

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-aula 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-aula 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-aula 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 kategoria berean bildu eta 2x2 kontingentzia-aula 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-aula 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:

$\frac{\text{enpiriko (O)}}{(O - E)^2}$	$\frac{\text{teoriko (E)}}{(O - E)} \sqrt{E}$	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:

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

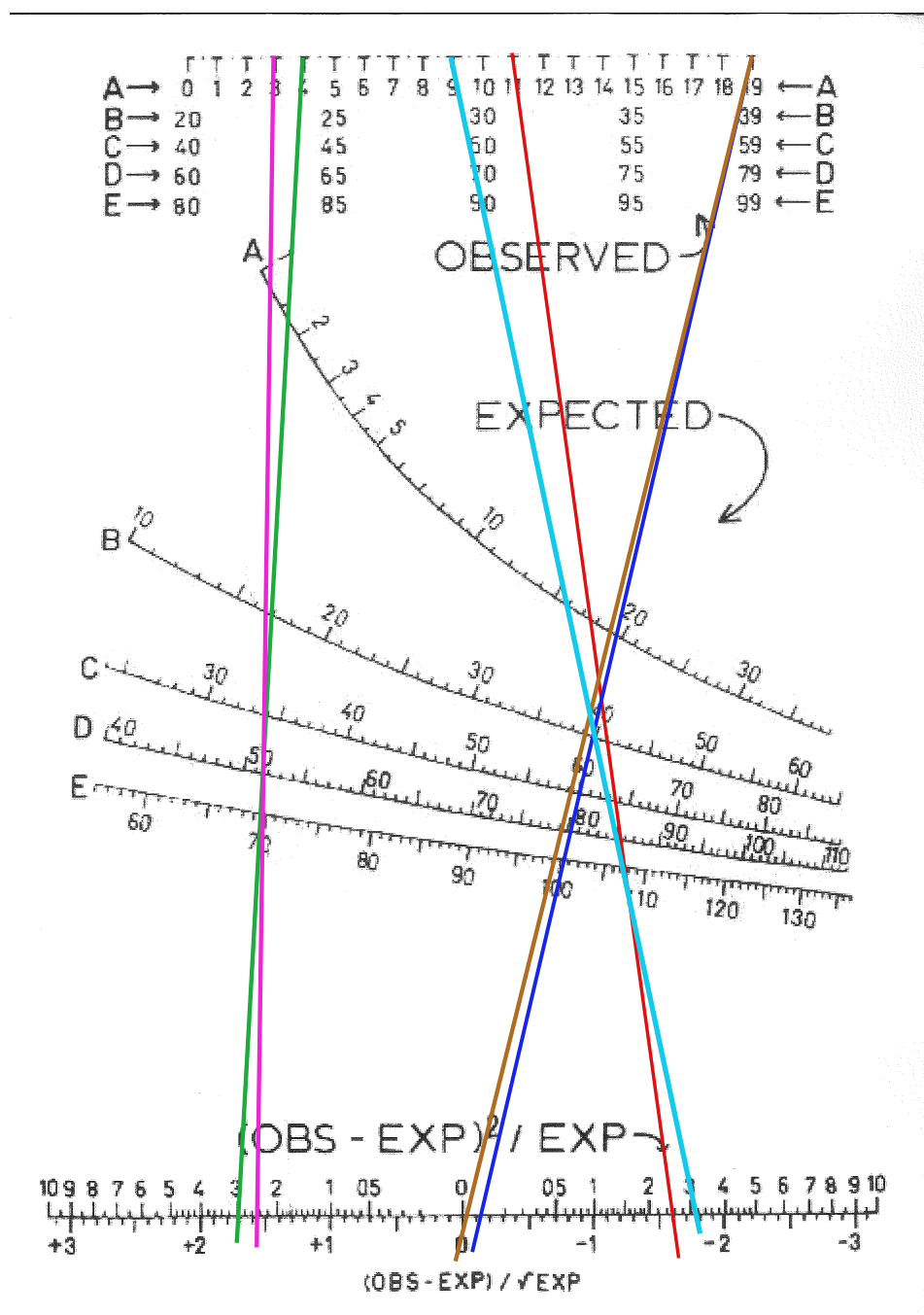
Normalizatuz:

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

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

Horrela, azpiko galtzen kolekzioa 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

11	19	24
23	19	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	0	43	43	0
0	0	23	0	23	0	0	23
0	1	28	0	28	0	0	28
		105	54	51	43	54	51

$$\bullet \bar{x} = \frac{54}{105} = 0.51 \quad ; \quad \bar{y} = \frac{51}{105} = 0.48$$

$$\bullet s_{xy} = \frac{43}{105} - 0.51 \times 0.48 = 0.165$$

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

$$\bullet s_y = \sqrt{\frac{51}{105} - 0.48^2} = 0.5$$

$$\bullet r_{xy} = \frac{0.165}{0.5 \times 0.5} = 0.66$$

Beraz, lotarako azpiko galtzak janztearen eta adinaren edo adin-taldearen arteko korrelazioa sendoa da (> 0.6), betiere lagin errearearen 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 sendoa 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 esanguratsua 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

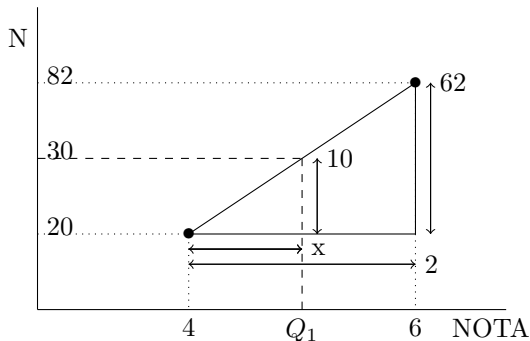
Egin beharreko atazak:

- (a) %25ean (alde bietatik) winsorizatutako batezbestekoa kalkulatu.
- (b) Jakinaenez, 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)

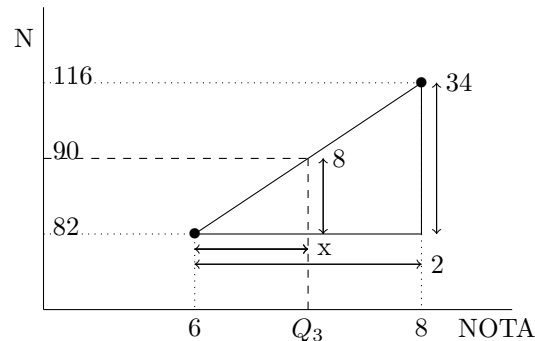
%25ean winsorizatutako batezbestekoa kalkulatzeko, $Q_1 = P_{25}$ eta $Q_3 = P_{75}$ kalkulatu behar dira. Datuak tartetan biduta daudenez, interpolazio linealez egin behar da kalkulua:

$$Q_1: 0.25 \times 120 = 30 \text{ gn.datua} \in [4, 6]$$



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

$$Q_3: 0.75 \times 120 = 90 \text{ gn.datua} \in [6, 8]$$



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

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 92% an 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 lituzke 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-aldaiketaren 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 "fami-liak"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 itzaropen 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	