

ESTADÍSTICA ETA DATUEN ANALISIA

Azterketa ebatziak

2022-2023 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea. EHU

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

gizapedia.org

ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2023ko urtarrilaren 23an, 16:00

Iraupena: 90 min

I. ebazkizuna (3 puntu) (Ebaluazio jarraitua: 2.5 puntu)

1995ean Jaizkibelen landatutako pinudi bateko pinuen enborreko diametroa jaso da (cm), 150cm-ko altueran. Guztira 81 pinu neurtu dira. Hona hemen jasotako datuak:

146 168 135 165 110 136 172 139 140
 134 137 145 144 126 111 137 124 160
 141 144 170 134 123 152 142 197 151
 152 144 144 178 142 143 174 163 146
 164 144 159 153 169 129 148 132 160
 136 140 143 177 98 166 146 154 212
 127 130 145 134 150 160 208 155 162
 140 118 165 138 173 137 119 164 149
 167 145 107 176 150 138 138 141 174

Egin beharreko atazak:

- Sturges-en erregela erabiliz, eta histogramak eratzeko irizpide egokiak baliatuz, datuen tartekako maiztasun-taula eratu, maiztasun absolutuak baino ez zehaztuz, eta dagokion histograma eratu (etzazu histograma koloreztatu).
- Kuartil arteko ibiltartea kalkulatu **tartekako maiztasun-taulatik** eta histograman bertan irudikatu ezazu, marrak eta geziak erabiliz eta kuartil arteko ibiltartearen barruan kokatzen diren datuak koloreztatuz edo marrakatu. Koloreztatutako azalera ehuneko zenbatekoa da berez histograma osoko azalerari buruz?
- Populazio-bariantza kalkulatu ezazu **tartekako maiztasun-taulatik**. Zenbat desbideratzen da orohar zuhaitza baikoitzaren diametroa batez besteko diametrotik?

a

Tarte-kopurua kalkulatzeko dugu: $k = \frac{\ln 81}{\ln 2} + 1 = 7.34$

Gehiegiz borobiltzen dugu: $7.34 \rightarrow 8$. Beraz, 8 tarteko histograma eratu behar da.

Datu txikiena: 98. Datu handiena: 212. Ibiltartea: $212 - 98 = 114$.

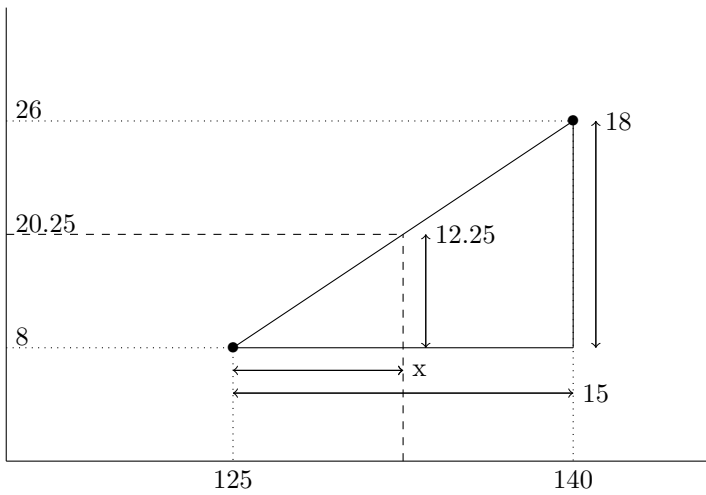
114ko ibiltartea 8 tartetan biltzeko, tarte bakoitzak $114/8 = 14.25$ hartu behar du. Tarteak biribilak izateko, gehiegiz borobiltzen dugu: $14.25 \times 8 = 114$. Beraz, 15 zabalerako 8 tarte eratu behar dira, guztira $15 \times 8 = 120$ -eko zabalera hartzen dutenak. 114tik 120erako soberan dagoen 6ko tarte hori bi muturretan erdibanatzen dugu, 3 azpitik eta 3 gaineratik. Beraz, 95etik hasi eta 215ean bukatuko dugu. Datuen zenbaketa egin eta maiztasun-taula eratzeko dugu:

Diametroa (cm)	Zenbaketa	Zuhaitzak (n)	N
95-110	//	2	2
110-125	////	6	8
125-140	////////	18	26
140-155	//////////	29	55
155-170	////////	15	70
170-185	////	8	78
185-200	/	1	79
200-215	//	2	81
		81	

b) Tartekako maiztasunetik, kuartilak interpolazioz kalkulatu behar dira.

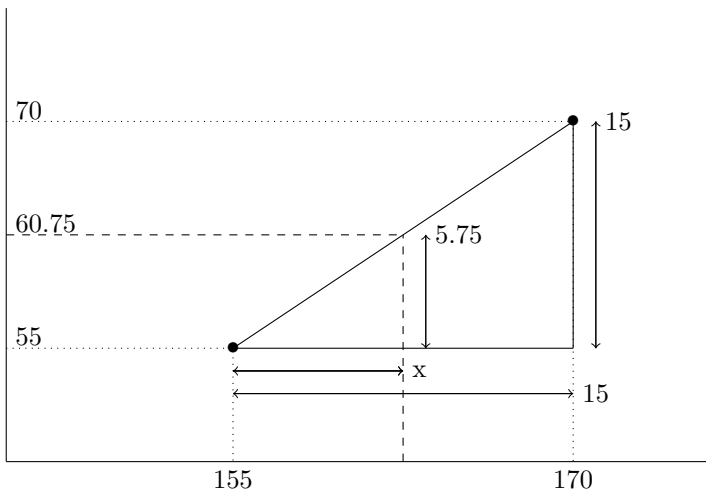
Lehen kuartila: Q_1 ? $81 \times 0.25 = 20.25$ gn datua, 125-140 tartean.

Interpolazioz:



$$\frac{x}{12.25} = \frac{15}{18} \rightarrow x = 10.21 \rightarrow Q_1 = 125 + 10.21 = 135.21$$

Hirugarren kuartila: Q_3 ? $81 \times 0.75 = 60.75$ gn datua, 155-170 tartean.



