

ESTADÍSTIKA ETA DATUEN ANALISIA

Azterketa ebatziak

2024-2025 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea. EHU

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

gizapedia.org

ESTADÍSTIKA ETA DATUEN ANALISIA (PRAKTIKA)

Irakaslea: Josemari Sarasola

Data: 2025eko urtarrilaren 14a, 15:00

Iraupena: 130 min

I. ebazkizuna (3 puntu)

SALARICH emakumeentzako barruko arropa ekologikoa ekoizten duen enpresa katalana da. Lo egiteko azpiko galtza anatomikoen kolekzio bat prestatzen ari da, baina aurretik merkatu ikerketa bat garatu du emakumeek azpiko galtzarik edo azpiko galtzekin lo egiten duten jakiteko. 95 emakumeri egin zaio inkesta, ohikoan lotarako azpiko galtzak janzten dituzten eta adina galdetuz (gaztea, 15-30 urte bitartean; heldua, 30-50 urte bitartean; eta pertsona nagusia 50 urtetik gora), eta datuekin kontingentzia-taula hau eratu du:

Azpiko galtzak? (↓)/ Adina (→)	Gaztea	Heldua	Nagusia	Guztira
Bai	8	18	24	50
Ez	22	17	6	45
Guztira	30	35	30	95

Aurreko kontingentzia-taula eratu ondoren, azken orduko inkestatzaile batek bildutako datuak jaso dira. Datu gehigarri horiek honako hauek dira, emakumez emakume puntu eta komaz bereizita (adin zehatza zenbakiz jaso da; b, azpiko galtzak bai; e, azpiko galtzak ez): 18b;20b;25b;22e;44e;36e;39b;57e;62e;68e.

Egin beharreko atazak:

- Datu gehigarri horiekin kontingentzia-taula berria eratu, kontingentzia-koefizientea kalkulatu eta modu egokian interpretatu, behar diren gelaxkak koloreztatuz eta horien esanahia azalduz baita ere. Emaizetan oinarrituta, azpiko galtzen kolekzioa zein adin-tarteko emakumeei begira diseinatu beharko litzateke bereziki? Zergatik?
- Aurreko atalean kalkulatu dituzun khi-karratu kontribuzioak eta horien erroak adierazi orri hauetan erantsita datorren khi-karratu nomograman.
- Pertsona helduak eta nagusiak kategoriatan berean bildu eta 2x2 kontingentzia-taula berria eratu. Bi aldagaien Pearson-en korrelazio-koefiziente lineala (dikotomiko-dikotomiko) kalkulatu eta eskematikoki interpretatu modu egokian, kodifikazio estandarra ezarriz.
- Etha edo korrelazio-ratioa kalkulatu eta interpretatu, datu gehigarriak baino ez harturik, aldagai independente gisa azpiko galtzak janzten diren hartuz.
- Etha edo korrelazio-ratioa kalkulatu eta interpretatu, datu guzti-guztiak harturik eta pertsona nagusien adin-muga egindako inkestan 70 urtekoa dela jakinik.

(a)

Kontingentzia-taula berria datu gehigarriekin:

Azpiko galtzak? (↓)/ Adina (→)	Gaztea	Heldua	Nagusia	Guztira
Bai	11	19	24	54
Ez	23	19	9	51
Guztira	34	38	33	105

Kontingentzia-aula horretatik, khi-karratua estatistikorako kalkuluak garatzen ditugu, khi-karratu kontribuzioen erroak ere kalkulatu, gero asoziazioaren norabidea aztertzeko:

$\frac{\text{enpiriko (O)}}{(O - E)^2}$	$\frac{\text{teoriko (E)}}{\sqrt{E}}$	Gazte		Heldu		Nagusia	
Azpiko galtzak bai		11 2.40	17.48 -1.55	19 0.015	19.54 -0.122	24 2.91	16.97 1.70
Azpiko galtzak ez		23 2.54	16.51 1.59	19 0.016	18.45 0.126	9 3.082	16.02 -1.75

Horrela khi-karratu estatistikoa hau izango da:

$$\mathbf{X}^2 = 2.40 + 0.015 + 2.91 + 2.54 + 0.016 + 3.082 = 10.97$$

Kontingentzia-koefizientea hau da:

$$C = \sqrt{\frac{10.97}{10.97 + 105}} = 0.3075$$

Normalizatzeko, koefiziente maximoa kalkulatu dugu:

$$C_{max} = \left(\frac{2}{3} \cdot \frac{3}{4} \right)^{\frac{1}{4}} = 0.84$$

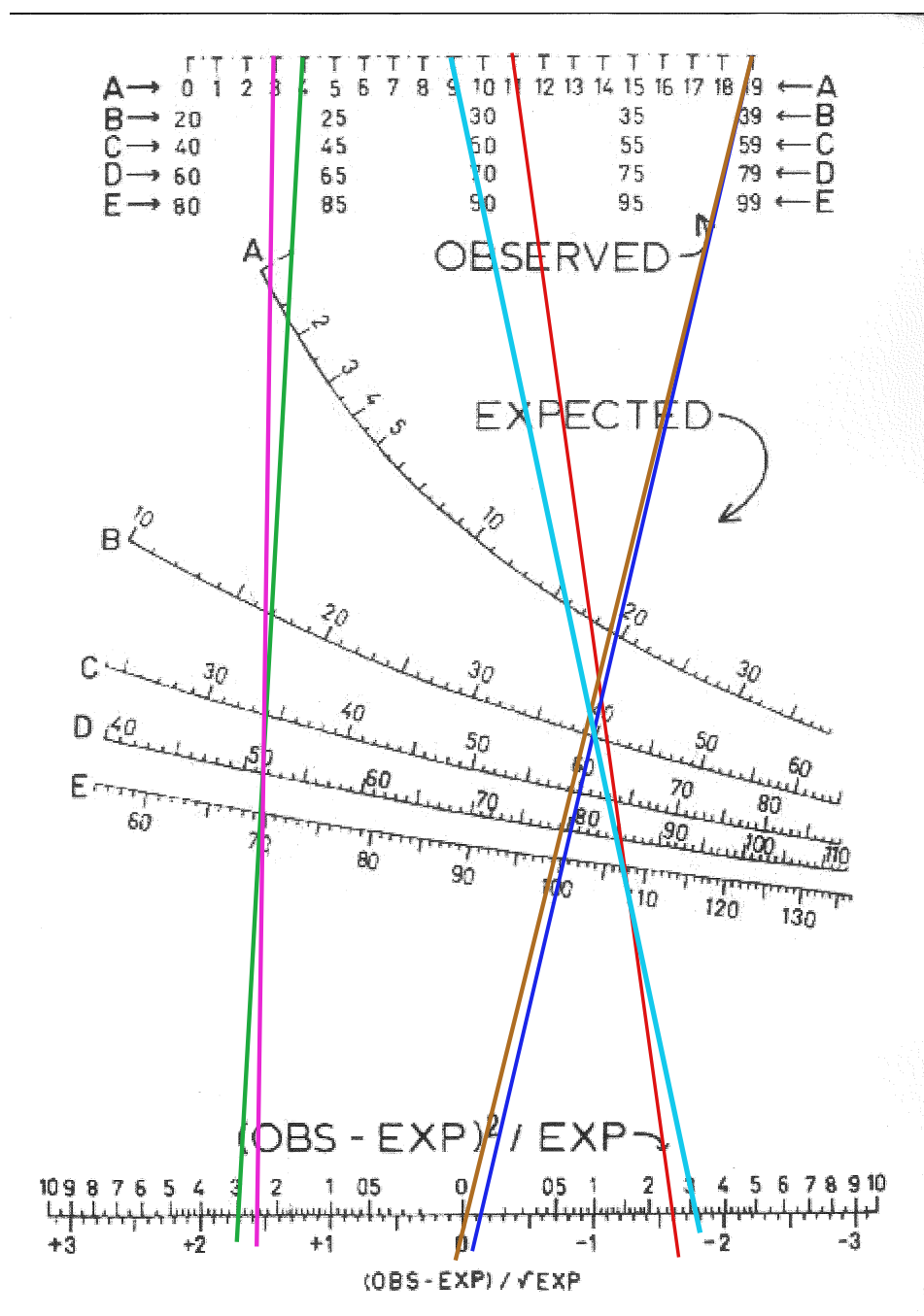
Normalizatuz:

$$C/C_{max} = 0.3075/0.84 = 0.366$$

Horrela, bi aldagaien arteko **asoziazioa ertaina** da, lagin errorearen eta antzeko ikerketen erreserbapean.

Horrela, azpiko galtzen kolezioa 50 urtetik gorako emakumeentzat diseinatu beharko litzateke, asoziazio positibo nabarmenena emakume nagusien eta lotarako azpiko galtzak janzten dituztenen artekoa delako.

