

ESTADÍSTIKA ETA DATUEN ANALISIA

Azterketa ebatziak.

2017-2018 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea.
EHU

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

gizapedia.hirusta.io

ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2018ko urtarrilaren 18a, 10:00

Iraupena: Ordu t'erdi.

Praktika: Adierazburu-orria

I. ebazkizuna (2.5 puntu)

2005 eta 2015 urteetan herri bateko bizilagunen errentak jaso dira :

- 2005: 3-5-7-10-15
- 2015: 3-5-12-14-16

Egin beharreko atazak

- Bi urteetarako Lorenz kurbak eratu eta adierazi, ahal bada, zein urtetan izan zen errenta kontzentratuago.
- Lorenz kurbak hartuta, zein urtetan izan ziren %20 aberatsenak aberatsago? Zehaztu eta azaldu zure erantzuna.
- Giniren indizeak kalkulatu bi urteetarako eta emaitzak interpretatu.
- Kalkulatu Robin Hood adierazlea bi urteetarako eta emaitzak interpretatu.
- Kalkulatu 25:25 ratioa (75. pertzentiletik gorako eta 25. pertzentiletik beherako errenten batezbestekoen arteko erlazioa) bi urteetarako eta emaitzak interpretatu.

II. ebazkizuna (2.5 puntu)

Merkataritzako langile batzuen artean sexua eta lortutako salmentak jaso dira:

Sexua	g	e	e	e	g	e	e	g	g	g	g
Salmentak	42	58	68	52	54	75	80	54	62	59	67

Egin beharreko atazak

- Grafiko egokia erabiliz, bi aldagaien arteko erlazioa esploratu.
- Kalkulatu batezbesteko salmenten diferentzia bi sexuen artean, eta horren efektuaren tamaina ebaluatu Cohen-en d kalkulatu.
- Eta korrelazio-ratioa kalkulatu eta interpretatu.
- Saltzaileak sailka itzazu 60tik gorako salmentak lortu dituzten ala ez kontuan harturik. Sexuaren eta $>60/<60$ sailkapenaren arteko korrelazio-koefiziente lineala kalkulatu eta interpretatu.

III. ebazkizuna (2.5 puntu)

Denda batek ostiraletik igandera bakarrik zabaltzen ditu atea. Hiru astetan zehar egun horietan zerbitzaturiko entsaladak jaso dira:

		Astea		
		1. astea	2. astea	3. astea
Eguna	Ostirala	46	47	45
	Larunbata	64	66	69
	Igandea	52	55	60

Egin beharreko atazak

- Erabaki denbora-serie horretarako eredu egokiena.
- Denbora osagaiak (J, U eta H) bereizi.
- U osagaia interpretatu ostiral, larunbat eta iganderako.
- Noiz izan zen salmentak igo edo jaitsiarazi egin zituen gertakari nabarmenena?
- Salmentak desestazionalizatu.
- Hurrengo asterako aurrean egin, *drift* metodoa baliatuz.

ESTADISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2018ko urtarrilaren 18a, 10:00

Iraupena: Ordu t'erdi.

Praktika: Ebazpena

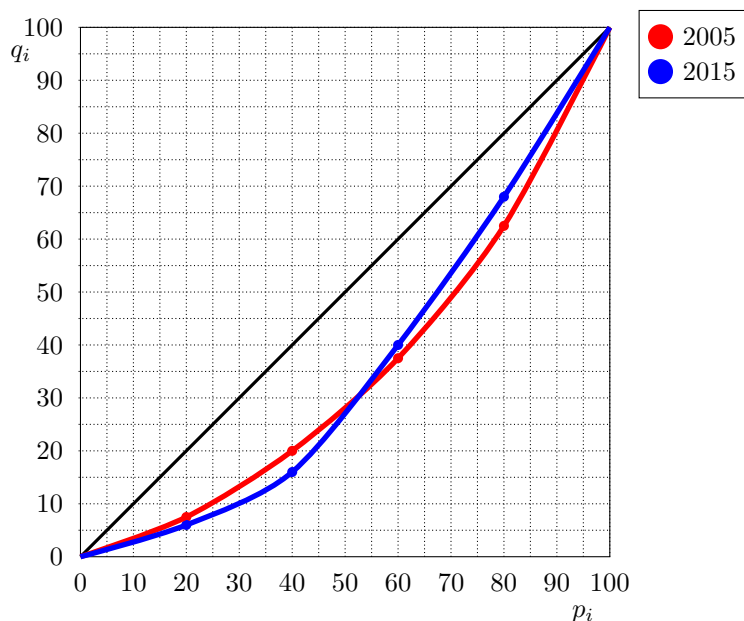
I. EBAZKIZUNA

(a)

Lorenz kurbarako kalkuluak egin ditzagun. Datu kopuruak bi urteetan berdinak direnez, p_i balioak berdinak dira bi urteetarako.

p_i	$X_{i,2005}$	$q_{i,2005}$	$X_{i,2015}$	$q_{i,2015}$
0.2	3	0.075	3	0.06
0.4	8	0.20	8	0.16
0.6	15	0.375	20	0.40
0.8	25	0.625	34	0.68
1	40	1	50	1

Lorenz kurbak:



Interpretazioa: bi kurbak ebaki egiten direnez, eta diagonalaren azpitik antzeko azalera uzten dutenez, ezin esan daiteke zein den banaketa kontzentratuena.

(b)

%20 aberatsenak errenta osoaren $\%100 - \%62.5 = \%37.5$ hartu zuen 2005 urtean; eta $\%100 - \%68 = \%32$ hartu zuen 2015 urtean. Beraz, erlatiboki 2005 urtean izan ziren %20 aberatsenak aberatsago.

(c)

Giniren indizeak:

p_i	2005 : $p_i - q_i$	2015 : $p_i - q_i$
0.2	$0.2 - 0.075 = 0.125$	$0.20 - 0.06 = 0.14$
0.4	0.200	0.24
0.6	0.225	0.20
0.8	0.175	0.12
1(ez)	0.000	0.00
2	0.725	0.7

$$G_{2005} = \frac{0.725}{2} = 0.3625 ; G_{2015} = \frac{0.7}{2} = 0.35$$

Hartara, 2005 urtean handiagoa izan zen errentaren kontzentrazioa.

