

ESTADÍSTIKA ETA DATUEN ANALISIA

Azterketa ebatziak.

2016-2017 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea. EHU.

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

gizapedia.hirusta.io

ESTATISTIKA ETA DATUEN ANALISIA (Praktika)

Data: 2017ko urtarrilaren 26a - Iraupena: Ordu t'erdi

Praktika: Adierazburu-orria

I. EBAZKIZUNA (3.25 puntu)

KM kulturuneko azken bi erakusketetan eguneko bisitari-kopuruak jaso dira:

A erakusketa: 44-67-86-102-45-54-78-56-92-75

B erakusketa: 66-77-68-86-97-106-92-89-96-82-79-83

Egin beharreko atazak:

1. Bi erakusketei dagoekin kaxa-diagramak marraztu eta interpretatu, zentroari nahiz sakabanatzeari buruz.
2. Kaxa-diagrama interpretatzeko hartu duzun sakabanatzearen erreferentzian oinarrituta, bi erakusketei dagokien sakabanatze erlatiboaren koefizientea kalkulatu eta interpretatu.
3. B erakusketa oraindik zabalik dago. Fotokopia-sorta berezi bat prestatu da erakusketan gidatzeko. Zenbat sorta prestatu behar dira egunean bisitarien %80ek sorta bana hartzeko aukera izan dezan?
4. Kalkulatu A erakusketako eguneko bisitari kopuruaren lagin-bariantza eta populazio-bariantza.

II. EBAZKIZUNA: Asoziazioa eta logit eredua (4.75 puntu)

Enpresa batek bizikleta-modelo berezi bat merkaturatu zuen duela urtebete, bateria elektrikoarekin zein bateriarik gabe. Bateriadun 400 bizikleta saldu zituen eta bateriarik gabe 800. Urtebete pasata, enpresak galdera-sorta bat bidali die bizikleta erosi zuten bezeroei, eta besteak beste epe horretan inoiz egun batean 50 km-tik gora egin duten galdetu zaie. Bateriadun bizikleta erosi zutenen artean 200 lagunek baietz erantzun dute, bateriarik gabe erosi zutenen artean 220 izan dira baieztatu erantzuna eman dutenak.

Egin beharreko atazak:

1. Khi-karratua oinarritutako neurriren bat erabiliz, bizikletak bateria izatearen artean eta txirrindulariak ibilbide luzeak egitearen artean asoziazioaren sendotasuna kalkulatu.
2. Aurreko bi aldagaien artean lambda kalkulatu eta interpretatu.
3. Zein da aurreko bi aldagaien arteko asoziazioaren norabidea?
4. Ahal bada, Goodman eta Kruskalen gamma kalkulatu eta interpretatu.
5. Ahal bada, bi aldagaien arteko Pearson korrelazio-koefizientea kalkulatu eta interpretatu.
6. Bateria beteta zenbat kilometro egin dituzten ere galdetu zaie bezeroei. 20 km-ra bizikleten %90ek ibiltzen jarraitzen zuten, 40 km-ra %50ek eta 60 km-ra %20k. Logit eredua erabiliz, aurrean ezazu zenbat bizikleta ibiliko liratekeen 50 km egin eta gero.

ESTATISTIKA ETA DATUEN ANALISIA (Praktika)

Data: 2017ko urtarrilaren 26a - Iraupena: Ordu t'erdi

Ebazpena

I. EBAZKIZUNA (3.25 puntu)

KM kulturuneko azken bi erakusketetan eguneko bisitari-kopuruak jaso dira:

A erakusketa: 44-67-86-102-45-54-78-56-92-75

B erakusketa: 66-77-68-86-97-106-92-89-96-82-79-83

Egin beharreko atazak:

1. Bi erakusketei dagoekin kaxa-diagramak marraztu eta interpretatu, zentroari nahiz sakabanatzeari buruz.
2. Kaxa-diagrama interpretatzeko hartu duzun sakabanatzearen erreferentzian oinarrituta, bi erakusketei dagokien sakabanatze erlatiboaren koefizientea kalkulatu eta interpretatu.
3. B erakusketa oraindik zabalik dago. Fotokopia-sorta berezi bat prestatu da erakusketan gidatzeko. Zenbat sorta prestatu behar dira egunean bisitarien %80ek sorta bana hartzeko aukera izan dezan?
4. Kalkulatu A erakusketako eguneko bisitari kopuruaren lagin-bariantza eta populazio-bariantza.

(a)
Lehen pausoa bi datu-multzoak ordenatzea da, txikienetik handienra, kuartilak kalkulatu ahal izateko:

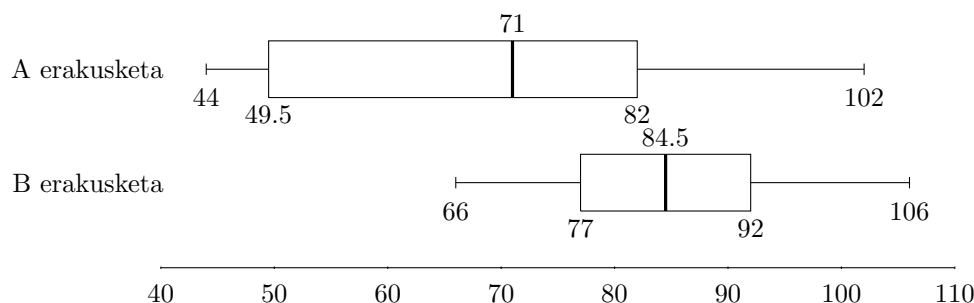
$$A_{ord.} : 44 - 45 - 54 - 56 - 67 - 75 - 78 - 86 - 92 - 102$$

$$B_{ord.} : 66 - 68 - 77 - 79 - 82 - 83 - 86 - 89 - 92 - 96 - 97 - 106$$

Zer	A	B
Me	$67+75/2=71$	$83+86/2=84.5$
Q_3	82	92
Q_1	49.5	77
$Q_3 - Q_1$	32.5	15
$1.5(Q_3 - Q_1)$	48.75	22.5
$Q_3 + 1.5(Q_3 - Q_1)$	130.75	114.5
$Q_1 - 1.5(Q_3 - Q_1)$	0.75	54.5
albo balioak goitik	102	106
albo balioak behetik	44	66
outlier behetik	-	-
outlier goitik	-	-

