
ESTADÍSTIKA ENPRESARA APLIKATUA (Praktika: Lehen zatia)

Irakaslea: Josemari Sarasola

Data: 2012ko azaroaren 7a

I. ebazkizuna (2.5 puntu)

Bi eskoletako ikasleei jakintzagai batzuetan egindako azterketak egin eta gainditu ez duten irakasgai kopurua jaso dira:

A eskola: 1-1-2-3-2-0-1-2-0-1-2-3-2-1-1-2-3-4-1-0-0-1-1-2-1-1-2-3-2-2-1-1-2-2

B eskola: 0-0-1-2-0-0-1-1-1-2-3-1-2-0-0-0-0-1-1-0-0-0-1-1-2-3-2-3-0-0-0-0-1-1

Egin beharreko atazak

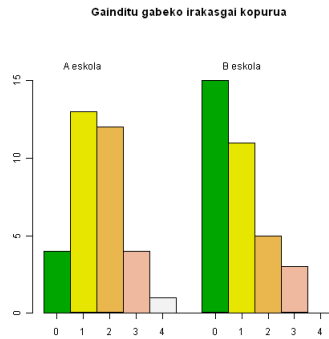
- (a) Taularaketa eta datu-diagrama egokia erabiliz, bi eskoletako ikasleen gainditu gabeko irakasgai-kopuruak alderatu behar dituzu, gutxi gorabehera zein eskolatan suspentitzen den gehien adieraziz. (1 puntu)
- (b) Aurreko emaitzak froga itzazu bi eskoletarako batezbesteko aritmetiko sinplea kalkulatu, aurreko atalean osaturiko taula oinarritzat harturik. (0.5 puntu)
- (c) Bi eskoletako datuetarako desbidazio estandarrak kalkulatu, behar den moduan alderatu eta emaitzak interpretatu. (1 puntu)

Ebazpena

(a)

Gainditu gabeko irakasgaiak	A eskola	B eskola
0	4	15
1	13	11
2	12	5
3	4	3
4	1	0
Guztira (ikasleak)	34	34

Aldagaiak balio ezberdin gutxi hartzen dituzenez, zutabe-diagrama da grafiko egokiena datuak irudikatzeko.



```
(Informazio huts moduan, ez baitabarrezkoa noski, hemen dago R kodea:
> dena=cbind(c(4,13,12,4,1),c(15,11,5,3,0))
> barplot(dena,col=terrain.colors(5),beside=T,ylim=c(0,18))
> axis(1,at=c(1.5,2.5,3.5,4.5,5.5,7.5,8.5,9.5,10.5,11.5),labels=c("0","1","2","3","4","0","1","2","3","4"))
> text(2.5,16,"A eskola")
> text(8.5,16,"B eskola")

> title("Gainditu gabeko irakasgai kopurua")
```

Diagraman argi ikusten da B eskolan gainditu gabeko irakasgai kopuru txikiagoa izaten dela, 0-1 gehienetan, A eskolan gehienetan 1-2 izaten den bitartean.

(b)

Batezbesteko aritmetiko sinplearen kalkulua azaltzen da jarraian, eskolaren arabera bereizita.

x	$n(A)$	$N(B)$	$n(A) \times x$	$n(B) \times x$
0	4	15	0	0
1	13	11	13	11
2	12	5	24	10
3	4	3	12	9
4	1	0	4	0
	34	34	53	30

$$\bar{x}_A = \frac{53}{34} = 1.56$$

$$\bar{x}_B = \frac{30}{34} = 0.88$$

A eskolan 1.56 irakasgai ez du gainditzen ikasle batek batez beste; B eskola, berriz, kopuru hori 0.88 da. B eskolan, beraz, gutxiago suspenditzen da batez bestez.