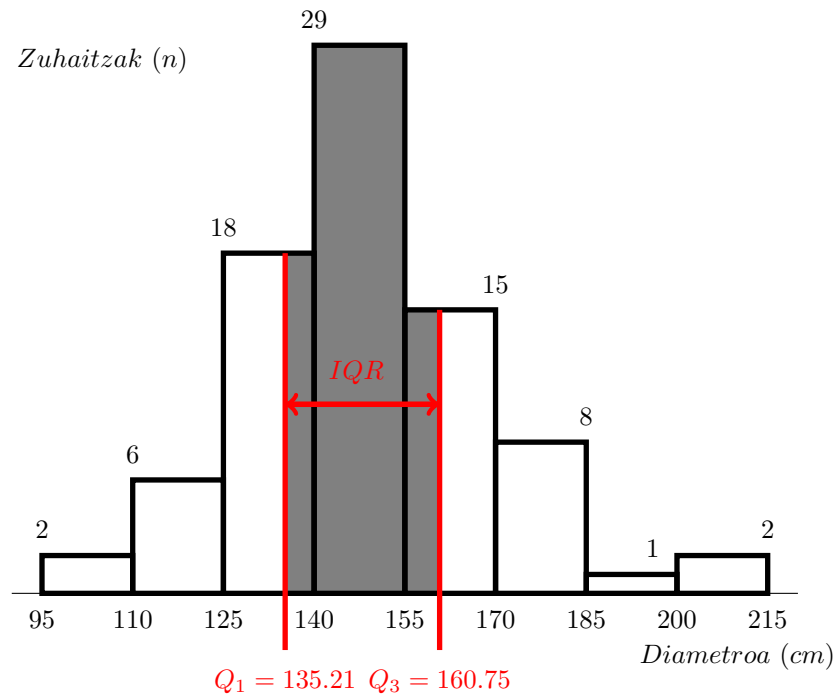
$$\frac{x}{5.75} = \frac{15}{15} \rightarrow x = 5.75 \rightarrow Q_3 = 155 + 5.75 = 160.75$$

Beraz, hauxe da kuartil arteko ibiltartea:

$$IQR = Q_3 - Q_1 = 160.75 - 135.21 = 25.54cm$$

Bi puntako gezi batez seinalatzen da histograman. **Tarte horretan datu guztietatik %50ak biltzen dira, histogramako azalerako %50a, alegia.**

Histograma, IQRrekin zein IQRaren baitan gertatzen den datuek hartzen duten azalerarekin batera, ondoren azaltzen da:



c

Bariantzaren kalkulua:

Tartea	x	n	nx	nx ²
95-110	102.5	2	205	21012.5
110-125	117.5	6	705	82837.5
125-140	132.5	18	2385	316012.5
140-155	147.5	29	4277.5	630931.25
155-170	162.5	15	2437.5	396093.75
170-185	177.5	8	1420	252050
185-200	192.5	1	192.5	37056.25
200-215	207.5	2	415	86112.5
		81	12037.5	1822106

$$\bar{x} = \frac{12037.5}{81} = 148.61$$

Bariantza jarraian zehazten dugu:

$$s_x^2 = \frac{1822106}{81} - 148.61^2 = 410.2$$

Bataz beste datu bakoitza 148.61eko batezbestekotik desbideratze estandararen neurria desbideratzen da:

$$s_x = \sqrt{410.2} = 20.25 \text{ cm}$$

II. ebazkizuna (2.5 puntu) (Ebaluazio jarraitua: 2.25 puntu)

Ligatzeko sare sozial batean, *match* egin duten bikote heterosexualen artean, gizona eta emakumea, hurrenez hurren, ederrak, arruntak edo itsusiak diren ebaluatu du talde batek. Hona hemen datuak, kontingentzia taula batean bildurik:

Gizon (↓) / Emakume (→)	Ederra	Arrunta	Itsusia
Ederra	45	36	32
Arrunta	30	46	29
Itsusia	6	17	35

- (a) Kontingentzia-koefizientea kalkulatu eta modu zorrotz eta egokian interpretatu.
- (b) Goodman eta Kruskal-en gamma kalkulatu eta interpretatu, interpretazioa modu honetan egin ahal izateko moduan: emakumea zenbat eta ederrago, ...
- (c) Sexu bakoitzeko selektibitate-neurri hau planteatu da: sexu bakoitzeko kategoria bakoitzean norbera baino ederragoa den beste sexuko pertsona bat aukeratzeko portzentajea kalkulatu behar da lehenbizi (**pertsona ederren kasuan ezik, non ederragoa aukeratu ordez, besterik gabe beste sexuko pertsona ederra aukeratzeko portzentajea kalkulatu den**); adibidez, **emakume itsusietarako** kalkulatu beharreko portzentajea hau da: $[(32 + 29)/(32 + 29 + 35)] \times 100$; portzentaje horiek kalkulaturik, SELEKTIBITATE-NEURRIA SEXU BAKOITZEKO KALKULATURIKO HIRU PORTZENTAJEEN BATEZBESTEKO ARITMETIKOA KALKULATUZ ZEHAZTEN DA. **Gizon nahiz emakumeetarako hiru portzentajeak kalkula itzazu (3 SEXU BAKOITZEKO, ALEGIA), eta gero bi sexu etarako selektibitate-neurriak eman. Zein da sexu selektiboena?**

a

Taulari bazter-maiztasunak erantsi behar dizkiogu:

Gizon (↓) / Emakume (→)	Ederra	Arrunta	Itsusia	Totalak
Ederra	45	36	32	113
Arrunta	30	46	29	105
Itsusia	6	17	35	58
Totalak	81	99	96	276

Khi-karratu kalkulatzeko taula eratzen dugu, maiztasun enpirikoak (goian ezkerrean), maiztasun teorikoak (goian eskuinean), khi-karratu ekarpenak (behean ezkerrean), eta horien erroak (behean eskuinean):

$\frac{\text{enpiriko (O)}}{(O - E)^2}$ E	$\frac{\text{teoriko (E)}}{(O - E)}$ \sqrt{E}	Emak. ederra		Emak. arrunta		Emak. itsusia	
Gizon ederra		45	33.1	36	40.5	32	39.3
		4.22	2.05	0.5	-0.7	1.35	-1.16
Gizon arrunta		30	30.8	46	37.6	29	36.5
		0.02	-0.14	1.84	1.35	1.55	-1.24
Gizon itsusia		6	17	17	20.8	35	20.2
		7.1	-2.7	0.7	-0.84	10.9	3.3

