

# ESTADÍSTIKA ETA DATUEN ANALISIA

Azterketa ebatziak

2015-2016 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: 2015eko azaroaren 3a (Partziala: Lehen zatia)

Iraupena: 90 min

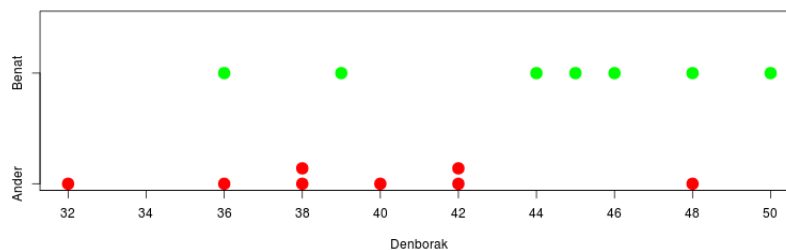
**I. ebazkizuna** (3.25 puntu)

Bi txoferrek autobus ibilbide berdina egiteko behar duten denbora jaso da hainbat alditan (minututan):

*Ander:* 32-38-42-48-36-40-42-38*Beñat:* 36-39-46-45-44-48-50**Egin beharreko atazak:**

1. Bi datu-multzoak alderatzeko diagrama egokia eratu. (0.25 puntu)

Aldagai kuantitatiboak dira, balio ezberdin asko har ditzaketenak, baina datu gutxi dadue aldi berean. Diagrama egokiena puntu-diagrama da egoera horretan.



2. Aurreko diagrama interpretatu, begi bistan, denbora zentralari eta denboren sakabanatzeari buruz, ahal bada. (0.25 puntu)

Anderren denborak 38-40 minutu inguruan biltzen dira. Benatenak 44 inguruan. Beraz, Anderrek ibilbideak azkarrago egiten ditu oro har.

Sakabanatzeari buruz ezin da ondorio ziurrik atera: biek antzeko sakabanatzea dute begi bistan, bistan muturreko izan daitezkeen datuak sartuta zein kenduta.

3. Populazio-bariantza kalkulatu bi datu-multzoetarako, eta horietatik abiatuta kalkulatu lagin-bariantza ere. (0.5 puntu)

Populazio bariantza:  $s_x^2 = \frac{\sum_i x_i^2}{n} - \bar{x}^2$

Kalkulurako taulak:

$x_A$	$x_A^2$	$x_B$	$x_B^2$
32	1024	36	1296
38	1444	39	1521
42	1764	46	2116
48	2304	45	2025
36	1296	44	1936
40	1600	48	2304
42	1764	50	2500
38	1444	-	-
316	12640	308	13698

Populazio bariantzak:

$$\bar{x}_A = \frac{316}{8} = 39.5 ; s_{xA}^2 = \frac{12640}{8} - 39.5^2 = 19.75$$

$$\bar{x}_B = \frac{308}{7} = 44 ; s_{xB}^2 = \frac{13698}{7} - 44^2 = 20.85$$

Lagin bariantzak:

$$\hat{s}_{xA}^2 = \frac{8}{7} s_{xA}^2 = 22.57$$

$$\hat{s}_{xB}^2 = \frac{7}{6} s_{xB}^2 = 24.33$$

4. Aurreko ataleko emaitzetatik abiatuz, zein da denbora sakabanatuenak dituen txoferra? (0.25 puntu)

Aldakortasun-koefizientea baliatu behar da:

$$A_{xA} = \frac{s_{xA}}{\bar{x}_A} = \frac{4.44}{39.5} = 0.11$$

$$A_{xB} = \frac{s_{xB}}{\bar{x}_B} = \frac{4.56}{44} = 0.10$$

Anderren denborak pixka bat sakabanatuagoak dira.

5. Bi datu-multzoen kaxa-diagramak marraztu eta diagrama bakar batean jarri, horretarako kalkuluak taula batean ezarrita. (1 puntu)

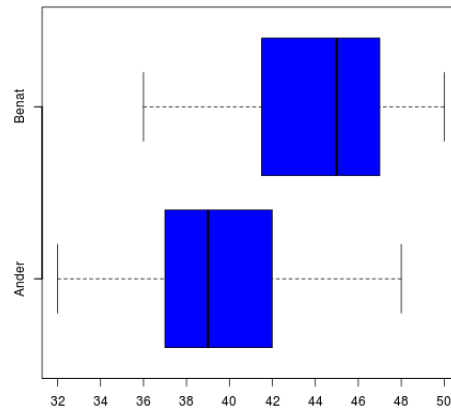
Kalkuluetarako datuak ordenatu egingo dira:

Ander: 32-36-38-38-40-42-42-48

Benat: 36-39-44-45-46-48-50

KAXA DIAGRAMAK	ANDER	BENAT
$Me$	39	45
$Q_1$	36	38.25
$Q_3$	42	46.5
$Q_3 - Q_1$	6	8.25
$1.5(Q_3 - Q_1)$	9	12.375
$Q_1 - 1.5(Q_3 - Q_1)$	27	25.875
Albo-balioak azpitik	32	36
Outlier balioak azpitik	-	-
$Q_3 + 1.5(Q_3 - Q_1)$	51	58.875
Albo-balioak gainetik	48	50
Outlier balioak gainetik	-	-

Kaxa diagramak R-z marraztu dira. Eskuz marraztuta, Anderren kaxa guttiz simetrikoa litzateke,  $Q_1=36$  baita.



6. Kaxa diagramak alderatu eta interpretatu. (0.25 puntu)

Medianari erreparatur, argi ikusten da Anderren denborak oro har laburragoak direla, 6 minutu laburragoak hain zuzen. Sakabanatzeari buruz, kaxaren zabalerari erreparatur, esan daiteke Benaten datuak zertxobait sakabanatuagoak direla.

7. Bowley koefizientea erabiliz, zein da datu-multzo alboratuena? (0.25 puntu)

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1}$$

$$A_B(\text{Ander}) = \frac{(42 - 39) - (39 - 36)}{42 - 39} = 0$$

$$A_B(\text{Benat}) = \frac{(46.5 - 45) - (45 - 38.25)}{46.5 - 38.25} = -0.63$$

Anderren denborak simetrikoak dira, alboragabeak alegia, %50ari zentralari erreparatuta behinik behin. Benaten datuak, berriz, ezkerrean daude alboraturik.

8. Kurtosi neurri jasankor bat erabili Anderren datuetarako eta Bowley koefizientearekin batera, aztertu datu horietarako banakuntza normala hartzea egokia litzatekeen. (0.5 puntu)

$$K_M = \frac{(P_{87.5} - P_{62.5}) + (P_{37.5} - P_{12.5})}{P_{75} - P_{25}}$$

12.5garren pertzentila?  $8 \times 0.125 = 1$ . Beraz, lehen datua da: 32.

37.5garren pertzentila?  $8 \times 0.375 = 3$ . Beraz, hirugarren datua da: 38.

62.5garren pertzentila?  $8 \times 0.625 = 5$ . Beraz, bosgarren datua da: 40.

87.5garren pertzentila?  $8 \times 0.875 = 7$ . Beraz, zazpigarren datua da: 42.

$$K_M = \frac{(42 - 40) + (38 - 32)}{42 - 36} = 1.33$$

1.33 baino zertxobait handiagoa da. Beraz, banakuntza leptokurtikoa da. Simetrikoa izan arren, mesokurtikoa ez denez, banakuntza normala elitzateke datu horietarako egokia lagin errorearen erreserbapean, mesokurtiko izatetik urrun ez badago ere (beraz, banakuntza normala egokia izan daitekeela esatea ere ez dago gaizki).

**II. ebazkizuna** (2.75 puntu)

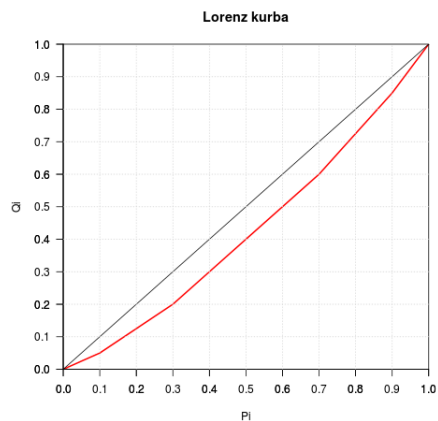
Herri batean familien errentari buruzko inkesta bat egin da. Errenta datu pribatua denez, errentaren hurbilketa soil bat eskatu zaie familiei. Emaitzak maiztasun-taula honetan bildu dira (errentak milaka eurotan):