(c)

Desbiderazio estandarrak kalkulatu dira jarraian:

x	$n(A)$	$n(B)$	$n(A) \times x$	$n(B) \times x$	$n(A) \times x^2$	$n(B) \times x^2$
0	4	15	0	0	0	0
1	13	11	13	11	13	11
2	12	5	24	10	48	20
3	4	3	12	9	36	27
4	1	0	4	0	16	0
	34	34	53	30	113	58

$$s_{xA} = \sqrt{\frac{113}{34} - 1.56^2} = 0.94$$

$$s_{xB} = \sqrt{\frac{58}{34} - 0.88^2} = 0.96$$

Alderatu ahal izateko, batezbesteko aritmetiko sinplearekin alderatu behar dira, aldakortasun koefizientearen bitartez:

$$A_{xA} = \frac{0.94}{1.56} = 0.60$$

$$A_{xB} = \frac{0.96}{0.88} = 1.09$$

Gaizto gabeko irakasgai kopuruaren sakabanatzea ikasleen artean handiagoa da B eskoletan.

II ebazkizuna (3.5 puntu)

Gros-Antiguo-Igara ibilbidea egiten duten DBUSeko bi txoferrek bidaiari eman duten denbora-datu batzuk jaso dira (minututan), txoferraren arabera bereizirik:

Kepa: 24-28-22-38-28-20-22-24-21-25-26-24-27-23

Ana: 28-29-30-26-27-24-25-39-42-29-32

Egin beharreko atazak:

- Sakabanatze-neurri jasankor bat erabiliz eta bi txoferren datuak alderatuz, adieraz ezazu zein den denbora *homogeneoagoak* dituen txoferra. (1 puntu)
- Bi txoferrei dagozkien kaxa-diagramak marraztu behar dituzu eta interpretatu zentroari eta sakabanatzeari buruz, horretarako kalkuluak taula batean aurretik adieraziz (besoen luzerak, ...). (1.5 puntu)
- Anaren datuetarako banakuntza normala egokia litzatekeen erabaki ezazu, horretarako Bowleyen eta Moorsen koefizienteak kalkulatu. (1 puntu)

(a) Bi txoferren datuak alderatzeko sakabanatze-neurri ERLATIBOA behar da eta horietan jasankorra hau da:

$$IQR_{erlatiboa} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Kuartilak kalkulatzeko datuak ordenatu egin behar dira:

Kepa: 20-21-22-22-23-24-24-24-25-26-27-28-28-38

Ana: 24-25-26-27-28-29-29-30-32-39-42

Keparen datuetarako kuartilak

lehen kuartila? $14 \cdot 0.25 = 3.5$ garren datua

3 datua: 22, 4. datua: 22

Interpolazioa egin beharrik ez dago : $Q_1 = 22$.

hirugarren kuartila? $14 \cdot 0.75 = 10.5$ garren datua

10. datua: 26, 11. datua: 27

$Q_3 = 0.5 \times 26 + 0.5 \times 27 = 26.5$

Anaren datuetarako kuartilak:

lehen kuartila? $11 \cdot 0.25 = 2.75$ garren datua

2 datua: 25, 3. datua: 26

$$Q_1 = 0.25 \times 25 + 0.75 \times 26 = 25.75$$

hirugarren kuartila? $11 \cdot 0.75 = 8.25$ garren datua

8. datua: 26, 9. datua: 27

$$Q_3 = 0.75 \times 30 + 0.25 \times 32 = 30.5$$

Orain, kuartil arteko ibiltarte erlatiboa bi txoferretarako kalkulatzeko soilik falta da:

$$IQR_{erlatiboa}(Kepa) = \frac{26.5 - 22}{26.5 + 22} = 0.092$$

$$IQR_{erlatiboa}(Ana) = \frac{30.5 - 25.75}{30.5 + 25.75} = 0.084$$

(b) Kaxa-diagrama eratzeko, kuartilak eta mediana behar dira. Mediana soilik kalkulatu behar da, kuartilak aurreko ataletik kalkulatu dauzkagutako:

