1,2951	1,2400	1,1686	1,0000

7. taula

Fisher-Snedecor-en F banaketa

$\alpha = 0{,}05$

n_2 \ n_1	1	2	3	4	5	6	7	8	9
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8868	8,8452	8,8123
4	7,7086	6,9443	6,5914	6,3883	6,2560	6,1631	6,0942	6,0410	5,9988
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2066	4,1468	4,0990
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767
8	5,3177	4,4590	4,0662	3,8378	3,6875	3,5806	3,5005	3,4281	3,3881
9	5,1174	4,2565	3,8626	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204
11	4,8443	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123	2,9480	2,8962
12	4,7472	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134	2,8486	2,7964
13	4,6672	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321	2,7669	2,7144
14	4,6001	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642	2,6987	2,6458
15	5,5431	3,6823	3,2874	3,0556	2,9013	2,7905	2,7066	2,6408	2,5876
16	4,4940	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572	2,5911	2,5377
17	4,4513	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143	2,5480	2,4943
18	4,4139	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767	2,5102	2,4563
19	4,3808	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435	2,4768	2,4227
20	4,3513	3,4928	3,0984	2,8661	2,7109	2,5900	2,5140	2,4471	2,3928
21	4,3248	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876	2,4205	2,3661
22	4,3009	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638	2,3965	2,3419
23	4,2793	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422	2,3748	2,3201
24	4,2597	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226	2,3551	2,3002
25	4,2417	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047	2,3371	2,2821
26	4,2252	3,3690	2,9751	2,7426	2,5868	2,4741	2,3883	2,3205	2,2655
27	4,2100	3,3541	2,9604	2,7278	2,5719	2,4591	2,3732	2,3053	2,2501
28	4,1960	3,3404	2,9467	2,7141	2,5581	2,4453	2,3593	2,2913	2,2360
29	4,1830	3,3277	2,9340	2,7014	2,5454	2,4324	2,3463	2,2782	2,2229
30	4,1709	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343	2,2662	2,2107
40	4,0848	3,2317	2,8387	2,6060	2,4495	2,3359	2,2490	2,1802	2,1240
60	4,0012	3,1504	2,7581	2,5252	2,3683	2,2540	2,1665	2,0970	2,0401
120	3,9201	3,0718	2,6802	2,4472	2,2900	2,1750	2,0867	2,0164	1,9588
∞	2,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,9384	1,8799

7.taula(jar.)