(d)

Robin Hood adierazlea, 2005 urtean:

- errenta osoa: 40;
- berdintasunezko banaketan, bakoitzak $40/5=8$ hartu beharko luke;
- horretarako, hortik gora irabazten dutenei $(10-8)+(15-8)=9$ kendu behar zaie;
- $RH_{2005} = \frac{9}{40} = 0.225$

Robin Hood adierazlea, 2015 urtean:

- errenta osoa: 50;
- berdintasunezko banaketan, bakoitzak $50/5=10$ hartu beharko luke;
- horretarako, hortik gora irabazten dutenei $(12-10)+(14-10)+(16-10)=12$ kendu behar zaie;
- $RH_{2015} = \frac{12}{50} = 0.24$

Beraz, Robin Hood adierazlearen arabera, 2015 urtean izan zen handiagoa kontzentrazioa.

(e)

25:25 ratioa kalkulatzeko, P_{25} eta P_{75} kalkulatu behar dira bi urteetako datuetarako. Datu kopurua berdina denez bi urteetan:

- $P_{25} = 0.25 \times 5 = 1.25gn \text{ datua}$
- $P_{75} = 0.75 \times 5 = 3.75gn \text{ datua}$

Ez da beharrezkoa pertzentilak zehatz kalkulatzea. Aski da 1.25gn datutik beherako eta 3.75gn datutik gorako datuak hartu, hurrenez hurren, eta datu horien batezbestekoen arteko zatiketa egitea:

2005 urtean:

- 1.25gn datutik behera, datu bakarra da, 3, eta horren batezbestekoa 3 da noski.
- 3.75gn datutik gora, datu bi daude, 10 eta 15, eta horien batezbestekoa 12.5 da.
- $R_{25:25} = \frac{12.5}{3} = 4.16$ (hau da, aberatsek pobreek baino 4.16 bider gehiago irabazten dute batezbestez).

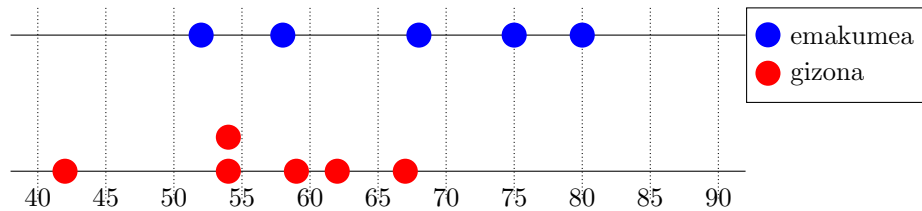
2015 urtean:

- 1.25gn datutik behera, datu bakarra da, 3, eta horren batezbestekoa 3 da noski.
- 3.75gn datutik gora, datu bi daude, 14 eta 16, eta horien batezbestekoa 15 da.
- $R_{25:25} = \frac{15}{3} = 5$ (hau da, aberatsek pobreek baino 5 bider gehiago irabazten dute batezbestez).

Beraz, adierazle horren arabera, kontzentrazioa handiagoa da 2015 urtean.

II. EBAZKIZUNA

(a)
Aldagai independentea sexua, eta beraz, horren arabera bereiziko dira datuak.
Salmenta aldagai jarraitua denez, eta sexu bakoitzeko datu gutxi direnez, grafiko egokiena puntu diagrama da:



Nahiko argi ikus daiteke emakumeek orokorrean salmenta handiagoak lortzen dituztela.

(b)
Cohen-en d

- Gizonen batez besteko salmenta:

$$\bar{s}_{giz} = \frac{42 + 54 + 54 + 62 + 59 + 67}{6} = 56.33$$

- Emakumeen batez besteko salmenta:

$$\bar{s}_{emak} = \frac{58 + 68 + 52 + 75 + 80}{5} = 66.66$$

- Bi sexuen arteko diferentzia salmentetan batezbestez $66.66 - 56.33 = 10.33$ da.
- Gizonen datuen kuasibariantza:

$$s_{giz}^2 = \frac{42^2 + 54^2 + 54^2 + 62^2 + 59^2 + 67^2}{6} - 56.33^2 = 61.93$$

$$\hat{s}_{giz}^2 = \frac{6}{6-1} \times 61.93 = 74.31$$

- Emakumeen datuen kuasibariantza:

$$s_{emak}^2 = \frac{58^2 + 68^2 + 52^2 + 75^2 + 80^2}{5} - 66.66^2 = 99.84$$

$$\hat{s}_{emak}^2 = \frac{5}{5-1} \times 99.84 = 124.8$$

- Bariantza globala:

$$s = \sqrt{\frac{(6-1) \times 74.31 + (5-1) \times 124.8}{6+5-2}} = 9.83$$

- Somers-en d:

$$d = \frac{10.33}{9.83} = 1.05$$

1 baino handiagoa denez, bi sexuen arteko salmenten diferentzia sendoa dela esan daiteke.

(c)

$$\text{Korrelazio-ratioa (eta): } \eta^2 = \frac{\sum_x n_x (\bar{y}_x - \bar{y})^2}{\sum_{x,i} (y_{xi} - \bar{y})^2}$$

- x , aldagai independentea: sexua; y , aldagai dependentea, salmentak.
- datu guztien batezbestekoa:

$$\bar{y} = \frac{42 + 54 + 54 + 62 + 59 + 67 + 58 + 68 + 52 + 75 + 80}{11} = 61$$

- $\sum_{x,i} (y_{xi} - \bar{y})^2 = (42 - 61)^2 + (54 - 61)^2 + (54 - 61)^2 + (62 - 61)^2 + (59 - 61)^2 + (67 - 61)^2 + (58 - 61)^2 + (68 - 61)^2 + (52 - 61)^2 + (75 - 61)^2 + (80 - 61)^2 = 1196$
- $\sum_x n_x (\bar{y}_x - \bar{y})^2 = 6 \times (56.33 - 61)^2 + 5 \times (66.66 - 61)^2 = 291.03$
- $\eta^2 = \frac{291.03}{1196} = 0.24$

Bi aldagaien arteko korrelazioa ahula da, η korrelazio-ratioaren arabera.