(b)



Khi-karratu kontribuzioak eta horien zeinuzko erroak zehazteko nomograma

red 11	blue 19	green!70 24
purple!70 23	brown!70 19	cyan!30 9

Taula 1: Lerro bakoitza zein gelaxkari dagokion

(c)

Hasierako kontingentzia-taulatik abiaturik, 2x2 kontingentzia taula berria eratzen da, adierazburuan adierazi bezala:

Azpiko galtzak? ($\downarrow$) / Adina ($\rightarrow$)	Gaztea	Ez gaztea	Guztira
Bai	11	43	54
Ez	23	28	51
Guztira	34	71	105

Pearson-en korrelazio linealeko koefizientea kalkulatzeko, kodifikazio estandarretik abiatzen gara (azpiko galtzak bai: 1, azpiko galtzak ez: 0; gaztea bai: 0 -adin txikia-, gaztea ez: 1 -adin handia-), eta prozesuan zehar balio bikote bakoitza (00, 01, 10 et 11) hainbat aldiz errepikatzen dela aintzakotzat harturik:

x (azpikogaltzak)	y (adina)	n	nx	ny	nxy	nx^2	ny^2
1	0	11	11	0	0	11	0
1	1	43	43	43	43	43	43
0	0	23	0	0	0	0	0
0	1	28	0	28	0	0	28
		105	54	71	43	54	71

$$\bullet \bar{x} = \frac{54}{105} = 0.51 \quad ; \quad \bar{y} = \frac{71}{105} = 0.67$$

$$\bullet s_{xy} = \frac{43}{105} - 0.51 \times 0.67 = 0.067$$

$$\bullet s_x = \sqrt{\frac{54}{105} - 0.51^2} = 0.5$$

$$\bullet s_y = \sqrt{\frac{71}{105} - 0.67^2} = 0.46$$

$$\bullet r_{xy} = \frac{0.067}{0.5 \times 0.46} = 0.2949$$

Beraz, lotarako azpiko galtzak janztearen eta adinaren edo adin-taldearen arteko korrelazioa ahula-ertaina da (≈ 0.3), betiere lagin errorearen eta antzeko ikerketen erreserbapean. Korrelazioa norabidea eskematikoki interpretatuz:

$x \uparrow: 0 \rightarrow 1$ $y \uparrow: 0 \rightarrow 1$

Beraz, ez janztetik janztera pasatzen garenean, adina igo egiten da. Horrek esan nahi du, emakume heldu gehiagok egiten dutela lo azpiko galtzak jantzita gazteek baino, eta esan bezala, hori korrelazio sendoarekin. Emaitzak bat datoz, espero zitekeen bezala, hasierako taulako asoziazioa aztertuta atera ditugun konklusioekin, oraingo honetan korrelazioa ahula-ertaina atera zaigula matizaturik.

(d)

Etha edo korrelazio-ratioa kalkulatzeko, datu gehigarriak aldagai independentearen arabera bereiziko dira, kasu honetan, adierazburuan esaten den bezala, azpiko galtzak janzten diren ala ez erreparatuz. Ondoren, guztien batezbestekoa ere eman behar da.

Azpiko galtzak bai: 18-20-25-39 $\rightarrow \bar{x}_{bai} = 25.5$ Azpiko galtzak ez: 22-44-36-57-62-68 $\rightarrow \bar{x}_{ez} = 48.16$

Emakume gehigarri horien guztien batezbestekoa, arestiko batezbestekoak haztatuz:

$$\bar{x} = \frac{25.5 \times 4 + 48.16 \times 6}{10} = 39.09$$

Orain, etha kalkulatzeko moduan gaude:

$$\eta^2 = \frac{4 \times (25.5 - 39.09)^2 + 6 \times (48.16 - 39.09)^2}{(18 - 39.09)^2 + (20 - 39.09)^2 + \dots + (68 - 39.09)^2} = \frac{1232.3}{3014.9} = 0.41$$

Balio horrekin, bi aldagaien arteko korrelazioa ertaina dela esan daiteke.

(e)

Taula osoko datuak baditugu, baiezko nahiz ezezkoetarako, adina tartetan dugu; beraz, talde horietako batezbestekoak kalkulatzeko klase-marka hartu behar da:

x	n(bai)	nx
[15, 30] $\rightarrow$ 22.5	11	247.5
[30, 50] $\rightarrow$ 40	19	760
[50, 70] $\rightarrow$ 60	24	1440
	54	2447.5

- Azpiko galtzak janzten dituztenen adinen batezbestekoa: $\bar{x}_{bai} = \frac{2447.5}{54} = 45.32$

x	n(ez)	nx
[15, 30] $\rightarrow$ 22.5	23	517.5
[30, 50] $\rightarrow$ 40	19	760
[50, 70] $\rightarrow$ 60	9	540
	51	1817.5

- Azpiko galtzak janzten ez dituztenen adinen batezbestekoa: $\bar{x}_{ez} = \frac{1817.5}{51} = 35.63$
- Emakume guztien batezbesteko adina, batezbesteko haztatua erabiliz:

$$\bar{x} = \frac{45.32 \times 54 + 35.63 \times 51}{105} = 40.6$$

Eta azkenik etha edo korrelazio ratioa eman dezagun, jakinda emakume guztien adin banaketa, azpiko galtzak janzten edo ez kontuan hartu gabe, ondokoa dela:

x	n
[15, 30] $\rightarrow$ 22.5	34
[30, 50] $\rightarrow$ 40	38
[50, 70] $\rightarrow$ 60	33

$$\eta^2 = \frac{54 \times (45.32 - 40.6)^2 + 51 \times (35.63 - 40.6)^2}{34 \times (22.5 - 40.6)^2 + 38 \times (40 - 40.6)^2 + 33 \times (60 - 39.09)^2} = 0.096$$

Emaitza honek, aitzitik, aurreko atalekoak ez bezala, korrelazio ahula adierazten du bi aldagaien artean. Azken hau esanguratsuagoa da noski, datu guztiak barne hartzen dituelako.

II. ebazkizuna (2.5 puntu)

EAZ graduko "Estatistika eta Datuen Analisia" irakasgaiko azterketa KUID akademian prestatu duten azken urteotako ikasle guztien notak batera jarri dira, irakasleak azken urteotan ebaluazio-sistema homogeneoa erabili duela aintzat harturik. Nota horiekin klase-tartekako maiztasun-taula hau eratu da:

Nota	Ikasleak
0-2	8
2-4	12
4-6	62
6-8	34
8-10	4
Guztira	120

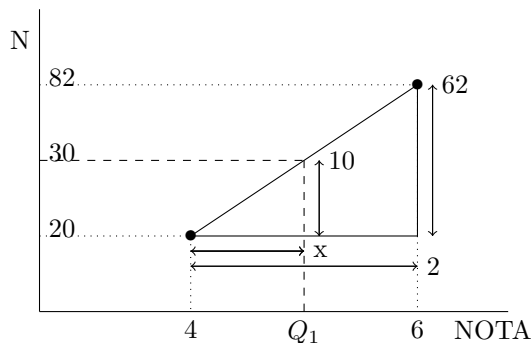
Egin beharreko atazak:

- (a) %25ean (alde bietatik) winsorizatutako batezbestekoa kalkulatu.
- (b) Jakina denez, winsorizatutako batezbestekoa bereziki erabilgarria da datu atipikoak daudenean. Datu atipikoak antzemateko kaxa-diagramaren irizpidea baliatuz (baina kaxa-diagrama eratu gabe), erabaki ezazu datu-banaketa horretarako winsorizatutako batezbestekoaren erabilera justifikatuta dagoen (hau da, datu atipikoak dauden). Zein aldetatik leudeke datu atipikoak? Zure erantzuna guztiz ziurra al da? Zergatik?
- (c) Lehiakidea den PEKOS akademiako ikasleen artean nota berdinak jaso dira, eta batezbesteko aritmetikoa sinplea 6.5 eta populazio-bariantza 2.25 suertatu dira, hurrenez hurren, 100 ikasleren datuak jasota. Kalkulatu eta interpretatu KUID eta PEKOS ikasleen estatistikako noten batezbestekoen diferentziaren efektuaren tamaina, Cohen-en d neurriaren bitartez, eta horretarako aurretik KUID akademiako ikasleen noten batezbestekoa eta bariantza kalkulatu, datu guztiekin noski.