Keparen datuetarako mediana:

datu kopurua : 14

mediana? datu kopurua bikoitia: erdiko bi datuak: 24-24 : Me=24

Anaren datuetarako mediana:

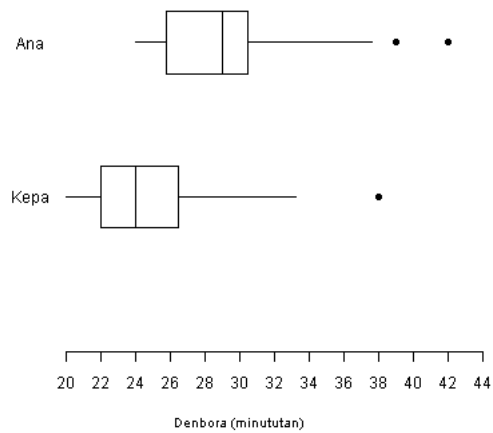
datu kopurua: 11

mediana? datu kopuru bakoitia: erdiko datua: 29

Orain, medianak, kuartilak, besoen luzerak, besoak mozteko muturreko datuak taula honetan jasoko dira:

Parametroa	Kepa	Ana
Me	24	29
Q_1	22	25.75
Q_3	26.5	30.5
$1.5(Q_3 - Q_1)$	6.75	7.125
$Q_1 - 1.5(Q_3 - Q_1)$	15.25	18.625
datuak azpitik?	ez	ez
min	20	24
$Q_3 + 1.5(Q_3 - Q_1)$	33.25	37.625
datuak gainetik?	bai	bai
zein?	38	39,42

Kaxa-diagramak



Zentro-joerari dagokionean, Anak oro hat denbora luzeagoa ematen du ($Me=29$, kaxako erdiko marra) Kepak baino ($Me=24$). Zehatzago $29-24=5$ minutu gehiago, mediana erabiltzen bada.

Kaxaren eta besoen zabalerari erreparatuz, bi txoferren denbora-datuak antzeko sakabanatzea dutela esan daiteke.

(c)

Anaren datuak hartuta, Bowleyen alborapen-koefizientea kalkulatzeko kuartilak eta mediana kalkulaturik ditut iada aurreko ataletik:

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} = \frac{(30.5 - 29) - (29 - 25.75)}{30.5 - 25.75} = -0.37$$

Ezker alderako alborapena dago itxuraz, oso nabarmena ez bada ere, Otik nahiko gertu baitago.

Moorsen koefizientea kalkulatzeko lau pertzentil berezi kalkulatu behar dira.

- $P_{87.5}$? $0.875 \cdot 11 = 9.625$ garren datua
9. datua: 32, 10. datua: 39

$$P_{87.5} = 32 \times (1 - 0.625) + 39 \times 0.625 = 36.375$$

- $P_{62.5}$? $0.625 \cdot 11 = 6.875$ garren datua
6. datua: 29, 7. datua: 39

$$P_{62.5} = 29 \times (1 - 0.875) + 39 \times 0.875 = 29$$

- $P_{37.5}$? $0.375 \cdot 11 = 4.125$ garren datua
4. datua: 27, 5. datua: 28

$$P_{37.5} = 27 \times (1 - 0.125) + 28 \times 0.125 = 27.125$$

- $P_{12.5}$? $0.125 \cdot 11 = 1.375$ garren datua
1. datua: 24, 2. datua: 25

$$P_{12.5} = 24 \times (1 - 0.375) + 25 \times 0.375 = 24.375$$

Moorsen koefizientea kalkulatzeko da jarraian:

$$K_2 = \frac{(P_{87.5} - P_{62.5}) - (P_{37.5} - P_{12.5})}{Q_3 - Q_1} = \frac{(36.375 - 29) + (27.125 - 24.375)}{30.5 - 25.75} = 2.13$$

Banakuntza normalerako erreferentzia moduan erabiltzen den 1.23 balioa baino handiagoa denez, banakuntza leptokurtikoa da.