(d)

Bigarren zutabean salmenta datuak bihurtuta, korrelazio-koefizientea kalkulatzeari ekiten diot hurrengo zutabeetan:

Kodeketa:

- x , sexua: gizon, 0; emakume, 1.
- y , salmentak: < 60,0; > 60,1.

x	y	xy	x^2	y^2
0	0	0	0	0
1	0	0	1	0
1	1	1	1	1
1	0	0	1	0
0	0	0	0	0
1	1	1	1	1
1	1	1	1	1
0	0	0	0	0
0	1	0	0	1
0	0	0	0	0
0	1	0	0	1
5	5	3	5	5

- $\bar{x} = \frac{5}{11} = 0.45$; $\bar{y} = \frac{5}{11} = 0.45$
- $s_{xy} = \frac{3}{11} - 0.45 \times 0.45 = 0.07$
- $s_x = \sqrt{\frac{5}{11} - 0.45^2} = 0.502$; $s_y = \sqrt{\frac{5}{11} - 0.45^2} = 0.502$
- $r_{xy} = \frac{0.07}{0.502 \times 0.502} = 0.28$

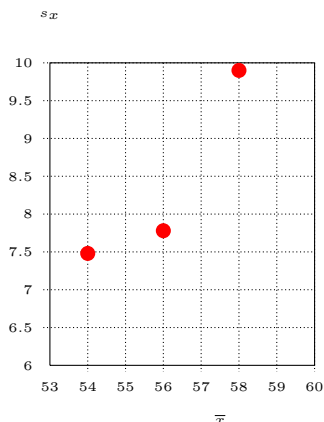
Bi aldagaien arteko korrelazioa ahula da, ertaina izateetik gertu badago ere. Korrelazioa izatekotan, norabide honetan genuke: sexua 0tik 1era pasatzean, gizonengandik emakumeengana, salmentetan ere 0tik 1era igarotzen da, 60 baino gutxiago saltzetik 60 baino gehiago saltzera; beraz, emakumeek 60 baino gehiago saltzeko joera handiagoa lukete, baina hala ere korrelazio hori ahula da, arestian aipatu bezala.

III. EBAZKIZUNA

(a)

Eredua hautatzeko, asteko datuen batezbestekoak eta desbideratzeak kalkulatu dira:

- 1. astea: $\bar{x} = \frac{46 + 64 + 52}{3} = 54$; $s_x = \sqrt{\frac{(46 - 54)^2 + (64 - 54)^2 + (52 - 54)^2}{3}} = 7.48$
- 2. astea: $\bar{x} = 56$; $s_x = 7.78$
- 3. astea: $\bar{x} = 58$; $s_x = 9.9$



Puntuen konfigurazioa ikusita, goraka alegia, biderkakorra da eredu egokiena.

(b)

Osagaiak bereizteko kalkuluak egiteko, serie historikoa, non osagaiak $J \times U \times H$ eran bateraturik dauden, formato estandarrean jarriko dugu (epe luzea, errenkadetan):

		Eguna		
		$J \times U \times H$	Ostirala	Larunbata
Astea	1. astea	46	64	52
	2. astea	47	66	55
	3. astea	45	69	60

Joera bereizteko, BH3 kalkulatu behar da:

		Eguna		
		J	Ostirala	Larunbata
Astea	1. astea	-	54	54.33
	2. astea	55	56	55.33
	3. astea	56.33	59	-

Serie historikoa joerarekin zatituz, $U \times H$ osagaia eskuratzen da, eta eguneko batezbestekoak kalkulatu U urtarokotasun soila:

		Eguna		
		$U \times H$	Ostirala	Larunbata
Astea	1. astea	-	1.185	0.957
	2. astea	0.855	1.179	0.994
	3. astea	0.799	1.190	-
Urtarokotasuna		0.827	1.185	0.975

$U \times H$ serieko balioei dagokien U balioekin zatituz, H bakantzen da:

		Eguna		
		H	Ostirala	Larunbata
Astea	1. astea	-	1	0.982
	2. astea	1.033	0.995	1.019
	3. astea	0.966	1.004	-

(c) Urtarokotasuna interpretatu: ostiraletan izateagatik soilik eskatzen diren entsaladak $100 - 82.7 = \%17.3$ murrizten dira, larunbatean izateagatik $\%18.5$ igo, eta igandean izateagatik $\%2.5$ jaitsi. Datu gutxiarekin lan egiteagatik ez da betetzen batura 3 izatea, teoriarik izan behar den bezala.

(d) Hondarra interpretatu: aparteko eta ustekabeko gertakaririk nabarmenenak 2. eta 3. asteko ostiraletan izan ziren (borobilketa erroregatik dira desberdinak balio absolutuan bi balioak), salmentak $\%3.3$ igoaraziz eta jaitsiaraziz hurrenez hurren.

(e) Salmentak desestazionalizatzeko, $J \times U \times H$ serie historikoko balioak dagokien U balioekin zatitu behar dira:

		Eguna		
		$J \times H$	Ostirala	Larunbata
Astea	1. astea	55.62	54.10	53.33
	2. astea	56.83	55.70	56.41
	3. astea	54.41	58.23	61.54

(f) Aurreanak serie desestazionalizatuan egiten ditugu lehenbizi, eta dagokien U osagaia txertatu (bidertuz, eredu biderkakorra denez).

Drift edo batezbesteko gehikuntza kalkulatzeko da:

$$\overline{\Delta y_t} = \frac{61.54 - 55.62}{9 - 1} = 0.74$$

- 4 asteko ostiralerako auresana = $(61.54 + 0.74) \times 0.827 = 51.50$
- 4 asteko larunbaterako auresana = $(61.54 + 0.74 \times 2) \times 1.185 = 74.67$
- 4 asteko iganderako auresana = $(61.54 + 0.74 \times 3) \times 0.975 = 62.17$

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2018ko 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.