(a)

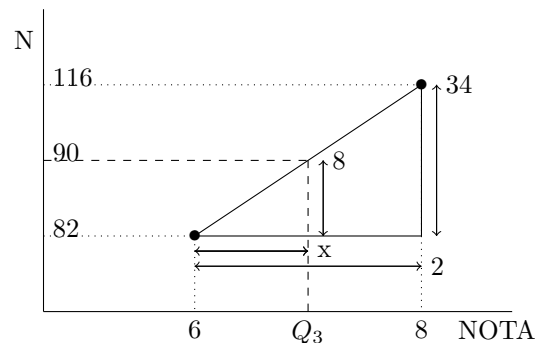
%25ean winsorizatutako batezbestekoa kalkulatzeko, $Q_1 = P_{25}$ eta $Q_3 = P_{75}$ kalkulatu behar dira. Datuak tartetan biduta daudenez, interpolazio linealez egin behar da kalkulua:

$$Q_1? 0.25 \times 120 = 30gn.datua \in [4, 6]$$



$$\frac{x}{10} = \frac{2}{62} \rightarrow x = \frac{20}{62} = 0.32 \rightarrow Q_1 = 4.32$$

$$Q_3? 0.75 \times 120 = 90gn.datua \in [6, 8]$$



$$\frac{x}{8} = \frac{2}{34} \rightarrow x = \frac{16}{34} = 0.47 \rightarrow Q_3 = 6.47$$

Orain, lehen eta azken 30 datuak kuartil horien balioekin berdintzen ditugu, eta bi kuartilen artean dauden erdiko datu horiek bere horretan utziz. Datuen banaketa winsorizatu berri horretatik batezbestekoa kalkulatzeko, klase-markak hartzen ditugu:

x	n	nX
4.32	30	129.6
$[4.32 - 6] \rightarrow 5.16$	$62-10=52$	268.32
$[6 - 6.47] \rightarrow 6.23$	$34-26=8$	49.84
6.47	30	194.1
Guztira	120	641.86

$$\bar{x}_{winsor} = \frac{641.86}{120} = 5.34$$

(b)

Kaxa-diagrama eratzen denean, datu atipikoak Q_1 eta Q_3 -tik alde banatara $1.5(Q_3 - Q_1)$ -eko distantziara baino haratago dauden datuak dira. Ikus dezagun gure kasuan:

$$Q_1 - 1.5(Q_3 - Q_1) = 1.15$$

$$Q_3 + 1.5(Q_3 - Q_1) = 9.6$$

Beraz, 1.15 baino txikiagoak eta 9.6 baino handiagoak diren notak datu atipikoak lirateke kaxa-diagraman.

[0,2] tartean 8 datu daude eta [8-10] tartean 4; tarte horietako datuak izan litezke 1.15 baino txikiagoak eta 9.6 baino handiagoak; horrenbestez, winsorizatutako batezbestekoaren erabilera justifikatuta egon liteke kasu horretan, baina ezin dugu guztiz ziurtatu, tarte horietako datuak datu atipikoen eremutik kanpo ere izan litezkeelako, ez baitakigu datuak tartearen barnean nola banatzen diren.

(c)

Cohen-en d kalkulatzeko KUID akademiako noten batezbesteko arimetiko sinplea eta bariantza falta zaizkigu. Datuak tartetan emanda daudenez, klase-markak hartzen ditugu:

x	n	nx	nx^2
1	8	8	8
3	12	36	108
5	62	310	1550
7	34	238	1666
9	4	36	324
Guztira	120	628	3656

$$\bar{x}_{winsor} = \frac{628}{120} = 5.23$$

$$s_x^2 = \frac{3656}{120} - 5.23^2 = 3.11$$

Goazen orain Cohen-en d kalkulatzera. Horretarako, lehenbizi *pooled deviation* edo desbideratze bateratua kalkulatzeko dugu:

$$s = \sqrt{\frac{120 \times 3.11 + 100 \times 2.25}{120 + 100 - 2}} = 1.65$$

Eta bukatzeko:

$$d = \frac{6.5 - 5.23}{1.65} = 0.77$$

Emaitza horrekin, akademiaren efektuaren tamaina, alegia batezbestekoen arteko diferentziaren adierazgarritasuna, handi samarra dela esan daiteke, PEKOS akademiako ikasleek nota handiagoak izaten dituztelako norabidean.

III. ebazkizuna: Egunkari saltzailearen problema (2.5 puntu)

ARGIEDER Gipuzkoako denda-kate handientzako erakusleihoak diseinatu eta apailatu egiten dituen enpresa txiki bat da. 2024ko Eguberrietako erakusleihoak prestatzeko, BASURKO basogintza-enpresari aldeztu aurretik zenbat pinu handi (zehatzago, 4m-6m bitartekoak) enkargatu behar dizkion erabaki behar du. Pinu handi bakoitzaren prezioa 2.000 eurokoa izango da, BASURKO enpresak berak horrela adierazita. Dendek egingo dizkioten pinu handien eskaria honela banatzen dela uste da, gutxienez 4 pinu ziurtatuta dituelarik:

Eskatuko diren pinuak	Probabilitatea
4	0.3
5	0.4
6	0.2
7	0.1
1	

ARGIEDER enpresak 3.000 eurotan fakturatuko du pinu bakoitza. Saltzen ez dituen pinuak hondakindegira eraman beharko ditu han botatzeko eta horregatik 500 euroko tasa kobratzen dio pinu bakoitzeko hondakinak kudeatzeko mankomunitateak.

Egin beharreko atazak:

- Beharrezko kalkuluak eginez, eman ezazu batez besteko mozkina maximotzeko ARGIEDER enpresak erosi beharko duen pinuen kopurua.
- 5 zuhaitz erosita**, kalkulatu gainera batez beste zenbateko salmenta galtzen den, batez beste zenbat saltzen den, saldu gabe eta botatzeko zenbat pinu geratzen diren batez beste, *fill rate* delakoa, *instock* probabilitatea eta *stockout* probabilitatea.
- Aurreko parametroen arteko hiru koadratzeak egiazta itzazu.
- Pinuen eskabidea egin baino lehenxeago, BASURKO enpresak eskaintza berezia egin dio ARGIEDER enpresari: 7 pinu erosiz gero, bakoitzagatik 1.500 euro kobratuko dio. Batez besteko mozkinari erreparatu, bere erabakia aldatu behar al du ARGIEDER enpresak? ARGIEDER enpresak bere erabakia aldatuko badu, zenbateko eskaintza egin behar dio BASURKO enpresak pinu bakoitzeko?

(a)

• 4 zuhaitz erosita

s	m	$p(m)$	$mp(m)$
4	$3000 \times 4 - 2000 \times 4 = 4000$	1	4000
		1	4000

• 5 zuhaitz erosita

s	m	$p(m)$	$mp(m)$
4	$3000 \times 4 - 2000 \times 5 = 1000 - 500 = 500$	0.3	150
5	$3000 \times 5 - 2000 \times 5 = 5000$	0.7	3500
		1	3950

• 6 zuhaitz erosita

s	m	$p(m)$	$mp(m)$
4	$3000 \times 4 - 2000 \times 6 = 1000 - 500 \times 2 = -1000$	0.3	-300
5	$3000 \times 5 - 2000 \times 6 = 5000 - 500 \times 1 = 2500$	0.4	1000
6	$3000 \times 6 - 2000 \times 6 = 6000$	0.3	1800
		1	2500

- 7 zuhaitz erosita

s	m	$p(m)$	$mp(m)$
4	$3000 \times 4 - 2000 \times 7 - 500 \times 3 = -3500$	0.3	-1050
5	$3000 \times 5 - 2000 \times 7 - 500 \times 2 = 0$	0.4	0
6	$3000 \times 6 - 2000 \times 7 - 500 \times 1 = 3500$	0.2	700
7	$3000 \times 7 - 2000 \times 7 = 7000$	0.1	700
		1	350

Honenbestez, batez beste mozkin handiena ematen duen eta hortaz erosi beharreko zuhaitz kopurua 4 da.

(b)

5 zuhaitz erosita

- Batez besteko salmenta galera

g	$p(g)$	$gp(g)$
0	0.7	0
1	0.2	0.2
2	0.1	0.2
	1	0.4

5 zuhaitz erosita, batez beste 0.4 zuhaitzen salmenta galtzen da.

- Batez besteko salmenta

s	$p(s)$	$sp(s)$
4	0.3	1.2
5	0.7	3.5
	1	4.7

5 zuhaitz erosita, batez beste 4.7 zuhaitz saltzen da.

- Batez beste bota beharreko zuhaitz kopurua

b	$p(b)$	$bp(b)$
0	0.4	0
1	0.3	0.3
	1	0.3

5 zuhaitz erosita, batez beste 0.3 zuhaitz bota behar da.

- Fill-rate

Fill-rate kalkulatzeko batez besteko eskaria kalkulatu behar da:

g	$p(g)$	$gp(g)$
4	0.3	1.2
5	0.4	2
6	0.2	1.2
7	0.1	0.7
	1	5.1

Fill-rate batez besteko salmenta zati batez besteko eskaria da. Beraz, $4.7/5.1=0.92=\%92$ koa da fill-rate, hots, eskaria %92an asebetetzen da 5 zuhaitz erosita.

- In-stock probabilitatea

5 zuhaitz erosita, eskatzen diren zuhaitz guztiak jarri ahal izango ditugu (hots, in-stock egongo gara), 4 edo 5 zuhaitzeko eskaria dagoenean. Beraz, in-stock probabilitatea $0.3+0.4=0.7$ da.

- Stock-out probabilitatea

5 zuhaitz erosita, eskatzen diren zuhaitzak ezingo dira jarri (hots, stock-out egongo gara), 6 edo 7 zuhaitzeko eskaria dagoenean. Beraz, stock-out probabilitatea $0.2+0.1=0.3$ da.