Fisher-Snedecor-en F banaketa

$\alpha = 0{,}05$

10	12	15	20	24	30	40	60	120	∞
241,88	243,91	245,95	248,01	249,05	250,09	251,14	252,20	253,25	254,32
19,396	19,413	19,429	19,446	19,454	19,462	19,471	19,479	19,487	19,496
8,7855	8,7446	8,7029	8,6602	8,6385	8,6166	8,5944	8,5720	8,5494	8,5265
5,9644	5,9117	5,8578	5,8025	5,7744	5,7459	5,7170	5,6878	5,6581	5,6281
4,7351	4,6777	4,6188	4,5581	4,5272	4,4957	4,4638	4,4314	4,3984	4,3650
4,0600	3,9999	3,9381	3,8742	3,8415	3,8082	3,7743	3,7398	3,7047	3,6688
3,6365	3,5747	3,5108	3,4445	3,4105	3,3758	3,3404	3,3043	3,2674	3,2298
3,3472	3,2840	3,2184	3,1503	3,1152	3,0794	3,0428	3,0053	2,9669	2,9276
3,1373	3,0729	3,0061	2,9365	2,9005	2,8637	2,8259	2,7872	2,7475	2,7067
2,9782	2,9130	2,8450	2,7740	2,7372	2,6996	2,6609	2,6211	2,5801	2,5379
2,8536	2,7876	2,7186	2,6464	2,6090	2,5705	2,5309	2,4901	2,4480	2,4045
2,7534	2,6866	2,6169	2,5436	2,5055	2,4663	2,4259	2,3842	2,3410	2,2962
2,6710	2,6037	2,5331	2,4589	2,4202	2,3803	2,3392	2,2966	2,2524	2,2064
2,6021	2,5342	2,4630	2,3879	2,3487	2,3082	2,2664	2,2230	2,1778	2,1307
2,5437	2,4753	2,4035	2,3275	2,2878	2,2468	2,2043	2,1601	2,1141	2,0658
2,4935	2,4247	2,3522	2,2756	2,2354	2,1938	2,1507	2,1058	2,0589	2,0096
2,4499	2,3807	2,3077	2,2304	2,1898	2,1477	2,1040	2,0584	2,0107	1,9604
2,4117	2,3421	2,2686	2,1906	2,1497	2,1071	2,0629	2,0166	1,9681	1,9168
2,3779	2,3080	2,2341	2,1555	2,1141	2,0712	2,0264	1,9796	1,9302	1,8780
2,3479	2,2776	2,2033	2,1242	2,0825	2,0391	1,9938	1,9464	1,8963	1,8432
2,3210	2,2504	2,1757	2,0960	2,0540	2,0102	1,9645	1,9165	1,8657	1,8117
2,2967	2,2258	2,1508	2,0707	2,0283	1,9842	1,9380	1,8895	1,8380	1,7831
2,2747	2,2036	2,1882	2,0476	2,0050	1,9605	1,9139	1,8649	1,8128	1,7570
2,2547	2,1834	2,1077	2,0267	1,9838	1,9390	1,8920	1,8424	1,7897	1,7331
2,2365	2,1649	2,0889	2,0075	1,9643	1,9192	1,8718	1,8217	1,7684	1,7110
2,2197	2,1479	2,0716	1,9898	1,9464	1,9010	1,8533	1,8027	1,7488	1,6906
2,2043	2,1323	2,0558	1,9736	1,9299	1,8842	1,8361	1,7851	1,7307	1,6717
2,1900	2,1179	2,0411	1,9586	1,9147	1,8687	1,8203	1,7689	1,7138	1,6541
2,1768	2,1045	2,0275	1,9446	1,9005	1,8543	1,8055	1,7537	1,6981	1,6377
2,1646	2,0921	2,0148	1,9317	1,8874	1,8409	1,7918	1,7396	1,6835	1,6223
2,0772	2,0035	1,9245	1,8389	1,7929	1,7444	1,6928	1,6373	1,5766	1,5089
1,9926	1,9174	1,8364	1,7480	1,7001	1,6491	1,5943	1,5343	1,4673	1,3893
1,9105	1,8337	1,7505	1,6587	1,6084	1,5543	1,4952	1,4290	1,3519	1,2539
1,8307	1,7522	1,6664	1,5705	1,5173	1,4591	1,3940	1,3180	1,2214	1,0000

