

ESTADÍSTIKA ETA DATUEN ANALISIA

Azterketa ebatziak

2021-2022 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea. EHU

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

gizapedia.org

ESTADÍSTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2022ko urtarrilaren 14an, 16:00

Iraupena: 90 min

I. ebazkizuna: Kontzentrazioa (3 puntu)

Lapurdikiko *commune* edo udalerriak bi sailetan banatu dira: kostaldekoak (Baiona, Angelu, Biarritz, Bidart, Getaria, Donibane, Ziburu, Sokoa, Urruña eta Hendaia) eta barnealdekoak, beste guztiak. Sail horietako bakoitzean familien errentari buruz *sondage* edo inkesta bat garatu eta familia-errentari buruz datu hauek eskuratu dira:

- Kostaldean: 90-30-30
- Barnealdean: 30-50

Egin beharreko atazak:

- (a) Eman ezazu errenta guzietan buruzko Gini indizea. Horren balioa interpretatzen ahal duzu, kontzentrazioa aski handia edo ttipia den erraiteko?
- (b) Arestian eman indizea deskonposatu behar duzu bi ataletan, kostaldea/barnealdea artean nahiz horien baitan (*talde (edo eskualde) arteko nahiz barneko Gini indizeak kalkulatu*). Emaitzak interpretatu: errenta desberdintasunak zergatik sortzen dira bereziki: bi eskualdeetan errentak orokorki aski desberdinak direlako ala eskualde bakoitzaren baitan ere desberdintasunak handiak direlako?

(a)

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/5=0.2	30	30	30/230=0.13	0.07
2/5=0.4	30	60	60/230=0.26	0.14
0.6	30	90	0.39	0.21
0.8	50	140	0.60	0.20
(1)	90	230	1	0
2	230			0.61

$$G = \frac{0.61}{2} = 0.305$$

$G = 0.305$ izanik, ez da erraiten ahal Lapurdiko kontzentrazioa ttipia edo handia den. Hortako, bertze erreferentzia bat behar da (bertze probintzia bat, kasu), alderatu ahal izateko.

(b)

Lehen gauza, emanen ditugu Gini indizeak bi eskualdeetarako:

- kostaldean,

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/3=0.33	30	30	30/150=0.20	0.13
0.66	30	60	60/150=0.40	0.26
(1)	90	150	1	0
1	150			0.39

$$G = \frac{0.39}{1} = 0.39$$

- barnealdean,

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/2=0.5	30	30	30/80=0.375	0.125
(1)	50	80	1	0
0.5	80			0.125

$$G = \frac{0.125}{0.5} = 0.25$$

Eta orain banakoen portzentajeak (p_k) eta errenta portzentajeak (q_k) eransten ditugu bi eskualdeetarako:

Eskualdea	p_k	q_k	G_{bk}
Kostaldea	3/5=0.6	150/230=0.65	0.39
Barnealdea	2/5=0.4	80/230=0.35	0.25

Emaitza horiekin eskualde barneko kontzentrazioa ematen ahal dugu:

$$G_b = \sum p_k q_k G_{bk} = 0.6 \times 0.65 \times 0.39 + 0.4 \times 0.35 \times 0.25 = 0.187$$

Orai, eskualde arteko kontzentrazioa eman behar dugu. Hortarako, behar ditugu eskualde bakoitzeko batez besteko errentak kalkulatu:

$$\bar{e}(kostaldea) = \frac{150}{3} = 50 ; \bar{e}(barnealdea) = \frac{80}{2} = 40$$

Eta orain batezbesteko hauek harturik Gini indizea emanen dugu Lapurdi osoarentzat:

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/5=0.2	40	40	40/230=0.17	0.03
2/5=0.4	40	80	80/230=0.34	0.06
0.6	50	130	0.56	0.04
0.8	50	180	0.78	0.02
(1)	50	230	1	0
2	230			0.15

$$G_a = \frac{0.15}{2} = 0.075$$

Egin dezagun deskonposizioa:

$$G(osea) = G(artekoa) + G(barnekoa) + hondarra = 0.075 + 0.187 + 0.043 = 0.305$$

Ikus daitekeenez, anitzez handiagoa eskualde barneko desberdintasuna eskualde artekoa baino. Ez da erraiten ahal, beraz, bi eskualdeetako errentak oso desberdinak direnik.

II. ebazkizuna: Korrelazioa (1 puntu)

COVID19-ko [aztarnari](#) batek bi aldagai hauek jaso ditu kontaktu batzuen artean: alde batetik, kontaktu hori kutsatuaren familiakoa, laguna edo lantokiko lankidea den; eta bestetik kontaktu hori azkenean kutsatu den. Datuak hauek dira (fam:familia, lag:laguna, lan:lantokia):

kontaktu-mota	fam	lag	lan	fam	fam	lag	lan	fam	lag	lag
kutsatu?	bai	ez	bai	bai	bai	ez	ez	bai	ez	bai

Kalkulatu kontaktua familiako kidea izatearen eta kontaktua azkenean kutsatuta suertatzearen arteko Pearson korrelazio-koefiziente lineala eta interpreta ezazu argi eta eskematikoki.