(c)

Koadratzeak:

- $p[\text{in-stock}] + p[\text{stock-out}] = 0.7 + 0.3 = 1$ ✓
- batez besteko salmenta + batez beste galtzen den salmenta = batez besteko eskaria $\rightarrow 4.7 + 0.4 = 5.1$ ✓
- batez besteko salmenta + batez beste bota = erositako kopurua $\rightarrow 4.7 + 0.3 = 5$ ✓

(d)

7 pinu erosita 1.500 eurotan:

- 7 zuhaitz erosita

s	m	$p(m)$	$mp(m)$
4	$3000 \times 4 - 1500 \times 7 - 500 \times 3 = 0$	0.3	0
5	$3000 \times 5 - 1500 \times 7 - 500 \times 2 = 3500$	0.4	1400
6	$3000 \times 6 - 1500 \times 7 - 500 \times 1 = 7000$	0.2	1400
7	$3000 \times 7 - 1500 \times 7 = 10500$	0.1	1050
		1	3850

7 zuhaitz erosita, 3.850 euro irabaziko lirateke batez beste; 4 zuhaitz erosita, 4.000 euro, gogora dezagun. Beraz, 4 zuhaitz erosi beharko litzuzke orduan ere.

Batez besteko mozkin berria kalkulatzeko, askoz ere errazagoa da kontuak honela egitea: 7 zuhaitz 2000 eurotan erosita, 350 irabazten da; 1500 eurotan erosita, $7 \times 500 = 3500$ gehiago irabazten da kasu guztietan. Beraz, batez besteko irabazi berria $350 + 3500 = 3850$ da; $Y = X + k \rightarrow E[Y] = E[X] + k$ jatorri-alaketaren propietatea erabiliz.

Erabakia aldatzeko, batez besteko irabazia 4000 igo beharko litzateke, 7 zuhaitzak erosita, hau da, $4000 - 3850 = 150$ euro gehiago gutxienez irabazi beharko lirateke 7 zuhaitzekin, eta bakoitzarekin $150/7 = 21.42$ gehiago. Beraz, deskontua $1500 - 21.42 = 1478.57$ izan beharko litzateke gutxienez.

Analitikoki ere egin daiteke:

s	m	$p(m)$	$mp(m)$
4	$3000 \times 4 - 7k - 500 \times 3 = 10500 - 7k$	0.3	3150-2.1k
5	$3000 \times 5 - 7k - 500 \times 2 = 14000 - 7k$	0.4	5600-2.8k
6	$3000 \times 6 - 7k - 500 \times 1 = 17500 - 7k$	0.2	3500-1.4k
7	$3000 \times 7 - 7k = 21000 - 7k$	0.1	2100-0.7k
		1	14350-7k=4000

$$14350 - 7k = 4000 \rightarrow k = 1478.57$$

Erantzuna berdina da, noski.

ESTADISTIKA ETA DATUEN ANALISIA (TESTA)

Irakaslea: Josemari Sarasola

Data: 2025eko urtarrilaren 14, 15:00

Iraupena: 40 minutu

Erantzun bakarra da zuzena galdera bakoitzean. Guztira testak 2 puntu balio du. Erantzun zuzenak 0.1 puntu balio du. Erantzun oker bakoitzak zuzenak ematen duenaren erdia kentzen du. Galderak erantzun gabe utz daitezke, punturik gehitu eta kendu gabe.

1. Zein da estatistikaren objektu formalak?
 - (a) Datu-multzoak.
 - (b) Fenomeno aldakorrak.
 - (c) Datu-multzoetan joerak antzematea.
 - (d) Kasu partikularrak auresatea.
2. Estatistikan banakako datuak edo kasu partikularrak adierazgarriak ez direla adierazten duen atsotitz egokiena zein da?
 - (a) Beltz guztiak ez dira ikatzak.
 - (b) Ahuntzak garia ikusi, baina ez hesia.
 - (c) Enara batek ez du udaberria egiten.
 - (d) Egin arren distira, denak urre ez dira.
3. Gipuzkoako familia ugarien batezbesteko errenta kalkulatu da, familia horien egoera orokorra ezagutzearen. Zein motako ikerketa da?
 - (a) Ikerketa nomotetikoa.
 - (b) Ikerketa idiografikoa.
 - (c) Ikerketa positibista.
 - (d) Ikerketa jeneralista.
4. $x_1 = 2$, $x_2 = 4$, $x_3 = 5$. Kalkulatu $\sum_{i=1}^3 (2x_i^2 + 5)$.
 - (a) 105
 - (b) 247
 - (c) 257
 - (d) 499
5. Denda bateko eguneko salmenta-datuak harturik, klase-tarte hauek eratu dira (tarte:maiztasuna): (0-100:60;100-500:80;500-1000:40) (hau da 60 egunetan 0tik 100 eurotara bitarteko salmentak daude). Zer dago gaizki klase-tartekako datu-maiztasun honetan:
 - (a) Tarte kopurua txikiegia da.
 - (b) Tarte kopurua txikiegia da eta tarte-zabalera desberdina ez dago justifikatuta.
 - (c) Tarte kopuru txikiegia, tarte-zabalera desberdina eta klase-tarteen mugako balioak.
 - (d) Tarte kopuru txikiegia, tarte-zabalera desberdina eta klase-tarteen mugako balioak eta tarteen solapamendua.

6. Andoni Maiza ekonomia-irakasleak hiru kontrol edo proba jarri ditu 15 asteko ikastaroan zehar. Lehenengo lau astetako gaiak jasotzen dituen horretan, ikasle batek 5 atera du; hurrengo sei astetako zatian, 6 eta azken bost astetako zatian, 3.5. Gai guztiek garrantzi berdina badute eta martxa edo abiadura beretsuan eman badira, zein izan beharko litzateke ikasle horren batezbesteko nota?
- 4.83
 - 4.9
 - Ezin da kalkulatu zehazki, baino argi dagoena da ez duela ikastaroa gainditzen (5 edo gehiago atera).
 - Ezin da kalkulatu zehazki modu egokian.
7. Zein neurri erabiliko zenuke ondoko hauetatik bi lantegi desberdinetako langileen eguneko ekoizpenen sakabanatzea alderatzeko, jakinda datu atipikoak badaudela?
- Desbideratze absolutuen mediana.
 - Aldakuntza-koefizientea
 - Kuartil arteko sakabanatze-koefizientea, hots, $(Q_3 - Q_1)/(Q_3 + Q_1)$
 - Lagin-bariantza.
8. Zein aldagaitarako ez da zilegi kontzentrazioa edo desparekotasuna aztertzea?
- Espainiako unibertsitateetako ikasle-kopuruak.
 - Laborantza-sailetako azalerak.
 - Zeruko beste gorputz batekin talkaz sorturiko asteroide-familia edo talde bateko asteroideen masak (gogoratu asteroideak Eguzkiaren inguruan orbitatzen duten "arroka" handiak direla eta jatorri berdina dutenean "familiak" osatzen dituztela).
 - Gipuzkoako enpresa txiki eta ertainetako likidezia-ratioak.
9. Enpresa batek lantegi bana ditu Euskal Herrian eta Txinan. Euskal Herriko eta Txinako langileen soldaten laginak jaso eta Gini-ren indizearen zatiketa eta deskonposizioa egin da, emaitza hauekin: $G(\text{artekoa})=0.4$; $G(\text{barnekoa})=0.1$; hon-darra oso txikia da, eta beraz baztergarria. Honako soldata-lagin hauetatik zein dagokio neurri egokienean aipatutako zatiketari? Oharra: lagin horien datu batzuk baino ez dira ematen, baina lagin osoaren adierazgarri dira.
- EH:100-110-120-105-...; Txina: 50-55-65-45-...
 - EH:100-180-200-300-...; Txina: 120-280-160-200-...
 - EH:100-110-120-105-...; Txina: 110-105-120-125-...
 - EH:100-180-200-300-...; Txina: 110-105-120-125-...
10. Familia batek ez du berokuntzarik etxean, ezin die ordaindu seme-alabei nahiko luketen hezkuntza, ez da azken bost urteotan joan oporretan inora, arazoak dituzte hilaren bukaerara heltzeko, ... Zein pobrezia-kontzeptutaz ari gara?
- Pobreziaren intentsitatea.
 - Pobrezia sintetikoa.
 - Pobrezia absolutua.
 - Pobrezia erlatiboa.
11. Zein tarteko balioak hartzen ditu khi-karratu estatistikoak?
- $[-1, 1]$
 - 0 edo positiboak
 - $[0, 1]$
 - Otik kontingentzia-taula bakoitzari dagokion balio maximora.

