

# ESTADÍSTICA ETA DATUEN ANALISIA

Azterketa ebatziak

2020-2021 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea. EHU

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

[gizapedia.hirusta.io](http://gizapedia.hirusta.io)

## ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2021eko urtarrilaren 14an, 15:00

Iraupena: 90 min

**I. ebazkizuna: Kontzentrazioa** (2.5 puntu)

2015 eta 2020 urteetan, herri bateko familien errentak jaso dira:

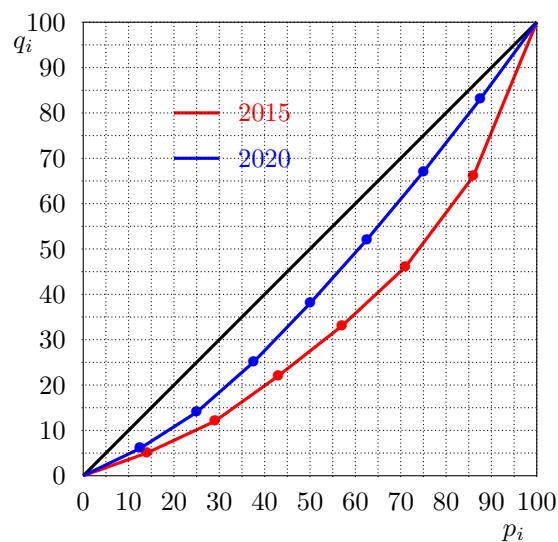
2015:	12	34	87	52	27	19	26	
2020:	25	31	56	64	60	46	52	67

**Egin beharreko atazak:**

- Bi urteetako Lorenzen kurbak eratu eta, horretan oinarrituta, azter ezazu zein urtetan izan zen errenten banaketa kontzentratuen.
- Giniren indizea kalkulatu eta interpretatu bi urteetarako.
- Kalkulatu mediana 2015 eta 2020 urteetarako, eta emaitzak interpretatu (zein urtetan izan zen kontzentrazio handiena).
- Kalkulatu Robin Hood adierazlea 2015 urterako.
- Kalkulatu  $Q_3$  eta  $D_4$  kuantilak 2015 urterako.

(a)

2015: $i$	2015: $p_i$	2015: $x_i$ ord.	2015: $x_i$ met	2015: $q_i$	2020: $i$	2020: $p_i$	2020: $x_i$	2020: $x_i$ met	2020: $q_i$
1	1/7=0.14	12	12	12/257=0.05	1	0.125	25	25	0.06
2	2/7=0.29	19	31	31/257=0.12	2	0.25	31	56	0.14
3	0.43	26	57	0.22	3	0.375	46	102	0.25
4	0.57	27	84	0.33	4	0.5	52	154	0.38
5	0.71	34	118	0.46	5	0.625	56	210	0.52
6	0.86	52	170	0.66	6	0.75	60	270	0.67
7	1	87	257	1	7	0.875	64	334	0.83
-	-	-	-	-	8	1	67	401	1
-	3 (*)	257	-	-	-	3.5(*)	401	-	-

(\*)  $p_i$  baturak (azken 1 balioa kanpo utziz) hurrengo atalean Gini indizea kalkulatzeko ematen dira.

2015eko Lorenz kurba 2020 urtekoaren azpitik doanez ibilbide osoan, 2015 urtean kontzentrazio handiagoa izan zela ondorioztatu behar da.

(b)

Gini indizea kalkulatzeko, aski da aurreko taulatik  $p_i - q_i$  diferentziak kalkulatu eta batu. Eta aurreko taulan,  $p_i$  balioen batuketa egiten dugu (azken 1 balioa ezik).

Aurreikus dezakegu 2015eko Gini indizea handiagoa izango dela, diagonalera bitarteko azalera handiagoa delako urte horretan.

2015: $p_i - q_i$	2020: $p_i - q_i$
0.14-0.05= 0.09	0.125-0.06=0.06
0.17	0.11
0.21	0.12
0.24	0.11
0.25	0.10
0.20	0.07
0	0.04
-	0
1.16	0.61

$$G_{2015} = \frac{1.16}{3} = 0.386 ; G_{2016} = \frac{0.61}{3.5} = 0.174$$

Aurreikus bezala, 2015 urtean kontzentrazioa handiagoa dela baieztatu daiteke.

(c)

Mediala kalkulatzeko, errentak metatu behar dira, ordenatu ondoren. Hori ja egina dago aurreko ataletan.

Gero, errenta total erdia kalkulatzeko: 2015ean,  $257/2=128.5$ , eta 2020an, 200.5.

Errenta metatuetan, errenta erdi horiek zein errentetarako eskuratzen diren ikusten dugu:

$$Ml(2015) = 52 ; Ml(2020) = 56$$

Kontzentrazio-neurri gisa hartzeko, medianarekin alderatzen dugu, modu erlatiboan. Medianak kalkulatzeko ditugu:

$$Me(2015) = 27 ; Me(2020) = \frac{52 + 56}{2} = 54$$

Eta orain mediana horiekin alderatzen ditugu medialak:

$$(Ml - Me)/Me_{2015} = (52 - 27)/27 = 0.92 ; (Ml - Me)/Me_{2020} = (56 - 54)/54 = 0.04$$

Beraz, argi dago 2015ean kontzentrazioa handiagoa izan zela.

(d)

Robin Hood adierazlea kalkulatzeko, batez besteko errenta kalkulatzeko:

$$\bar{x}_{2015} = \frac{257}{7} = 36.7$$

Eta orain errentak handienetik txikienera ordenatzen ditugu (aberatsenei kentzeko), batez besteko errentara heldu arte:

2015: $x_i - \bar{x}$
87-36.7=50.3
52-36.7=15.3
65.6

Beraz,

$$RH(2015) = \frac{65.6}{257} = 0.255$$

(e)

- $Q_3 = P_{75} = 0.75 \times 7 = 5.25$ garren datua

5. eta 6. datuak hartzen ditugu:

$$Q_3 = 34 \times 0.75 + 52 \times 0.25 = 38.5$$

- $D_4 = P_{40} = 0.40 \times 7 = 2.8$ garren datua

2. eta 3. datuak hartzen ditugu:

$$D_4 = 19 \times 0.2 + 26 \times 0.8 = 24.6$$

## II. ebazkizuna: Korrelazioa eta Cronbach-en alfa (2.25 puntu)

Matematika-test batean 5 galdera daude. Erantzun zuzenek puntu bana ematen dute, eta okerrak 0 puntu. Ondoren, ikasle batzuen erantzunak agertzen dira.