- (1) 200 ikaslek Mikroekonomia irakasgaiari izandako notak jaso dira. Sexuaren arabera diferentziak hautemateko, zein grafiko da egokiena?
 - (a) Puntu-diagrama bikoitza
 - (b) Back-to-back edo elkarren ondoko histogramak
 - (c) Bi maiztasun-poligono, bata bestearen gainean.
 - (d) (b) eta (c) dira zuzenak
- (2) Txanpu baten pH 7 izan behar da. 4 lote jaso eta pH balio hauek jaso dira haietarako: 7.2, 6.9, 6.8 eta 7. Zenbat da batez besteko errorea, datu horietarako formula egokiena baliatuz?
 - (a) 0.025
 - (b) 0.125
 - (c) 0.023
 - (d) 0.173
- (3) Datu hauek dituzu: 2-4-6-8. Kalkulatu 3garren oktila.
 - (a) 3
 - (b) 3.25
 - (c) 3.5
 - (d) 3.75
- (4) 20 datuen batezbestekoa 4 da. Beste 10 datuen batezbestekoa 8 da. Zenbat da 30 datu horien batezbestekoa?
 - (a) 5
 - (b) 5.33
 - (c) 5.5
 - (d) 6
- (5) Honako hauetatik zein da neurri ez jasankorra?
 - (a) Moors koefizientea.
 - (b) Bowley koefizientea.
 - (c) Pearson kurtsoi-koefizientea.
 - (d) Kuartil arteko ibiltartea.
- (6) Zenbat datu daude D_4 eta Q_3 kuantilen artean?
 - (a) %10
 - (b) %25
 - (c) %35
 - (d) %40
- (7) Moors koefizientea 1.8 suertatu da. Nola interpretatu behar da?
 - (a) Banaketa leptokurtikoa da.
 - (b) Banaketa mesokurtikoa da.
 - (c) Banaketa platikurtikoa da.
 - (d) Banaketa uniforme da.

- (8) Lorenz kurba bateko puntuak dira hauek: $p_1 = 0.1, q_1 = 0.04$ eta $p_2 = 0.2, q_2 = 0.10$. Har dezagun orain: $p_3 = 0.3, q_3$. Zenbat izan behar da q_3 ?
- 0.14 edo txikiagoa izan behar da.
 - 0.3 edo txikiagoa izan behar da.
 - 0.16 edo handiagoa izan behar da.
 - (b) eta (c) dira zuzenak.
- (9) X errentetarako $Y = 2X + 100$ aldagai-aldaketa egin behar da. Nola aldatzen da kontzentrazioa X -etik Y -ra?
- Kontzentrazioa igo egiten da.
 - $2X$ aldaketaren ondorioz jaitsi egiten da, eta igo +100 aldaketaren ondorioz.
 - $2X$ aldaketaren ondorioz igo egiten da, eta jaitsi +100 aldaketaren ondorioz.
 - Kontzentrazioa jaitsi egiten da.
- (10) Familien errentei buruzko datuak jaso dira Euskal Herriko 3 lurraldetan:

<i>Errenta</i>	<i>HegoEH</i>	<i>IparEH</i>	<i>Nafarroa</i>
\bar{x}	20	30	10
s_x	10	15	8

Hego EH, Ipar EH eta Nafarroako hiru familiek 24, 38 eta 16-ko errenta izan zuten, hurrenez hurren. Erlatiboki, jar itzazu errenta handienetik txikienerako ordenan.

- Nafarroa-Ipar EH-Hego EH
 - Ipar EH-Nafarroa-Hego EH
 - Ipar EH-Hego EH-Nafarroa
 - Beste bat da erantzun zuzena.
- (11) Abere-etxalde batean bi pentsu-mota erabiltzen dira: A eta B. Ondoren, bi pentsu horietatik kontsumituriko q kopuruak (tonatan) eta horien p prezioak (tonako) zehazten dira.

Pentsua	A		B	
	p	q	p	q
2015	5	10	6	5
2016	6	12	8	4

Zenbat igo ziren prezioak 2016 urtean 2015eko oinarria harturik Laspeyres indizea erabiliz?

- %10
 - %15
 - %20
 - %25
- (12) Datu berdinekin jarraituz, zenbat igo ziren prezioak 2016 urtean 2015eko oinarria harturik Paasche indizea erabiliz?
- %13.8
 - %18.8
 - %23.8
 - %28.8
- (13) Datu berdinekin jarraituz, zenbat izan zen A pentsuaren partaidetza 2016ko prezio-igoeran 2015eko oinarria harturik?
- %25
 - %50
 - %75
 - Beste bat da erantzuna.
- (14) Datu berdinekin jarraituz, zenbat pentsu gehiago kontsumitu zen *fisikoki* 2016 urtean 2015eko oinarria harturik, Laspeyres indizea erabiliz?
- %5
 - %10
 - %15
 - %20

- (15) Zer adierazten du zenbaki indize baten zirkulartasun-propietateak?
- Ondasunak elkar aldatzen badira, zenbaki indizean ez dela aldaketarik izango.
 - Oinarria aldatzeko, aski dela oinarri berriko indizea 100 bihurtu, eta beste indize-balioak proportzionalki aldatzea.
 - Prezio-igoerak daudenean, ondasunaren kontsumoa txikiagoa dela, eta ondorioz prezio jaisteko joera dagoela ondoren, eta prezio-indizeek behar baino igoera handiagoa adierazten dutela.
 - Laspeyres indize batetik Paasche indize batera pasatzeko, eta alderantziz, aski dela aldi bakoitza oinarri berrizat hartuz eta indizeak prportzioan aldatzea.

- (16) Energiaren prezio-indizearen bilakaera eta enpresa batek erabiltzen duen energia iturri ezberdinetan izandako gastu nominalak jasotzen dira ondoren:

Urtea	2015	2016	2017
Prezio-indizeak	100	120	150
Gastua	1000	1500	2500

Zenbat gastatu zuen enpresak termino errealetan 2017 urtean 2016ko prezio konstantetan?

- 1700
 - 1800
 - 1900
 - 2000
- (17) Leunketa esponenzialari buruz, zein da egiazkoa honako hauetatik?
- α handia denean, leunketa handiagoa egiten da.
 - Azken aldian izandako aldaketa historikoak jaso nahi badira aurreanak egiterakoan, α handia komeni da.
 - α txikiarekin, aurreanak hobeto zuzentzen dira izandako *shock*-en ondorioz.
 - Joera gorakorra dagoenean, hobe da α txikia jartzea.
- (18) 10 ikasleen artean, ikasketa-orduen eta matematika noten arteko karratu txikiaren erregresio batean, errore karratu batura 100 suertatu eta aurreanen bariantza 40. Zenbat da mugatze-koefizientea?
- 0.8
 - 0.6
 - 0.4
 - Ezin da kalkulatu, y balioen bariantza jakin gabe.
- (19) Selektibitate aurkeztu diren gazteen artean, selektibitate egunetan sexu-harremanik izan ote zuten, selektibitate egunetako stres-indize bat (batez besteko taupada kopurua) eta selektibitateko nota jaso dira. Sexu-harremanak izatea nota altuarekin loturik dagoela ikusi da. Nola esplikatuko zenuke fenomenoak?
- Korrelazio partzialaren bitartez.
 - Sasiko korrelazioa edo korrelazio faltsua da.
 - Sexu harremanak izateak sortzen duen interakzio-efektua da.
 - Autokorrelazioa dago bi aldagaien artean.
- (20) Zer dira mikrodatuak?
- Pertsonen buruzko inkestetan, pertsonaren izena eta datu pertsonalak. Orokorrean, ikerketa-unitateen identifikazio-datuak.
 - Datuak noiz eta nork jaso zituen adierazten duten datuak (inkestazailearen izena, ordua).
 - Banako edo ikerketa-unitateen mailan jasotako datuak, horiek agregatu gabe, gordinik.
 - Datuak ahalik eta xehetasun handienaz aztertuz eskuratzen den informazioa.