12. Bi aldagaien arteko kobariantza 0.87 suertatu da. Zein ondorio atera daiteke hortik?
- (a) Korrelazio lineala zuzena dela, besterik gabe.
 - (b) Korrelazioa zuzena eta sendoa dela, antzeko ikerketen erreserbapean.
 - (c) Korrelazio zuzena, sendoa eta lineala dela, antzeko ikerketen erreserbapean.
 - (d) Ezin da ondorioz atera, bi aldagaien desbideratzeak ere behar ditugu eta.
13. $x = 1, 2, 3, 4, 5, 6 \dots$ balioak hartzen dituen probabilitate-banaketa bati buruz hau dakigu: $F(x = 1) = 0.1; F(x = 2) = 0.2; F(x = 3) = 0.25; F(x = 4) = 0.5; F(x = 5) = 0.6; F(x = 6) = 0.9$. Kalkulatu $P[3 < X < 6]$.
- (a) Ez da zuzena emandako informazioa, banaketa funtzioaren gehikuntzak ez direlako gero eta handiagoak.
 - (b) 0.25
 - (c) 0.3
 - (d) 0.35
14. $F(x) = x^2$; $0 \leq x \leq 1$. Zenbat da $P[0.25 < X < 0.75]$?
- (a) 0.25
 - (b) 0.4325
 - (c) 0.5
 - (d) 0.5625
15. $f(x) = \frac{x}{3}$; $0 \leq x \leq a$. Zenbat da a (gutxigorabehera)?
- (a) $a = 1.35$
 - (b) $a = 1.65$
 - (c) $a = 2.45$
 - (d) $a = 3.33$
16. $F(x) = 2x$; $0 \leq x \leq 0.5$. Kalkulatu $VaR(0.1)$.
- (a) 0.01
 - (b) 0.05
 - (c) 0.1
 - (d) 0.2

17. $f(x) = \frac{2x}{k^2}$; $0 \leq x \leq k$. Kalkulatu itxaropen matematikoa, k parametroa zehaztu gabe.

- (a) $k/3$
- (b) $2k/3$
- (c) k
- (d) $4k/3$

18. X zorizko aldagai bat honela banatzen da:

x	p(x)
-1	1/3
0	1/3
1	1/3

Kalkulatu bariantza.

- (a) $1/3$
- (b) $2/3$
- (c) 0
- (d) 1

19. $\mu = 10$; $\sigma = 2$. Hurbildu $P[X < 13]$ Txebixev-en desberdintza erabiliz.

- (a) $4/9$ baino handiagoa edo berdina.
- (b) $5/9$ baino txikiagoa edo berdina.
- (c) $4/9$ baino txikiagoa edo berdina.
- (d) Aurreko guztiak faltsuak dira.

20. A eta B inbertsio biren errentagarritasunari buruzko informazioa da honako hau: $\mu_A = 10$, $\sigma_A = 2$; $\mu_B = 8$, $\sigma_B = 1$. Zein da egiazkoa? Oharra: Beharko balitz, utilitate-funtzio gisa har ezazu μ/σ .

- (a) Epe laburrera hobe da A eta epe luzera B.
- (b) Epe laburrera nahiz luzera hobe da A.
- (c) Epe laburrera hobe da B eta epe luzera A.
- (d) Epe laburrera nahiz luzera hobe da B.

Estatistika eta datuen analisia

2025eko urtarrilaren 14

Izena eta abizenak:_____

Galdera	Erantzuna
1	C
2	C
3	A
4	A
5	A
6	B
7	C
8	D
9	A
10	C
11	B
12	A
13	D
14	C
15	C
16	B
17	B
18	B
19	D
20	C

KOPURUA

ONGI	
GAIZKI	
ERANTZUN GABE	

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2025eko ekainaren 23a, 10:00

Iraupena: 130 min

I. ebazkizuna (3 puntu)

Lapurdiko lurralde-kontseiluak lurralde horretako familia-unitateen artean errentari buruzko inkesta zabal bat egin du. Datuen esplorazio azkarra egiteko, errenta-datuen azpilagin bat aukeratu da (errenta urteko milaka eurotan):

6 14 18 22 23 25 27 28 28 30

31 32 33 35 37 38 39 39 39 39

40 40 41 41 42 44 45 45 46 46

47 47 48 49 51 56 57 58 59 60

62 76 77 83 84 85 89 93 96 99 125

Errenta-datu hauetan oinarrituz, gizarte-klaseak honela banatu dira: behe-mailako klasea, 38ko errenta eta hortik behera; ertaineko klasea, 39tik 60ra arte (60 barne); goi-mailako klasea, 60ko errentatik gorakoak.

Egin beharreko atazak:

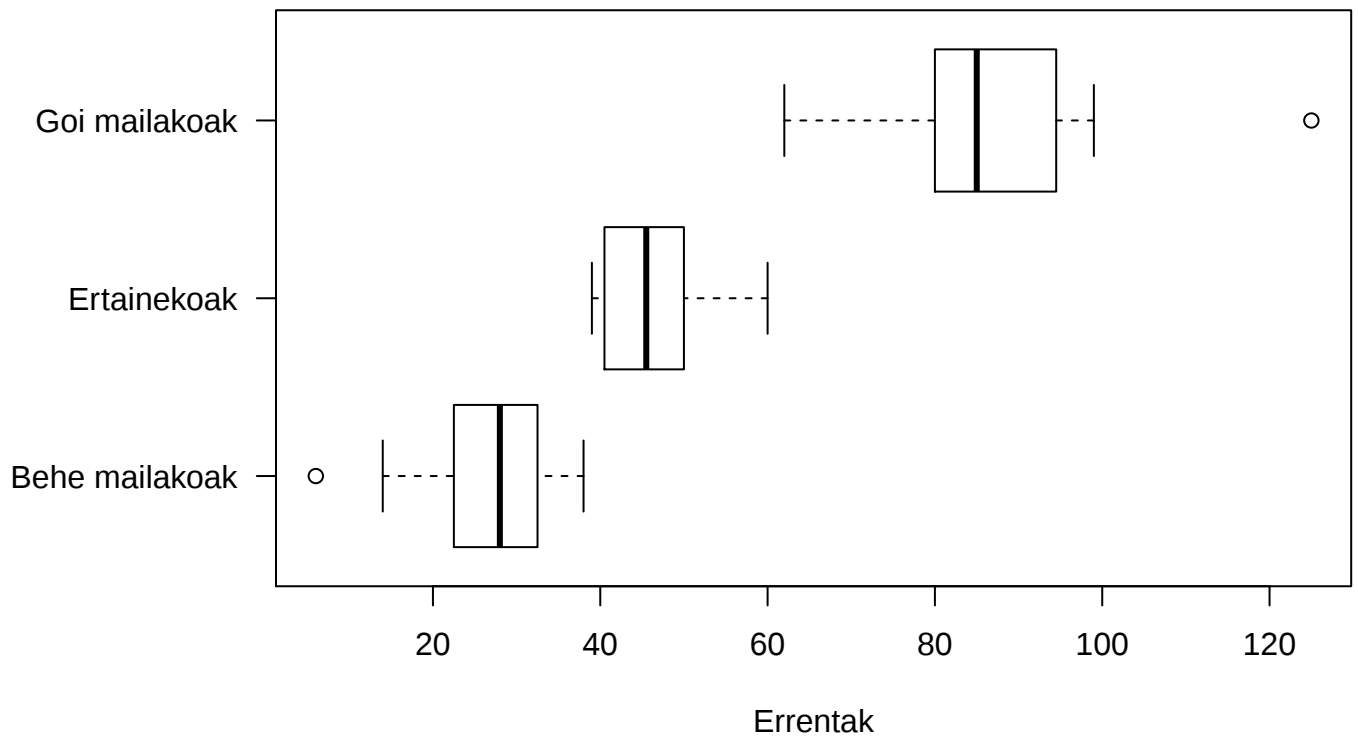
- Eratu ezazu hiru gizarte klase horietako bakoitzari dagokion kaxa-diagrama, eraketa-taula osoa garatuz. Hiru kaxa-diagramak grafikoki batera irudikatu itzazu.
- Aurrekoan oinarrituta, orokorrean non dago errenta-tarte handiagoa: behe-mailako klasetik ertaineko klasera ala ertaineko klasetik goi-mailako klasera?
- Kuartil arteko ibiltartetik abiatuta behar bezala garatuz, azter ezazu zein klasetan dauden errenta sakabanatuenak. Proposatu zaizun neurria egokia dela deritzozu egoera honetarako? Zergatik?
- Klase guztietako errenta guztiak bateratu berriz ere, ordenaturik, eta haietatik 51gna, 46gna, 41gna, 36gna, 31gna, 26gna, 21gna, 16gna, 11gna eta 6gna harturik, kalkula itzazu Giniren indizea (ez da beharrezkoa Lorenzen kurba eratzea) eta Palma ratioa. Emaitzak ikusita, errenten kontzentrazioa handia dela esan al daiteke? Zergatik?
- Datu guzti-guztiak hartuz berriz ere eta pobrezia-mugatzat 24ko errenta harturik, kalkula ezazu pobrezia-aren adierazle sintetikoa (adierazita utzi, ez da beharrezkoa azken emaitza kalkulatzeko). Zer adierazten du adierazle horrek: pobrezia-tasa ala pobrezia-aren intentsitatea?

a

KAXA-DIAGRAMAK: ERAKETARAKO TAULA

Lehen eta hirugarren kuartilak datu kopurua bider dagokion portzentajea ematen duen datu-ordenari dagokion datua emanaz kalkulatu dira; datu-ordena hori zenbaki osoa ez den kasuetan interpolazio linealeko formula erabili da. Mediana erdiko datu edo erdiko datuen batezbesteko gisa kalkulatu da.