Khi-karratu ekarpenak (behean ezkerrean) gehituz, khi-karratu estatistikoa eskuratzen dugu:

$$\mathbf{X}^2 = 4.22 + 0.5 + \dots + 10.9 = 28.23$$

Kontingentzia-koefizientea kalkulatu dugu:

$$C = \sqrt{\frac{28.23}{28.23 + 276}} = 0.304$$

Kontingentzia-koefizientearen maximoa kalkulatu dugu:

$$C_{max} = \sqrt{\frac{3 - 1}{3}} = 0.81$$

Koefizientea normalizatu egin dugu: $0.304/0.81 = 0.375$

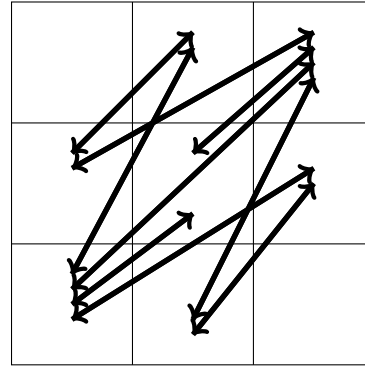
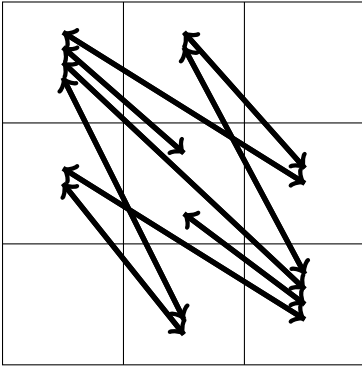
Beraz, sortutako bikoteen edertasun mailen artean asoziazio ertaina dago, lagin errorearen eta antzeko ikerketen erreserba-pean. Asoziazioaren norabide nagusiak hauek dira: gizon itsusiak emakume itsusiekin elkartzen direla, eta emakume ederrak eta gizon itsusiak ez direla inola ere elkartzen.

b

Emakumea zenbat eta ederragoa, ... erako interpretazioa egiteko, aldagaiak edertasun txikienetik handienara ordenatu behar dira. Beraz, taula berregin behar da:

Gizon (↓) / Emakume (→)	Itsusia	Arrunta	Ederra	Totalak
Itsusia	35	17	6	58
Arrunta	29	46	30	105
Ederra	32	36	45	113
Totalak	96	99	81	276

Gamma koefizientea kalkulatzeko, konkordantziak (hego ekialderako norabidean) eta diskordantziak (hego mendebalderako norabidean) zehaztu behar dira ($4+2+2+1=9$ gezi, kasu bakoitzean, eta beraz 9 biderketa egin behar):



$$k = 35 \times 46 + 35 \times 30 + 35 \times 36 + 35 \times 45 + 29 \times 36 + 29 \times 45 + 46 \times 45 + 17 \times 30 + 17 \times 45 = 11189$$

$$d = 6 \times 29 + 6 \times 46 + 6 \times 32 + 6 \times 36 + 30 \times 32 + 30 \times 36 + 46 \times 32 + 17 \times 29 + 17 \times 32 = 5407$$

$$\gamma = \frac{11189 - 5407}{11189 + 5407} = 0.3484$$

Elkartutako emakumearen eta gizonaren arteko asoziazio ordinala ertaina, norabide honetan: emakumea zenbat ederragoa, gizona ere orduan eta ederragoa.

c

Taldea	Portzentajea (%)
Emakume itsusiak	$(32 + 29)/96 \times 100 = 63.5$
Emakume arruntak	$36/99 \times 100 = 36.3$
Emakume ederrak	$45/81 \times 100 = 55.5$
Gizon itsusiak	$(17 + 6)/58 \times 100 = 39.6$
Gizon arruntak	$30/105 \times 100 = 28.5$
Gizon ederrak	$45/113 \times 100 = 39.8$

Selektibitate-neurriak eman ditzagun:

$$s(\text{emakume}) = \frac{63.5 + 36.3 + 55.5}{3} = 51.76$$

$$s(\text{gizon}) = \frac{39.6 + 28.5 + 39.8}{3} = 35.96$$

Emaitza hauen arabera, emakumeak askoz ere selektiboak dira.

Iruzkun gisa, ordea, adierazi behar da selektibitate-neurri hau kritikagarria dela guztiz, hiru portzentajeak berdin haztatzen dituelako, total desberdinetan oinarritzen badira ere. Egokiena kalkulu hau litzateke:

$$s(\text{emakume}) = \frac{32 + 29 + 36 + 45}{276} = 51.45$$

$$s(\text{gizon}) = \frac{17 + 6 + 30 + 45}{3276} = 35.91$$

Emaitzak, kasu honetan, aurrekoen oso antzekoak dira, eta beraz interpretazioa berdina izango da.

III. ebazkizuna (2.5 puntu) (Ebaluazio jarraitua: 1.75 puntu)

Enpresa batean zorizkoak dira eguneko ekoizpenak nahiz eguneko salmentak. Eguneko ekoizpena 1 edo 2 izan daiteke, 0.3 eta 0.7ko probabilitateekin, hurrenez hurren. Eguneko salmentak 1 edo 2 izan daitezke 0.6 eta 0.4 probabilitateekin, hurrenez hurren. Salmenta-eskaerak egunean zehar heltzen dira, baino egunaren amaieran bideratzen dira, nahiko unitate badago biltegian, noski; bestela, atzerapenez bada ere, biharamuneko ekoizpenarekin beteko dira, lehentasun edo prioritate osoarekin. Egunaren amaieran saldu gabeko unitateak hurrengo egun baterako gordetzen dira biltegian. Egun desberdinetako ekoizpenak eta salmentak elkarrekiko independenteak dira.

- (a) Eratu ezazu taula moduan bi egunetako ekoizpen eta salmentei buruzko probabilitate-zuhaitza, bi egunetako ekoizpen eta salmenta posible guztiak konbinatuz.
- (b) Bi egunen buruan biltegian geratzen diren batez besteko unitateen kopurua kalkulatu (itxaropen matematikoa, alegia), **suposatuz lehen egunaren hasieran biltegia hutsik dagoela**, eta baita horien **bariantza** ere, emaitzak bilduz aparteko taula batean.
- (c) Aurreko suposizio berdinekin, bi egunen buruan egindako salmenta-eskaeretatik, atzerapenez zerbitzatu edo bideratuko diren batez besteko unitateen kopurua kalkulatu, emaitzak bilduz aparteko taula batean.

e1	p(e1)	s1	p(s1)	e2	p(e2)	s2	p(s2)	p(e1 eta s1 eta e2 eta s2)	b2/b0=0	atzeratuta
1	0.3	1	0.4	1	0.3	1	0.4	0.0144	0	0
1	0.3	1	0.4	1	0.3	2	0.6	0.0216	0	1
1	0.3	1	0.4	2	0.7	1	0.4	0.0336	1	0
1	0.3	1	0.4	2	0.7	2	0.6	0.0504	0	0
1	0.3	2	0.6	1	0.3	1	0.4	0.0216	0	1
1	0.3	2	0.6	1	0.3	2	0.6	0.0324	0	2
1	0.3	2	0.6	2	0.7	1	0.4	0.0504	0	1
1	0.3	2	0.6	2	0.7	2	0.6	0.0756	0	1
2	0.7	1	0.4	1	0.3	1	0.4	0.0336	1	0
2	0.7	1	0.4	1	0.3	2	0.6	0.0504	0	1
2	0.7	1	0.4	2	0.7	1	0.4	0.0784	2	0
2	0.7	1	0.4	2	0.7	2	0.6	0.1176	1	0
2	0.7	2	0.6	1	0.3	1	0.4	0.0504	0	0
2	0.7	2	0.6	1	0.3	2	0.6	0.0756	0	1
2	0.7	2	0.6	2	0.7	1	0.4	0.1176	1	0
2	0.7	2	0.6	2	0.7	2	0.6	0.1764	0	0
								1		

b

b2	p(b2)	b2p(b2)	b2 ² p(b2)
2	0.0784	0.1568	0.3136
1	0.0336+0.0336+0.1176+0.1176=0.3024	0.3024	0.3024
0	1-0.3024-0.0784=0.6192	0	0
	1	0.4592	0.616

$$\mu_{b2} = E[b2] = \alpha_1 = 0.4592 \text{ unit}$$

Bi egunen buruan, biltegian batz beste 0.4592 pieza izango dira.