Kodifikazioa modu estandarrean egingo dugu:

- familiako kidea izateari buruz (x), familiakoa bada, 1; ez bada, 0.
- kontaktua kutsatu den ala ez (y): kutsatuta 1; kutsatu gabea, 0.

(x)	y	xy	x ²	y ²
1	1	1	1	1
0	0	0	0	0
0	1	0	0	1
1	1	1	1	1
1	1	1	1	1
0	0	0	0	0
0	0	0	0	0
1	1	1	1	1
0	0	0	0	0
0	1	0	0	1
4	6	4	4	6

$$\bar{x} = \frac{4}{10} = 0.4 \quad ; \quad \bar{y} = \frac{6}{10} = 0.6$$

$$s_{xy} = \frac{4}{10} - 0.4 \times 0.6 = 0.16$$

$$s_x = \sqrt{\frac{4}{10} - 0.4^2} = 0.49 \quad ; \quad s_y = \sqrt{\frac{6}{10} - 0.6^2} = 0.49$$

$$r_{xy} = \frac{0.16}{0.49 \times 0.49} = 0.66$$

Interpretazio eskematikoa

$$|r_{xy}| > 0 \rightarrow \text{korrelazio positiboa dago} \quad \begin{array}{l} \longrightarrow x \uparrow \Rightarrow y \uparrow \\ \longrightarrow familia : 0 \rightarrow 1 \Rightarrow kutsatu : 0 \rightarrow 1 \end{array}$$

Dekodifikatu eta esan dezakegu familiakoak diren kontaktuek joera handiagoa dutela kutsatzeko, lagun girokoek eta lankideek baino. Korrelazio hori sendoa da gainera (balio absolutua 0.6 baino handiagoa izanik), lagin errearen eta antzeko ikerketen erreserbapean.

III. ebazkizuna: Egunkari saltzailearen problema (3 puntu)

Konpainia batek ikerketa-laborategientzako zehaztasun handiko mikroskopio elektronikoak erosi, probatu eta doitzen ditu. 2022 urtean zehar izango dituen eskaerak betetzeko, zenbat mikroskopio erosi behar dituen erabaki behar du une honetan. Eskaria honela banatzen dela uste da:

x	2	3	4
p(x)	0.3	0.5	0.2

Mikroskopioak 0.8M eurotan erosten ditu eta 1.1M eurotan saldu. Urtean saltzen ez dituen mikroskopioak 0.7M eurotan saldu beharko ditu, obsoleszentzia teknologikoa dela eta.

Egin beharreko atazak:

- Batez besteko mozkina maximotzeko erosi beharreko mikroskopio kopurua kalkulatu behar duzu.
- Kalkulatu ondorengo parametroak, erosten den mikroskopio kopurua 3 den kasuan, eta haien arteko koadratzeak egiaztatu: batez besteko salmenta, batezbeste galtzen den salmenta, batez besteko soberako unitateak, batez besteko eskaria.
- Kalkulatu *fill rate* delakoa eta *stockout* probabilitatea, 3 mikroskopio erosten diren kasuan betiere.

(a)

- 2 mikroskopio erosten direnean:

eskaria	salmentak	mozkina (x)	p(x)	xp(x)
2,3,4	2	0.6	1	0.6
				0.6

- 3 mikroskopio erosten direnean:

eskaria	salmentak	mozkina (x)	p(x)	xp(x)
2	2	0.6-0.1=0.5	0.3	0.15
3,4	3	0.9	0.7	0.63
				0.78

- 4 mikroskopio erosten direnean:

eskaria	salmentak	mozkina (x)	p(x)	xp(x)
2	2	0.6-0.2=0.4	0.3	0.12
3	3	0.9-0.1=0.8	0.5	0.40
4	4	1.2	0.2	0.24
				0.76

Batezbesteko mozkina maximotzen da 3 mikroskopio erosita.

(b) 3 mikroskopia erositako,

• batezbesteko salmenta:

eskaria	salmenta (x)	p(x)	xp(x)
2	2	0.3	0.6
3,4	3	0.7	2.1
2.7 gailu			

• batezbeste galtzen den salmenta:

eskaria	galdu (x)	p(x)	xp(x)
2,3	0	0.8	0
4	1	0.2	0.2
0.2 gailu			

• batez besteko soberako unitateak:

eskaria	soberan (x)	p(x)	xp(x)
2	1	0.3	0.3
3,4	0	0.7	0
0.3 gailu			

• batez besteko eskaria:

eskaria (x)	p(x)	xp(x)
2	0.3	0.6
3	0.5	1.5
4	0.2	0.8
2.9 gailu		

KOADRATZEAK:

- $E(\text{saldu}) + E(\text{galdu}) = E(\text{eskaria}) \rightarrow 2.7 + 0.2 = 2.9 \checkmark$
- $E(\text{saldu}) + E(\text{soberan}) = \text{erosi} \rightarrow 2.7 + 0.3 = 3 \checkmark$

(c)

FILL RATE: $E(\text{saldu})/E(\text{eskaria}) = 2.7/2.9 = \%93$

3 gailu erositako, eskariaren %93 estaltzen da batezbeste.