Urratsa	Behe-mailako klasea	Ertaineko klasea	Goi-mailako klasea
Q_1	4gn datua=22	6gn datua=40	2.75gn datua=76.75
Q_3	12gn datua=32	18gn datua=49	8.25gn datua=93.75
Me	$(8gn\ datua + 9gn\ datua)/2=28$	$(12gn\ datua + 13gn\ datua)/2=45.5$	6gn datua=85
$IQR = Q_3 - Q_1$	10	9	17
$1.5IQR$	15	13.5	25.5
$Q_1 - 1.5IQR$	7	26.5	51.25
$Q_3 + 1.5IQR$	47	62.5	119.25
Albo-balioa azpitik	14	39	62
Albo-balioa gainetik	38	60	99
Outlier azpitik	6	-	-
Outlier gainetik	-	-	125



b

Kaxako erdiko marrak, medianak alegia, begiz jota, argi dago errenta-tarte handiena klase ertainetik goi-mailako klasera doala. Zehazki, hauxe da tartearen orokorrean ertaineko errentetatik goi-mailako errentetara, medianak alderatuz: $85 - 45.5 = 39.5$.

c

Sakabanatzeak alderatzeko, sakabanatze-neurri erlatibo bat behar da. IQRri dagokion neurri erlatiboa kuartil arteko sakabanatze-koefizientea da:

$$CD_{IQR}(behe) = \frac{IQR}{Q_3 + Q_1} = \frac{10}{22 + 32} = 0.18$$

$$CD_{IQR}(ertain) = \frac{IQR}{Q_3 + Q_1} = \frac{9}{40 + 49} = 0.10$$

$$CD_{IQR}(goi) = \frac{IQR}{Q_3 + Q_1} = \frac{17}{76.75 + 93.75} = 0.10$$

Tarte handiarekin sakabanatze-handiena behe mailako errenten artean dago.

IQR-aren eta dagokion neurri erlatiboaren erabilera guztiz justifikatuta dago, outlier edo datu atipikoak daudelako; izan ere, IQR neurri sendoa da eta beraz, datu atipikoek ez dute nabarmen erasaten.

d

Kontzentrazioa aztertzeo aukeratu diren datuak 125, 85, 62, 56, 47, 44, 40, 38, 31 eta 25 da. Ohartu behar da Giniren indizea kalkulatzeko datuak txikienetik handienera ordenatu behar ditugula aurrez:

x_{ord}	p_i	x_{met}	q_i	$p_i - q_i$
25	0.1	25	0.0452	0.0548
31	0.2	56	0.1013	0.0987
38	0.3	94	0.1700	0.1300
40	0.4	134	0.2423	0.1577
44	0.5	178	0.3215	0.1785
47	0.6	225	0.4076	0.1924
56	0.7	281	0.5081	0.1919
62	0.8	343	0.6206	0.1794
85	0.9	428	0.7736	0.1264
125	(1.0)	553	1.0000	0.0000
	4.5			1.1098

Kalkula dezagun Giniren indizea:

$$G = \frac{\sum(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = \frac{1.1098}{4.5} = 0.2466$$

Palma ratioa %10 aberatsenak hartzen duen errenta zatia (1-0.7736=0.2264) zati %40 pobreenak hartzen duen errenta zatia (0.1577) da:

$$Palma = \frac{0.2264}{0.1577} = 1.43$$

Beste errenta-banaketa batzuekin alderatzen ez diren bitartean, ezin da baieztatu balio hauek kontzentrazio handia adierazten duten.

e

24 pobrezia-muga izanik, pobretzat hartzen direnen errentak hauek dira: 7, 14, 18, 22, 23.

Pobreziaren adierazle sintetikoa eman dezagun:

$$S = \frac{2 \times [(24 - 7) \times 5 + (24 - 14) \times 4 + (24 - 18) \times 3 + (24 - 22) \times 2 + (24 - 23) \times 1]}{(5 + 1) \times 51 \times 24}$$

Adierazle honek aipaturiko bi ezaugarriak, pobrezia-tasa nahiz pobreziaren intentsitatea, neurtzen ditu batera; izan ere, horregatik deitzen zaio adierazle sintetikoa.

II. ebazkizuna (2.5 puntu)

Espainian, *choni* hitzak emakume estereotipo bat adierazten du: klase sozial baxuko emakumea, orokorrean gaztea, hipersexualizatua eta estetikoki nabarmena, konplexurik gabe bere buruaz harro dagoena. Horren eredu gisa Rosalía abeslaria hartu izan da. Gizonezkoen artean, esan liteke *cani* dela estereotipo baliokidea. Ikerketa soziologikorako institutu baten eskutik, Donostiako Altza auzoko neska gazteen artean behaketa zuzeneko galdetegiak bete dituzu neska gazte *choni*-ak identifikatzeko. Galdetegiko itemen artean honako hauek daude:

- (1) Izen *choni*-a du.
- (2) Txandal edo *leggin*-ak (praka elastiko estu-estuak) janzten ditu.
- (3) Eratzun formako belarritakoak janzten ditu.
- (4) Ageriko tatuajeak ditu.
- (5) Ile luzea du eta motots batean bilduta darama.

Aurreko item-ekin betetako galdetegi batzuetan jasotako datuak ematen zaizkizu jarraian. Izena *choni*-a den erabakitzea, eta beraz item horri puntuazioa ematea, zure esku uzten da. *Choni* izatearen puntuazio totala 5 itemetan dauden baieztokak zenbatuz (bai: 1; ez: 0) kalkulatzeko da:

Neskak (↓)/Galdera (→)	(2)	(3)	(4)	(5)
Jenni	1	1	1	1
Silvia	1	1	1	1
Vanessa	1	1	0	1
Yoli	1	0	1	1
Susi	1	0	1	1
Jessi	1	1	0	1
Eugenia	0	0	0	1

Egin beharreko atazak:

- (a) (Galdera honek ez du punturik ematen, baino ebazkizuna ulertzen eta garatzen lagundu dezake.) Zure iritziz, zein da *choni* estereotipoarekin zerikusi txikiena duen itema?
- (b) Emandako datuekin kalkulatu Cronbach-en alpha eta interpreta ezazu. (Gogoratu izena *choni*-a den zure esku uzten dela.)
- (c) Auzoan behatutako neska guztiak harturik, item bakoitzaren (zenbakiz adierazita) eta testeko puntuazio totalaren (T) arteko korrelazio-koefizienteak ematen zaizkizu ondoren:

$$r_{2T} = 0.74; r_{3T} = 0.67; r_{4T} = 0.53; r_{5T} = -0.42$$

Kalkulatu eman zaizkizun datuekin (7 neska horiek 5 itemetan eman duten erantzunekin), izenari buruzko itemaren eta 5 item horietan lortutako puntuazio totalaren arteko korrelazio-koefizientea, eta aurrekoekin batera jarri, erabaki ezazu testaren koherentzia gehitzeko ezabatuko zenukeen itema.

a

Nire iritzi apalean, motots batean bilduta ile luzea izatea ez da *choni*-en ezaugarri eskusiboa. Neska asko daude *choni* izan gabe, ilea motots batean bilduta daramatenak, eta neska *choni* asko ile luzea ez dutenak.

b

Cronbach-en alfa kalkulatzeko, item guztien eta puntuazio totalaren bariantzak kalkulatu behar dira. Horretarako lehen itemerako, neska bakoitzaren izena *choni*-a den kontsideratu behar da. Nire iritziz, neska guztien izenak dira *choni* izen tipikoak, Eugeniarena ezik. Beraz, item honetan 1 balioa esleituko diegu neska guztiei, Eugeniari ezik; azken honi, 0 balioa emango diegu.