Kalkuluak:

- A erakusketa
 $Q_3? 0.75 \times 10 = 7.5gn. \text{ datua} \rightarrow Q_3 = 78 \times (1 - 0.5) + 86 \times 0.5 = 82$
 $Q_1? 0.25 \times 10 = 2.5gn. \text{ datua} \rightarrow Q_1 = 45 \times (1 - 0.5) + 54 \times 0.5 = 49.5$
- B erakusketa
 $Q_3? 0.75 \times 10 = 7.5gn. \text{ datua} \rightarrow Q_3 = 78 \times (1 - 0.5) + 86 \times 0.5 = 82$
 $Q_1? 0.25 \times 10 = 2.5gn. \text{ datua} \rightarrow Q_1 = 45 \times (1 - 0.5) + 54 \times 0.5 = 49.5$



Zentroari dagokionean, B erakusketara orokorrean bisitari gehiago joaten dela ikus daiteke, medianari erreparatuta, $84.5-71=13.5$ bisitariko aldearekin.

Sakabanatzeari dagokionean, A erakusketan bisitari kopuruak sakabanatuagoak dira, kaxaren zabalerari erreparatuta.

(b)

Sakabanatzea interpretatzeko $IQR = Q_3 - Q_1$ neurrian, kaxaren zabaleran alegia, oinarritu garez, bi datu-multzoak erlatiboki alderatzeko IQR/Me kalkulatu beharko da:

$$IQR/Me_A = \frac{32.5}{71} = 0.45$$

$$IQR/Me_B = \frac{15}{84.5} = 0.17$$

Hartara, modu erlatiboan ere, sakabanatze handiena A erakusketan gertatzen da.

(c)

P_{80} , 80garren pertzentila, kalkulatu behar da:

P_{80} ? $0.8 \times 12 = 9.6gn. \text{ datua} \rightarrow P_{80} = 92 \times (1 - 0.6) + 96 \times 0.6 = 94.4 \rightarrow 95$ sorta prestatu beharko dira helburua betetzeko.

(d)

- Populazio-bariantza: $s_x^2 = \frac{\sum x^2}{n} - \bar{x}^2$
- Lagin-bariantza: $\hat{s}_x^2 = \frac{n}{n-1} s_x^2$

x_i	x_i^2
44	1936
45	2025
54	2916
56	3136
67	4489
75	5625
78	6084
86	7396
92	8464
102	10404
699	52775

$$s_x^2 = \frac{52475}{10} - 69.9^2 = 361.49$$

$$\hat{s}_x^2 = \frac{10}{10-1} \times 361.49 = 401.65$$

II. EBAZKIZUNA: Asoziazioa eta logit eredia (4.75 puntu)

Enpresa batek bizikleta-modelo berezi bat merkaturatu zuen duela urtebete, bateria elektrikoarekin zein bateriarik gabe. Bateriadun 400 bizikleta saldu zituen eta bateriarik gabe 800. Urtebete pasata, enpresak galdera-sorta bat bidali die bizikleta erosi zuten bezeroei, eta besteak beste epe horretan inoiz egun batean 50 km-tik gora egin duten galdetu zaie. Bateriadun bizikleta erosi zutenen artean 200 lagunek baietz erantzun dute, bateriarik gabe erosi zutenen artean 220 izan dira baiezeko erantzuna eman dutenak.

Egin beharreko atazak:

1. Khi-karratua oinarritutako neurriren bat erabiliz, bizikletak bateria izatearen artean eta txirrindulariak ibilbide luzeak egitearen artean asoziazioaren sendotasuna kalkulatu.
2. Aurreko bi aldagaien artean lambda kalkulatu eta interpretatu.
3. Zein da aurreko bi aldagaien arteko asoziazioaren norabidea?
4. Ahal bada, Goodman eta Kruskalen gamma kalkulatu eta interpretatu.
5. Ahal bada, bi aldagaien arteko Pearson korrelazio-koefizientea kalkulatu eta interpretatu.
6. Bateria beteta zenbat kilometro egin dituzten ere galdetu zaie bezeroei. 20 km-ra bizikleten %90ek ibiltzen jarraitzen zuten, 40 km-ra %50ek eta 60 km-ra %20k. Logit eredia erabiliz, auresan ezazu zenbat bizikleta ibiliko liratekeen 50 km egin eta gero.

(a)

Osa dezagun kontingentzia-etaula:

		50 km baino gehiago?		
		Ez	Bai	
Bateria?	Ez	580	220	800
	Bai	200	200	400
		780	420	1200

Orain, taulako maiztasun teorikoak, khi-karratuko ekarpenak (adibidez, lehen gelaskakoa: $780 \times 800/1200 = 520$) eta horien erroak, hurrenez hurren, ematen dira:

		50 km baino gehiago?		
		Ez	Bai	
Bateria?	Ez	520-6.92-2.63	280-12.86-(-3.59)	800
	Bai	260-13.8-(-3.72)	140-25.7-5.07	400
		520	680	1200

Hartara,

$$X^2 = 6.92 + 12.86 + 13.8 + 25.7 = 59.28$$

Taula 2×2 denez, phi da sendotasuna zehazteko koefiziente egokia:

$$\varphi = \sqrt{\frac{59.28}{1200}} = 0.22$$

0.3 baino txikiagoa denez, bi aldagaien arteko asoziazioa ahula dela esan daiteke, lagin errorearen erreserbapean eta antzeko ikerketak kontuan hartu gabe.

(b)

Lambda kalkulatzeko aldagai independentea eta dependentea zehaztu behar dira:

- aldagai independentea: bateria? bai/ez
- aldagai dependentea: 50 kilometro? bai/ez

- Aldagai independentea kontuan hartu gabe:
 - auresana: 50 km ez (780 handiago 420)
 - erroreak: 420

- Aldagai independentea kontuan hartuta:
 - bateria badu,
 - auresana: 50 km bai (edo ez) (200=200)
 - erroreak: 200

- bateria ez badu,
 - auresana: 50 km ez (580 handiago 220)
 - erroreak: 220
- aldagai independentea kontuan hartuta, $200+220=420$ errore guztira.