$P(\text{STOCKOUT}) = P(\text{eskaria}=4) = 0.2$

IV. ebazkizuna: Markoven kateak eta xurgapen-denborak (1 puntu)

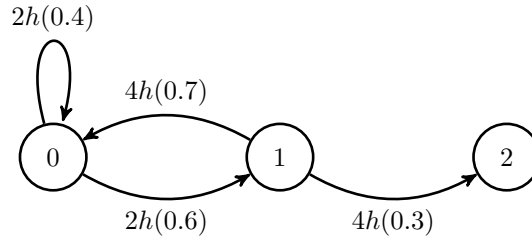
Produktu bat bi fasetan ekoizten da: lehenengo faseak 2 ordu irauten du, eta arrakastatsua bada, bigarren fasera igarotzen gara, 4 ordu irauten duena; bigarren fasea arrakastatsua bada, produktu bukatua eskuratzen da. Lehenengo fasean lana arrakastatsua izateko probabilitatea 0.6 eta bigarren fasean 0.3. Fasea arrakastatsua ez bada, izan lehena edo bigarrena, prozesua hasiera hasieratik abiarazi behar da berriz ere. Zenbat denbora behar da batezbeste produktu bukatu bat ekoizteko? Oharra: azken soluziora iritsi behar da.

0 egoerak adierazten du ekoizpen prozesuaren hasieran gaudela, lehen fasea hastear dagoela alegia.

1 egoerak adierazten du, lehen fasea arrakastatsu bukatu dela eta bigarren fasea hastear dagoela.

2. egoerak adierazten du produktua bukatua dela.

h hizkiak trantsizioko ordu kopurua adierazten du. Parentesien artean, trantsizioaren probabilitatea.



(Oharra: notazio minimalista erabiltzen da itxaropenak izendatzeko)

T_0	$p(T_0)$	$T_0 p(T_0)$
$2 + E[T_0]$	0.4	$0.8 + 0.4E[T_0]$
$2 + E[T_1]$	0.6	$1.2 + 0.6E[T_1]$
$2 + 0.4E[T_0] + 0.6E[T_1]$		

Beraz, laburbilduz:

$$E[T_0] = 2 + 0.4E[T_0] + 0.6E[T_1]$$

T_1	$p(T_1)$	$T_1 p(T_1)$
4	0.3	1.2
$4 + E[T_0]$	0.7	$2.8 + 0.7E[T_0]$
$4 + 0.7E[T_0]$		

Beraz, laburbilduz:

$$E[T_1] = 4 + 0.7E[T_0]$$

Bigarren ekuazioa lehenengoan ordeztuz:

$$E[T_0] = 2 + 0.4E[T_0] + 0.6(4 + 0.7E[T_0]) = 4.4 + 0.82E[T_0] \rightarrow 0.18E[T_0] = 4.4 \rightarrow E[T_0] = 24.44h$$

24.44 ordu behar dira batezbeste produktua bukatzeko.

ESTADISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2022ko urtarrilaren 14a, 16:00

Iraupena: 40 minutu

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

1. Autobus gidari batek geltoki bakoitzean zenbat bidaiari sartzen diren jasotzen du. Zein da behaketa-unitatea?
 - (a) Autobusa
 - (b) Gidaria
 - (c) Geltokia
 - (d) Bidaiaria
2. Honako hauetatik zein da faltsua estatistika deskribatzaileari buruz?
 - (a) Ez du lagin-errorea kontuan hartzen.
 - (b) Lagin nahiz populazio osoko bateko datuak azter ditzake.
 - (c) Inferentzia estatistikoa baino fidagarriagoa da.
 - (d) Probabilitateak ez ditu kontuan hartzen.
3. 10 egunetan zehar tenperatura maximoak jaso dira Donostian nahiz Aralarren kokatutako estazio meteorologikoetan. Bi toki horietako datuak alderatzeko, zein grafiko da egokiena?
 - (a) Zutabe diagrama
 - (b) Puntu diagrama
 - (c) Kartograma
 - (d) Histograma
4. Zenbat datu daude portzentajea 2garren kuartiletik 4garren kintilera?
 - (a) %20
 - (b) %30
 - (c) %40
 - (d) %50
5. Pote baten eduki nominal edo estandarra 200gr. Jasotako neurriak (gr): 192-194-202. Kalkulatu batezbesteko errorea.
 - (a) 5.28gr
 - (b) 5.48gr
 - (c) 5.68gr
 - (d) 5.88gr
6. Internet-en irakurria: *Si una persona ha estado expuesta al nuevo coronavirus y se ha contagiado, los síntomas suelen aparecer entre los cinco y seis días posteriores al contacto. Sin embargo, estos tienen un rango de aparición entre 1 y 14 días, según explican los expertos.* Zein da egiazkoa sintomen agerpen-denborari buruz?
 - (a) Zentroa 1 eta 14 artean dago.
 - (b) Zentroa 5.5 dela esan liteke.
 - (c) Ibiltartea 13 da.
 - (d) Aurreko guztiak dira egiazkoak.