Errentak	Familiak
10	1
15	2
20	4
25	2
30	1

**Egin beharreko atazak:**

1. Lorenz kurba marraztu. (0.75 puntu)

$x$	$x_{met}$	$p_i$	$q_i$
10	10	0.1	0.05
15	25	0.2	0.125
15	40	0.3	0.20
20	60	0.4	0.30
20	80	0.5	0.40
20	100	0.6	0.5
20	120	0.7	0.6
25	145	0.8	0.725
25	170	0.9	0.85
30	200	1	1



2. Kontzentrazio neurri gisa ondoko hau proposatu da:

$$K = \frac{Q_3 - Q_1}{\bar{x}}$$

Kalkulatu neurri hori emandako datuetarako. (0.75 puntu)

3. Emandako datuak  $X$  izendatuta,  $Y = 2X$  eta  $Z = X - 5$  aldagai-aldaketak burutu, hurrenez hurren, eta aurreko ataleko neurria kalkulatu bi datu-multzo berrietarako. Emaizak ikusita, azter ezazu neurri horrek Pigou-Dalton baldintzak betetzen dituen. (0.75 puntu)

10 15 15 20 20 20 20 25 25 30

$y=2x$

20 30 30 40 40 40 40 50 50 60

$y=x-5$

5 10 10 15 15 15 15 20 20 25

aldagaia	$\bar{x}$	$Q_1$	$Q_3$	$K$
$x$	20	15	22.5	0.375
$2x$	40	30	45	0.375
$x - 5$	15	10	17.5	0.5

Eskala-aldaketaren ondorioz  $K$  ez da aldatzen. Jatorri aldaketaren ondorioz, ordea, bai, igo egiten da. Bi kasuetan, hala behar du, Pigou-Dalton baldintzen arabera. Beraz, kontzentrazio-neurri egokia litzateke.

4. Pobrezia-muga 17.000€-tan ezarririk, kalkulatu pobreziaren intentsitatea eta pobreziaren neurri orokor bat. (0.50 puntu)

Intentsitatea:

$$I = \frac{(17 - 10) + (17 - 15) - (17 - 15)}{3 \times 17} = 0.52$$

Pobrezia neurri orokorra:

$$P = \frac{2[(17 - 10)(3 + 1 - 1) + (17 - 15)(3 + 1 - 2) - (17 - 15)(3 + 1 - 3)]}{(3 + 1) \times 10 \times 17} = 0.079$$

Bestelako erreferentzia ezean, ezin da esan handiak edo txikiak diren.

### III. ebazkizuna: R softwarea (1 puntu)

notak izeneko datu-multzoa duzu R kontsolan sartuta. Ondokoak burutzeko aginduak idatzi:

1. Datuen histograma marraztu, maiztasun erlatiboak erabiliz,  $x$  ardatzean 10-20, 20-30, ..., 90-100 tartek jarritz (sekuentzia egokia jarrita), izenburu eta ardatz-izen edo etiketa egokiekin.
2. Aurrekoaren gainean, maiztasun-poligonoa marraztu, berdez eta lodiera nabarmen batekin.
3. Orain, maiztasun-poligono bera marraztu, baina azpian histograma azaldu gabe.

Oharra: Beharrezko balitz, aurreko aginduetarako behar den paketea deskargatu, instalatu eta aktibatzeke aginduak ere idatzi.

```
>install.packages("agricolae")
>library(agricolae)
>histograma=plot.graph.freq(datuak,breaks=seq(10,100,by=10),
frequency=2,col="pink", main="Histograma",xlab="Aldagaia",ylab="Portzentajeak")
>polygon.freq(histograma,frequency=2,lwd=4,col=c("green"))
>histograma=plot.graph.freq(datuak,breaks=seq(10,100,by=10),
frequency=2,border=F, main="Histograma",xlab="Aldagaia",ylab="Portzentajeak") #poligonoa soilik egiteko,
erantsi border=F eta kolorerik gabe behar du izan
>polygon.freq(histograma,frequency=2,lwd=4,col=c("green")) #poligonoa soilik
```

## ESTADÍSTIKA ETA DATUEN ANALISIA (Testa: Lehen zatia)

Irakaslea: Josemari Sarasola - Data: 2015eko azaroaren 4a

Iraupena: 30 minutu

**Erantzun bakarra da zuzena galdera bakoitzean. Guztira testak 1.5 puntu balio du. Erantzun zuzenak 0.1 puntu balio du. Erantzun okerrak 0.05 puntu kentzen du. Galderak erantzun gabe utz daitezke, punturik gehitu eta kendu gabe.**