Beraz, aldagai independentea kontuan hartuta, guztira $420-420=0$ errore gutxiago egiten dira. Eta hortik:

$$\lambda = \frac{0}{420} = 0$$

Hartara, lambdaaren arabera bi aldagaien artean ez dago asoziaziorik inondik ere.

(c)

Asoziaziorik ez dagoenez, edo izatekotan ahila denez, ez du zentzurik norabidea aztertzea, baina norabiderik izatekotan, bateria edukita 50 km egiteko aukera handiagoak daudela aipatu behar da. Horretarako, khi-karratuko ekarpenen erroei erreparatu behar zaie eta horietan balio handienari eta txikienari erreparatu: balio handiena bateria bai/50 km bai gelaskan suertatzen denez, lotura nabarmenena hor dagoela esan daiteke.

(d) Bi aldagaiak ordinaltzat har daitezke, "bai" "ez" baino gehiago dela pentsatuz. Hasierako taula ordena horri jarraiki dugunez, konkordantziak eta diskordantziak zuzenean eman ditzakegu taula aldatu gabe:

- konkordantziak: $580 \times 200 = 116000$
- diskordantziak: $220 \times 200 = 44000$

$$\gamma = \frac{116000 - 44000}{116000 + 44000} = 0.45$$

Beraz, gammaren arabera erlazio estatistikoa ertaina da, 0.3-0.6 tartean dagoenez, lagin errorearen eta antzeko ikerketen erreserbapean. Positiboa denez, bateria edukitzeak 50 km baino gehiago egiteko aukera handiagoak ematen dituela ondorioztatzen da.

(e)

Korrelazio-koefiziente lineala kalkulatzen, datuak kodetu egin behar ditugu. Ohiko eran egingo dugu:

- Bateria? $ez \rightarrow 0, bai \rightarrow 1$
- 50 km? $ez \rightarrow 0, bai \rightarrow 1$

Hortik,

x	y	n	nx	ny	nx^2	ny^2	nxy
0	0	580	0	0	0	0	0
0	1	220	0	220	0	220	0
1	0	200	200	0	200	0	0
1	1	200	200	200	200	200	200
		1200	400	420	400	420	200

$$\bar{x} = \frac{400}{1200} = 0.33$$

$$\bar{y} = \frac{420}{1200} = 0.35$$

$$s_x^2 = \frac{400}{1200} - 0.33^2 = 0.47$$

$$s_y^2 = \frac{420}{1200} - 0.35^2 = 0.48$$

$$s_{xy} = \frac{200}{1200} - 0.33 \times 0.35 = 0.05$$

$$r_{xy} = \frac{0.05}{\sqrt{0.47}\sqrt{0.48}} = 0.11$$

Positiboa denez, bateria edukitzeak (0tik 1era pasatzeak) 50 km baino gehiago egiteko aukerak (0tik 1era) handitzen ditu. Korrelazio hori ahula da, ordea, 0.3 baino txikiagoa denez, betiere lagin errorearen eta antzeko ikerketen erreserbapean.

(f)

- x : distantzia
- p : ibiltzen ez jarraitzeko probabilitatea

Logit ereduak garatzeko, beharrezkoa probabilitateak dosiarekin batera gora egitea. Horretarako, 50 kmra ibiltzen jarraitzeko probabilitatearen ordean, ibiltzen ez jarraitzeko probabilitatea hartu behar da.

x (distantzia)	p	$y = \ln \frac{p}{1-p}$	x^2	xy
20	0.1	-2.2	400	-44
40	0.5	0	1600	0
60	0.8	1.39	3600	83.4
120		-0.81	5600	39.4

$$\bar{x} = \frac{120}{3} = 40$$

$$\bar{y} = \frac{-0.81}{3} = -0.27$$

$$s_x^2 = \frac{5600}{3} - 40^2 = 266.67$$

$$s_{xy} = \frac{39.4}{3} - 40 \times (-0.27) = 23.9$$

Hartara,

$$b = \frac{s_{xy}}{s_x^2} = \frac{23.9}{266.67} = 0.0896$$

$$a = \bar{y} - b\bar{x} = -3.85$$

Eta estimatutako kurba logistikoa honela geratzen da:

$$\ln \frac{\hat{p}}{1-\hat{p}} = -3.85 + 0.0896x$$

50 kmra ibiltzen jarraitzen duten bizikleten auresana egin dezagun:

$$\ln \frac{\hat{p}}{1-\hat{p}} = -3.85 + 0.0896 \times 50 = 0.63$$

p probabilitatea bakan dezagun:

$$\ln \frac{\hat{p}}{1-\hat{p}} \rightarrow 0.63 \rightarrow \frac{\hat{p}}{1-\hat{p}} = e^{0.63} \rightarrow \hat{p} = \frac{e^{0.63}}{1+e^{0.63}} = 0.65$$

Horrela, 50 kmra ibiltzen ez jarraitzeko probabilitatea 0.65 dela estimatzen da, eta ibiltzen ez jarraitzekoa, beraz, $1-0.65=0.35$.

ESTATISTIKA ETA DATUEN ANALISIA (Testa)

Irakaslea: Josemari Sarasola

Data: 2017ko urtarrilaren 26a

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 0.05 puntu kentzen du. Galderak erantzun gabe utz daitezke, punturik gehitu eta kendu gabe. Praktika baliozkoa izateko, 7 galdera zuzen erantzun behar dira, gutxienez.