8. taula

Fisher-Snedecor-en F banaketa

$P(F \geq F_{\alpha; n_1, n_2}) = 0{,}01$

$\alpha = 0{,}01$

n_2 \ n_1	1	2	3	4	5	6	7	8	9
1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	5981,6	6022,5
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158
6	13,745	10,925	9,7795	9,1483	8,7459	8,4661	8,2600	8,1016	7,9761
7	12,246	9,5466	8,4513	7,8467	7,4604	7,1914	6,9928	6,8401	6,7188
8	11,259	8,6491	7,5910	7,0060	6,6318	6,3707	6,1776	6,0289	5,9106
9	10,561	8,0215	6,9919	6,4221	6,0569	5,8018	5,6129	5,4071	5,3511
10	10,044	7,5194	6,5523	5,9943	5,6363	5,3858	5,2001	5,0567	4,9424
11	9,6400	7,2057	6,2167	5,6683	5,3160	5,0692	4,8861	4,7445	4,6315
12	9,3302	6,9266	5,9526	5,4119	5,0643	4,8206	4,6395	4,4994	4,3875
13	5,0708	6,7010	5,7394	5,2053	4,8616	4,6204	4,4410	4,3021	4,1911
14	8,8616	6,5149	5,5639	5,0354	4,6950	4,4558	4,2779	4,1399	4,0297
15	8,6831	6,3589	5,4170	4,8932	4,5556	4,3183	4,1415	4,0045	3,8948
16	8,5310	6,2262	5,2922	4,7726	4,4374	4,2016	4,2059	3,8896	3,7804
17	8,3997	6,1121	5,1850	4,6690	4,3359	4,1015	3,9267	3,7910	3,6822
18	8,2854	6,0129	5,0919	4,5790	4,2479	4,0146	3,8406	3,7054	3,5971
19	8,1850	5,9259	5,0103	4,5003	4,1704	3,9386	3,7653	3,6305	3,5225
20	8,0960	5,8489	5,9382	4,4307	4,1027	3,8714	3,6987	3,5644	3,4567
21	8,0166	5,7804	4,8740	4,3688	4,0421	3,8117	3,6396	3,5056	3,3981
22	7,9454	5,7190	4,8166	4,3134	3,9880	3,7583	3,5867	3,4530	3,3458
23	7,8811	5,6637	4,7649	4,2635	3,9392	3,7102	3,5390	3,4057	3,2986
24	7,8229	5,6136	4,7181	4,2184	3,8951	3,6667	3,4959	3,3679	3,2560
25	7,7698	5,5680	4,6755	4,1774	3,8550	3,6272	3,4568	3,3239	3,2172
26	7,7213	5,5263	4,6166	4,1400	3,8183	3,5911	3,4210	3,2884	3,1818
27	7,6767	3,4881	4,0009	4,1056	3,7848	3,5580	3,3882	3,2558	3,1494
28	7,6356	5,4529	4,5681	4,0740	3,7539	3,5276	3,3581	3,2259	3,1195
29	7,5976	5,4205	4,5378	4,0449	3,7254	3,4995	3,3302	3,1982	3,0920
30	7,5625	5,3904	4,5097	4,0179	3,6990	3,4735	3,3045	3,1726	3,0665
40	7,3141	5,1785	4,3126	3,8283	3,5138	3,2910	3,1238	2,9930	2,8876
60	7,0771	4,9774	4,1259	3,6491	3,3389	3,1187	2,9530	2,8233	2,7185
120	6,8510	4,7865	3,9493	3,4796	3,1735	2,9559	2,7918	2,6629	2,5586
∞	6,6349	4,6052	3,7816	3,3192	3,0173	2,8020	2,6393	2,5113	2,4073

8. taula (jar.)