Alboratua eta leptokurtikoa denez, banakuntza normala ez da egokia datu hauek modelizatzeko.

III ebazkizuna (1.5 puntu)

errenta izeneko datu-bilduma osatu da R softwarean non herri bateko errentak jasotzen diren (436, 563, ...). Errenta txikiena 106 da eta handiena 967.

Egin beharreko atazak:

- (a) Histograma eratu behar duzu, datuak 100-200, 200-300, 300-400, tartetan bilduz, horretarako aurretik *sekuentzia* egokia sortuz eta horri izen bat emanaz. Histogramari izenburu bat jarri behar diozu eta ardatz bertikalean maiztasun absolutuak azaldu behar dira. Ardatz bertikaleko balioak horizontalean agertu behar dira. (0.3 puntu)
- ```
> tarteak=seq(100,1000,by=100)
> hist(errenta,breaks=tarteak,main="Errenten banaketa",las=1,ylab="Maiztasun absolutuak")
```
- (b) Aurrekoaren berdina, baina maiztasun erlatiboekin. (0.2 puntu)
- ```
> library(agricolae)
> plot.graph.freq(errenta,breaks=tarteak,frequency=2,main="Errenten banaketa",ylab="Portzentajeak")
```
- (c) Maiztasun poligonoa eratu, nahi dituzun maiztasunekin (absolutuekin edo erlatiboekin), lodiera jakin bat zehaztuz. (0.3 puntu)
- ```
> library(agricolae)
>errentahist=plot.graph.freq(errenta,breaks=tarteak,frequency=2,main="Errenten banaketa",ylab="Portzentajeak")
> polygon.freq(errentahist,col="green",lwd=3)
```
- (d) 500 edo gutxiago irabazten dutenak soilik aukeratu datu-multzo osotik, horiei izen bat eman eta batezbestekoa kalkulatu. (0.3 puntu)
- ```
> txikiak=errentak[errentak<=500]
> mean(txikiak)
```
- (e) Dezil guztiak kalkula itzazu, agindu bakar baten bitartez, horretarako aurretik *sekuentzia* bat sortuz eta horri izen bat emanaz. (0.2 puntu)
- ```
> dezil=seq(0,1,by=0.1)
> quantile(errenta,prob=dezil)
```
- (f) Lorenzen kurbaren grafikoa osatu. (0.1 puntu)
- ```
> library(ineq)
> plot(Lc(errentak))
```
- (g) Lorenzen kurbako p_i eta q_i balioak eskuratu itzazu. (0.1 puntu)
- ```
> library(ineq)
> Lc(errentak)
```



---

ESTADÍSTIKA ENPRESARA APLIKATUA (Testa: Lehen  
partziala).Irakaslea: Josemari Sarasola. Data: 2012ko azaroaren 7a

Iraupena: 20 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.**

**Erantzun zuzenak: BABAD DBACA ABCDA**

- (1) Zein izan zen XIX. mendean estatistika gizarte zientzietara zabaldu zuena?
  - (a) John Graunt
  - (b) Adolphe Quetelet
  - (c) John Tukey
  - (d) Udney Bowley
- (2) Zer da "data mining"?
  - (a) Datu-multzo handien ikerketarako teknika-multzoa.
  - (b) Datu-multzoetatik muturreko datuak hauteman eta eraztea.
  - (c) Datu-multzoak teknika grafikoak erabiliz aztertzea.
  - (d) Datu-multzoen azterketa esploratzailea.
- (3) 50 bidaietan autobus batek ibilbidea egiteko behar duen denbora jaso da minututan. Zein da datu horietarako diagrama egokiena?
  - (a) Puntu-diagrama.
  - (b) Histograma.
  - (c) Zutabe-diagrama, datuak taula batean bildu ondoren.
  - (d) Zutabe-diagrama, bidaia bakoitzeko zutabe bat altxatuz.
- (4) Zein da neurri jasankorra?
  - (a)  $Q_3 - Q_1$
  - (b) Desbidazio estandarra
  - (c) Fisherren alborapen koefizientea.
  - (d) Aldakortasun-koefizientea.
- (5)  $x$  datuen desbiderazio estandarra 5 da. Datu guztiak bider 2 eta, ondoren, gehi 4 egiten badira, zenbat da datu berrien desbiderazioa?
  - (a) 9
  - (b) 10
  - (c) 14
  - (d) 22