- 20 ikasleren kalifikazioen banaketa honela abiatzen da: 0-1 tartean, 2 ikasle; 1-2 tartean, 3 ikasle; 2-3 tartean 6 ikasle, ... Zenbat da mediana?
 - 2.33
 - 2.66
 - 2.83
 - Datuak falta dira kalkulua egiteko.
- Honako neurri hauetatik zein *ez* da neurri jasankorra?
 - Kuartil arteko ibiltartea
 - Ibiltartea
 - Desbideratze estandarra
 - Ibiltartea zein desbideratze estandarra dira jasankorrak ez direnak.
- Honako egoera hauetatik zein da egokiena banakuntza normala eredu gisa baliatzeko?
 - Bowley koefizientea -0.2 da eta Moors koefizientea 3.2.
 - Bowley koefizientea -0.2 da eta Pearson kurtosi-koefizientea 3.2.
 - Fisher koefizientea 0.3 eta Moors koefizientea 3.2.
 - Fisher koefizientea 0.3 eta Pearson kurtosi-koefizientea 3.2.
- Honako egoera hauetatik zein kasutan ez da zilegi kontzentrazioaren azterketa?
 - Gipuzkoako Aldunditik udalek zenbat jasotzen duten.
 - Donostiako ikastetxeetako ikasle-kopuruak.
 - Gipuzkoako museoetarako sarreraren prezioak.
 - Esne-kooperatiba batek bere bazkideengandik jasotako esne-kantitateak.
- Lorenz kurba batetik informazio hau eskuratu da: $p_1 = 0.4, q_1 = 0.1, p_2 = 0.9, q_2 = 0.6$. Zenbat da Palma ratioa?
 - Ezin da kalkulatu emandako datuekin.
 - 6
 - 4
 - 2
- Honako datu-multzo hauetatik zein ez da barra-diagrama batez irudikatu behar?
 - 0-5 urteko 1000 haurren ganean, urtean ospitaleko larrialdi-zerbitzura joandako aldi-kopurua.
 - 200 ikasle unibertsitariak gradu bat bukatzeko egin behar izan duten ikasturte-kopurua.
 - 30 emakumek darabilten antisorgailu-mota.
 - 500 pertsonak hilean telefono mugikorretik hitz egin duten minutu-kopurua.
- Honako hauetatik zein da faltsua?

- (a) Datu gehiago hartzen badira, errealitatea argitu baino, nahasi agertzeko arriskua dago.
- (b) Populazio bati buruz datu guztiak jasota ere, errore estatistikoa egon daiteke, fenomenoaren berezko zorrikotasunagatik.
- (c) Inkesta batean atzemandako heterogeneotasunaren sorburuak ikerketaren denbora-eremuaren definizioezarekin zerikusirik izan dezake.
- (d) Errenta aldagai kuantitatiboa da beti.
8. Zer adierazten dute Pigou-Dalton baldintzek?
- (a) Kontzentrazio-neurriak egokiak izan daitezten haiek bete beharreko baldintzen multzoa.
- (b) Pobrezia murrizteko egin beharrekoak: zergak proportzionalki gehiago igo gehiago dutenei, besteak beste.
- (c) Giniren indizea eta beste kontzentrazio neurriak murrizteko baldintzak.
- (d) Giniren indizea eta beste kontzentrazio neurriak gehitzeko baldintzak.
9. Errenta-banaketa batean informazio hau eskuratu da: $(p_1 = 0.1, q_1 = 0.01)$, $(p_9 = 0.9, q_9 = 0.84)$. Zein da zuzena honako hauetatik?
- (a) %10 aberatsenak errenta osoaren %84a hartzen du.
- (b) %90 aberatsenak errenta osoaren %16a hartzen du.
- (c) %10 aberatsenak errenta osoaren %16a hartzen du.
- (d) %90 aberatsenak errenta osoaren %84a hartzen du.
10. Merkatu bat hiru enpresaren artean kuota berdinetan banatzen da. Zenbat da aniztasun-adierazlea? Oharra: $H = -\sum p_i \ln p_i$.
- (a) 1.09 gutxigorabehera.
- (b) 2.09 gutxigorabehera.
- (c) 3.09 gutxigorabehera.
- (d) Beste bat da erantzuna.
11. Karratu txikiaren erregresio batean, y eta \hat{y} balioen informazioa daukagu, ezezagun bakar batekin: $(y_1 = 1, \hat{y}_1 = 2)$, $(y_2 = 2, \hat{y}_2 = 3)$, $(y_3 = 3, \hat{y}_3 = ?)$. Zenbat da \hat{y}_3 ?
- (a) 1
- (b) 4
- (c) -5
- (d) Ez da posible datu horiek izatea, erroreen batura edonola ere 0 ez delako.
12. Nola linealizatzen da $\hat{y} = ke^{mx}$?
- (a) $\ln \hat{y} = \ln k + mx$
- (b) $\hat{y} = k + mx$
- (c) $\ln \hat{y} = \ln k + m \ln x$
- (d) $\hat{y} = k + m \ln x$
13. Emakumeek humanitate-gradua ikasteko joera handiagoa izaten dute unibertsitatean, eta gizonek zientzia edo teknologikoak. Ikasten ez duten ikasleen artean, berriz, langabezia da nagusi. Zein da interakzioa sorrarazten duen aldagai kualitatiboa?
- (a) Gradu-mota.
- (b) Sexua.
- (c) Langabezia.
- (d) Azaldu den egoera horretan ez du zentzurik interakzioaz galdetzeak.

14. Enpresa bateko zuzendaria zara eta enpresarentzat erabakigarriak diren aldagai batzuk aztertu behar dituzu. Zein kasutan ez da zilegi aurreanak egitea?
- (a) Enpresak erosten dituen prezioen lehengaiak.
 - (b) Langileen absentismoa.
 - (c) Enpresak ordaindu behar dituen interes-tipoak.
 - (d) Enpresaren inbertsioen zenbatekoa.
15. Salmenten serie historiko batean urtarrileko salmentak 144 izan dira. Serierako eredu biderkakorra ezarri da eta urtarrileko urtarokotasun-indizearen arabera, urtarrilean izateagatik salmentak %20 jaisten dira. Zenbat dira salmenta desestazionalizatuak?
- (a) 124
 - (b) 164
 - (c) 115.2
 - (d) 180
16. Karratu txikiaren erregresio batean, aurreanen bariantza 100 da. Erroreen bariantza 200 da. Zenbat da mugatze-koefizientea?
- (a) 0.33
 - (b) 0.5
 - (c) Emandako informazioa inkoherentea da.
 - (d) Beste bat da erantzuna.
17. Karratu txikiaren erregresio batean, x eta e balioen informazioa daukagu: $(x_1 = 1, e_1 = -1)$, $(x_2 = 2, e_1 = 2)$, $(x_3 = 3, e_3 = 1)$, $(x_4 = 4, e_4 = -2)$. Errore-diagrama ikusita, zein diagnostiko egin behar da?
- (a) Zehaztapen-errorea.
 - (b) Autokorrelazioa.
 - (c) Heteroskedastikotasuna.
 - (d) Erregresioaren emaitzak ontzat har daitezke.
18. 2012eko salmentak 144 izan ziren. Prezio-indizeen bilakaera hau izan da oinarri urte gisa hartutako 2009tik: (2009:100), (2010:105), (2011:115), (2012:125). Zenbat dira 2012ko salmentak errealak 2010eko prezioetan?
- (a) 116.96
 - (b) 118.96
 - (c) 120.96
 - (d) 122.96
19. Salmenten serie historikoa duzu honako hau: 36-39-46. $\alpha = 0.4$ izanik, zenbat da \hat{y}_3 ? Laguntza: $\hat{y}_t = \hat{y}_{t-1} + \alpha(y_{t-1} - \hat{y}_{t-1})$.
- (a) 36.8
 - (b) 37
 - (c) 37.2
 - (d) Beste bat da erantzuna.
20. Zertarako erabiltzen da leunketa esponentzial bikoitza?
- (a) Gorabehera handiak dituzten serieetan are eta gehiago leuntzeko.
 - (b) Aurrean fidagarriagoak egiteko, haiek bi aldiz kalkulatzeko direnez formulaz.
 - (c) Joera duten serieetarako.
 - (d) Bi parametro baliatzen dituzenez, aurrean fidagarriagoak egiteko.