$$\sigma_{b2}^2 = 0.616 - 0.4592^2 = 0.4051 \text{ unit}^2$$

c

atz	p(atz)	atzp(atz)
2	0.0324	0.0628
1	0.0216+0.0216+0.0504+0.0756+0.0504+0.0756=0.2952	0.2952
0	1-0.0324-0.2952=0.6724	0
	1	0.4592 0.358

$$\mu_{atz} = E[atz] = 0.358 \text{ unit}$$

2 egun horietan eskatutako piezetatik 0.358 zerbitzatu dira batz beste atzerapenez.

ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2023ko urtarrilaren 23a, 16:00

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 Jorge Luis Borgesek kartografiari buruz idatzi zuen ipuintxoaren irakaspena estatistikari dagokionean?
 - (a) Zenbat eta datu gehiago bildu fenomeno bati buruz, hari buruzko irudikapen hainbat eta zehatzagoa izango dugu.
 - (b) Behar diren datuak baino gehiago biltzen badira, eskuratzen den informazioa nahasia izango da, errealitatearen kopia hutsa.
 - (c) Zentsua, ikergai ditugun datu guztien bilduma, fenomeno baten irudikapen perfektua dela, lagin errorerik ez duelako.
 - (d) Maparen irudikapena edo datu bilketa baino, garrantzitsuena datu horietatik ateratzen den ondorioa dela, mapa hartuta nora goazen alegia.
2. Zein arrazoitze-prozedurarekin lotzen duzu estatistika?
 - (a) Dedukzioarekin, datuetatik ondorioak ateratzearekin alegia.
 - (b) Indukzioarekin.
 - (c) Kasu partikularren azterketatik lege orokorrak eratzearekin.
 - (d) (b) eta (c) batera.
3. 30 ikasleko ikasgela batean, azken ebaluaketan ikasle bakoitzak izan duen suspentsio-kopurua jaso da. Zein da adierazpide grafiko egokiena?
 - (a) Sektore-diagrama.
 - (b) Barra-diagrama.
 - (c) Puntu-diagrama.
 - (d) Histograma.
4. Nola moldatu behar da histograma tarte-zabalerak desberdinak direnean?
 - (a) Maiztasun-poligonoa eratuz.
 - (b) Ondoz-ondoko histogramak eratuz.
 - (c) Dentsitateak kalkulaturik, haiek altueratzat hartuz.
 - (d) Tarte-kopurua gehiegiz borobilduz.
5. A makinaren errendimendua 4 unit/h da; B makinarena, berriz, 6 unit/h. A makinarekin 6 orduz eta B makinarekin 4 orduz arituta, zenbat da batezbesteko errendimendua orduko, horretarako batezbesteko egokiena erabiliz?
 - (a) 4.66 unit/h
 - (b) 4.8 unit/h
 - (c) 5 unit/h
 - (d) 5.33 unit/h
6. 44 datu ordenatu ditugu handienetik txikienera. Zerrenda horretatik lehenengo 8 datuak hauek dira: 88-80-74-70-66-62-60-58. Kalkulatu 9. dezila.
 - (a) 64.4
 - (b) 63.8
 - (c) 62.2
 - (d) 60.6
7. Zein da faltsua honako hauetatik?
 - (a) DAME sakabanatze-neurri sendoa da.
 - (b) Koartil arteko ibiltartea sakabanatze-neurri sendoa da.
 - (c) Populazio-bariantza beti da handiagoa lagin-bariantza baino.

- (d) Cohen-en d kalkulatu ahal izateko, bi datu-multzoetako bariantzak populazioan berdinak direla suposatu behar da.
8. Neurri batzuk proposatu dira sakabanatze-neurri gisa. Zein EZ da, inola ere ez, sakabanatze-neurria?
- (a) $D_8 - D_2$
 (b) $(Q_3 + Q_1)/2$
 (c) $\frac{\sum |x_i - \bar{x}|}{n}$
 (d) Datu-multzoko datu bikote guztien arteko batez besteko diferentzia.
9. Honako familia-errenta hauek jaso dira, eta batura kalkulatu:
 $16+8+8+8+8+7+7+6+6+6+6+5+5+5+4+4+4+3+2=112$.
 Zenbat da familia-errenten mediala?
- (a) 5
 (b) 6
 (c) 7
 (d) 8
10. Herri batean bi familia bizi dira. Aldanondotarrak 2 heldu eta 2 haur dira, eta familia-errenta 1.000 euro da. Basurkotarrak 2 heldu dira eta 600 euroko errenta dute. Pobrezia-muga 350 euro da. Zenbat da pobrezia-tasa?
- (a) 0
 (b) 0.4
 (c) 0.6
 (d) 1
11. Urpeko arrantzale batek harrapatu dituen arrainak sailkatu ditu: A espeziekoak, 3; B espeziekoak 2; eta C espeziekoa, 1 baino ez. Zenbat da, Hill zenbakia deskonposatuz, harrapatutako arrain-espezieen parekotasunaren neurria gutxi gorabehera?
- (a) 2.7
 (b) 1.8
 (c) 1.2
 (d) 0.9
12. Honako Lorenz kurbetatik, haien puntuen arabera definituta, zein da zuzena edo akatsik ez duena?
- (a) $(p_1 = 0.1, q_1 = 0.02); (p_2 = 0.2, q_2 = 0.05); (p_3 = 0.3, q_3 = 0.10)$
 (b) $(p_1 = 0.1, q_1 = 0.03); (p_2 = 0.2, q_2 = 0.07); (p_3 = 0.3, q_3 = 0.10)$
 (c) $(p_1 = 0.2, q_1 = 0.14); (p_2 = 0.4, q_2 = 0.38); (p_3 = 0.6, q_3 = 0.58)$
 (d) $(p_1 = 0.2, q_1 = 0.22); (p_2 = 0.4, q_2 = 0.32); (p_3 = 0.6, q_3 = 0.52)$
13. Honako Lorenz kurbetatik, ondoren zehazten diren haien puntuen arabera definituta (batekotan), $(0,0)$ puntuan hasiz $(1,1)$ puntuak bukatuz, zein da banaketa parekoena (kontzentrazio txikieneko) ematen duena?
- (a) $(p_1 = 0.1, q_1 = 0.02); (p_2 = 0.2, q_2 = 0.05); (p_3 = 0.3, q_3 = 0.20)$
 (b) $(p_1 = 0.1, q_1 = 0.02); (p_2 = 0.2, q_2 = 0.06); (p_3 = 0.3, q_3 = 0.20)$
 (c) $(p_1 = 0.1, q_1 = 0.04); (p_2 = 0.2, q_2 = 0.09); (p_3 = 0.3, q_3 = 0.20)$
 (d) $(p_1 = 0.1, q_1 = 0.04); (p_2 = 0.2, q_2 = 0.11); (p_3 = 0.3, q_3 = 0.20)$
14. Inkesta batean jasotako datuen arabera, immigranteetatik 22 ezkondu daude eta 14 ezkondu gabe. Bertakoen artean, 16 ezkondu daude, eta 20 ezkondu gabe. Lambda kalkulatzera, ezkontza egoera aldagai independentetzat harturik, zenbat dira erroreak aldagai independente hori kontuan harturik?
- (a) 30
 (b) 34
 (c) 36
 (d) 42