- 
- (6) Noiz da bariantza 0?
- (a) Datu guztiak berdinak direnean.
  - (b) Datu bakarra daukagunean.
  - (c) Inoiz ez da 0.
  - (d) (a) eta (b) dira egiazkoak.
- (7) Makina baten ekoizpen edo errendimendu estandarra 100 da orduko. Praktikan, ordea, kopuru horretatik aldentzen da, gora nahiz behera. Desbiderazio horiek kalkulatu dira. Zein da estandar horretatik desbideratzen den kopuruaren batezbestekoa kalkulatzeko egokiena:
- (a) Batezbesteko harmonikoa.
  - (b) Batezbesteko koadratikoa.
  - (c) Errendimenduen desbiderazio estandarra.
  - (d) (a) eta (c) dira egiazkoak.
- (8) Noiz dator bat Lorenzen kurba diagonalarekin?
- (a) Datu guztiak berdinak direnean.
  - (b) Total osoa banako bakar batek hartzen duenean.
  - (c) Inoiz ez, beti aldentzen da, piska bat bada ere.
  - (d) Aurreko guztiak faltsuak dira.
- (9) Hauetatik, zein aldagaitarako ezin da kontzentrazio aztertu?
- (a) Kate bateko denden salmentak, dendaz denda.
  - (b) Hirietako biztanleriak, hiriz hiri.
  - (c) Ikasleen adinak, ikaslez ikasle.
  - (d) Unibertsitateko ikasle kopuruak, fakultatez fakultate.
- (10) Zenbatgarren pertzentila da 2. kintila?
- (a) 40.a
  - (b) 10.a
  - (c) 50.a
  - (d) 30.a
- (11) Zein da egiazkoa?
- (a)  $P_{30} - P_{70}$  beti da negatiboa.
  - (b)  $P_{30} - P_{70}$  positiboa balitz, eskubirako alborapena dagoela adieraziko luke.
  - (c) Sakabanatze-neurri absolututzat har daiteke, kuartil arteko ibiltartea bezalaxe.
  - (d) Aurreko guztiak faltsuak dira.

- 
- (12) R-n emaitza hauek eskuratu dira:  $\text{skewness}(x)=3.06$ ,  $\text{kurtosis}(x)=1.23$ . Nola interpretatu behar dira?
- (a)  $x$  datuetarako banakuntza normala erabil daiteke.
  - (b) Heterogeneotasuna egon daiteke.
  - (c) Aurreko biak faltsuak dira.
  - (d) Lehenengo biak egiazkoak dira.
- (13) Errenta hauek jaso dira inkesta batean: 30, 20, 50, 100. Pobrezia-muga 40 da. Pobreziaren intentsitatearen neurria kalkulatu.
- (a) 0.125
  - (b) 0.275
  - (c) 0.375
  - (d) 0.425
- (14) Sektore bat horrela banaturik dago enpresen artean: A:%20, B:%50, C:%30. Sektorearen kontzentrazioa neurtzen duen Shannonen indizea kalkulatu behar da, era absolutuan.
- (a) 0.73
  - (b) 0.83
  - (c) 0.93
  - (d) 1.03
- (15) Errenta hauek jaso dira inkesta batean: 30, 20, 50, 100. Robin Hood adierazlea kalkulatu.
- (a) 0.25
  - (b) 0.35
  - (c) 0.45
  - (d) 0.55