7. Datu hauek jaso dira aldagai bati buruz: 4-9-6-17-13-22. Kalkulatu 6gn dezila.
- 11.4
 - 12.6
 - 14.6
 - Aurreko guztiak faltsuak dira.
8. Lorenz kurbako puntuak dituzu hauek: $(p_1 = 0.1, q_1 = 0.04)$, $(p_2 = 0.4, q_2 = 0.2)$, $(p_3 = 0.8, q_3 = 0.67)$, $(p_4 = 0.9, q_4 = 0.8)$. Kalkulatu Palma ratioa.
- 5
 - 2
 - 1
 - Aurreko guztiak faltsuak dira.

9., 10. eta 11. galderak kontingentzia-aula honi buruzkoak dira:

sexua/botoa	ezkerra	eskuina
gizon	20	60
emakume	40	60

9. Bi aldagaien artean erabateko independentzia balitz, zenbat litzateke gizon ezkertiarren maiztasuna?
- 20, zeren bi sexuetarako joera politiko probableenak berdina izanik, eskuina alegia, taulako bi aldagaiak independenteak dira.
 - 30
 - 40
 - Aurreko guztiak faltsuak dira.
10. Lambda kalkulatzera, zenbat errore egiten dira aurreko egitean pertsona gizona dela badakigu?
- 20
 - 30
 - 40
 - Aurreko guztiak faltsuak dira.
11. Asoziazio ordinala aztertu nahi da. Zenbat da gamma?
- 0.5
 - 0.33
 - 0.33
 - Aurreko guztiak faltsuak dira.
12. Korrelazioari buruz, honako hauetatik zein da faltsua?
- Kobariantzak edozein balio har dezake, positibo zein negatibo, 0 ezik.
 - Pearson koefizienteak -1 eta 1 artean dago, biak barne.
 - Korrelazioa aztertzerakoan, atributu dikotomiko bati 0 eta 1 balioak esleitzeko modu bat baino gehiago dago.
 - 5 aldagai kuantitatiboren artean, binaka harturik, 10 korrelazio koefiziente kalkulatu daitezke.
13. Cronbach-en alfarako zein baliotarako esango zenuke galderen artean erredundantzia dagoela?
- 0.4
 - 0.4
 - 0.9
 - Aurrekoetatik, ezin da inongo kasutan esan erredundantzia dagoenik.

14. Probabilitate-funtzio hau definitu da: $P[X = x] = x/12; x = 1, 2, 3, 4, 5$. Zein da
- (a) Guztiz egoki edo zuzena da.
 - (b) Desegokia da.
 - (c) $F[X = 3] = 0.5$
 - (d) $P[X \geq 4] = 0.75$
15. Banaketa jarraitua definitu dugu: $F(x) = x^2/16; 0 \leq x \leq 4$. Kalkulatu $P[1 < X < 3]$.
- (a) 0.25
 - (b) 6/8
 - (c) 0.5
 - (d) 7/8
16. Honako hauetatik zein ez da, banaketa jarraitu bati buruz, banaketa-funtzio egoki edo zuzena?
- (a) $F(x) = x^2/9; 0 \leq x \leq 3$
 - (b) $F(x) = 4 - x^2/3; 0 \leq x \leq 1$
 - (c) $F(x) = 1 - e^{-2x}; x \geq 0$
 - (d) $F(x) = x^3 - 1/26; 1 \leq x \leq 3$
17. 10 alde dituen dado berezi bateko aldeak honela daude banatuta: 1-1-1-1-2-2-2-3-3-4. Kalkulatu puntuazioaren itxaropen matematikoa.
- (a) 1.5
 - (b) 2
 - (c) 2.5
 - (d) 2.75
18. Akzio batek bihar 1 puntu edo 2 puntu igoko da, 0.8 eta 0.2 probabilitateekin hurrenik hurren. Kalkulatu bariantza.
- (a) 0.16
 - (b) 0.36
 - (c) 1.16
 - (d) 1.6
19. A inbertsioaren errendimenduaren itxaropena eta bariantza: 4 eta 1, hurrenik hurren. B inbertsioarenak: 6 eta 3. C inbertsioarenak: 8 eta 2. Zein da egiazkoa?
- (a) Epe laburrera, dilema dago A eta C artean
 - (b) B inbertsioak A inbertsioa menderatzen du epe laburrera.
 - (c) Epe laburrera, onena C da.
 - (d) C inbertsioak A inbertsioa menderatzen du epe laburrera.
20. $F(x) = x^2/16; 0 \leq x \leq 4$. Kalkulatu $Var(0.1)$.
- (a) 0.44
 - (b) 0.66
 - (c) 1.26
 - (d) 1.6