Estatistika eta datuen analisia

-
2017ko ekainaren 27a**Izena eta abizenak:**

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

KOPURUA

ONGI	
GAIZKI	
ERANTZUN GABE	

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2017ko ekainaren 27a, 15:00 - Iraupena: Ordu t'erdi.

Praktika: Adierazburu-orria

I. ebazkizuna (2.5 puntu)

Kontxako hondartzan bainu-denboraldian zehar jasotako zabor kantitatea jaso da kilotan:

$$45 - 62 - 58 - 23 - 87 - 38 - 48 - 52 - 86 - 145$$

Egin beharreko atazak:

- Langile bakar batek bere lanaldian 50 kilo jaso ditzakeela uste bada, estimatu ezazu egun guztietatik ehuneko zenbatetan den langile bakarra nahikoa.
- Alborapen- eta kurtosi-neurri sendo edo jasangorak erabiliz, erabaki ezazu hondakin-kopururako banakuntza normala eredu egokia den.
- Lorenz kurba eratzeko kalkuluak egin itzazu, baina Lorenz kurba marraztu gabe. Zabor gehieneko egun guztietatik %20tan, zabor osoaren ehuneko zenbat jasotzen da?

II. ebazkizuna (2.5 puntu)

Zoriz aukeraturiko 19 urteko zenbait gazteri sexua (g: gizon, e: emakume), unibertsitateko ikasketak egiten ari diren (bai: b, ez: e) eta sexu-harreman osoak izan dituzten (bai: b, ez: e) galdetu zaizkie:

Sexua	g	g	g	g	g	g	g	e	e	e	e	e	e	e	e	e
Unib.?	e	e	e	e	b	b	b	b	b	b	b	b	b	b	e	e
Sexu-har.?	e	b	e	b	e	e	e	b	b	b	e	b	b	b	b	e

Egin beharreko atazak:

- Sexuaren eta unibertsitari izatearen arteko kontingentzia-taula eratu. Berdin, sexuaren eta sexu-harremana izatearen artean, eta unibertsitari izatearen eta sexu-harremana izatearen artean.
- Sexuaren eta unibertsitari izatearen arteko kontingentzia-taulatik, kalkulatu khi-karratu eta dagokion asoziazio-neurria. Interpretatu.
- Sexuaren eta sexu-harremana izatearen arteko kontingentzia-taulatik, kalkulatu lambda. Interpretatu.
- Unibertsitari iatearen eta sexu-harremana izatearen arteko kontingentzia-taulatik, kalkulatu Pearson korrelazio-koefiziente lineala. Interpretatu.

III. ebazkizuna (2.5 puntu)

Serotonina dosi ezberdinak eman zaizkie produktu batean insomnia jasaten duten pertsona batzuei. Dosi bakoitzeko hobekuntza nabaritu duten pertsonen portzentajea (hobetzeko probabilitatea, alegia) jaso da:

Dosia (mg)	0	1	2	4	6
Prob.	%10	%15	%45	%60	%85

Egin beharreko atazak:

- Logit ereduak garatu.
- 5 mgko dosi baterako, zenbat izango dira portzentajea hobetzen direnak?
- Hobetzen diren %90era heltzeko, zenbateko dosia eman behar da?
- Mugatze-koefizientea kalkulatu. Hori ikusita, aurreko ataletako auresanak fidagarriak izan al daitezke?

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2017ko ekainaren 27a, 15:00 - Iraupena: Ordu t'erdi.

Ebazpena

I. ebazkizuna (2.5 puntu)

Kontxako hondartzan bainu-denboraldian zehar jasotako zabor kantitatea jaso da kilotan:

$$45 - 62 - 58 - 23 - 87 - 38 - 48 - 52 - 86 - 145$$

Egin beharreko atazak:

1. Langile bakar batek bere lanaldian 50 kilo jaso ditzakeela uste bada, estimatu ezazu egun guztietatik ehuneko zenbatetan den langile bakarra nahikoa.
2. Alborapen- eta kurtosi-neurri sendo edo jasankorrak erabiliz, erabaki ezazu hondakin-kopururako banakuntza normala eredu egokia den.
3. Lorenz kurba eratzeko kalkuluak egin itzazu, baina Lorenz kurba marraztu gabe. Zabor gehieneko egun guztietatik %20tan, zabor osoaren ehuneko zenbat jasotzen da?

(a) 50 kilo edo gutxiago jasotzen den egun kopurua 4 da, 10 egunetarik. Beraz, egunen %40tan behar da langile bakarra.

Erantzun zehatzagoa eman daiteke, pertzentiletarako erabiltzen dugun formulatik. Planteamendua hau da: 50 balioa zein pertzentilari dagokion. 50 balioa 48 eta 52 datuen artean dago, justu erdian; beraz, 4.5garren datua litzateke. $4.5/10=0.45$ dugunez, 50 balioa 45. pertzentila litzateke eta horrela, egun guztietatik %45etan izango litzateke nahikoa langile bakarra.