Ikaslea →	A	B	C	D	E	F	G	H	I	J
1. itema	1	1	1	1	1	0	1	0	0	1
2. itema	1	1	1	0	1	0	0	0	0	0
3. itema	0	1	0	1	1	1	0	1	0	1
4. itema	1	0	1	0	1	1	1	0	1	1
5. itema	1	1	1	1	1	1	1	1	0	0

Item-test korrelazioak (ongi:1;gaizki:0) ere kalkulatu dira:

Item zenbakia	1	2	3	4	5
r(item,puntuazioa)	0.73	0.80	?	0.12	0.56

### Egin beharreko atazak:

- Kalkulatu eta interpretatu 3. itemeko erantzunen eta testeko puntuazio totalaren arteko korrelazio koefizientea. Galdera egokia al da testerako?
- Kalkulatu Cronbach-en alfa eta interpreta ezazu.
- Testetik galdera bat ezabatu behar izatekotan zein ezabatuko zenuke?

(a)

Jar ditzagun lehen bi zutabeetan bi aldagaien datuak: lehenean, itemeko datuak (enuntziatutik) eta bigarrean puntuazio totala (itemeko puntuazioak gehituz, besterik gabe, enuntziatuan esaten den bezala). Ondoren, korrelazio koefizientea kalkulatzeko dugu:

$x$ item	$y$ test	$xy$	$x^2$	$y^2$
0	4	0	0	16
1	4	4	1	16
0	4	0	0	16
1	3	3	1	9
1	5	5	1	25
1	3	3	1	9
0	3	0	0	9
1	2	2	1	4
0	1	0	0	1
1	3	3	1	9
6	32	20	6	114

$$\bar{x} = \frac{6}{10} = 0.6 ; \bar{y} = \frac{32}{10} = 3.2$$

$$s_{xy} = \frac{20}{10} - 0.6 \times 3.2 = 0.08$$

$$s_x = \sqrt{\frac{6}{10} - 0.6^2} = 0.49$$

$$s_y = \sqrt{\frac{114}{10} - 3.2^2} = 1.07$$

$$r_{xy} = \frac{0.08}{0.49 \times 1.07} = 0.1525$$

Ez da galdera egokia testerako, zuzen erantzuteak ez baitago loturik testean puntuazio altua lortzearekin, korrelazioa ahula baita.

(b)

Cronbach-en alfa kalkulatzeko behar ditugu: testeko bariantza (kalkulatuta daukagu aurreko ataletik) eta itemeko bariantzak (kalkulatu behar ditugu, 3gna ezik, aurreko atalean kalkulatu dugulako). Beraz, 1., 2., 4. eta 5. itemetako bariantzak kalkulatu behar ditugu:

$x_1$	$x_2$	$x_4$	$x_5$	$x_1^2$	$x_2^2$	$x_4^2$	$x_5^2$
1	1	1	1	1	1	1	1
1	1	0	1	1	1	0	1
1	1	1	1	1	1	1	1
1	0	0	1	1	0	0	1
1	1	1	1	1	1	1	1
0	0	1	1	0	0	1	1
1	0	1	1	1	0	1	1
0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	0
1	0	1	0	1	0	1	0
7	4	7	8	7	4	7	8

$$s_{x1}^2 = \frac{7}{10} - \left(\frac{7}{10}\right)^2 = 0.21$$

$$s_{x2}^2 = \frac{4}{10} - \left(\frac{4}{10}\right)^2 = 0.24$$

$$s_{x4}^2 = \frac{7}{10} - \left(\frac{7}{10}\right)^2 = 0.21$$

$$s_{x5}^2 = \frac{8}{10} - \left(\frac{8}{10}\right)^2 = 0.16$$

Eta aurreko ataletik:  $s_{x3} = 0.49^2 = 0.24$  ;  $s_T^2 = 1.07^2 = 1.14$

Eta orain, Cronbach-en alfa (kontuz, aurreko ataleko desbideratzeak bariantza bihurtu behar ditugu, ber 2 eginez):

$$\alpha = \frac{5}{5-1} \left[ 1 - \frac{0.21 + 0.24 + 0.24 + 0.21 + 0.16}{1.14} \right] = 0.09$$

0.7 baino txikiagoa denez, testak orokorrean barne koherentziarik ez duela baieztatu daiteke.

(c)

Galderak ezabatzen hasita, korrelazio balio txikieneko ezabatu genuke, kasu honetan 4. galdera.

### III. ebazkizuna: Probabilitate-zuhaitza (2.25 puntu)

Lantegi batean makina bakarra aritzen da egun batean. 1, 2 edo 3 pieza ekoitzi ditzake 0.2, 0.5 eta 0.3ko probabilitateaz hurrenez hurren, uniformeki eguneko 24 orduetan zehar. Pieza bat edo bi ekoizten baditu, biharamonean makina gehigarri bat jarriko da martxan, bezperako makinaren kondizio berdinetan. 3 pieza ekoizten badira azkenik lehen egunean, biharamonean ez da makina berririk jarriko eta ekoizpenak lehen eguneko emaitza berdina izan ditzake, probabilitate berdinekin.

#### Egin beharreko atazak:

- Probabilitate zuhaitza taula moduan eratu, adar (lerro) bakoitzeko probabilitateekin batera. Oharra: Taula ematen da 36 lerroekin, baina horrek ez du esan nahi denak erabili behar dituzunik. Taulako lerro kopurua 36 edo gutxiago da.
- Lerro bakoitzaren bukaeran 3 pieza egiteko behar den denbora adieraz ezazu
- 1-2-3 konbinazioan 3 pieza egiteko behar den denboraren kalkulua azaltzeko eskema simple bat egin ezazu.
- 3 pieza egiteko behar den batezbesteko denbora kalkulatu, taulako emaitzak taula trinko batean bilduz.