Fisher-Snedecor-en F banaketa

$\alpha = 0{,}01$

10	12	15	20	24	30	40	60	120	∞
6055,8	6106,3	6157,3	6708,7	6234,6	6260,7	6286,8	6313,0	6339,4	6366,0
99,399	99,416	99,432	99,449	99,458	99,466	99,474	99,483	99,491	99,501
27,229	27,052	26,872	26,690	26,598	26,505	26,411	26,316	26,221	26,125
14,546	14,374	14,198	14,020	13,929	13,838	13,745	13,652	15,558	13,463
10,051	9,8883	9,7222	9,5527	9,4665	9,3793	9,2912	9,2020	9,1118	9,0204
7,8741	7,7183	7,5590	7,3958	7,3127	7,2285	7,1432	7,0568	6,9690	6,8801
6,6201	6,4691	6,3143	6,1554	6,0743	5,9921	5,9084	5,8236	5,7372	5,6495
5,8143	5,6668	5,5151	5,3591	5,2793	5,1980	5,1156	5,0316	4,9460	4,8588
5,2565	5,1114	4,9621	4,8080	4,7290	4,6486	4,5667	4,4831	4,3978	4,3105
4,0492	4,7059	4,5582	4,4054	4,3269	4,2469	4,1653	4,0819	3,9965	3,9090
4,5393	4,3974	4,2509	4,0990	4,0209	3,9411	3,8596	3,7761	3,6904	3,6025
4,2961	4,1553	4,0096	3,8584	3,7805	3,7008	3,6192	3,5355	5,4494	3,3608
4,1003	3,9603	3,8154	3,6646	3,5868	3,5070	3,4253	3,3413	3,8548	3,1654
3,9394	3,8001	3,6557	3,5052	3,4274	3,3476	3,2656	3,1813	3,0942	3,0040
3,8049	3,6662	3,5222	3,3719	3,2940	3,2141	3,1319	3,0471	2,9595	2,8684
3,6909	3,5527	3,4089	3,2588	3,1808	3,1007	3,0182	2,9330	2,8447	2,7528
3,5931	3,4552	3,3117	3,1615	3,0835	3,0032	2,9205	2,8348	2,7459	2,6530
3,5082	3,3706	3,2273	3,0771	2,9990	2,9185	2,8354	2,7493	2,6597	2,5660
3,4338	3,2965	3,1533	3,0031	2,9249	2,8442	2,7608	2,6742	2,5839	2,4893
3,3682	3,2311	3,0880	2,9377	2,8594	2,7785	2,6947	2,6077	2,5168	2,4212
3,3098	3,1729	3,0299	2,8796	2,8011	2,7200	2,6359	2,5484	2,4568	2,3603
3,2576	3,1209	2,9780	2,8274	2,7488	2,6675	2,5831	2,4951	2,4029	2,3055
3,2106	3,0740	2,9311	2,7805	2,7017	2,6202	2,5355	2,4471	2,3542	2,2559
3,1681	3,0316	2,8887	2,7380	2,6591	2,5773	2,4923	2,4035	2,3099	2,2107
3,1294	2,9931	2,8502	2,6993	2,6203	2,5383	2,4530	2,3637	2,2695	2,1694
3,0941	2,9579	2,8150	2,6640	2,5848	2,5026	2,4170	2,3273	2,2325	2,1315
3,0618	2,9256	2,7827	2,6316	2,5522	2,4699	2,3840	2,2938	2,1984	2,0965
3,0320	2,8959	2,7530	2,6017	2,5223	2,4397	2,3535	2,2629	2,1670	2,0642
3,0045	2,8685	2,7256	2,5742	2,4946	2,4118	2,3253	2,2344	2,1378	2,0342
2,9791	2,8431	2,7002	2,5487	2,4689	2,3860	2,2992	2,2079	2,1107	2,0062
2,8005	2,6648	2,5216	2,3689	2,2880	2,2034	2,1162	2,0194	1,9172	1,8047
2,6318	2,4961	2,3523	2,1978	2,1154	2,0285	1,9360	1,8363	1,7263	1,6006
2,4721	2,3363	2,1915	2,0346	1,9500	1,8600	1,7628	1,0557	1,5530	1,3805
2,3209	2,1848	2,0385	1,8783	1,7908	1,6964	1,5923	1,4730	1,3246	1,0000

H I Z T E G I A

A

adierazgarri = significativo/a
adierazgarritasun-maila = nivel de significación
adierazpen = expresión
adierazpide grafikoa = representación gráfica
ahalmen = potencia
ahaltsu = potente
ahurtasun = concavidad
alboragabe = insesgado/a
alborapen = sesgo
alboratu = sesgado/a
aldagai = variable
aldakor = variable(adj.)
aldakuntza = variación
aldebateko = unilateral
aldebiko = bilateral
aldeko kasu = caso favorable
ale = unidad, individuo
arrakasta = éxito
arrazoi = razón
arrisku = riesgo
askatasun-graduak = grados de libertad
aske = independiente
ateraldi = extracción
azterketa estatistiko = análisis estadístico