(b) Neurri jasankorrak behar ditugunez Bowley alborapen-koefizientea eta Moors kurtosi koefizientea behar ditugu. Kalkula ditzagun horietarako pertzentilak, horretarako datuak ordenatuz lehenbizi:

$$23 - 38 - 45 - 48 - 52 - 58 - 62 - 86 - 87 - 145$$

$P_{12.5}$?

$$0.125 \times 10 = 1.25 \rightarrow P_{12.5} = 1.25\text{garren datua} \rightarrow 1. \text{ datua} = 23, 2. \text{ datua} = 38$$

$$P_{12.5} = 23 \times 0.75 + 38 \times 0.25 = 26.75$$

$Q_1 = P_{25}$?

$$0.25 \times 10 = 2.5 \rightarrow P_{25} = 2.5\text{garren datua} \rightarrow 2. \text{ datua} = 38, 3. \text{ datua} = 45$$

$$P_{25} = 38 \times 0.5 + 45 \times 0.5 = 41.5$$

$P_{37.5}$?

$$0.375 \times 10 = 3.75 \rightarrow P_{37.5} = 3.75\text{garren datua} \rightarrow 3. \text{ datua} = 45, 4. \text{ datua} = 48$$

$$P_{37.5} = 45 \times 0.25 + 48 \times 0.75 = 47.25$$

$Me = P_{50}$?

$$0.5 \times 10 = 5 \rightarrow P_{50} = 5 \text{garren datua} = 52$$

$$P_{50} = 52$$

$P_{62.5}$?

$$0.625 \times 10 = 6.25 \rightarrow P_{62.5} = 6.25 \text{garren datua} \rightarrow 6. \text{ datua} = 58, 7. \text{ datua} = 62$$

$$P_{62.5} = 58 \times 0.75 + 62 \times 0.25 = 59$$

$Q_3 = P_{75}$?

$$0.75 \times 10 = 7.5 \rightarrow P_{75} = 7.5 \text{garren datua} \rightarrow 7. \text{ datua} = 62, 8. \text{ datua} = 86$$

$$P_{75} = 62 \times 0.5 + 86 \times 0.5 = 74$$

$P_{87.5}$?

$$0.875 \times 10 = 8.75 \rightarrow P_{87.5} = 8.75 \text{garren datua} \rightarrow 8. \text{ datua} = 86, 4. \text{ datua} = 87$$

$$P_{87.5} = 86 \times 0.25 + 87 \times 0.75 = 86.75$$

Alborapen- eta kurtosi-koefizienteak eman ditzagun orain:

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} = \frac{(74 - 52) - (52 - 41.5)}{74 - 41.5} = 0.35$$

Alborapen positiboa dago (oso nabarmena ez dirudien arren). Simetriari buruzko erabaki ziurragoa hartzeko, froga estatistiko inferentziala garatu beharko litzateke.

$$K_M = \frac{(P_{87.5} - P_{62.5}) + (P_{37.5} - P_{12.5})}{P_{75} - P_{25}} = \frac{(86.75 - 59) + (47.25 - 26.75)}{74 - 41.5} = 1.48$$

Badirudi banakuntza leptokurtikoa dela, $K_M > 1.23$ enez, balio horretatik asko aldentzen ez den arren.

Beraz, ez dago argi banakuntza normala egokia denik datu horietarako; badirudi ezetz, baina inferentzia baliatu beharko litzateke erantzun ziurra emateko.

(c)

p_i	x_i metatua	q_i
0.1	23	0.035
0.2	61	0.094
0.3	106	0.164
0.4	154	0.239
0.5	206	0.319
0.6	264	0.410
0.7	326	0.506
0.8	412	0.639
0.9	499	0.774
1	644	1

Zabor gehieneko egun guztietatik %20tan, zabor osoaren ehuneko $1 - 0.639 = 0.361 = \%36.1$ jasotzen da.

II. ebazkizuna (2.5 puntu)

Zoriz aukeraturiko 19 urteko zenbait gazteri sexua (g: gizon, e: emakume), unibertsitateko ikasketak egiten ari diren (bai: b, ez: e) eta sexu-harreman osoak izan dituzten (bai: b, ez: e) galdetu zaizkie:

Sexua	g	g	g	g	g	g	g	e	e	e	e	e	e	e	e	
Unib.?	e	e	e	e	b	b	b	b	b	b	b	b	b	b	e	e
Sexu-har.?	e	b	e	b	e	e	e	b	b	b	e	b	b	b	b	e

Egin beharreko atazak:

1. Sexuaren eta unibertsitari izatearen arteko kontingentzia-taula eratu. Berdin, sexuaren eta sexu-harremana izatearen artean, eta unibertsitari izatearen eta sexu-harremana izatearen artean.
2. Sexuaren eta unibertsitari izatearen arteko kontingentzia-taulatik, kalkulatu khi-karratu eta dagokion asoziazio-neurria. Interpretatu.
3. Sexuaren eta sexu-harremana izatearen arteko kontingentzia-taulatik, kalkulatu lambda. Interpretatu.
4. Unibertsitari iatearen eta sexu-harremana izatearen arteko kontingentzia-taulatik, kalkulatu Pearson korrelazio-koefiziente lineala. Interpretatu.

(a)

		Unibertsitatea?		
		Ez	Bai	
Sexua	Gizona	4	3	7
	Emakumea	2	7	9
		6	10	16

		Sexu harremanik?		
		Ez	Bai	
Sexua	Gizona	5	2	7
	Emakumea	2	7	9
		7	9	16

		Sexu harremanik?		
		Ez	Bai	
Unib.?	Ez	3	3	6
	Bai	4	6	10
		7	9	16

(b)

Ondoko taulan azaltzen dira bi aldagaiak biltzen dituen kontingentzia-taulako maiztasunak (enpirikoak) (lehen kopuruak: 4-3-2-7), maiztasun teorikoak, eta khi-karratu estatistikorako ekarpena ($\frac{(enpiriko-teoriko)^2}{teoriko}$):

		Unibertsitatea?		
		Ez	Bai	
Sexua	Gizona	4-2.625-0.720	3-3.375-0.560	7
	Emakumea	2-4.375-0.432	7-5.625-0.336	9
		6	10	16

$$X^2 = 0.720 + 0.560 + 0.432 + 0.336 = 2.048$$

Khi-karratuan oinarrituriko asoziazio-neurria 2×2 tauletarako phi da:

$$\varphi = \sqrt{\frac{2.048}{16}} = 0.35$$

Koefizientearen emaitzak bi aldagaien arteko asoziazio ertaina erakusten du, lagin errorearen eta antzeko ikerketen erreserbapean.