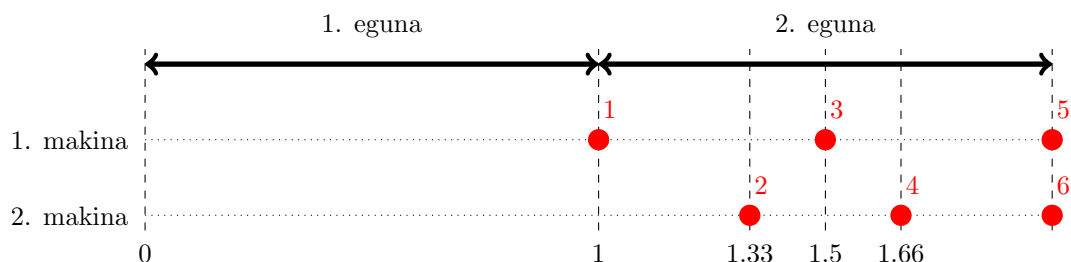
(a),(b)

konbin.	1.eguna:1.makina	prob	2.eguna:1.makina	prob	2.eguna:2.makina	prob	prob osoa	3 piezako denbora
1	1	0,2	1	0,2	1	0,2	0,008	2
2	1	0,2	1	0,2	2	0,5	0,02	2
3	1	0,2	1	0,2	3	0,3	0,012	1.66
4	1	0,2	2	0,5	1	0,2	0,02	2
5	1	0,2	2	0,5	2	0,5	0,05	1,5
6	1	0,2	2	0,5	3	0,3	0,03	1,5
7	1	0,2	3	0,3	1	0,2	0,012	1,66
8	1	0,2	3	0,3	2	0,5	0,03	1,5
9	1	0,2	3	0,3	3	0,3	0,018	1,33
10	2	0,5	1	0,2	1	0,2	0,02	2
11	2	0,5	1	0,2	2	0,5	0,05	1,5
12	2	0,5	1	0,2	3	0,3	0,03	1,33
13	2	0,5	2	0,5	1	0,2	0,05	1,5
14	2	0,5	2	0,5	2	0,5	0,125	1,5
15	2	0,5	2	0,5	3	0,3	0,075	1,33
16	2	0,5	3	0,3	1	0,2	0,03	1,33
17	2	0,5	3	0,3	2	0,5	0,075	1,33
18	2	0,5	3	0,3	3	0,3	0,045	1,33
19	3	0,3	1	0,2	-	-	0,06	1
20	3	0,3	2	0,5	-	-	0,15	1
21	3	0,3	3	0,3	-	-	0,09	1

1

(c)

Galdera: 1-2-3 sekuentzian, zenbat egunetara egiten da 3. pieza?



Erantzuna: 3. pieza egun t'erdira egiten da.

(d)

---

3 pieza ekoizteko egunak (x)	prob p(x)	xp(x)
1	0.06+0.15+0.09=0.30	0.30
1.33	0.273	0.36309
1.5	0.335	0.5025
1.66	0.024	0.03984
2	0.068	0.136
	1	1.3414 egun

---



#### IV. ebazkizuna (ebaluazio ez jarraitua) (1 puntu)

Maiztasun-banaketa honi buruz:

Tarteak	Maiztasun absolutuak
0-10	4
10-50	14
50-100	22
100-200	40

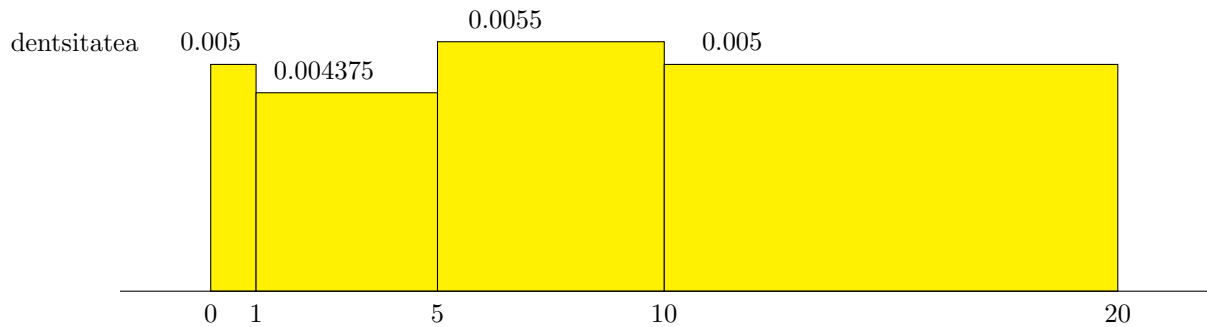
**Egin beharreko atazak:**

- Histograma marraztu eta interpretatu zentroari buruz.
- Maiztasun poligonoa marraztu.
- Ojiba marraztu.

(a): Histograma

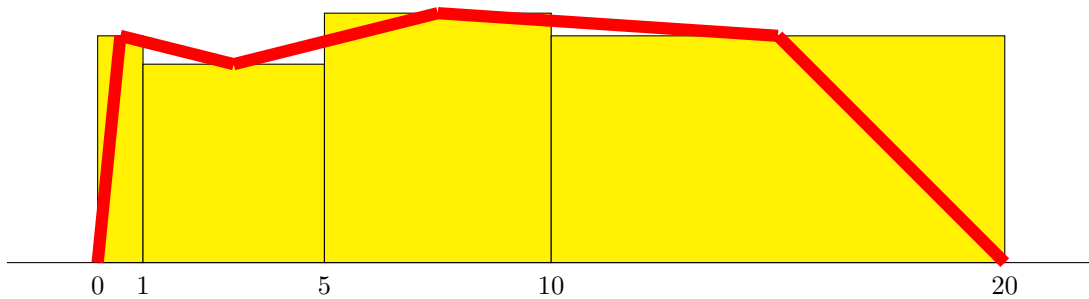
Tarte zabalerak desberdinak direnez, histogramako zutabeak tarteko dentsitatearen arabera altxatuko dira. Eman ditzagun bada, dentsitate horiek, maiztasun erlatiboak ( $f$ ) tarte zabalerarekin ( $z$ ) zatituz:

Tarteak	n	f	$d = f/z$
0-10	4	0.05	$0.05/10=0.005$
10-50	14	0.175	$0.175/40=0.004375$
50-100	22	0.275	$0.275/50=0.0055$
100-200	40	0.50	$0.50/100=0.005$
	80		



Banaketa bimodala izan arren, esan daiteke zentroa 10 inguruan dagoela, 0-20 tartean, dentsitatea nahiko konstantea delako tarte batetik bestera.