Estatistika eta datuen analisia

-
2022ko urtarrilaren 14a

Josemari Sarasola

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

KOPURUA

ONGI	
GAIZKI	
ERANTZUN GABE	

ESTADÍSTICA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola Ledesma

Data: 2022ko ekainaren 22, 16:00

I. ebazkizuna (*Ebaluazio ez jarraitua: 3 puntu*)

Paris hiria bi zatitan banatu da, Sena ibaiak bereizita: *rive gauche* eta *rive droite*, ezker eta eskuin erriberak alegia. Erribera bietan errenta datuen laginak jaso dira:

- *Rive gauche*: 20-30
 - *Rive droite*: 10-28-40
- (a) Hiri osoko errenten nahiz erribera bakoitzeko errenten Gini indizeak kalkulatu. Zein erriberetan da errenten kontzentrazioa handiagoa?
- (b) Hiri osoko errentei buruzko Gini indizea bi ataletan deskonposatu, erriberen arteko kontzentrazioa nahiz erriberen barneko kontzentrazioa bereiziz. Emaizak interpretatu: errenta desberdintasunak zergatik sortzen dira bereziki, bi aldeetan errentak orokorki aski desberdinak direlako ala erribera bakoitzaren baitan ere desberdintasunak handiak direlako?

(a)

- rive gauche-n,

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/2=0.5	20	20	20/50=0.40	0.10
(1)	30	50	50/50=1	0
0.5	50			0.1

$$G = \frac{0.1}{0.5} = 0.2$$

- rive droite-n,

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/3=0.33	10	10	10/78=0.13	0.2
2/3=0.66	28	38	38/78=0.48	0.18
(1)	40	78	1	0
1	78		0.38	

$$G = \frac{0.38}{1} = 0.38$$

- hiri osoan (datu denak ordenatu behar dira!)

p_i	$x_i(ord)$	$x_i(met)$	q_i	$p_i - q_i$
1/5=0.2	10	10	10/128=0.07	0.13
2/5=0.4	20	30	30/128=0.23	0.17
3/5=0.6	28	58	58/128=0.45	0.15
4/5=0.8	30	88	88/128=0.68	0.12
(1)	40	128	128/128=1	0
2	128			0.57

$$G = \frac{0.57}{2} = 0.285$$

Erribera bakoitzeko Gini indizeei erreparatuz, rive droite izenekoan kontzentrazioa handiagoa dago.

(b)

rive gauche

- $G = 0.2$
- $p = 2/5 = 0.4$
- $q = (20 + 30)/128 = 0.39$

erriberen barneko kontzentrazioa

rive droite

- $G = 0.38$
- $p = 3/5 = 0.6$
- $q = (10 + 28 + 40)/128 = 0.61$

$$\sum p_k q_k G_k = 0.2 \times 0.4 \times 0.39 + 0.38 \times 0.6 \times 0.61 = 0.17$$

Erriberen arteko kontzentrazioa aztertzeko, erribera bakoitzeko batezbesteko errentak kalkulatu dira:

- rive gauche-n, $(20+30)/2=25$
- rive droite-n, $(10+28+40)=26$

Batezbestekoak oso antzekoak dira. Beraz, aurreratu daiteke erriberen arteko kontzentrazioa eskasa izango dela, eta desberdintasunak erriberen baitan sortzen direla.

Baina gara dezagun zehatz jakiteko:

p_i	$x_i(Ord)$	$x_i(Met)$	q_i	$p_i - q_i$
1/5=0.2	25	25	25/128=0.19	0.01
2/5=0.4	25	50	50/128=0.39	0.01
3/5=0.6	26	76	76/128=0.59	0.01
4/5=0.8	26	102	102/128=0.79	0.01
(1)	26	128	1	0
2	128			0.04

erriberen arteko kontzentrazioa

$$G_a = \frac{0.04}{2} = 0.02$$

deskonposaketa

$$G = G_a + G_b + h$$

$$0.285 = 0.02 + 0.17 + 0.095$$

Erriberen barneko kontzentrazioa askoz ere handiagoa artekoa baino (0.17;0.02). Beraz, errenten kontzentrazioa erriberen baitan sortzen da, eta ez erribera batetik bestera. Ezin da esan, hartara, erribera bat bestea baino aberatsagoa denik, erribera bakoitzaren barnean sortzen baitira desberdintasunak, *mixité sociale* edo nahasketa soziala adieraziz horrela erribera bakoitzean.

II. ebazkizuna (Ebaluazio ez jarraitua: 3 puntu)

Jarrera nazionalista espainiarra ebaluatzen test bat eratu da, bi aukerako 4 galderarekin, non aukera bati puntu bat ematen zitzaion nazionalismoaren eskalan, eta besteari zero. Hauek dira galderak:

- ¿Has votado alguna vez a Vox?
- ¿Posees algun objeto con la bandera española, o la misma bandera?
- ¿Te gusta especialmente la música folklórica española, especialmente la andaluza?
- ¿Vas a misa regularmente?

Baiezkoan, 1 puntu ematen da jarrera nazionalistaren alde, eta ezezkoan 0.