Jarraian, bariantzak kalkulatzeko taula ematen dugu (kalkuluak oso azkarrik dira, eta kasu batzuetan errepikatu egiten dira; hain zuzen ere, lehen eta bigarren itemek, alde batetik, eta hirugarrenak eta laugarrenak, bestetik, batezbesteko eta bariantza berdinak izango dituzte, datuak berdinak direlako haietan):

x_1	x_2	x_3	x_4	x_5	x_T	x_1^2	x_2^2	x_3^2	x_4^2	x_5^2	x_T^2
1	1	1	1	1	5	1	1	1	1	1	25
1	1	1	1	1	5	1	1	1	1	1	25
1	1	1	0	1	4	1	1	1	0	1	16
1	1	0	1	1	4	1	1	0	1	1	16
1	1	0	1	1	4	1	1	0	1	1	16
1	1	1	0	1	4	1	1	1	0	1	16
0	0	0	0	1	1	0	0	0	0	1	1
6	6	4	4	7	27	6	6	4	4	7	115

Orain, goazen bariantzak (eta horretarako lehenago batezbestekoak) kalkulatzera:

$$\bar{x}_1 = \bar{x}_2 = \frac{6}{7} = 0.86; \quad \bar{x}_3 = \bar{x}_4 = \frac{4}{7} = 0.57; \quad \bar{x}_5 = \frac{7}{7} = 1; \quad \bar{x}_T = \frac{27}{7} = 3.85$$

$$s_1^2 = s_2^2 = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{6}{7} - 0.86^2 = 0.12$$

$$s_3^2 = s_4^2 = \frac{4}{7} - 0.57^2 = 0.24$$

$$s_5^2 = \frac{7}{7} - 1^2 = 0 \quad (\text{ez dago kalkulatu beharrik: datu guztiak berdinak direnean, bariantza 0 da})$$

$$s_T^2 = \frac{115}{7} - 3.85^2 = 1.60$$

Cronbach-en alfa kalkulatu dugu orain:

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right] = \left[\frac{5}{5-1} \right] \left[1 - \frac{0.12 + 0.12 + 0.24 + 0.24 + 0}{1.60} \right] = 0.68$$

Gutxigatik bada ere, Cronbach-en alfa ez da iristen testak barne koherentzia duela baieztatu ahal izateko behar den gutxieneko mailara.

c

Kalkula dezagun lehen itemaren eta testeko puntuazioaren arteko korrelazioa:

x_1	x_T	x_1^2	x_T^2	x_1x_T
1	5	1	25	5
1	5	1	25	5
1	4	1	16	4
1	4	1	16	4
1	4	1	16	4
1	4	1	16	4
0	1	0	1	0
6	27	6	115	22

Lehen itemaren eta testeko puntuazioaren arteko korrelazio-koefizientea kalkulatzeko behar ditugun desbideratzeak aurreko ataleko bariantzetatik ateratzen ditugu:

$$s_1 = \sqrt{0.12} = 0.34; \quad s_T = \sqrt{1.60} = 1.265$$

Kobariantza kalkulatzeko orain:

$$s_{1T} = \frac{22}{7} - 0.86 \times 3.85 = -0.168$$

Azkenik, lehen itemaren eta testaren arteko korrelazio-koefizientea ematen dugu:

$$r_{1T} = \frac{-0.168}{0.34 \times 1.265} = -0.39$$

Hau kalkulaturik, item guztietatik testarekin korrelazio txikiena duena azkenekoa (ile luzea-mototsa bai/ez) dela ikusten dugu berehala, hasieran susmatzen genuen bezalaxe; beraz, berau da ezabatuko genukeen lehena barne koherentzia gehitzeko.

III. ebazkizuna (2.5 puntu)

BIZIO online bizikleta-dendako biltegian eskaria handiko modelo baten 4 bizikleta daude. Eguneko eskaria 1, 2 edo 3 bizikleta izan daiteke, 0.2, 0.5 eta 0.3 probabilitateekin hurrenez hurren. Egun ezberdinetako eskariak elkarrekiko independenteak dira. Bizikleta stockik ez badago, hots, biltegiko bizikleta guztiak konprometituta daudenean, bezeroari online ohartarazten zaio ez dagoela stock-ik eta beraz salmenta ez da gauzatzen. Suposatu ezazu eguneko eskariak eta ondoriozko salmentak egunean zehar uniformeki gertatzen direla, eta beraz, biltegia agortu bitarteko denbora egun kopuru oso bat gehi egunaren frakzio bat izan daitekeela (adibidez 1-2-2 badira eskariak, biltegia bi egun t'erdia husten da).

Egin beharreko atazak:

- Bigarren egunaren bukaeran biltegian geratzen diren bizikleta kopuruaren itxaropena eman ezazu, horretarako probabilitate-zuhaitza taula formatuan eta modu sistematikoan osatuz.
- Bizikletak guztiz agortu bitarteko denboraren itxaropena eman ezazu, horretarako probabilitate-zuhaitza zuhaitz gisa (taula moduan ez) irudikatuz.

a

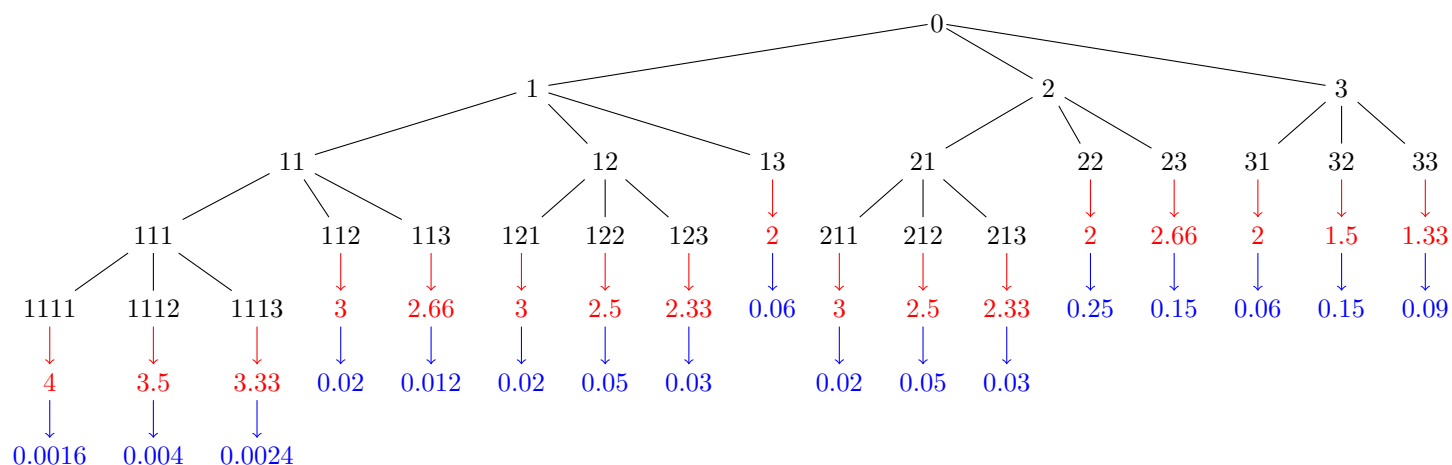
Egunero hiru salmenta posible daude: 1, 2 eta 3. Beraz, bi egunetan $3 \times 3 = 9$ aukera izango dira, taula honetako lehen 4 zutabeetan zehazten direnak:

1.eg.sal.	p(1.eg.sal.)	2.eg.sal.	p(2.eg.sal.)	bz=biltegian zenbat=4-1.eguna-2.eguna	p(bz)	bz $\times$ p(bz)
1	0.2	1	0.2	4-1-1=2	$0.2 \times 0.2 = 0.04$	0.08
1	0.2	2	0.5	4-1-2=1	$0.2 \times 0.5 = 0.10$	0.10
1	0.2	3	0.3	4-1-3=0	$0.2 \times 0.3 = 0.06$	0
2	0.5	1	0.2	4-2-1=1	$0.5 \times 0.2 = 0.10$	0.10
2	0.5	2	0.5	4-2-2=0	$0.5 \times 0.5 = 0.25$	0
2	0.5	3	0.3	4-2-3=0 (stock-haustura)	$0.5 \times 0.3 = 0.15$	0
3	0.3	1	0.2	4-3-1=0	$0.3 \times 0.2 = 0.06$	0
3	0.3	2	0.5	4-3-2=0 (stock-haustura)	$0.3 \times 0.5 = 0.15$	0
3	0.3	3	0.3	4-3-3=0 (stock-haustura)	$0.3 \times 0.3 = 0.09$	0
					1	0.28

- Bosgarren zutabea bi egunen ondoren biltegian zenbat pieza geratzen diren kalkulatzeko da: 4 ken 1. eguneko salmenta ken 2. eguneko salmenta.
- Aurreko balioa negatiboa denean, stock-haustura gertatzen dela esan nahi du: bezero batek bizikleta eskatu du baino ezin zaio zerbitzatu.
- Adar bakoitzeko probabilitatea adar barruko probabilitateen biderketa da.
- Biltegian geratzen den pieza kopuru bakoitza bider probabilitate horiek eginez eta emaitza guztiak batuz, bi egun igarota biltegian geratzen den pieza kopuruaren itxaropena kalkulatzeko dugu: 0.28 pieza.

b

Probabilitate-zuhaitza goitik behera marraztu dugu. Adar bakoitzak biltegia hustu arteko egun bakoitzean zenbat saldu den adierazten du, eta adarraren bukaeran, gezi baten ondoren eta gorriz, biltegia hustu arteko egun kopurua adierazten da, enuntziatuan azaldu bezala. Gezi urdin baten ondoren, adar bakoitzeko probabilitatea (urdinez baita ere) adierazten da:



Aurrera jarraitu baino lehen, probabilitate guztien batura 1 dela egiaztatu behar dugu. Eta hala da. Beraz, aurrera!