(b): Maiztasun-poligonoa

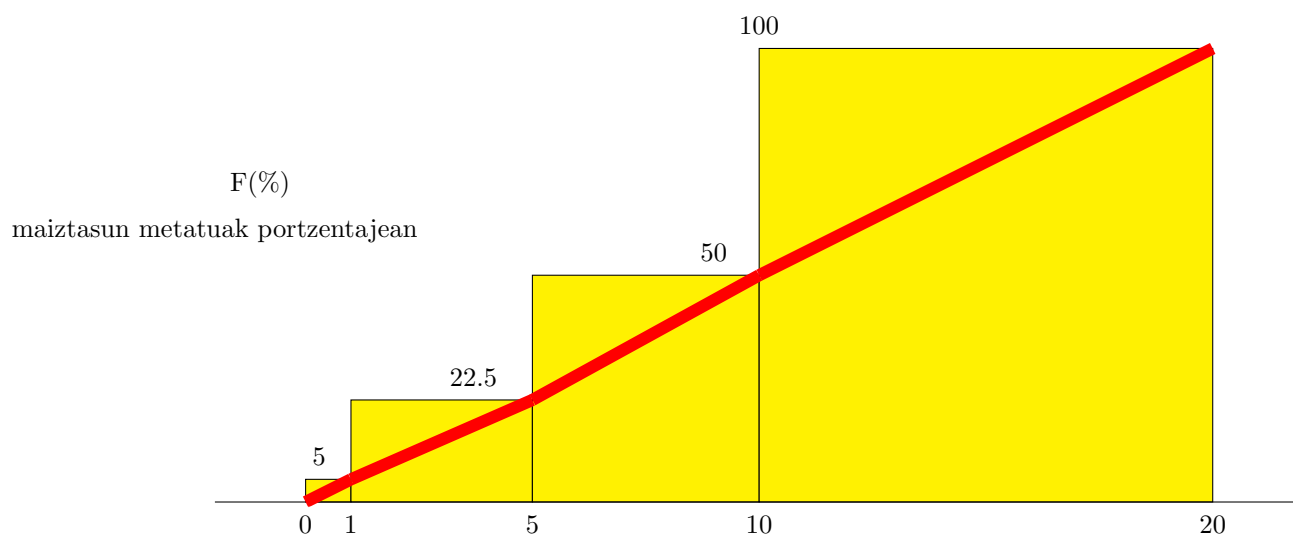


Gutxigatik ez zait atera piperpotoa - ;)

(c): Ojiba

Kasu honetan ez dira dentsitateak baliatu behar, maiztasun absolutu handiagoak eragina tarte luzeagoan zehar gehitzean "desegin" egiten delako, nolabait esateko.

Tarteak	n	f	F
0-10	4	0.05	0.05
10-50	14	0.175	0.225
50-100	22	0.275	0.500
100-200	40	0.50	1
	80		



Hemen oso ondo ikusten da, datuen banaketa nahiko konstantea dela, datuak nahiko uniformeki banatzen direla alegia, malda nahiko konstantea delako. Horrela, baieztatu dezakegu lehen atalean esandakoa: zentroa datuen eremuen erdian egongo da, 10 inguruan alegia.

## ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2021eko urtarrilaren 14a, 15:00

Iraupena: 40 minutu

**TESTA**

**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.**

- Adin eta egoera patologiko bereko bi pertsonen farmako baten dosi bat eman eta erreakzio kuantitatibo zertxobait desberdinak izan dituzte. Nola deitzen zaio teknikoki diferentzia horri?
  - Errore estokastikoa.
  - Zorizko errorea.
  - Faktore kontrolagaitzek eragindako diferentzia da.
  - Aurreko hirurak dira zuzenak.
- Honako neurri hauetatik zein da neurri sendoa?
  - Datu handiena ken datu txikiena
  - Kuartil arteko ibiltartea
  - Desbideratze estandarra
  - Aldakortasun (edo aldakuntza) koefizientea
- Zer egin behar da bi datu multzo desberdinetako unitateak alderatzeko?
  - Haien datuak estandartu.
  - Haien sakabanatze erlatiboak alderatu.
  - Cohen-en  $d$  kalkulatu.
  - Aurreko guztiak faltsuak dira.
3. kuartila 44. datua da, 30-40 tartean dago, 30 bitarteko maiztasun metatua 20 da, eta 30-40 tartean 40 datu daude. Zenbat da 3. kuartila?
  - 32
  - 34
  - 36
  - 38
- 11 datuen zuzendu gabeko desbideratzea 4 da. Zenbat da desbideratze zuzendua?
  - 3.69 (borobilduz)
  - 3.89 (borobilduz)
  - 4.19 (borobilduz)
  - 4.29 (borobilduz)
- Datuak: 1-2-3-4-4-4-5-6-7. Kalkulatu DAME.
  - 1
  - 2
  - 3
  - 4
- Kalkulurik ez duzu egin behar galdera hau erantzuteko: honako hauetatik zein da aniztasun handieneko maiztasun-banaketa?
  - A:4; B:4; C:4
  - A:3; B:3; C:3, D:3
  - A:8; B:3; C:1
  - A:6; B:2; C:2; D:2

8.  $2 \times 2$  taula batean, kontingentzia-koefizientea 0.44 suertatu da. Ikastaroan zehar emandako interpretazio erregelaren arabera eta antzeko ikerketak kontuan hartu gabe, nolako asoziazioa dago bi aldagaien artean?
- Ahula
  - Ertaina
  - Sendoa
  - Ezin da esan, taula  $2 \times 2$  izanda,  $\phi$  kalkulatu behar delako.
9. Hiritarren artean jasotako bi aldagai ordinalen arteko korrelazioa aztertzeko, datuak lurraldearen arabera banatu dira: Hego Euskal Herria eta Ipar Euskal Herrian, lagin tamaina berdinekin. Hego Euskal Herrian, gamma ( $\gamma$ ) neurria 0.87 suertatu da. Zenbat suertatu beharko litzateke gamma Euskal Herri osoan interakzio garbia eta erabatekoa dagoela baieztatu ahal izateko?
- 0.87 inguruko balio bat.
  - 0.87 inguruko balio bat.
  - 0 inguruko balio bat.
  - Ipar Euskal Herriari dagokion gamma jakin beharko genuke.
10. Aldagai biko datuak edo datu bidimentsionalak:  $(x_1 = 4.6, y_1 = 8.7); (x_2 = 6.4, y_2 = 14.7); (x_3 = 9.6, y_3 = 10.7)$ . Datu horietarako Spearman korrelazio-koefizientea kalkulatzeko eskatzen zaizu:
- 0.33
  - 0.5
  - 0.66
  - Datu horietarako ez da zuzena Spearman koefizientea kalkulatzea.
11. Bi aldagaien artean korrelazio sendoa dago, baina bien arteko erlazio kausala ez dago batere egiazaturik. Nola deitzen diozu fenomeno horri?
- Sasiko korrelazioa.
  - Interakzioa.
  - Korrelazio monotonikoa.
  - Korrelazio lerromakurra.
12. Zorizko aldagai bat honela banatzen da:  $f(x) = \frac{x}{50}$ ;  $0 < x < 10$ . Kalkulatu  $P[X > 4]$ .
- 0.54
  - 0.64
  - 0.74
  - 0.84
13. Zorizko aldagai bat honela banatzen da:  $F(x) = \frac{x^2}{100}$ ;  $0 < x < 10$ . Kalkulatu  $P[2 < X < 4]$ .
- 0.10
  - 0.12
  - 0.14
  - 0.16
14. Elur-prezipitazioa jasotzen duen gailu batek 2mm-ko zehaztasuna du (hots, 0-2-4-6-... erregistroak ematen ditu). Erregistroa aldagai jarraitutzat hartu bada, nola kalkulatu behar da erregistroa 10 baino txikiagoa izateko probabilitatea?
- $P[X < 10]$  kalkulatu.
  - $P[X < 9]$  kalkulatu.
  - $P[X < 11]$  kalkulatu.
  - $P[9 < X < 11]$  kalkulatu.
15. Zorizko aldagai bat honela banatzen da:  $f(x) = \frac{x}{50}$ ;  $0 < x < 10$ . Kalkulatu itxaropen matematikoa.
- 5.33