Estatistika eta datuen analisia

2018ko urtarrilaren 26a

Izena eta abizenak: Josemari Sarasola

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

KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2018ko ekainaren 28a, 10:00

Iraupena: Ordu t'erdi.

Praktika: Adierazburu-orria

I. ebazkizuna (2.5 puntu)

KILAUEA surf dendako azken bi urteotako eguneko salmenta datuekin taula hau eratu da (salmentak milaka eurotan):

Salmentak	Egunak
0-2	80
2-4	160
4-6	114
6-8	98
8-10	43

Azken 10 egunetako salmentak ere jaso dira:

3.4-4.5-2.1-6.6-5.4-2.3-3.8-4.5-4.1-3

Egin beharreko atazak:

- Azken 10 egunetan batezbestez azken bi urteotan baino gehiago edo gutxiago saldu den aztertu.
- Desbideratze estandarretik abiatu, aztertu azken 10 egunetako salmentak azken bi urteotakoak baino sakabana-tuagoak diren aztertu.
- Aurreko ataleko ondorioei buruz, haietan egon daitezkeen erroreak aipa itzazu.
- Azken 10 egunetako salmenten populazio-bariantza eta lagin-bariantza kalkulatu, eta azaldu zein den bien arteko diferentzia.

II. ebazkizuna (2.5 puntu)

Jostailu denda batean, datu hauek jaso dira bezeroei buruz:

- jostailua erosten dutenak aita-amak (X) edo aiton-amona diren (XX);
- erosi duten jostailu mota: puzzleak eta mahai jokoak (A), eraikuntza jokoak (B) eta panpinak eta beste jostailu sinbolikoak (C).

Jasotako datuak hauek dira:

X/XX?	X	X	X	X	X	X	X	X	X	X
A/B/C?	A	A	A	A	B	B	B	C	C	C
X/XX?	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX
A/B/C?	C	C	C	C	C	B	B	B	A	A

Egin beharreko atazak:

- Kontingentzia-koefizientea kalkulatu eta interpretatu.
- Lambda kalkulatu eta interpretatu.
- Ahal izanez gero, gamma kalkulatu eta interpretatu.
- B eta C jostailuak D kategoria bakar batean bilduz, bi aldagaien arteko korrelazio-koefiziente lineala kalkulatu eta interpretatu.

III. ebazkizuna (2.5 puntu)

Enpresa batek erosten dituen A, B eta C lehengaien prezioak (tonako) eta haietatik kontsumitutako kopuruak (tonatan) jaso dira urte batzuetan zehar:

Urtea ↓ / Prezioa eta kopuruak →	p_A	q_A	p_B	q_B	p_C	q_C
2011	4	10	2	22	6	18
2012	5	12	2.3	24	7	20
2013	7	11	3	26	8	22
2014	6	14	3.5	25	9	21

Egin beharreko atazak:

- Kalkulatu Laspeyres prezio-indizeak 2011 oinarri urtea harturik.
- Kalkulatu Paasche prezio-indizeak 2011 oinarri urtea harturik.
- Kalkulatu 2012 urtean Laspeyres-en araberako prezio igoeran hiru lehengaiak izan duten partaidetza.
- Deflaktazioa*: 2015 eta 2016 urteetarako Laspeyres-en prezio-indizeak 170 eta 190 izango direla aurreikusten bada, zenbat izan behar da kontsumo nominala urte horietan kontsumo erreala 2014ko kontsumoaren berdina izateko?

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2018ko ekainaren 28a, 10:00

Iraupena: Ordu t'erdi.

Praktika: Ebazpena

I. EBAZKIZUNA

KILAUUA surf dendako azken bi urteotako eguneko salmenta datuekin taula hau eratu da (salmentak milaka eurotan):

Salmentak	Egunak
0-2	80
2-4	160
4-6	114
6-8	98
8-10	43

Azken 10 egunetako salmentak ere jaso dira:

3.4-4.5-2.1-6.6-5.4-2.3-3.8-4.5-4.1-3

Egin beharreko atazak:

- Azken 10 egunetan batezbestez azken bi urteotan baino gehiago edo gutxiago saldu den aztertu.
- Desbideratze estandarretik abiatu, aztertu azken 10 egunetako salmentak azken bi urteotakoak baino sakabanatuagoak diren aztertu.
- Aurreko ataleko ondorioei buruz, haietan egon daitezkeen erroreak aipa itzazu.
- Azken 10 egunetako salmenten populazio-bariantza eta lagin-bariantza kalkulatu, eta azaldu zein den bien arteko diferentzia.

(a)

Azken bi urteotako batezbestekoa kalkula dezagun:

x	n	nx
1	80	80
3	160	480
5	114	570
7	98	686
9	43	387
	495	2203

$$\bar{x}_{2 \text{ urte}} = \frac{2203}{495} = 4.45$$

Azken 10 egunetako batezbestekoa, orain:

$$\bar{x}_{10 \text{ egun}} = \frac{3.4 + 4.5 + 2.1 + 6.6 + 5.4 + 2.3 + 3.8 + 4.5 + 4.1 + 3}{10} = 3.97$$

Beraz, azken 10 egunetan azken bi urteotan baino gutxiago saldu batezbestez, zehazkiago $4.45 - 3.97 = 0.48 = 480\text{€}$ gutxiago.