Galdetegia balidatzeko, lagun talde bat inkesta pasa zaio:

Pertsonak (↓) / Galderak (→)	1. galdera	2. galdera	3. galdera	4. galdera
Manolo	1	1	1	1
Macarena	0	0	1	1
Juan	1	1	0	0
Carmen	0	1	1	1
Rafael	1	1	1	0
Lola	0	0	1	1

Cronbach-en alfa kalkulatu eta **INTERPRETATU**.

Galdera bakoitzeko erantzunen bariantzak kalkulatzeko ditugu, galdetegiko x_T puntuazio osoaren bariantzarekin batera:

x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2	x_T	$x_1 T^2$
1	1	1	1	1	1	1	1	4	16
0	0	0	0	1	1	1	1	2	4
1	1	1	1	0	0	0	0	2	4
0	0	1	1	1	1	1	1	3	9
1	1	1	1	1	1	0	0	3	9
0	0	0	0	1	1	1	1	2	4
3	3	4	4	5	5	4	4	16	46

- $s_1^2 = \frac{3}{6} - \frac{3_2}{6} = 0.250$
- $s_2^2 = \frac{4}{6} - \frac{4_2}{6} = 0.222$
- $s_1^2 = \frac{5}{6} - \frac{5_2}{6} = 0.138$
- $s_1^2 = \frac{4}{6} - \frac{4_2}{6} = 0.222$
- $s_1^2 = \frac{46}{6} - \frac{16_2}{6} = 0.555$

Eman dezagun Cronbach-n alfa:

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right] = \frac{4}{4-1} \times \left[1 - \frac{0.250 + 0.222 + 0.138 + 0.222}{0.555} \right] = -0.665$$

Testa koherentea izateko, Cronbach-en alfa 0.7-0.9 tartean behar da izan. Urrun dago, eta are, negatiboa da. Beraz, esan daiteke testeko galderak ez direla batera koherenteak beraien artean (zehazkiago, barne konsistentziarik ez dagoela esango genuke).

Hala ere, galderei erreparatuta, aurretik esango genuke galderak loturik daudela, eta bai neurtzen dutela, guztiek batera, espainiar jarrera nazionalista, frankismotik bereziki nazionalkatolizismoa eta folklorismo espainiarra sendotu zirelako espainiar gizarte sektore zabaletan. Baina ez da horrela, datuok adierazten baitute hori ez dela horrela, badela jendea (laginean, Lola, Macarena eta agian Carmen) musika folkloriko espainiarra dena eta mezatara joaten dena, nazionalista izan gabe (Vox-i botoa ematea eta espainiar bandera izatea lirateke horren erakusgarri argienak).

III. ebazkizuna (Ebaluazio ez jarraitua: 2 puntu)

Bideojoko bat diseinatzean, jokalariai aurrera pauso bat egiteko behar duen saialdi kopurua 1, 2, 3 edo 4 izan daitekeela uste da, 0.3, 0.4, 0.2 eta 0.1 probabilitateekin.

Bideojokoaren diseinua bukatuta, jokalariai batzuei eman zaie probatzeko: 8 jokalarik saialdi bakarra behar izan dute, 7 jokalarik 2 saialdi, 3 jokalarik 3 saialdi eta 2 jokalarik 4 saialdi.

Kolmogorov-Smirnov proba garatu ezazu, taula moduan hasierako ustea egiazkotzat jo daitekeen erabakitzeko, %5eko adierazgarritasun- edo esangura-mailarekin. Marraztu itzazu banaketa funtzio teorikoa nahiz enpirikoa eta bertan adierazi distantzia maximoa.

Laginari dagozkion maiztasun erlatibo bakunak eta metatuak kalkulatzeko, probabilitate bakun eta metatuekin batera:

x	$p(x)$	$F(x) = P[X \leq x]$	$f(x)$	$\hat{F}(x)$
1	0.3	0.3	8/20=0.4	0.4
2	0.4	0.7	7/20=0.35	0.75
3	0.2	0.9	3/20=0.15	0.9
4	0.1	1	2/20=0.1	1