15. Puntu-hodei batean, x eta y aldagaiei buruzko puntu edo datu-bikote hauek ditugu: $(x,y):(0,2),(2,0),(1,1)$. Zenbat da kobariantza?
- (a) -0.66
 - (b) -0.33
 - (c) 0.33
 - (d) 0.66
16. $P[X = x] = \frac{x}{15}; x = 1, 2, 3, 4, 5$. Zenbat da $F(x = 3)$?
- (a) 0.2
 - (b) 0.3
 - (c) 0.4
 - (d) 0.5
17. Zorizko aldagai jarraitu batean $F(x) = \frac{x^2}{9}; 0 \leq x \leq 3$. Zenbat da $P[1 < X < 2]$?
- (a) 0.22
 - (b) 0.33
 - (c) 0.44
 - (d) Aurreko guztiak faltsuak dira.
18. Inbertsio baten errendimendua honela banatzen da: $F(x) = \frac{x^2}{9}; 0 \leq x \leq 3$. Zenbat $\text{VaR}(0.05)$?
- (a) 0.67
 - (b) 0.87
 - (c) 1.07
 - (d) Aurreko guztiak faltsuak dira.
19. A eta B inbertsioren itzaropenak eta bariantzak estimatu dira: $\mu_A = 10; \sigma_A = 2; \mu_B = 15; \sigma_B = 4$. Utilitate-funtzio gisa, μ/σ erabili.
- (a) Epe luzera hobe da B, baina epe laburrera A.
 - (b) Epe luzera hobe da A, baina epe laburrera B.
 - (c) Epe laburrera nahiz luzera hobe da A.
 - (d) Epe laburrera nahiz luzera hobe da B.
20. $r_{xy} = 0.8; r_{xz} = 0.7; r_{yz} = 0.6$. Zenbat da gutxigorabehera y eta z aldagaien arteko Pearson korrelazio koefizientea, x aldagaiaren efektua deuseztuz edo baztertuz?
- (a) 0.093
 - (b) 0.393
 - (c) 0.593
 - (d) 0.793

Estatistika eta datuen analisia

2023ko urtarrilaren 23a

Izena eta abizenak: Josemari Sarasola

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

KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0

ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2023ko ekainaren 23a, 16:00

Iraupena: 90 min

I. ebazkizuna (3 puntu 10etik)

1995ean Jaizkibelgo ipar eta hego maldetan landatutako pinudi bateko pinuen enborreko perimetroa jaso da (cm), 150cm-ko altueran. Hona hemen jasotako datuak:

Perimetroa	Pinuak (iparreko malda)	Pinuak (hegoko malda)
100-110	37	22
110-120	44	34
120-130	65	66
130-140	36	48
140-150	18	30
	200	200

Egin beharreko atazak:

- Desbideratze estandarrak kalkulatu ipar nahiz hegoko pinuen perimetroetarako.
- Desbideratze estandar horietatik abiaturik, azter ezazu zein maldatan dira perimetroak sakabanatuen.
- Kalkulatu mediana bi pinu multzoetarako, eta mediana horien arteko aldea interpretatu.

(a)

x	n_I	n_H	$n_I x$	$n_H x$	$n_I x^2$	$n_H x^2$	N_I	N_H
105	37	22	3885	407925	2310	242550	37	22
115	44	34	5060	581900	3910	449650	81	56
125	65	66	8125	1015625	8250	1031250	146	122
135	36	48	4860	656100	6480	874800
145	18	30	2610	378450	4350	630750
	200	200	24540	3040000	25300	3229000		

$$\bar{x}_I = \frac{24540}{200} = 122.7cm ; \bar{x}_H = \frac{25300}{200} = 126.5cm$$

$$s_{x_I} = \sqrt{\frac{3040000}{200} - 122.7^2} = \sqrt{144.7} = 12.03cm ; s_{x_H} = \sqrt{\frac{3229000}{200} - 126.5^2} = \sqrt{142.75.7} = 11.94cm$$

Iparreko pinu bakoitza 12.03cm desbideratzen da 122.7ko batezbestekotik; Hegoko pinu bakoitza, berriz, 11.94cm desbideratzen da 126.5eko batezbestekotik.

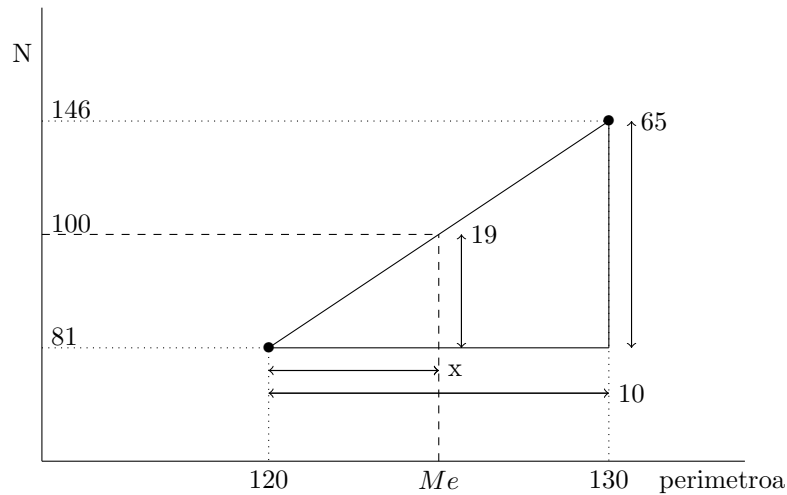
(b)