(c)

		Sexu harremanik?		
		Ez	Bai	
Sexua	Gizona	5	2	7
	Emakumea	2	7	9
		7	9	16

Aldagai independentea sexua (gizon/emakume) da. Aldagai independentea kontuan hartu gabe, gazte batek sexu-harremana izango duela aurreikus daiteke 8edo ez, berdinketa dagoenez); edonola ere, errore kopurua 7 da.

Aldagai dependentea kontuan harturik:

- gaztea gizona bada, sexu-harremanik ez duela izan aurreikusi behar da (5) eta errore kopurua 2 izango da;
- gaztea emakumea bada, sexu-harremanik izan duela aurreikusi behar da (7) eta errore kopurua 2 izango da.

Aldagai dependentea kontuan harturik, beraz, guztira 2+2=4 errore egiten dira, kontuan hartu gabe baino 7-4=3 errore gutxiago.

Beraz, $\lambda = 3/7 = 0.42$, eta orduan bi aldagaien arteko asoziazioa ertaina dela esan daiteke, lagin errorearen eta antzeko ikerketen erreserbapean.

(d)

Kodetzea: Unib.?: 0, ez; 1,bai. Sexu-har.?: 0, ez; 1,bai.

x (unib.)	y (sexu-har.)	n	nx	ny	nxy	nx^2	ny^2
0	0	3	0	0	0	0	0
0	1	3	0	3	0	0	3
1	0	4	4	0	0	4	0
1	1	6	6	6	6	6	6
-	-	16	10	9	6	10	9

$$\bar{x} = \frac{10}{16} = 0.625$$

$$\bar{y} = \frac{9}{16} = 0.5625$$

$$s_{xy} = \frac{6}{16} - 0.625 \times 0.5625 = 0.0234$$

$$s_x^2 = \frac{10}{16} - 0.625^2 = 0.2343$$

$$s_y^2 = \frac{9}{16} - 0.5625^2 = 0.246$$

$$r_{xy} = \frac{0.0234}{\sqrt{0.2343}\sqrt{0.246}} = 0.09$$

Pearson korrelazio-koefizientearen arabera, bi aldagaien arteko asoziazioa ahula da, lagin errorearen eta antzeko ikerketen erreserbapean.

III. ebazkizuna (2.5 puntu)

Serotonina dosi ezberdinak eman zaizkie produktu batean insomnia jasaten duten pertsona batzuei. Dosi bakoitzeko hobekuntza nabaritu duten pertsonen portzentajea (hobetzeko probabilitatea, alegia) jaso da:

Dosia (mg)	0	1	2	4	6
Prob.	%10	%15	%45	%60	%85

Egin beharreko atazak:

1. Logit ereduak garatu.
2. 5 mgko dosi baterako, zenbat izango dira portzentajearen hobetzen direnak?
3. Hobetzen diren %90era heltzeko, zenbateko dosia eman behar da?
4. Mugatze-koefizientea kalkulatu. Hori ikusita, aurreko ataletako auresanak fidagarriak izan al daitezke?

(a)

Datuak formato egokian daude: dosia gora, probabilitatea gora. Bere horretan uzten ditugu, beraz.

x (dosia)	p	$y = \ln \frac{p}{1-p}$	xy	x^2	y^2	\hat{y}	\hat{y}^2
0	0.1	-2.19	0	0	4.81	-2.09	4.36
1	0.15	-1.73	-1.73	1	3.008	-1.44	2.07
2	0.45	-0.20	-0.40	4	0.040	-0.79	0.62
4	0.60	0.40	1.62	16	0.164	0.51	0.26
6	0.85	1.73	10.40	36	3.008	1.81	3.27
13	-	-2	9.89	57	11.050	-2	10.60

$$\bar{x} = \frac{13}{5} = 2.6$$

$$\bar{y} = \frac{-2}{5} = -0.4$$

$$s_{xy} = \frac{9.89}{5} - 2.6 \times -0.4 = 3.02$$

$$s_x^2 = \frac{57}{5} - 2.6^2 = 4.64$$

$$b = \frac{3.02}{4.64} = 0.65$$

$$a = -0.4 - 0.65 \times 2.6 = -2.09$$

$$\hat{y} = -2.09 + 0.65x$$

(b)

$$\hat{y} = -2.09 + 0.65 \times 5 = 1.16$$

$$\ln \left(\frac{p}{1-p} \right) = 1.16 \rightarrow \frac{p}{1-p} = e^{1.16} = 3.19 \rightarrow p = 3.19 - 3.19p \rightarrow p = 0.76$$

(c)

$$p = 0.9 \rightarrow \ln \left(\frac{0.9}{1-0.9} \right) = 2.19 \rightarrow 2.19 = -2.09 + 0.65x = 6,58mg$$

(d)

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{10.60/5 - (-0.4)^2}{11.05/5 - (-0.4)^2} = 0,956$$

Mugatze-koefizienteak doikuntza oso ona adierazten du. Beraz, auresanak fidagarriak lirateke alde horretatik.