Egun kopuru posibleak bilduz eta kopuru bakoitzari dagozkion adar desberdinetako probabilitateak gehituz, taula laburra sortzen dugu, itxaropena kalkulatzeko:

egun kopuruak biltegia hustu arte (e)	p(e)	ep(e)
1.33	0.09	0.1197
1.5	0.15	0.225
2	0.37	0.74
2.33	0.06	0.1398
2.5	0.10	0.25
2.66	0.162	0.43092
3	0.06	0.18
3.33	0.0024	0.007992
3.5	0.004	0.0014
4	0.0016	0.0064
	1	2.113812 egun

Beraz, biltegia guztiz hustu arte batez 2.11 egun igarotzen dira.

ESTADÍSTIKA ETA DATUEN ANALISIA (TESTA)

Irakaslea: Josemari Sarasola

Data: 2025eko urtarrilaren 14, 15:00

Iraupena: 40 minutu

Erantzun bakarra da zuzena galdera bakoitzean. Guztira testak 2 puntu balio du. Erantzun zuzenak 0.1 puntu balio du. Erantzun oker bakoitzak zuzenak ematen duenaren erdia kentzen du. Galderak erantzun gabe utz daitezke, punturik gehitu eta kendu gabe.

1. Nola deitzen zaio kasu berezi bat aztertzen duen ikerketari?
 - (a) Ikerketa nomotetikoa.
 - (b) Ikerketa idiografikoa.
 - (c) Ikerketa atipikoa.
 - (d) Ikerketa partikularra.
2. Noiz egin zen lehen zentsua?
 - (a) Antzinako zibilizazioetatik egin dira zentsuak.
 - (b) Erdi Aroan.
 - (c) John Graunt-ek egin zuen lehen aldiz Londres-en XVII. mendean.
 - (d) XIX. mendean, estatistika demografiara aplikatzen hasi zenean.
3. Zer aztertzen du estatistikak bereziki?
 - (a) Kasu partikularretan dauden berezitasunak.
 - (b) Fenomeno aldakorretan dauden joerak.
 - (c) Zorizko fenomenoetan aurki daitezkeen erroreak.
 - (d) Zenbakizko datuen aldagaiak.
4. Zein egoeratan eratu behar da puntu-diagrama?
 - (a) Aldagai kuantitatibo diskretua denean, datu gutxirekin.
 - (b) Aldagai kuantitatibo diskretua denean, datu askorekin.
 - (c) Aldagai kuantitatibo jarraitua denean, datu gutxirekin.
 - (d) Aldagai kuantitatibo jarraitua denean, datu askorekin.

5. Irakasle batek %20, %50 eta %30eko haztaperen-koefizienteak ezarri dizkie hurrenez hurren berak ematen dituen hiru ikasgaiei. Irakasle batek 7 eta 6ko notak lortu ditu lehen bi ikasgaietan. Zenbateko nota lortu behar du hirugarren ikasgaian bere batez besteko nota 6.5 izateko?
- 7.5
 - 7.4
 - 7.2
 - 7
6. 50 ikaslek osatzen duten talde batean hauek dira estatistikako notak, 0tik 10erako tartetan bilduak (ez dira tarte guztiak ematen):

Notak	Ikasleak
9-10	8
8-9	7
...	...

Kalkulatu 80garren pertzentila.

- 8.56
 - 8.61
 - 8.66
 - 8.71
7. Datu hauek dituzu: 3-9-5-7. Sakabanatze-neurri gisa, neurri hau proposatu da, batezbestekorako desbideratze absolutuetan oinarritzen dena:

$$BDA = \frac{\sum_i |x_i - \bar{x}|}{n}$$

Kalkula ezazu neurri hori emandako datuetarako.

- 2
 - 2.5
 - 3
 - 3.5
8. Noiz dator bat Lorenz-en kurba diagonalarekin?
- Ez da posible Lorenz-en kurba diagonalarekin guztiz bat etortzea, diagonalarek erreferentzia hutsa baita.
 - Errenten artean erabateko parekotasuna dagoenean.
 - Errentek linealki egiten dutenean gora, eta ez kurba baten arabera.
 - Batek errenta osoa hartu eta beste guztiek 0 hartzen dutenean.
9. Zein kintilekin dator bat 8garren dezila?
- Erdiko kintila gehi 30. pertzentila eginez sortzen den emaitzarekin.
 - Laugarren kintilarekin.
 - Azken kintilarekin.
 - Aurreko biak dira zuzenak.
10. Bi familia alderatu behar dira: alde batetik, bakarrik bizi den bikote bat, haurrik gabe, hilean 2.500 euroarekin; bestetik, bikote bat bi haurrekin, non gizonak 2.000 euro jasotzen dituen hilerok. Bi familiak errentaz berdina izateko, zenbat jaso beharko luke bigarren familiako emakumeak? Kalkuluak egiteko OCDE erakundearen baliokidetasun-eskala erabili.
- 2.040 euro gutxi gorabehera.
 - 1.970 euro gutxi gorabehera.
 - 1.840 euro gutxi gorabehera.
 - 1.770 euro gutxi gorabehera.

Hurrengo hiru galderak kontingentzia-taula honi buruzkoak dira:

Soldata ($\downarrow$)/Hezkuntza-maila ($\rightarrow$)	Handia	Txikia
Handia	42	68
Txikia	76	36

-
11. Bi aldagaien artean erabateko independentzia balitz, zenbatekoa litzateke handia/handia gelaxkako maiztasuna?
- $(118 \times 110)/222$
 - $222/(118 \times 110)$
 - 118×110
 - $42/222$
12. Aldagai independentea kontuan hartu gabe, zenbat errore egiten dira aldagai dependentearen aurreanari buruz?
- 104
 - 118
 - 110
 - 112
13. Zenbat errore egiten dira aldagai dependentearen aurrean aldagai independentearen balioa "handia" denean?
- 76
 - 68
 - 42
 - 36
14. Sei hilabeteko haur batek egunean erabiltzen duen pixoiak kopurua honela banatzen dela uste da:

$$P[X = x] = \frac{x+1}{27}; \quad x = 1, 2, 3, 4, 5, 6$$

Kalkulatu $F(x = 3)$.

- $6/27$
 - $7/27$
 - $15/27$
 - $1/3$
15. Eguneko salmenta denda batean honela banatzen dela uste da:

$$F(x) = \frac{x^2}{100}; \quad 0 \leq x \leq 10$$

Kalkulatu salmenta 4 eta 6 artean izateko probabilitatea.

- 0.18
 - 0.20
 - 0.36
 - 0.40
16. Zorizko aldagai batek -1 eta +1 balioak $1-p$ eta p probabilitateekin har ditzake. Kalkulatu bariantza.
- $4p(1-p)$
 - $4p^2 - 1$
 - $4p - 1$
 - $4p^2 - p + 1$

17. Bi inbertsioen asteko errentagarritasunari buruzko informazioa da honako hau: $\mu_A = 4; \mu_B = 6; \sigma_A = 2; \sigma_B = 4$. Honako hauetatik zein da zuzena? Behar izanez gero, $U = \frac{\mu}{\sigma}$ utilitate-funtzioa erabili.
- (a) Epe laburrera B da hobe.
 - (b) Epe laburrera A da hobe.
 - (c) Epe luzera B da hobe.
 - (d) Aurreko biak dira zuzenak.
18. Zorizko aldagai batek -a eta +a balioak hartzen ditu, biak 0.5eko probabilitateekin. Kalkulatu desbideratze estandarra.
- (a) a
 - (b) $\sqrt{a}$
 - (c) $\sqrt{a^2 - 1}$
 - (d) $\sqrt{a - 1}$
19. Bizikleta-denda batek zenbat bizikleta erosi behar dituen erabaki behar du datorren hilerako. Salmenta 3, 4, 5 edo 6 izan daiteke, 0.2, 0.4, 0.3 eta 0.1eko probabilitateekin hurrenez hurren. 4 bizikleta erosten baditu, zenbat da biltegian saldu gabe geratuko diren salmenten itzaropena?
- (a) 0.2
 - (b) 0.4
 - (c) 0.6
 - (d) 0.8
20. Zorizko aldagai jarraitu bati buruz itzaropena eta desbideratzea hurrenez hurren 20 eta 3 direla baino ez dakigu. Hurbildu $P[X > 25]$.
- (a) 3/25 baino txikiagoa.
 - (b) 9/25 baino txikiagoa.
 - (c) 9/25 baino handiagoa.
 - (d) 3/25 baino handiagoa.

Estatistika eta datuen analisia

2025eko ekainaren 23a

Izena eta abizenak:_____

Galdera	Erantzuna
1	B
2	A
3	B
4	C
5	D
6	D
7	A
8	B
9	D
10	B
11	A
12	C
13	C
14	D
15	B
16	A
17	D
18	A
19	A
20	B

KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0