ZUZENAK: AADDB-BABDC-ABBDD

- 2007ko hasiera-hasieratik 2011ko bukaerara salmentak 144tik 369ra pasa dira. Zenbat da urteko batez besteko hazkundea?
  - %20.7 gutxi gorabehera.
  - %21.7 gutxi gorabehera.
  - %22.7 gutxi gorabehera.
  - %23.7 gutxi gorabehera.
- Zertarako erabiliko zenuke  $\frac{P_{10}+P_{90}}{2}$  estatistikoa?
  - Zentroa neurtzeko.
  - Zentroa ez den beste kokapen bat emateko.
  - Sakabanatzea neurtzeko.
  - Alborapena neurtzeko.
- Honako hauetatik zein da neurri jasankorra?
  - Ibiltarte erlatiboa.
  - Aldakortasun-koefizientea.
  - Fisher koefizientea.
  - Bowley koefizientea.
- Merkatu bat bi enpresaren artean banatzen da erdibanan. Zenbat da Shannon adierazlea?
  - 0.39
  - 0.49
  - 0.59
  - 0.69
- 10-20-30-140 dira herri bateko errentak. Zenbat da Robin Hood adierazlea?
  - %35
  - %45
  - %55
  - %65
- Kaxa-diagrama batean, kaxaren zabalera 50 eta hirugarren kuartila 100 izanik, zein baliotik aurrerakoak izango dira muturreko datuak?
  - 150
  - 175
  - 200
  - Ezin da jakin.
- 100 datu izanik, gehienez zenbat dira goi aldetikako muturreko datuak kaxa-diagraman oinarrituta?
  - 25
  - 20
  - 15
  - Ezin da jakin.
- Frantziako filosofia-fakultateetako ikasle kopuruak jaso dira. Zein da ikerketa-unitatea?



- 
- (a) Frantzia  
(b) Filosofia-fakultatea.  
(c) Ikasleetako bakoitza.  
(d) Ikasle-kopurua.
9. Datu-multzo bat kurtosiari buruz modu uniforme banatzen den jakin nahi da. Horretarako, Fisker eta Pearson koefizienteak kalkulatu behar dira. Ondoko hauetatik, zein emaitza legokioke uniformetasunari?
- (a)  $A_F = 2, K_P = 3$   
(b)  $A_F = 0, K_P = 3$   
(c)  $A_F = 2, K_P = 1.6$   
(d)  $A_F = 0, K_P = 1.6$
10. 100 daturen batezbesteko aritmetikoa sinplea 20 suertatu da. Beste 200 daturen batezbestekoa 50 suertatu da. Zenbat da 300 datuen (datu guztien) batezbestekoa?
- (a) 30  
(b) 35  
(c) 40  
(d) Ezin da kalkulatu.
11. Noiz da desbideratze estandarra 0?
- (a) Datu guztiak berdinak direnean.  
(b) Datu guztiak 0 direnean.  
(c) Batezbestekotik gorako datuak batezbestekotik beherako datuekin konpentsatzen direnean ( $s_x$  batezbesteko koefizientea da).  
(d) Inoiz ez.
12. Lorenz kurbaren ibilbidean, noiz gertatzen da kurba zuzen bihurtzea edo malda konstante izatea?
- (a) Soilik errenta guztiak berdinak direnean.  
(b) Errenta-datu batzuk berdinak direnean.  
(c) Banako guztiek, batek izan ezik, 0 irabazten dutenean.  
(d) Inoiz ez.
13. Errenten batezbestekoak eta desbideratze estandarrak Gipuzkoan eta Nafarroan:  $\bar{x}_{Gip} = 100, s_{Gip} = 10, \bar{x}_{Naf} = 50, s_{Naf} = 10$ . Gipuzkoan pobrea 80tik behera eskuratzen duena izanik, batezbestekoa eta desbideratzea kontuan harturik, zein izango litzateke pobrezia-muga Nafarroan?
- (a) 10  
(b) 30  
(c) 40  
(d) Beste bat da erantzuna.
14. Zeri buruz izan zen lehen azterketa estatistikoa?
- (a) Datu astronomikoei buruz.  
(b) Probabilitateei buruz.  
(c) Datu antropometrikoei buruz (altuerak, pisuak, ...).  
(d) Datu demografikoei buruz.
15. Zenbat datu daude bigarren oktiletik laugarren kintilera?
- (a) %40.  
(b) %45.  
(c) %50.  
(d) %55.