(b)

Desbideratze estandarra kalkulatu behar da bi datu-multzoetarako. Azken 2 urteotako datuetarako:

x	n	nx	nx^2
1	80	80	80
3	160	480	1440
5	114	570	2850
7	98	686	4802
9	43	387	3483
	495	2203	12655

$$s_{2 \text{ urte}} = \sqrt{\frac{12655}{495} - 4.45^2} = 2.4$$

Azken 10 egunetarako:

$$s_{10 \text{ egun}} = \sqrt{\frac{3.4^2 + 4.5^2 + 2.1^2 + 6.6^2 + 5.4^2 + 2.3^2 + 3.8^2 + 4.5^2 + 4.1^2 + 3^2}{10}} - 3.97^2 = 1.45$$

Sakabanatzeak alderatzeko, desbideratze estandarrean oinarriturik, aldakortasun koefizientea kalkulatu behar da:

$$A_{2 \text{ urte}} = \frac{2.4}{4.45} = 0.54 ; A_{10 \text{ egun}} = \frac{1.45}{3.97} = 0.36$$

Sakabanatzea txikiagoa azken 10 egunetan, eta beraz salmentak homogenoagoak direla esan daiteke.

(c)

Tartetan bidutako datu-multzoari buruz, tarteko erdipuntua hartzean errore bat sortzen da, jatorrizko datuak orokorrean ez baitatoz bat erdiko puntu horrekin. Horri lagin-errorea gehitu behar zaio, datu multzo batean zein bestean, hartutako datuek laginak osatzen dituztenez, populazio osoa erabateko zehaztasunez islatzen ez dutelako.

(d)

Populazio-bariantza hau da:

$$s_{10 \text{ egun}}^2 = \frac{3.4^2 + 4.5^2 + 2.1^2 + 6.6^2 + 5.4^2 + 2.3^2 + 3.8^2 + 4.5^2 + 4.1^2 + 3^2}{10} - 3.97^2 = 2.11$$

Lagin-bariantza hau da:

$$\hat{s}_{10 \text{ egun}}^2 = \frac{10}{9} \times 2.11 = 2.34$$

Lagin bariantzak (\hat{s}^2) bariantzaren kalkuluan dagoen errorea kontuan hartu eta zuzendu egiten du, populazioaren bariantzaren estimazioa zehatzagoa eginez. Populazio bariantzak (s^2) eskuartean ditugun datuak populazioa balira bezala kalkulatu du bariantza, lagin errorea kontuan hartu gabe.

II. ebazkizuna (2.5 puntu) Jostailu denda batean, datu hauek jaso dira bezeroei buruz:

- jostailua erosten dutenak aita-amak (X) edo aiton-amona diren (XX);
- erosi duten jostailu mota: puzzleak eta mahai jokoak (A), eraikuntza jokoak (B) eta panpinak eta beste jostailu sinbolikoak (C).

Jasotako datuak hauek dira:

X/XX?	X	X	X	X	X	X	X	X	X	X
A/B/C?	A	A	A	A	B	B	B	C	C	C
X/XX?	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX
A/B/C?	C	C	C	C	C	B	B	B	A	A

Egin beharreko atazak:

- Kontingentzia-koefizientea kalkulatu eta interpretatu.
- Lambda kalkulatu eta interpretatu.
- Ahal izanez gero, gamma kalkulatu eta interpretatu.
- B eta C jostailuak D kategoria bakar batean bilduz, bi aldagaien arteko korrelazio-koefiziente lineala kalkulatu eta interpretatu.

(a)

Lehenbizi datuei dagokien kontingentzia-taula eratu behar da:

Nor? (↓) / Jostailua (→)	A	B	C	Guztira
X	4	3	3	10
XX	2	3	5	10
Guztira	6	6	8	20

Hemendik khi karratu eman eta interpretatzeko kalkuluak egiten dira:

- Lehenik maiztasun teorikoak (t) ematen dira:

Nor? (↓) / Jostailua (→)	A	B	C	Guztira
X	3	3	4	10
XX	3	3	4	10
Guztira	6	6	8	20

- Orain khi-karratu ekarpenak ($\frac{(e-t)^2}{t}$) kalkulatu dira:

Nor? (↓) / Jostailua (→)	A	B	C
X	0.33	0	0.25
XX	0.33	0	0.25

- Segidan, $\frac{(e-t)}{\sqrt{t}}$ balioak, khi karraturen eta bere koefizientearen interpretazionerako:

Nor? (↓) / Jostailua (→)	A	B	C
X	0.57	0	-0.5
XX	-0.57	0	0.5

Orain, khi-karratu kalkulatu dugun:

$$\chi^2 = 0.33 + 0 + 0.25 + 0.33 + 0 + 0.25 = 1.16$$

Orain, kontingentzia koefizientea:

$$C = \sqrt{1.331.33 + 20} = 0.25$$

Interpretazionerako, koefiziente maximoa bilatu behar da:

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

Hortik, $0.25/0.84 = 0,297$ eta beraz asoziazioa ahula dela esan behar da, antzeko ikerketen eta lagin errearen erreserbapean, ertaina izatetik gertu dagoen arren.

Asoziazioaren norabideri buruz, esango genuke, azken taula ikusirik, gurasoek puzzleak eta mahai-jokoak erosten dituztela, eta aiton-amonek jostailu sinbolikoak.

(b)

Aldagai dependentea jostailua da. Eroslearen aldagai kontuan hartu gabe, erosi beharrekoa jostailu sinbolikoa dela aurreikusiko genuke (8), baina horrela $6+6=12$ errore egingo genituzke.

Eroslea kontuan harturik berriz:

- Erosleak gurasoak badira, jostailua A dela esango genuke (4) baina horrela $3+3=6$ errore egingo genituzke.
- Erosleak aiton-amonak badira, jostailua C dela esango genuke (5) baina horrela $2+3=5$ errore egingo genituzke.

Beraz, eroslea kontuan harturik guztira $6+5=11$ errore egingo genituzke.

Eta horrela, eroslea kontuan harturik kontuan hartu gabe baino $12-11=1$ errore gutxiago egingo genuke.

Hartara, $\lambda = 1/12 = 0.083$, eta horrekin asoziazioa ahula dela esango genuke, arestiko erreserba berdinekin.

(c)

Goodman eta Kruskal-en gamma ezin da kalkulatu aldagaiak ordinalak ez direlako.

(d)

Kontingentzia-taula berregiten dut:

Nor? (\downarrow) / Jostailua (\rightarrow)	A	D	Guztira
X	4	6	10
XX	2	8	10
Guztira	6	14	20

Honela kodetzen dugu: X gurasoak: 0; XX aiton-amonak: 1, A puzzleak eta mahai-jokoak: 0, D beste jostailuak: 1.