B

bakun = simple
baldin eta soilik baldin = si y sólo si
baldintzazko probabilitate = probabilidad condicional
baliokide = equivalente
banaketa = distribución
banaketa-funtzio = función de distribución
bariantz analisi = análisis de varianza
bateraezin = incompatible, excluyente
bateragarri = compatible

batezbesteko = media
batezbesteko astatu = media ponderada
bazter- = marginal
behaketa = observación
beherakor = decreciente
bidezko joko = juego equitativo
bildura = unión

D

dentsitate-funtzio = función de densidad
doikuntza = ajuste
doikuntz egokitasun = bondad de ajuste
doitu = ajustar

E

ebakidura = intersección
egiantz = verosimilitud
egiantz handieneko estimatzaile = estimador de másima verosimilitud
egokitu = adecuar
egonkortasun = estabilidad
elkar askeak = mutuamente independientes
emaitza = resultado
emendatu = aumentar
emendio = aumento
eragiketa = operación
eragile = operador
eraginkortasun = eficiencia
eredu = modelo
eremu = dominio
erizpide = criterio
erregresio-koefiziente = coeficiente de regresión
erregresio-kurba = curva de regresión
erregresio-lerro = línea de regresión
erregresio-zuzen = recta de regresión
errenkada = fila
errepikatu gabeko aldakuntzak = variaciones sin repetición
errepikatuzko aldakuntzak = variaciones con repetición

errore = error
eskualde = región
estimatzaile = estimador
estimazio = estimación
ezinezko = imposible

G

gehikuntza = incremento
gehitu = incrementar
gertaera = suceso
gertaera-sistema oso = sistema completo de sucesos
gorakor = creciente

H

halabehartu = implicar
hein = rango
hipotesia baztertu = rechazar una hipótesis
hipotesia onartu = aceptar una hipótesis
hipotesi-kontraste = contraste de hipótesis
hoberen = óptimo
homogenotasun-probak = pruebas de homogeneidad
hondar-bariantza = varianza residual
hurbilketa = aproximación
hurrenez hurren = respectivamente

I

ibilbide = trayectoria
idazkera = notación
irudikapen = representación gráfica
itxaropen = esperanza
itxarotako balio = valor esperado
itxarotako maiztasun = frecuencia esperada
itzulera = reemplazamiento
izaera = carácter

J

jarrai = continuo/a
jazo = suceder
joko zuzen = juego equitativo
joera = tendencia

K

karratu latindar = cuadrado latino
karratu txikien metodo = método de los mínimos cuadrados
klasearteko bariantza = varianza interclases
klasebarneko bariantza = varianza intraclases
komunztadura = concordancia
konfidantza-maila = nivel de confianza
konfidantza-tarte = intervalo de confianza
kontingentzi taulak = tablas de contingencia
kontrako gertaera = suceso contrario

L

lagin = muestra
lagin-banaketa = distribución muestral
lagin-espazio = espacio muestral
laginketa = muestreo

M

maiztasun = frecuencia
menpeko = dependiente
menpekotasun = dependencia
metatu = acumulado/a
metakor = acumulativo/a
mendel = franja

N

naharo = frecuente
nulu = nulo

O

oinarrizko gertaera = suceso elemental
onarpen = aceptación
ondorioztatu = deducir
oreka = equilibrio
osagai = componente
osagarri = complementario

P

porrot = fracaso
probabilitate oso = probabilidad total
puntu-estimazio = estimación puntual
puntu-hodei = nube de puntos

S

sakabanapen = dispersión
sailkapen = clasificación

T

testik uniformeki ahaltsuen = test uniformemente más potente
txikien = mínimo(adj.)

Z

zehatz = exacto
zehaztasun = exactitud
zenbakigarri = numerable
ziurgabetasun = incertidumbre
zori = azar
zorizko = aleatorio/a
zutabe = columna
zuzenketa = corrección