- (b) 6.16  
(c) 6.66  
(d) 8.33
16. Zorizko aldagai batek bi balio hartzen ditu: -1 eta 1, bakoitza 0.5eko probabilitatearekin. Zenbat da horren bariantza?  
(a) 0  
(b) 0.5  
(c) 1  
(d) 2
17. A inbertsio-aukeraren itzaropena 4 da eta bariantza 1, B inbertsio-aukeraren itzaropena 6 da eta bariantza 2. Zein da egiazkoa? Behar denean,  $\mu/\sigma^2$  erabili utilitate-funtzio gisa.  
(a) Epe luzera nahiz laburrera hobe da A  
(b) Epe luzera nahiz laburrera hobe da B  
(c) Epe luzera hobe da A eta epe laburrera B  
(d) Epe luzera hobe da B eta epe laburrera A
18. Ikasle batek 0.13ko probabilitateaz gainditzen du azterketa da. Zenbat azterketa egin behar du batezbeste gainditu arte?  
(a) 7.69  
(b) 9.69  
(c) 11.69  
(d) 13.69
19. (Egunkari saltzailearen problema) Eskari diskretu bat honela banatzen da: 2, 0.2 probabilitateaz; 3, 0.5eko probabilitateaz; 4, 0.3ko probabilitateaz. 3 unitate erosten badira, zenbat da galduko diren salmenten itzaropena?  
(a) 0.3  
(b) 0.6  
(c) 0.9  
(d) 1
20. Probabilitate-banaketa bati buruz hau baino ez dakigu:  $\mu = 120, \sigma = 10$ . Hurbildu  $P[100 < X < 150]$ .  
(a)  $\geq 0.666$   
(b)  $\geq 0.8$   
(c)  $\geq 0.833$   
(d) Beste bat da erantzun zuzena.

## Estatistika eta datuen analisisa

2021eko urtarrilaren 14a

**Izena eta abizenak:** Josemari Sarasola Ledesma

Galdera	Erantzuna
1	d
2	b
3	a
4	c
5	c
6	a
7	b
8	c
9	c
10	d
11	a
12	d
13	b
14	b
15	c
16	c
17	d
18	a
19	a
20	b

## KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0

## ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2021eko ekainaren 21ean

Iraupena: 90 min

**I. ebazkizuna** (2.5 puntu)

Hiri batean COVID19 pandemiako bigarren eta hirugarren olatuetan izandako kasu positiboen adinak jaso dira:

Adina	Positiboak (2. olatua)	Positiboak (3. olatua)
0-15	26	40
15-30	45	96
30-50	55	79
50-60	49	62
60-80	45	78
80-100	22	26

**Egin beharreko atazak:**

- Zein izan da kasu positiboen batez besteko adinaren bilakaera bigarren olatutik hirugarren olatura bitartean?
- Kalkulatu adinaren desbideratze estandarra bigarren nahiz hirugarren olatuan.
- Desbideratze estandarrean oinarrituz, zein olatutan izan zen adina sakabanatuago?
- Aztertu bi olatuetako kasu positiboen batez besteko adinaren efektuaren tamaina, Cohen-en  $d$  kalkulatu. Zein hipotesi onartu behar duzu Cohen-en  $d$  kalkulatu eta bere emaitza aintzat hartzeko?

(a)

Datuak tartetan emanak daudenez, klase-marka hartzen da kalkuluak egiteko:

$x$	$n(2)$	$n(3)$	$n(2)x$	$n(3)x$
7,5	26	40	195	300
22,5	45	96	1012,5	2160
40	55	79	2200	3160
55	49	62	2695	3410
70	45	78	3150	5460
90	22	26	1980	2340
	242	381	11232,5	16830

$$\bar{x}(2. olatua) = \frac{11232,5}{242} = 46,41 ; \bar{x}(3. olatua) = \frac{16830}{381} = 44,17$$

Kasu positiboen batez besteko adina jaitsi egin da 2. olatutik 3. olatura, 2 urte pasatxo, 46.41 urtetik 44.17 urtera.

(b)

Aurreko kasuan bezalaxe, klase-markak harturik abiatzen ditugu kalkuluak:

$x$	$n(2)$	$n(3)$	$n(2)x^2$	$n(3)x^2$
7,5	26	40	1462,5	2250
22,5	45	96	22781,25	48600
40	55	79	88000	126400
55	49	62	148225	187550
70	45	78	220500	382200
90	22	26	178200	210600
	242	381	659168,75	957600

$$s(2. olatua) = \sqrt{\frac{659168,75}{242} - 46,41^2} = 23,87 \text{ urte}$$

$$s(3. olatua) = \sqrt{\frac{957600}{381} - 44,17^2} = 23,71 \text{ urte}$$

(c)

Sakabanatzeak alderatzeko, neurri erlatibo bat eratu behar da. Desbideratze estandarretatik abiatutik, aldakuntza-koefizientea da neurri hori:

$$A(2.olatua) = \frac{23.87}{46.41} = 0.51$$

$$A(3.olatua) = \frac{23.71}{44.17} = 0.53$$

Adinen sakabanatzea zertxobait handiagoa da 3. olatuan.