Sakabanatzeak alderatzeko neurri erlatiboa behar da; kasu honetan, desbideratzeak erarortzen den neurri erlatiboa aldakuntza-koefizientea da:

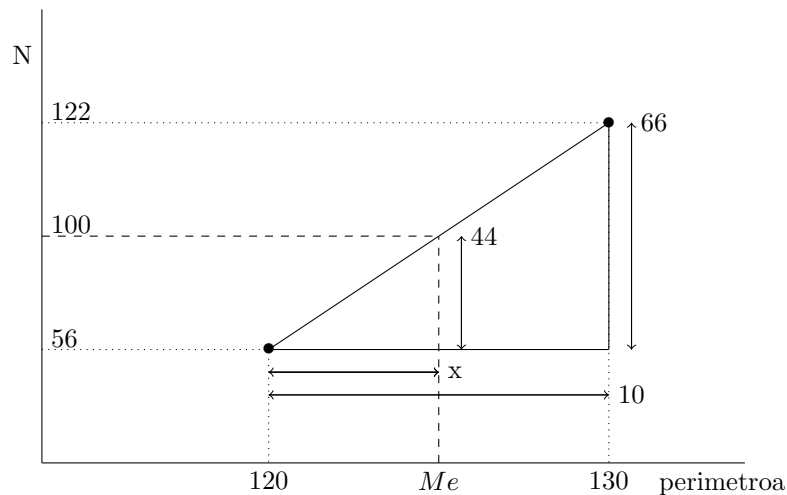
$$A_{x_I} = \frac{12.03}{122.7} = 0.098 = \%9.8 ; A_{x_H} = \frac{11.94}{126.5} = 0.094 = \%9.4$$

Iparreko pinuak dira portzentajearen batezbestekotik gehien desbideratzen direnak (%9.8, hain zuzen). Beraz, iparreko zuhaitzak dira sakabanatuena perimetroari buruz, logikoa ere bada bestalde, kondizio meteorologiko ezegonkorragoak dituztelako. Dena den, diferentzia nahiko txikia dela dirudi.

(c)



$$\frac{x}{19} = \frac{10}{65} \rightarrow x = 2.92 \rightarrow Me = 122.92$$



$$\frac{x}{44} = \frac{10}{66} \rightarrow x = 6.66 \rightarrow Me = 126.66$$

Hegoko pinuek perimetro handiagoa dute orokorrean, zehatzago 126.66-122.92=3.74cm handiagoa. Diferentzia ez dirudi oso handia denik, lagin errearen erreserbapean.

II. ebazkizuna (2.5 puntu 10etik)

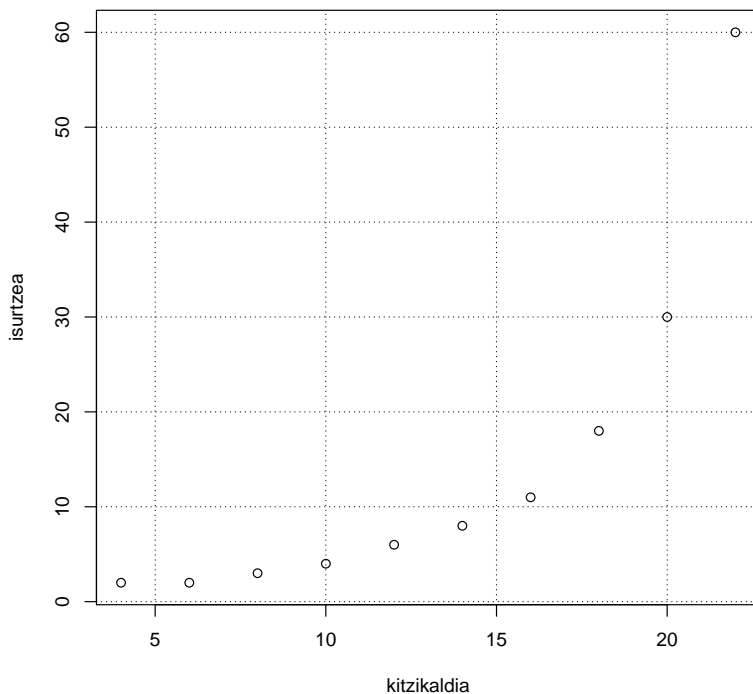
20-24 urte bitarteko emakume heterosexual batzuei euren ohiko bikotekidearekin hurrengo sexu harremanean kitzikatze edo eszitazio faseko denbora modulatzeko eskatu zaie, bakoitzari kitzikatze-denbora bat esleituz, ondoren sarpena gertatzen denetik bikotekideak isuri arteko denbora jasotzeko (homogeneotasun arrazoiengatik, kondoirik gabeko sarpenak bakarrik jasotzen dira). Hona hemen aurrez jarritako kitzikaldietarako lortutako datuak (parentesien arteko datuak adierazten du koitua denbora horretan isuri gabe bukatu zela -azken datua, hain zuzen- eta marren artekoak geroago izan zela, *coitus interruptus* baten bitartez -azken aurreko datua-):

Kitzikaldia (min)	4	6	8	10	12	14	16	18	20	22
Isuri arteko denbora	2	2	3	4	6	8	11	18	-28-	(40)

Egin beharreko atazak:

- Datuak puntu-hodei batean irudikatu itzazu, *coitus interruptus* kasuan agertzen den denbora baino handiagoa hartuz gutxi gorabehera eta isurtzea gertatu ez zen kasuan oso balio handia hartuz. Esango zenuke grafikoa ikusita bi aldagaien arteko erlazioa garbia eta sendoa dela? Gomendatuko zenuke, beraz, koitua luzatzeko kitzikaldi luzea izatea?
- Datu horiek hartuta, espermaren isurtzea koituan zehar gertatu ez zen kasuetarako arestian aipatutako denbora batzuk esleituz betiere, Pearson korrelazio koefizientea 0.28 suertatu zen. Nola esplikatuta daiteke horren balio txikia?
- Korrelazio-koefizienteak korrelazioaren ageriko eta egiazko sendotasuna adieraz dezan egin beharreko aldaketak eta kalkuluak egin itzazu. Emaitza interpretatu. Nola deitzen zaio horrela kalkulaturiko koefiziente berriari? Laguntza: izenak badu antza enuntziatuan agertzen den hitz batekin.

(a)



Isuri arteko denborarako coitus interruptusaren kasuan 30min-ko balioa ezarri da, eta isurtzerik gabeko kasuan 60min. Bestelakoak ere izan litezke.