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2017ko ekainaren 27a

Iraupena: 35 minutu

Erantzun bakarra da zuzena galdera bakoitzean. Guztira testak 2.5 puntu balio du. Erantzun zuzenak 0.125 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 da gaixotasunen intzidentzia eta medikuntzako beste alderdi batzuk estatistikoki aztertzen duen arloa?
 - (a) Medikometria
 - (b) Epidemiologia
 - (c) Biometria
 - (d) Patologia
2. Hurrenez hurren gradu honetako 100 eta 50 ikasle ebaluatu dituzten Finantza Matematikako bi irakasleen nota globalak alderatu nahi dira. Zein da grafiko egokiena?
 - (a) Back-to-back histograma
 - (b) Puntu-diagrama bikoitza
 - (c) Barra-diagrama bikoitza
 - (d) Aurreko guztiak dira okerrak.
3. Zein ez da neurri jasankorra?
 - (a) Ibiltartea
 - (b) Koartil arteko ibiltartea
 - (c) DAME
 - (d) Bowley koefizientea
4. Datuak: 1-2-2-3. Kalkulatu populazio-bariantza.
 - (a) 1
 - (b) 0.6
 - (c) 0.5
 - (d) Aurreko guztiak faltsuak dira.
5. 4 datuekin populazio-bariantza 10 suertatu da. Kalkulatu lagin-bariantza.
 - (a) 7.5
 - (b) 11.66
 - (c) 13.33
 - (d) Ezin da kalkulatuz informazio horrekin soilik.
6. Errenta osoaren %40a hartzen du errenta handiena duten pertsonen %10ek. Zenbat da (p_i, q_i) puntu horretan?
 - (a) (0.4, 0.1)
 - (b) (0.1, 0.4)
 - (c) (0.9, 0.6)
 - (d) (0.9, 0.4)
7. Errentak: 4-6-10-12-18. Kalkulatu Gini indizea.
 - (a) 0.21

- (b) 0.25
- (c) 0.29
- (d) Beste bat da erantzuna.

8. Errentak: 4-6-10-12-18. Pobrezia-muga: 8. Kalkulatu pobreziaren intentsitatea.

- (a) 0.275
- (b) 0.375
- (c) 0.475
- (d) Beste bat da erantzuna.

9. Errentak: 4-6-10-12-18. Pobrezia-muga: 8. Kalkulatu pobreziaren neurri orokorra.

- (a) 0.16
- (b) 0.26
- (c) 0.36
- (d) Beste bat da erantzuna.

10. Errenta guztiak %10 igotzekoak badira, zenbat izatera igarotzen da ($p_i = 0.4, q_i = 0.2$) balio-bikotea?

- (a) ($p_i = 0.4, q_i = 0.2$)
- (b) ($p_i = 0.5, q_i = 0.2$)
- (c) ($p_i = 0.4, q_i = 0.3$)
- (d) ($p_i = 0.5, q_i = 0.3$)

11. Kontingentzia taula honetan zein da egiazkoa?

Aprobata? (↓) / Sexua (→)	Gizona	Emakumea
Bai	12	8
Ez	38	32

- (a) Gizonetatik %16 dira aprobatuak.
- (b) Emakumeetatik %82 dira suspendituak.
- (c) Aprobaturtatik %60 dira emakumeak.
- (d) Suspendituetatik %54 dira gizonak.

12. Kontingentzia-aula batean $\mathbf{X}^2 = 0.9$. Nola interpretatzen da?

- (a) Asoziazioa sendoa da, beste ikerketan eta lagin-errorearen erreserbapean.
- (b) Independentek behintzat ez dira bi atributuak.
- (c) Datuak falta dira asoziazioaren sendotasunari buruz erabakitzeke.
- (d) Aurreko guztiak faltsuak dira.

13. Enpresei egindako inkesta batean azken urtean mozkinak izan dituzten (bai/ez) eta salmentak igo dituzten (bai/ez) jaso dira. Datu horietarako, zer kalkula daiteke?

- (a) Khi-karratu eta horretan oinarritutako datu guztiak.
- (b) Lambda.
- (c) Pearson korrelazio-koefiziente lineala.
- (d) Aurreko guztiak egiazkoak dira.

14. ($x_1 = 0, e_1 = -2.5$), ($x_2 = 1, e_2 = 2$), ($x_3 = 2, e_3 = 2$), ($x_4 = 3, e_4 = -1.5$). Erroreen diagrama marraztuta, zer diagnostiko egiten duzu?

- (a) Zehaztapen-errorea dago.
- (b) Autokorrelazioa da.

- (c) Heteroskedastikotasuna dago.
- (d) Dena ongi dabil.
15. Denbora-seriea: 4-5-3. Leunketa esponentziala burutuz, $\alpha = 0.8$ harturik, eman ezazu $\hat{y} = 5$.
- (a) 3.36
- (b) 3.86
- (c) 4.36
- (d) Aurreko guztiak faltsuak dira.
16. Zer da egia leunketa esponentzialari buruz?
- (a) Joerarik gabeko serieetarako erabiltzen da.
- (b) Aurreanak ez dira eguneratzen etorkizunean, konstante dira alegia.
- (c) α zenbat eta handiagoa, orduan eta garrantzi handiagoa ematen zaio azken aldiko datu historikoari aurreanak egiterakoan.
- (d) Aurreko guztiak egiazkoak dira.
17. Ana lehen aldiz ezkondu da gaur. Beraz, atzo bere senar-kopurua 0 zen, eta gaur 1. Beraz, astebete barru 7 izango da. Nola deitzen zaio arrazonomendu oker honi?
- (a) Alegiazko korrelazioa.
- (b) Estrapolazioa.
- (c) Interakzioa.
- (d) Elkarrekintza.
18. Banku bateko uneko egoera nahasiak bere akzioen %55eko beherakada izan du azken 5 burtsa-saioretan. Zein aldaketa-motaz ari gara?
- (a) Asterokotasuna.
- (b) Joera laburra.
- (c) Xoka edo shock-a.
- (d) 5 eguneko zikloa.
19. Sare elektriko batean kontsumoa jasotzen da une oro. 13:00-16:00 bitartean kontsumoa jaitsi egiten da orokorrean. Zein aldaketa-motaz ari gara egunaren baitan?
- (a) Egunerokotasuna.
- (b) Zikloa.
- (c) Hondarra.
- (d) Urtarokotasunaren parekoa, baina egunaren baitan.
20. Euri-kopurua toki batean aldatzen da urtaroaren arabera (udan, udazkenean, neguan zein udaberrian izan) eta horrekin batera bi urteroko ziklo baten arabera. Urtaroaren araberrako datuak izanda azken 20 urtetarako, zein tamainako batezbesteko higikorra kalkulatu behar da bi gorabehera mota horiek batera ezabatzeko?
- (a) 3
- (b) 4
- (c) 6
- (d) 8

Estatistika eta datuen analisia

-
2017ko ekainaren 27a**Izena eta abizenak:** _____

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

KOPURUA

ONGI	
GAIZKI	
ERANTZUN GABE	