(d)

Cohen-en d kalkulatzeko, eman dezagun lehenbizi desbideratze bateratua:

$$s = \sqrt{\frac{242 \times 23.87^2 + 381 \times 23.71^2}{242 + 381 - 2}} = 23.81$$

Eta orain, eman dezagun Cohen-en d:

$$d = \frac{46.41 - 44.17}{23.91} = 0.093$$

Efektuaren tamaina txikia da. Beraz, printzipioz batez besteko adinen arteko diferentzia ez da oso nabarmena.

Cohen-en d erabili ahal izateko, bi taldeetako bariantzak berdinak izan behar dira populazio mailan. Gure kasuan, baldintza hori bete egiten dela dirudi, bi bariantzak oso antzekoak direlako.



## II. ebazkizuna (2 puntu)

Administrazio publikoan deitutako oposizio batean hautagaien sexua eta oposizioko lehen fasea gainditu duten ala ez jaso da:

Gainditu? (↓) / Sexua (→)	Gizona	Emakumea
Ez	45	26
Bai	18	44

### Egin beharreko atazak:

- Kalkulatu lambda eta interpreta ezazu.
- Kalkulatu bi aldagaien arteko Pearsonen korrelazio-koefiziente lineala eta interpreta ezazu.

(a)

Aldagai independentea sexua dela kontuan harturik:

<u>sexua</u> (ald. indep.) kontuan hartu gabe	<u>sexua</u> kontuan hartuta	errore-aldea (amaiera)
→ auresana: <u>Ez</u> → erroreak: <u>62</u>	<ul style="list-style-type: none"> <li>gizona dela jakinda: → auresana: <u>Ez</u> → erroreak: <u>18</u></li> <li>emakumea dela jakinda: → auresana: <u>Bai</u> → erroreak: <u>26</u></li> <li>erroreak guztira: <u>18+26=44</u></li> </ul>	Zenbat errore gutxiago? <u>62-44=18</u>  Batekotan? <u>λ=18/62=0.29</u>

Beraz, bi aldagaien arteko asoziazioa ahula/ertaina da, lagin errorearen eta beste ikerketen erreserbapean.

(b)

Kodifikazioa egiten dugu: gizona → 0; emakumea → 1; ez → 0; bai → 1.

Kalkulu-taula honela garatzen dugu batezbestekoak, bariantzak eta kobariantzak kalkulatzeko ( $x$ , sexua;  $y$ , lehen fasea gainditu duten):

$x$	$y$	$n$	$nx$	$ny$	$nx^2$	$ny^2$	$nxy$
0	0	45	0	0	0	0	0
0	1	18	0	18	0	18	0
1	0	26	26	0	26	0	0
1	1	44	44	44	44	44	44
		133	70	62	70	62	44

$$\bar{x} = \frac{70}{133} = 0.52 ; \bar{y} = \frac{62}{133} = 0.47$$

$$s_x^2 = \frac{70}{133} - 0.52^2 = 0.25 ; s_y^2 = \frac{62}{133} - 0.47^2 = 0.25$$

$$s_{xy} = \frac{44}{133} - 0.52 \times 0.47 = 0.086$$

$$r_{xy} = \frac{0.086}{\sqrt{0.25}\sqrt{0.25}} = 0.34$$

Beraz, koefiziente honen arabera bi aldagaien arteko korrelazioa ertaina da, lagin errorearen eta beste ikerketen erreserbapean betiere.

Norabideari buruz, korrelazio positiboa da, beraz sexuan 0 → 1 eginez, gainditu? aldagaian ere 0 → 1 egiten dugu. Hots, emakumeek gainditzeko joera handiagoa dute gizonek baino.

### III. ebazkizuna (2.5 puntu)

HIRUSTA enpresak laborategietarako zentrifugagailu aurreratuak saltzen ditu. COVID19 pandemia dela eta, eskaria handiko gailua da, eta enpresa ekoizleak eskaerak kontratuz egin behar direla ezarri du, atzera egiteko posibilitaterik gabe. HIRUSTA enpresak bere eskariari buruz probabilitate-banaketa hau finkatu du, merkatua arakatu eta gero:

Eskaria	Probabilitatea
3	0.1
4	0.15
5	0.25
6	0.2
7	0.15
8	0.10
9	0.05

Saldutako zentrifugagailu bakoitzeko 1000€ irabaziko ditu, eta aurrez eskatu baina azkenean salduko ez den gailu bakoitzeko 400€ galduko ditu.

#### Egin beharreko atazak:

- (a) Kalkulatu itxarondako mozkinak maximotzeko eskatu beharreko gailu kopurua.

#### Aurreko atalean kalkulaturako kopururako, kalkulatu:

- (b) Galduko diren salmenten itxaropena  
 (c) Itxarondako salmentak  
 (d) Biltegian saldu gabe eta likidatu aurretik geratuko diren unitateen itxaropena  
 (e) Stockout probabilitatea

(a)

Itxarondako mozkina taula bakoitzeko azken lerroan adierazten da.

3 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000	1	3000
			3000

4 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000-400=2600	0.1	260
4	4000	0.9	3600
			3860

5 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000-800=2200	0.1	220
4	4000-400=3600	0.15	540
5	5000	0.75	3750
			4510

6 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000-1200=1800	0.1	180
4	4000-800=3200	0.15	480
5	5000-400=4600	0.25	1150
6	6000	0.5	3000
			4810

7 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000-1600=1400	0.1	140
4	4000-1200=2800	0.15	420
5	5000-800=4200	0.25	1050
6	6000-400=5600	0.2	1120
7	7000	0.3	2100
			4830

8 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000-2000=1000	0.1	100
4	4000-1600=2400	0.15	360
5	5000-1200=3800	0.25	950
6	6000-800=5200	0.2	1040
7	7000-400=6600	0.15	990
8	8000	0.15	1200
			4640

9 gailu erosiz

Salmenta	Mozkina $x$	Probabilitatea $p(x)$	$xp(x)$
3	3000-2400=600	0.1	60
4	4000-2000=2000	0.15	300
5	5000-1600=3400	0.25	850
6	6000-1200=4800	0.2	960
7	7000-800=6200	0.15	930
8	8000-400=7600	0.10	760
9	9000	0.05	450
			4310

Itxarondako mozkina maximotzen duen gailu kopurua 7 da.