Kalkulurako taula:

x	y	n	nx	ny	nx^2	ny^2	nxy
0	0	4	0	0	0	0	0
0	1	6	0	6	0	6	0
1	0	2	2	0	2	0	0
1	1	8	8	8	8	8	8
		20	10	14	10	14	8

$$\bar{x} = \frac{10}{20} = 0.5 ; \bar{y} = \frac{14}{20} = 0.7$$

$$s_x^2 = \frac{10}{20} - 0.5^2 = 0.25 ; s_y^2 = \frac{14}{20} - 0.7^2 = 0.21$$

$$s_{xy} = \frac{8}{20} - 0.5 \times 0.7 = 0.05$$

$$r_{xy} = \frac{0.4}{\sqrt{0.25}\sqrt{0.21}} = 0.218$$

Korrelazio ahula da, baina bere ahuldade horretan positiboa, hau da zeroak zeroekin doaz (gurasoak puzzleekin) eta batekoak batekoekin (aiton-amonak beste jostailuekin).

III. ebazkizuna (2.5 puntu) Enpresa batek erosten dituen A, B eta C lehengaien prezioak (tonako) eta haietatik kontsumitutako kopuruak (tonatan) jaso dira urte batzuetan zehar:

Urtea ↓ / Prezioa eta kopuruak →	p_A	q_A	p_B	q_B	p_C	q_C
2011	4	10	2	22	6	18
2012	5	12	2.3	24	7	20
2013	7	11	3	26	8	22
2014	6	14	3.5	25	9	21

Egin beharreko atazak:

- (a) Kalkulatu Laspeyres prezio-indizeak 2011 oinarri urtea harturik.
- (b) Kalkulatu Paasche prezio-indizeak 2011 oinarri urtea harturik.
- (c) Kalkulatu 2012 urtean Laspeyres-en arabera prezio igoeran hiru lehengaiak izan duten partaidetza.
- (d) Deflaktazioa: 2015 eta 2016 urteetarako Laspeyres-en prezio-indizeak 170 eta 190 izango direla aurreikusten bada, zenbat izan behar da kontsumo nominala urte horietan kontsumo erreala 2014ko kontsumoaren berdina izateko?

(a)

- $L_{p,2011}^{2011} = 100$
- $L_{p,2011}^{2012} = \frac{5 \times 10 + 2.3 \times 22 + 7 \times 18}{4 \times 10 + 2 \times 22 + 6 \times 18} = \frac{226.6}{192} = 1.18$
- $L_{p,2011}^{2013} = \frac{7 \times 10 + 3 \times 22 + 8 \times 18}{4 \times 10 + 2 \times 22 + 6 \times 18} = \frac{280}{192} = 1.45$
- $L_{p,2011}^{2014} = \frac{6 \times 10 + 3.5 \times 22 + 9 \times 18}{4 \times 10 + 2 \times 22 + 6 \times 18} = \frac{299}{192} = 1.55$

Horren arabera, 2012, 2013 eta 2014 urteetan prezioak %18, %45 eta %55 igo dira hurrenez hurren 2011 urteari buruz.

(b)

- $P_{p,2011}^{2011} = 100$
- $P_{p,2011}^{2012} = \frac{5 \times 12 + 2.3 \times 24 + 7 \times 20}{4 \times 12 + 2 \times 24 + 6 \times 20} = \frac{255.2}{216} = 1.18$
- $P_{p,2011}^{2013} = \frac{7 \times 11 + 3 \times 26 + 8 \times 22}{4 \times 11 + 2 \times 26 + 6 \times 22} = \frac{331}{228} = 1.45$
- $P_{p,2011}^{2014} = \frac{6 \times 14 + 3.5 \times 25 + 9 \times 21}{4 \times 14 + 2 \times 25 + 6 \times 21} = \frac{360.5}{232} = 1.55$

Horren arabera, 2012, 2013 eta 2014 urteetan prezioak %18, %45 eta %55 igo dira hurrenez hurren 2011 urteari buruz.

Laspeyres eta Paasche indizeek antzeko balioak, dezimalak gorabehera ia berdinak, ematen dituzte prezio eta kopuru horietarako.

(c)

Erreperkutsioak eman ditzagun:

- $R_{A,p,2011}^{2012} = \frac{(5 - 4) \times 10}{4 \times 10 + 2 \times 22 + 6 \times 18} = \frac{10}{192} = 0.052$
- $R_{B,p,2011}^{2012} = \frac{(2.3 - 2) \times 22}{4 \times 10 + 2 \times 22 + 6 \times 18} = \frac{10}{192} = 0.034$
- $R_{C,p,2011}^{2012} = \frac{(7 - 6) \times 18}{4 \times 10 + 2 \times 22 + 6 \times 18} = \frac{10}{192} = 0.09375$

Eta haietatik partaidetzak, kontuan harturik igoera orokorra 18 puntukoa izan dela:

- $P_{A,p,2011}^{2012} = \frac{0.052}{0.18} = 0.29 = \%29$
- $P_{B,p,2011}^{2012} = \frac{0.034}{0.18} = 0.19 = \%19$
- $P_{C,p,2011}^{2012} = \frac{0.09375}{0.18} = 0.52 = \%52$

2012ko prezio igoera orokorrean hiru produktuek izan dituzten partaidetzak %29, %19 eta %52koa izan dira, hurrenez hurren.

(d)

2014ko kontsumo nominala $6 \times 14 + 3.5 \times 25 + 9 \times 21 = 360.5$ izan da.

2015 eta 2016ko prezio igoerak 2014 urteari buruz $170/155=1.096$ (9.6 puntu) eta $190/155=1.22$ (22 puntu) dira.

Kontsumo erreala berdina izateko 2015 eta 2016ean, 2014ko urteari buruz, nahikoa dugu 2014ko kontsumoari prezio igoera horiek aplikatu, ez gehiago ez gutxiago:

- 2015 urterako, $360.5 \times 1.096 = 395.1\text{€}$
- 2015 urterako, $360.5 \times 1.22 = 439.8\text{€}$

ESTATISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2018ko ekainaren 28a, 10:00

Testa

Iraupena: 40 minutu

Erantzun bakarra da zuzena galdera bakoitzean. Guztira testak 2.5 puntu balio du. Erantzun zuzenak 0.125 puntu balio du. Erantzun oker bakoitzak 0.0625 puntu kentzen du. Galderak erantzun gabe utz daitezke, punturik gehitu eta kendu gabe.