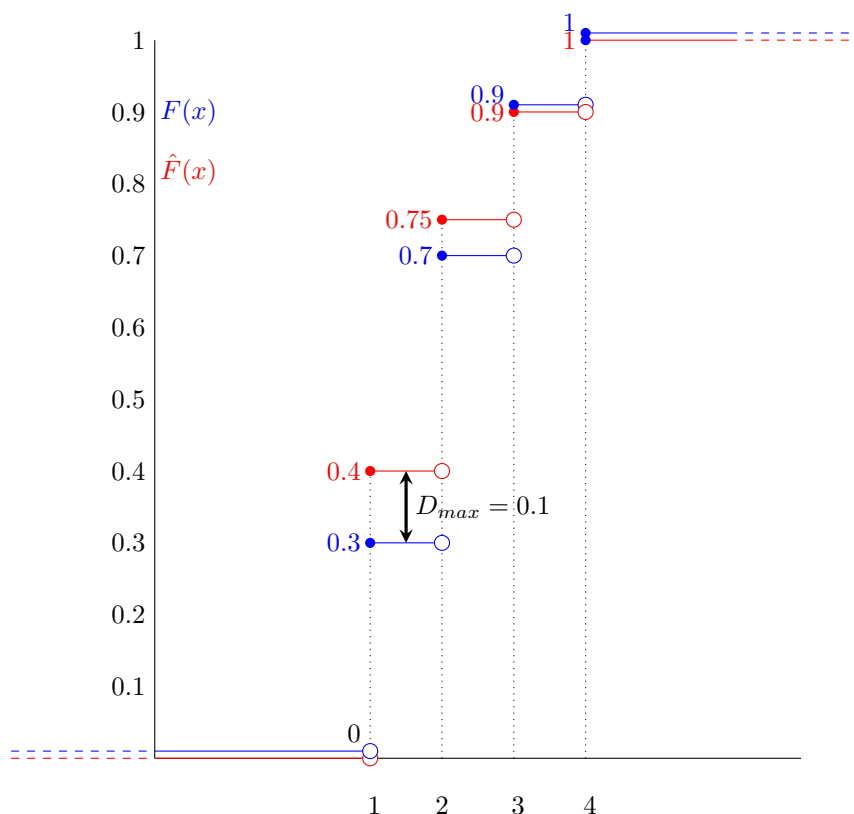
Banaketa diskretua izanda beharrezkoa ez den arren, prozedura orokorrari jarraitu eta x^- eta x balioak bereizten ditugu distantzia maximoa kalkulatzeko:

x	$F(x) = P[X \leq x]$	$\hat{F}(x)$	$D = F(x) - \hat{F}(x) $
1^-	0	0	0
1	0.3	0.4	0.1
2^-	0.3	0.4	0.1
2	0.7	0.75	0.05
3^-	0.7	0.75	0.05
3	0.9	0.9	0
4^-	0.9	0.9	0
4	1	1	0

Distantzia maximoa banaketa funtzio enpirikoaren eta teorikoaren artean $D_{max} = 0.1$ da.

Distantzia horren balio kritikoa banaketa teorikoa baztertze aldera, 20 datuetarako eta $\alpha = 0.05$ esangura-mailarako, $D_{max}^* = 0.294$. Lagineko distantzia maximoa ez da balio kritikora heltzen, eta beraz banaketa teorikoa egokia dela baieztatu daiteke.

Grafikoki:



ESTADISTIKA ETA DATUEN ANALISIA

Irakaslea: Josemari Sarasola

Data: 2022ko urtarrilaren 14a, 16:00

Iraupena: 40 minutu

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

1. Jorge Luis Borges kartografiari buruz ari da “Del rigor en la ciencia” ipuinean. Zein da ipuinean baieztatzen duen ideia nagusia?
 - (a) Datuen analisia konklusio egokiekin amaitu behar dela.
 - (b) Datuen sintesia egin beharra.
 - (c) Estatistika da zientziaren tresna gorena.
 - (d) Erabaki egokiak hartzeko ahalik eta datu gehienak bildu beharra.
2. Noiztik hasi zen historian zentsuak egiten?
 - (a) Antzinako zibilizazioetatik, hala nola Txinan eta Egipton.
 - (b) Antzinako Erroman, Herodesek Jesus aurkitzeko Israelen agindu zuen zentsuarekin.
 - (c) Al-Andalusen, arabiarren eskutik.
 - (d) John Graunt kapitainari esker, Londresen XVII. mendean.
3. Nola deitu zitzaion XIX. mendera arte European datu geografiko, ekonomiko eta politikoen bildumari?
 - (a) Staat Statistik, Alemanian.
 - (b) State Politics Data, John Lockeren eskutik Ingalaterran.
 - (c) Gobernantza kuantitatiboa.
 - (d) Aritmetika politikoa
4. Ospitale batean batezbeste 18 gaixo egoten da astero Zainketa Berezien Geletan COVIDagatik ingresaturik. Pairatzen dituzten sukarraldietan gorputzean izan duten tenperatura maximoa jasotzen da. Zein da grafiko egokiena datu horiek sexuaren arabera bereizita jasotzeko?
 - (a) Histograma
 - (b) Kaxa-diagrama.
 - (c) Puntu-diagrama.
 - (d) Aurreko biak dira egokiak.
5. Zertarako erabiltzen da klase-marka?
 - (a) Tartetan emandako maiztasun-banaketa batean egindako hurbilketen errorea estimatzeko.
 - (b) Tartean emandako maiztasun-banaketa batean batezbesteko aritmetikoa eta bestelako estatistikoak emateko.
 - (c) Maiztasun poligonoa modu egokian eratzeko.
 - (d) Aurreko biak dira zuzenak.
6. Datu-multzo batean, $Q_3 = 35, Q_1 = 25$, datu handiena 55 da, datu txikiena 18. Kaxa-diagrama eratzerakoan, zein da egiazkoa honako hauetatik?
 - (a) Gutxienez datu atipiko bat dago goitik. Albo-balioa azpitik 18 da.
 - (b) Ez dago datu atipikorik ez goitik ez behetik.
 - (c) Albo-balioak behetik eta goitik 10 eta 50 dira, hurrenik hurren.
 - (d) Datu atipikoak goitik nahiz behetik 55 eta 18 dira.
7. Datu-multzo batean, $Q_3 = 60, Q_1 = 40$. Datu ordenatuak: 8-14-22-36-...-54-62-66-78-88-92-94. Kaxa-diagrama eratuz gero, zenbat datu atipiko agertzen dira?
 - (a) 1
 - (b) 2
 - (c) 3
 - (d) 3 baino gehiago