(b)

7 gailu erosita, galtzen diren salmentak, eskariaren araberakoak, ondoko taula honetan adierazten den bezala, honela banatzen dira:

Eskaria	Galdutako salmentak $x$	Probabilitatea $p(x)$	$xp(x)$
7 edo gutxiago	0	0.85	0
8	1	0.1	0.1
9	2	0.05	0.1
			0.2

Batezbeste galtzen diren salmental 0.2 gailukoak dira batezbeste.

(c)

7 gailu erosita, salmentak honela banatzen dira:

Salmentak $x$	Probabilitatea $p(x)$	$xp(x)$
3	0.1	0.3
4	0.15	0.60
5	0.25	1.25
6	0.2	1.2
7	0.30	2.1
		5.45

Batezbeste 5.45 gailu salduko dira 7 gailu erosita.

(d)

7 gailu erosita, likidatu beharreko gailu kopurua honela banatzen da, eskariaren arabera:

Likidatu beharrekoak	Noiz (eskaria) $x$	Probabilitatea $p(x)$	$xp(x)$
0	7,8,9	0.3	0
1	6	0.2	0.2
2	5	0.25	0.5
3	4	0.15	0.45
3	0.1	0.4	
			1.55

**Batezbeste 1.55 gailu beharko dira likidatu.** Askoz ere laburrago: 7 gailu erosi baditugu, eta batez besteko salmenta 5.45 gailukoa bada, likidatu beharreko gailuen kopurua 1.55 gailua.

(e)

Stockout geratzen gara, eskariari erantzun ezin zaionean. Beraz, 7 gailu erosita stockout gerttuko gara eskaria 8 edo 9 denean. Horrela, stockout probabilitatea  $0.10+0.05=0.15$  da.

#### IV. ebazkizuna (ebaluazio ez jarraitua) (1 puntu)

Administrazio publikoan deitutako oposizio batean hautagaien sexua eta oposizioko lehen fasea gainditu duten ala ez jaso da:

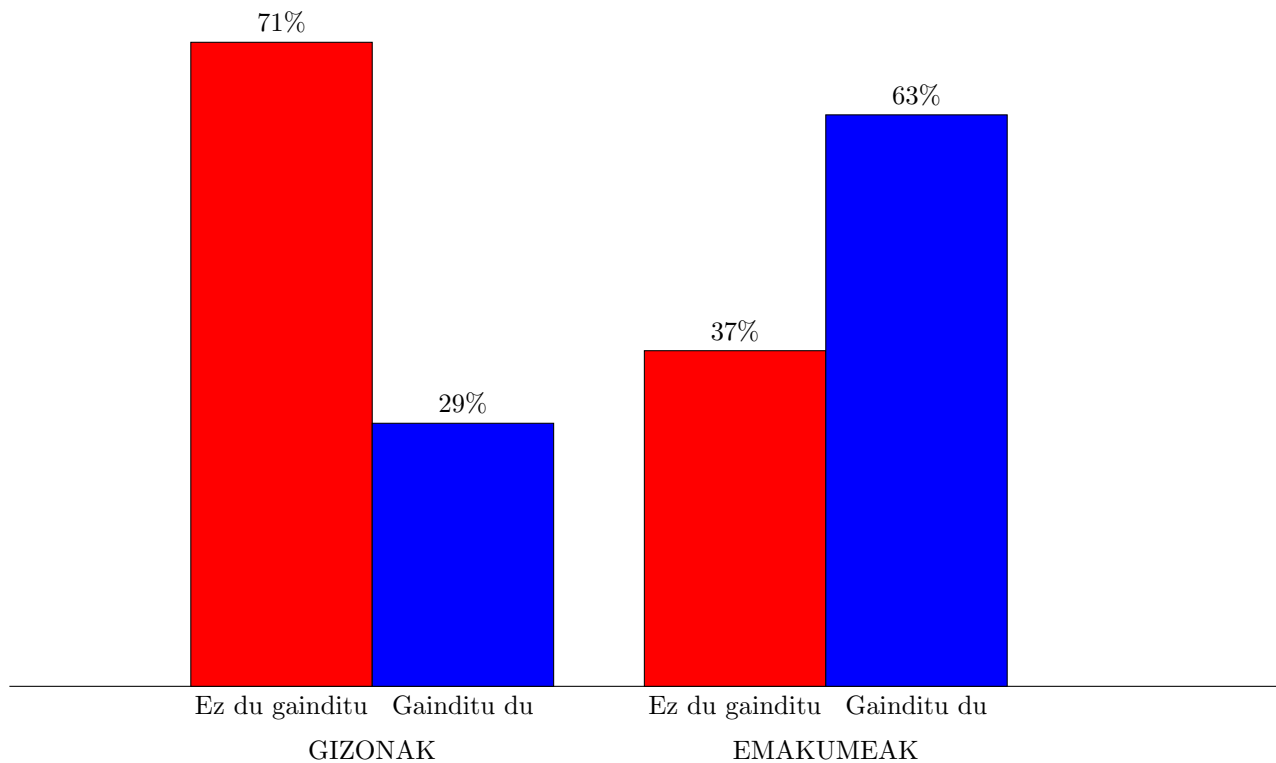
Gainditu? (↓) / Sexua (→)	Gizona	Emakumea
Ez	45	26
Bai	18	44

**Egin beharreko ataza:** Sexuaren araberako diferentziak erakusten dituen grafiko egokiena eratu ezazu.

Gizon eta emakumeen arteko diferentziak atzemateko, maiztasun erlatiboak kalkulatzeko ezinbestekoa da, gizon eta emakumeen lagin-tamainen efektua deusezteko. Eman ditzagun bada, maiztasun erlatibo horiek:

Gainditu? (↓) / Sexua (→)	Gizona	Emakumea
Ez	%71	%37
Bai	%29	%63

Portzentaje horiek irudikatzeko, egokiena barra-diagrama marraztea da:



Argi ikusten da emakumeek maizago gainditzen dutela.

## ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2021eko ekainaren 21a

Iraupena: 40 minutu

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

1. Nola deitu zitzaion XIX. mendera arte, estatistika datu geografiko, politiko eta ekonomikoen datu bilduma soiltzat hartu zuenari?
  - (a) Zentsaria
  - (b) Aritmetika politikoa
  - (c) Ekonometria
  - (d) Soziopolitika kuantitatiboa
2. Nola deitzen zaio gaixotasunei buruzko datu estatistikoak (zenbat gaixo, ...) aztertzen dituen arloari?
  - (a) Epidemiologia
  - (b) Etiologia
  - (c) Patogenia
  - (d) Patologia estatistikoa
3. 2 makinak 10na pieza ekoitzi dute 4 ordutan (ordu horietako bakoitzean). 3 makinak 6na pieza ekoitzi dute 2 ordutan. Eman batez besteko errendimendua.
  - (a) 8.18 pieza/h
  - (b) 8.28 pieza/h
  - (c) 8.38 pieza/h
  - (d) 8.48 pieza/h
4. Zein da zuzena bi gailurreko datu-banaketa bati buruz?
  - (a) Lehen mailako eta bigarren mailako deitzen zaie bi moda horiei.
  - (b) Absolutu eta erlatibo deitzen zaie bi moda horiei.
  - (c) Bi gailur badaude, altuena baino ez dagokio modari.
  - (d) U itxurako banaketa izango da, eta kasu horietan moda ez dago, gailurrak muturretan daudelako.
5. Honako hauetatik, zein sakabanatze-neurri erabiliko zenuke muturreko datuak daudenean?
  - (a) Desbideratze estandarra.
  - (b) Kuartil arteko ibiltartea.
  - (c) DAME
  - (d) (b) eta (c) dira egokiak.
6. Honako hauetatik, zein kasutan daude aukera gutxien datuen banaketan heterogeneotasuna sortzeko?
  - (a) Neska eta mutilen altuerak biltzean.
  - (b) Selektibitateko notak ikastetxe desberdinetarako biltzean.
  - (c) Bi irakaslek dagokien ikasle-taldean zuzendutako test motako azterketa bereko notak biltzean.
  - (d) Gradua bukatu eta lana aurkitu arteko denborari buruz, emakumeak eta gizonak batera biltzean.
7. Zein baliotik gora esan daiteke Giniren indizeak kontzentrazio handia erakusten duela?
  - (a) 0.5etik gora
  - (b) 0.6tik gora
  - (c) 0.8tik gora
  - (d) Ezin da balio jakinik eman, interpretazioa erlatiboa baita, beste Giniren indizeekin alderaturik.

- 
8. Errenta hauek jaso dira: 10-20-30-40. Kalkulatu Robin Hood adierazlea.
- (a) 0.1
  - (b) 0.2
  - (c) 0.3
  - (d) 0.4
9. Errenta hauek jaso dira: 10-20-30-40. Kalkulatu mediala.
- (a) 20
  - (b) 25
  - (c) 30
  - (d) Beste bat da erantzuna.
10. Errenta hauek jaso dira: 10-20-30-40. Pobrezia-atalasea 25 da. Zenbat da pobrezia-arraila?
- (a) 0.1
  - (b) 0.2
  - (c) 0.3
  - (d) 0.4
11. Soldata hauek jaso ditu langile kategoria batek azken urteotan: 2018:40€;2019:42€;2020:45€. Kontsumorako prezioen indizearen bilakaera urte horietan 123-129-137 izan da. Zenbat da soldata deflaktatua 2020an 2018ko euro konstanteetan?
- (a) 40.40
  - (b) 40.80
  - (c) 41.40
  - (d) 41.80
12. Zereal kopuru hauek, adierazitako prezioetan, kontsumitu dira baserri batean: (Artoa:2019:q:10;p:50) (Artoa:2020:q:12;p:60) (Garia:2019:q:20;p:20) (Garia:2020:q:20;p:25). Kalkulatu gariaren erreperkutsioa 2020ko Laspeyres indizean (oinarria:2019).
- (a) %20
  - (b) %30
  - (c) %40
  - (d) %50
13. X zorizko aldagai bat honela banatzen da:  $f(x) = 2x/9$  ;  $0 < x < 3$ . Kalkulatu  $P[X > 2]$ .
- (a) 4/9
  - (b) 5/9
  - (c) 6/9
  - (d) 7/9
14. X zorizko aldagai bat honela banatzen da:  $f(x) = 2x/9$  ;  $0 < x < 3$ . Kalkulatu horren itxaropen matematikoa.
- (a) 1.66
  - (b) 2
  - (c) 2.33
  - (d) 2.66

15.  $X$  zorizko aldagai bat honela banatzen da:

$x$	$p(x)$
-1	0.3
0	0.4
1	0.3

Kalkulatu bariantza.

- (a) 0.4
  - (b) 0.6
  - (c) 0.9
  - (d) 1.2
16.  $X$  zorizko aldagai baten bariantza 4 da.  $Y = 2X + 6$  aldagai-aldaketa egiten da. Zenbat da  $Y$  aldagaiaren bariantza?
- (a) 8
  - (b) 14
  - (c) 16
  - (d) 22
17. A inbertsioaren etekinen itzaropena 5 da, eta bariantza 2. B inbertsioaren etekinen itzaropena 3 da, eta bariantza 1. Zein da zuzena honako hauetatik?
- (a) Epe luzera eta laburrera hobe da A.
  - (b) Epe luzera eta laburrera hobe da B.
  - (c) Epe luzera hobe da A, eta laburrera B.
  - (d) Utilitate-funtziorik gabe ezin da zehaztu zein den hobe epe laburrera.
18.  $X$  zorizko aldagai bati buruz itzaropena 10 eta desbideratzea 1 direla baino ez dugu ezagutzen. Hurbildu  $P[X \leq 8]$ .
- (a)  $P[X \leq 8] \leq 0.2$
  - (b)  $P[X \leq 8] \leq 0.25$
  - (c)  $P[X \leq 8] \leq 0.4$
  - (d)  $P[X \leq 8] \leq 0.6$
19.  $X$  zorizko aldagai bati buruz itzaropena 10 dela baino ez dugu ezagutzen. Hurbildu  $P[X \geq 12.5]$ .
- (a)  $P[X \geq 12.5] \leq 0.6$
  - (b)  $P[X \geq 12.5] \leq 0.7$
  - (c)  $P[X \geq 12.5] \leq 0.8$
  - (d)  $P[X \geq 12.5] \leq 0.9$
20.  $X$  zorizko aldagai bati buruz itzaropena 10 eta desbideratzea 1 direla eta moda bakarrekoa eta simetrikoa dela baino ez ditugu ezagutzen. Hurbildu  $P[X \leq 8]$ .
- (a)  $P[X \leq 9] \leq 0.11$
  - (b)  $P[X \leq 9] \leq 0.22$
  - (c)  $P[X \leq 9] \leq 0.33$
  - (d)  $P[X \leq 9] \leq 0.66$



## Estatistika eta datuen analisisa

2021eko ekainaren 21a

**Izena eta abizenak:** Josemari Sarasola

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

KOPURUA

ONGI	
GAIZKI	
ERANTZUN GABE	