- (1) Unibertsitate bateko ERASMUS ikasleek aukeratutako herrialdeak jaso behar dira. Zein da ikerketa-unitatea?
- Ikaslea
 - Herrialdea
 - Unibertsitatea
 - Aurreko guztiak faltsuak dira.
- (2) Unibertsitate bateko ERASMUS ikasleek aukeratutako herrialdeak jaso dira. Zein dira datu horiek irudikatzeko aukeran ditugun diagramak?
- Histograma
 - Barra-diagrama
 - Sektore-diagrama
 - (b) eta (c) dira zuzenak
- (3) 2016 eta 2017 urteetan izandako eguneko ekoizpenak jaso dira lantoki batean. Zein da horiek alderatzeko grafiko egokiena?
- Puntu-diagrama bikoitza
 - Barra-diagrama bikoitza
 - Sektore-diagrama bikoitza
 - Ondoz ondoko edo back-to-back histogramak.
- (4) 80 ikasleren artean honako datu hauek jaso dira:

Kalifikazioak	Ikasleak
0-2	8
2-4	16
...	...

Kalkulatu 2. dezila.

- 3.25
 - 3
 - 2.75
 - Emandako datuekin ezin da kalkulat.
- (5) 2001 urtean salmentak 1200 izan ziren. 2005 urtean 1600 izan ziren. Bitarte horretan, zenbat da urteko batez besteko hazkunde tasa?
- 6.45
 - 7.45
 - 8.45
 - 9.45

- (6) Gaileta baten pisu estandarra: 50 gr. Hiru gaileta hauen pisuak jaso dira: 48-55-60. Kalkulatu batez besteko errorea.
- (a) 6.55
 - (b) 7.55
 - (c) 8.55
 - (d) 9.55
- (7) Zein da batez besteko errendimenduak kalkulatzeko batezbesteko egokiena?
- (a) Geometrikoa
 - (b) Koadratikoa
 - (c) Harmonikoa
 - (d) Haztatua, pieza kopuruaren arabera.
- (8) Muturreko datuen eragina mugatu nahiko bazenu, neurri hauetatik zein baztertuko zenuke?
- (a) Desbideratze estandarra
 - (b) Koartil arteko ibiltartea
 - (c) DAME
 - (d) Bowley koefizientea
- (9) Lorenz kurba batean, $p_{20} = 0.2, q_{20} = 0.05; p_{80} = 0.8, q_{80} = 0.7$. Zenbateko errenta hartzen du %20 aberatsenak?
- (a) %30
 - (b) %40
 - (c) %80
 - (d) %95
- (10) Errenta hauek jaso dira: 2-4-6-8-10. Mediala kalkulatu.
- (a) 4
 - (b) 6
 - (c) 8
 - (d) 10
- (11) Errenta hauek jaso dira: 2-4-6-8-10. Robin Hood adierazlea kalkulatu.
- (a) 0.8
 - (b) 0.6
 - (c) 0.4
 - (d) 0.2
- (12) Zein tarteko balioak hartzen ditu \mathbf{X}^2 estatistikoak?
- (a) 0 eta 1 artean
 - (b) -1 eta 1 artean
 - (c) Balio ez negatiboak.
 - (d) Edozein balio.
- (13) 10 egunetan zehar lantoki bateko ekoizpena eta kostu totala jaso dira. Behaketa horietatik informazio hau ematen da: $\sum y_i = 50, \sum y_i^2 = 1000$ eta $\sum e_i^2 = 200$. Karratu txikiaren erregresio eredu lineal batean, zenbat da aurrean bariantza?
- (a) 55
 - (b) 80
 - (c) 150
 - (d) 800
- (14) 10 egunetan zehar lantoki bateko ekoizpena eta kostu totala jaso dira. Behaketa horietatik informazio hau ematen da: $\sum y_i = 50, \sum y_i^2 = 1000, \sum e_i^2 = 200$. Karratu txikiaren erregresio eredu lineal batean, zenbat da mugatze-koefizientea?
- (a) 0.20
 - (b) 0.25
 - (c) 0.266
 - (d) 0.275

- (15) Zer da faltsua karratu txikiaren erregresio zuzenari buruz?
- (a) Aurresanen bariantza bariantza totala baino txikiagoa da.
 - (b) Aurresanen bariantza erroreen bariantza baino handiagoa da.
 - (c) Erroreen batura 0 da.
 - (d) Aurresanen batezbestekoa eta y balioen batezbestekoa berdinak dira.
- (16) Eredu batukor batean, urtarrileko urtarokotasun-koefizientea 1.25 suertatu da. Urtarril batean salmentak 100 izan ziren. Salmenta horiek desestazionalizatu.
- (a) 75
 - (b) 80
 - (c) 98.75
 - (d) 125
- (17) Zer da egiazkoa leunketa esponentzial batean?
- (a) Alfa parametroa zenbat eta handiagoa, orduan eta motelago eguneratzen dira leunketa-balioak.
 - (b) Serie egonkor batean gorakada nabarmena izan da, mantentzeko joera duena. Gorakada ondorengo aurresanak egoki egiteko alfa handia behar da.
 - (c) Iraunkorrak izango ez diren xokak daudenean, hobe da alfa handia izatea.
 - (d) Alfa zenbat eta handiagoa, orduan eta serie leunagoa izango da.
- (18) Kredituaren bolumen global handienak duela 12, 25, 35 eta 47 urte izan ziren. Zein osagairi dagozkio mugimendu horiek?
- (a) Joerari
 - (b) Zikloari
 - (c) Urtarokotasunari
 - (d) Hondarrari
- (19) Zein kasutan gaude eredu normaletik gertuago?
- (a) Bowley koefizientea=1, Pearson kurtosi-koefizientea=2
 - (b) Bowley koefizientea=0.5, Pearson kurtosi-koefizientea=3
 - (c) Fisher koefizientea=-0.5, Pearson kurtosi-koefizientea=3.16
 - (d) Fisher koefizientea=-0.5, Pearson kurtosi-koefizientea=1.6
- (20) Zerekin lotu behar da Sturges izena?
- (a) Kaxa-diagramarekin.
 - (b) Aniztasun-indize batekin.
 - (c) Lehen azterketa demografikoarekin.
 - (d) Histogramako tarte kopurua kalkulatzeko erregela batekin.

Estatistika eta datuen analisia

-
2018ko ekainaren 28a**Izena eta abizenak:** Josemari Sarasola

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

KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0