-
8. Datu-multzo batean, $Q_3 = 60, Q_1 = 40$. Datu ordenatuak: 8-14-22-36-42-...-54-62-88-92-94. Kaxa-diagrama eratuz gero, zeintzuk dira albo-balioak?
- (a) 10 eta 90
 - (b) 8 eta 92
 - (c) 14 eta 88
 - (d) 36 eta 62
9. Datu ordenatuak: 8-14-22-36-42-54-62-88-89-90-100-110. Kalkulatu 9. dezila.
- (a) 92
 - (b) 98
 - (c) 102
 - (d) 108
10. Errenta-datuak: 8-14-22-36-42-54-62-88-89-90-100-110. Pobrezia-muga: 25. Kalkulatu pobreziaren intentsitatea.
- (a) 0.41
 - (b) 0.51
 - (c) 0.61
 - (d) 0.71
11. Lorenz kurbako puntuak: $(p_1 = 0.1, q_1 = 0.02); (p_2 = 0.4, q_2 = 0.29); (p_3 = 0.6, q_3 = 0.54); (p_4 = 0.9, q_4 = 0.87)$. Kalkulatu Palma ratioa.
- (a) 0.33
 - (b) 0.44
 - (c) 1.61
 - (d) 3
12. 6 tamainako lagin batean, lagin-bariantza 18 suertatu da. Zenbat da populazio-bariantza?
- (a) 25.92
 - (b) 18.76
 - (c) 21.6
 - (d) 15
13. Zertarako erabiltzen da etha korrelazio-ratioa?
- (a) Bi aldagai kuantitatiboen arteko korrelazioa 0-1 tartera mugatzeko.
 - (b) Bi aldagai kuantitatibo kualitatibo bihurtzen diren kasuan, haien arteko asoziazio maila kuantifikatzeko.
 - (c) Bi aldagai kuantitatibo tarte batzuetan banatzen diren kasuan, haien arteko korrelazioa estimatzeko.
 - (d) Aldagai kuantitatibo eta aldagai kualitatibo baten arteko korrelazioa kuantifikatzeko.
14. $F(x) = k\sqrt{x}; 0 < x < 4$. Zenbat izan behar da k benetan banaketa-funtzioa izan dadin?
- (a) 1/4
 - (b) 1/2
 - (c) 2
 - (d) 4

15. $F(x) = 2x; 0 < x < 0.5$. Zenbat da $P[X > 0.4]$?
- (a) 0.2
 - (b) 0.4
 - (c) 0.8
 - (d) 1
16. $P[X = x] = x/6; x = 1, 2, 3$. Zenbat da $P[X \geq 2]$?
- (a) 0.16
 - (b) 0.33
 - (c) 0.66
 - (d) 0.83
17. $P[X = x] = x/6; x = 1, 2, 3$. Kalkulatu itxaropen matematikoa.
- (a) 1.33
 - (b) 1.66
 - (c) 2.33
 - (d) 2.66
18. Probabilitate-banaketa batean bi balio dira posible: 1 eta 2, 0.5eko probabilitatearekin biak. Kalkulatu bariantza.
- (a) 0.25
 - (b) 0.5
 - (c) 1
 - (d) 1.5
19. A inbertsioaren etekinak: itxaropena 25EUR, desbideratzea 5EUR. B inbertsioaren etekinak: itxaropena 40EUR, desbideratzea 10EUR. Zein da komenigarriena? Beharrezkoa bada, utilitate funtzio gisa itxaropena/desbideratzea baliatu.
- (a) Epe luzera A, epe laburrera A
 - (b) Epe luzera A, epe laburrera B
 - (c) Epe luzera B, epe laburrera A
 - (d) Epe luzera B, epe laburrera B
20. Eguneke ekoizpena 1 edo 2 izan daiteke gaur, biak 0.5eko probabilitatearekin. 1 bada, bihar 1 edo 2 izan daiteke, biak 0.5eko probabilitatearekin. 2 bada, bihar 0 edo 1 izan daiteke, biak 0.5eko probabilitatearekin. Kalkulatu bi egunetako batezbesteko ekoizpena.
- (a) 2
 - (b) 2.25
 - (c) 2.5
 - (d) 3

Estatistika eta datuen analisia

2022ko ekainaren 22a

Izena eta abizenak: **Josemari Sarasola**

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

KOPURUA

ONGI	10
GAIZKI	0
ERANTZUN GABE	0