Korrelazio lerromakur sendoa dago, positiboa, aldagai batak gora egitean besteak ere gora egiten du. Beraz, koitua luzatzeko bai gomendatzen dela kitzikaldi luzea.

(b)

Pearson korrelazio-koefizienteak korrelazio lineala neurtzen du, eta ez kasu honetan oso nabarmena korrelazio makurra. Kasu honetan, garbi dago puntuak ez diotela lerro zuzen bati jarraitzen eta horregatik ateratzen da koefiziente lineala txikia.

(c)

Korrelazio lerromakurraren sendotasuna eta norabidean neurtzeko egokiena, datuen ordeztatu multzoan haiek dituzten heinak edo ordenak hartzea da, eta haiekin Pearson-korrelazio koefizientea kalkulatzeko. Era honetan kalkulaturiko korrelazio-koefizienteari **Spearman-en korrelazio koefiziente** (Spearman-en rho: ρ) deritzen.

Presta ditzagun datuak eta egin ditzagun kalkuluak, zehaztu gabeko datuen ordeztatu enuntziatua adierazi bezala beste balio batzuk esleituz:

x	4	6	8	10	12	14	16	18	20	22	
y	2	2	3	4	6	8	11	18	30	60	
$X = m(x)$	1	2	3	4	5	6	7	8	9	10	55
$Y = m(y)$	1.5	1.5	3	4	5	6	7	8	9	10	55
XY	1.5	3	9	16	25	36	49	64	81	100	384.5
X^2	1	4	9	16	25	36	49	64	81	100	385
Y^2	2.25	2.25	9	16	25	36	49	64	81	100	384.5

$$\bullet \bar{X} = \frac{55}{10} = 5.5 ; \bar{Y} = \frac{55}{10} = 5.5$$

$$\bullet s_{XY} = \frac{384.5}{10} - 5.5 \times 5.5 = 8.2$$

$$\bullet s_X = \sqrt{\frac{385}{10} - 5.5^2} = \sqrt{8.2}$$

$$\bullet s_Y = \sqrt{\frac{384.5}{10} - 5.5^2} = \sqrt{8.25}$$

$$\bullet \rho = r_{XY} = \frac{8.2}{\sqrt{8.2} \times \sqrt{8.25}} = 0.9969$$

Korrelazio lerromakurra positiboa da (kitzikaldia zenbat eta luzeagoa, isuri arteko denbora orduan eta luzeagoa) eta sendotasun handikoa (> 0.6), laginketa erroreaken erreserbapean.

III. ebazkizuna (2.5 puntu 10etik)

Inbertsio bat egiteko bi akzio daude aukeran. A akzioko eguneko kotizazioak -1, 0, +1 edo +2 aldaketak izan ditzake egunean 0.1, 0.2, 0.4 eta 0.3 probabilitateekin hurrenez hurren. B akzioko eguneko kotizazioak, berriz, 0, +1, +2 edo +3 egin dezake egunean, probabilitate hauen arabera: 0 eta +1 izateko probabilitateak berdinak dira, eta +2 eta +3 izatekoak ere bai, eta aldi berean 0 eta +1 izateko probabilitateak +2 eta +3 izateko probabilitateen erdiak dira.

Egin beharreko atzak: (Oharra: beharrezkoa denean μ/σ utilitate-funtzioa baliatu.)

- B akzioan 0, +1,+2, +3 izateko probabilitateak zehaztu, arestian emandako informazioaren arabera.
- Epe luzera zein da akzio komenigarriena? Erantzuna arrazoitu.
- Eta epe laburrera? Erantzuna arrazoitu. Epe laburrerako erabakian, zein irizpide gailendu da azkenean: batez besteko errentagarritasuna ala arriskua?

a

$$P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] = p/2 + p/2 + p + p = 1 \rightarrow p = 1/3 = 0.33$$

Beraz, B inbertsioan:

$$P[X = 0] = P[X = 1] = 0.166 ; P[X = 2] = P[X = 3] = 0.333$$

b

x_A	$p(x_A)$	$x_A p(x_A)$	$x_A^2 p(x_A)$	x_B	$p(x_B)$	$x_B p(x_B)$	$x_B^2 p(x_B)$
-1	0,1	-0,1	0,1	0	0,1666	0	0
0	0,2	0	0	1	0,166	0,166	0,166
1	0,4	0,4	0,4	2	0,333	0,666	1,333
2	0,3	0,6	1,2	3	0,333	1	3
	1	0,9	1,7		1	1,833	4,5

$$\mu_A = 0.9 ; \mu_B = 1.83$$

Epe luzera μ itzaropenari erreparatu behar zaio, eta horren arabera B inbertsioa da hobea, batez beste etekin handiagoak ematen dituelako.

c

Epe laburrera **arriskua ere** kontuan hartu behar da, **itzaropenaz gainera**, eta hori neurtzeko bariantza kalkulatu dugu:

$$\sigma_A^2 = 1.7 - 0.9^2 = 0,89 ; \sigma_B^2 = 4.5 - 1.833^2 = 1,139$$

Beraz, epe laburrera:

- itzaropenari buruz, nahiago dugu B inbertsioa;
- arriskuari (bariantzari) buruz, nahiago dugu A inbertsioa, bariantza txikiena duena alegia.

Dilema sortzen da, eta dilema ebazteko utilitate-funtzioa baliatzen dugu:

$$U_A = \frac{0.9}{\sqrt{0.89}} = 0.954 ; U_B = \frac{1.83}{\sqrt{1.139}} = 1,714$$

Epe laburrera utilitate handiagoa ematen du B inbertsioa; hartara, horixe aukeratzen dugu.

Bide batez, B inbertsioa batezbesteko edo itzaropen handiena duena denez, irizpide hori gailentzen zaio beste irizpideari, arriskuari alegia.

ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2023ko ekainaren 23a, 16:00

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. XIX. mendean estatistika giza eta gizartet zientzietara zabaltzen hasi zen. Nor izan zen prozesu horretako abiapuntu gisa *batez besteko gizakiaren ideia* asmatu zuena?
 - (a) John Graunt
 - (b) Friedrich Gauss
 - (c) John Galton
 - (d) Adolphe Quetelet
2. Zeintzuk dira datu eta emaitza estatistikoetan dagoen aldakortasunaren eta errorearen sorburu nagusiak?
 - (a) Inkestatzatzailearen joera edo alborapena eta ikerketa-eremuaren zehazgabetasuna.
 - (b) Fenomenoen zorizkotasuna eta laginaren errorea, emaitzak populazioetara zabaltzean.
 - (c) Datu-multzoak batezbesteko eta bestelako estatistikoen balioekin ordezteak.
 - (d) Neurketak eta datuak une desberdinetan pertsona desberdinek jasotzean sortzen diren gorabeherak.
3. Zein grafiko eratortzen da bestetik zuzenean?
 - (a) Barra-diagramatik, histograma.
 - (b) Histogramatik, maiztasun metatuen poligonoa
 - (c) Maiztasun metatuen poligonotik, ojiba.
 - (d) (b) eta (c) dira egiazkoak.
4. 40 urteko 2000 gizonen altuerak jaso dira. Datuak zabalera desberdineko tartetan bildu dira. 170-175 tartean 500 gizon daude. Zenbatekoa da tarte horri dagokion zutabearen altuera histograman?
 - (a) 0.025
 - (b) 0.05
 - (c) 0.1
 - (d) 0.25
5. Enpresa batean salmentak 2000 izan ziren 2019 urtean, eta 3000 2022 urtean. Zenbat da salmenten batez besteko hazkundera urteko?
 - (a) %14.4
 - (b) %15.4
 - (c) %16.4
 - (d) %17.4
6. 20 urteko 400 mutilen zakilen luzerak erabateko erekzioan jaso dira. 18cm-20cm azken tartean, 50 mutil daude. Zein luzeratik gora dute zakilen %10 luzeenak?
 - (a) 18.2cm
 - (b) 18.4cm
 - (c) 18.6cm
 - (d) 18.8cm
7. Zeintzuk dira sakabanatze-neurri sendoak?
 - (a) IQR eta DAME
 - (b) Desbideratze estandarra eta bariantza
 - (c) Ibiltartea eta kuartil arteko ibiltartea
 - (d) Ibiltartea eta DAME

8. 10 daturen populazio bariantza 49 da. Zenbat da desbideratze estandar zuzendua edo laginaren desbideratze estandarra?
- 6.640
 - 7
 - 7.111
 - 7.378
9. Errenta-datuak pertsonako: 12-15-18-27-45. Pobrezia-muga: 20. Kalkulatu pobreziaren intentsitatea:
- 0.15
 - 0.20
 - 0.25
 - 0.30
10. Lorenz kurba bateko hasierako puntuak ematen zaizkizu: $(p_1 = 0.1, q_1 = 0.02)$; $(p_2 = 0.2, q_2 = 0.06)$. Zer esan dezakezu ahalik eta zehazkien $(p_3 = 0.3, q_3)$ puntuko q_3 balioari buruz?
- Ezin da izan 0.06 baino txikiagoa.
 - Ezin da izan 0.10 baino txikiagoa.
 - Ezin da izan 0.12 baino txikiagoa.
 - Ezin da izan 0.14 baino txikiagoa.
11. Kontzentrazioari buruz, hauetatik zein da faltsua?
- Gini indizea 0.6tik gorakoa denean, kontzentrazioa handia edo sendoa dela esan daiteke.
 - Bi Lorenz kurba ezin dira ebaki edo gurutzatu puntu batean baino gehiagotan.
 - Banaketa berdintsu batean, mediana eta mediala berdinak dira.
 - Lau pertsonaren arteko banaketa batean batek dena hartzen badu eta besteek hutsa, Robin Hood adierazlea %75 da.
12. Dibertsitateari buruz, zein da horren osagaia den parekotasuna kalkulatzeko formula?
- H
 - e^H
 - S
 - e^H/S
13. 2x2 kontingentzia-taula batean, lagin-tamaina 100 da eta diagonal nagusian dauden gelasketako maiztasun absolutuak 2 dira. Ez dakigu besterik. Zer da egia errore kopuruari buruz? Laguntza: muturreko egoerak aztertu erantzun egokia aurkitzeko.
- Aldagai independentea kontuan harturik 4 errore izan daitezke, eta kasu horretan aldagai independentea kontuan hartu gabe ere 4 izango dira.
 - Aldagai independentea kontuan harturik 2 errore izan daitezke, eta kasu horretan aldagai independentea kontuan hartu gabe 4 errore izango dira.
 - Aldagai independentea kontuan harturik 2 errore izan daitezke, eta kasu horretan aldagai independentea kontuan hartu gabe ere 2 errore izango dira.
 - Aldagai independentea kontuan harturik 0 errore izan daitezke, eta kasu horretan aldagai independentea kontuan hartu gabe 4 errore izango dira.
14. 2x2 kontingentzia-taula batean, lehen zutabeko bazter-maiztasuna 4 da. Lagin-tamaina 16 da. Ez dakigu besterik. Zer esan daiteke ahalik eta zehazkien lehenengo zutabeari eta lehenengo errenkadari dagokion gelaskako maiztasun teorikoari buruz?
- 0 da gutxienez eta 0.25 da gehienez.
 - 0 da gutxienez eta 1 da gehienez.
 - 1 da gutxienez eta 2 da gehienez.
 - 1 da gutxienez eta 4 da gehienez.

15. $F(x) = 0.4 + 0.15x$; $x = 1, 2, 3, 4$. Kalkulatu $P[X = 2]$.
- (a) 0.15
 - (b) 0.55
 - (c) 0.60
 - (d) Aurreko guztiak dira faltsuak.
16. $F(x) = x^2/100$; $0 \leq x \leq 10$. Kalkulatu $P[4 < X < 6]$.
- (a) 0.12
 - (b) 0.16
 - (c) 0.20
 - (d) 0.36
17. $f(x) = x/k$; $0 < x < 10$. Kalkulatu k dentsitate-funtzioa zan dadin.
- (a) 5
 - (b) 10
 - (c) 20
 - (d) 50
18. Zorizko aldagai batek 1 eta 2 balioak har ditzake probabilitate berdinekin. Kalkula ezazu desbideratze estandarra.
- (a) 0.05
 - (b) 0.25
 - (c) 0.5
 - (d) 1
19. Eguneko ekoizpena 1 edo 2 izan daiteke probabilitate berdinekin. 1 izan bada, hurrengo egunean 2 edo 3 izan daiteke probabilitate berdinekin. 2 izan bada, hurrengo egunean 0 edo 1 izan daiteke probabilitate berdinekin. Zenbat da bi egun horietarako batez besteko ekoizpen totala?
- (a) 1.5
 - (b) 2
 - (c) 2.5
 - (d) 3
20. Zein da faltsua itxaropen matematikoari buruz?
- (a) Finkoa da.
 - (b) Epe luzera betetzen den batez besteko balioa da.
 - (c) Ezezaguna izaten da praktikan, eta beraz zenbatetsi egin behar da.
 - (d) Batezbestekoaren estimatzaile gisa erabili ohi da.

Estatistika eta datuen analisia

-
2023ko ekainaren 23a**Izena eta abizenak: Josemari Sarasola**

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

KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0