

Autoevaluazioa

4. ARIKETA

(10 PUNTU)

FSA paketea kontsideratuko dugu.

- a) Instalatu eta kargatu FSA paketea. (0.5 puntu)

FSA paketearen barnean dagoen **ChinookArg** datu-markoa kontsideratuko dugu. Datu-marko honetan Chinook motako izokinen luzera (cm), pisua (kg) eta arrantzatutako kokalekua ageri dira.

- b) **ChinookArg** datu-markoaren egitura lortu. (0.5 puntu)
- c) Irudikatu arrantzatutako izokinen sektore-diagrama arrantzatutako kokalekuaren arabera. Errepikatu gutxienez 80 cm-ko luzera duten izokinekin. (puntu 1)
- d) Argentinan arrantzatutako izokinen artean aurkitu gehienez 60 cm-ko luzera edo gehienez 6 kg-ko pisua duten izokinak. (puntu 1)
- e) Ondorengo sailkapen erregela aplikatuz, izokinak berain pisuaren arabera sailkatzen duten zorizko aldagaia datu markoan sartu: (puntu 1)

$$k_i = \begin{cases} \text{Baxua} & w_i < \text{Mediana} \\ \text{Altua} & w_i \geq \text{Mediana} \end{cases}$$

- f) %1 adierazgarritasun-mailaz aurreko atalean definitutako izokinen sailkapena eta arrantzatutako lekua independenteak diren aztertu. (puntu 1)
- g) Argentinan arrantzatutako izokinen luzera histograma baten bidez grafikoki adierazi eta hipotesi kontraste ez-parametrikoko bat eginez, %5 adierazgarritasun-mailaz, aldagai horren normala aztertu. (puntu 1)
- h) Puyehuen arrantzatutako izokinen luzera histograma baten bidez grafikoki adierazi eta hipotesi kontraste ez-parametrikoko bat eginez, %5 adierazgarritasun-mailaz, aldagai horren normala aztertu. (puntu 1)
- i) Argentinan arrantzatutako izokinen batezbestekoaren puntu-estimazioa eta %5eko tarte estimazio egin. %5 adierazgarritasun-mailaz, batezbesteko luzera 85 cm dela onar al daiteke? (1.5 puntu)
- j) Argentinan arrantzatutako izokinen luzeraren bariantzaren puntu-estimazioa eta %5eko tarte estimazio egin. %5 adierazgarritasun-mailaz, luzeraren bariantza 195 cm² dela onar al daiteke? (1.5 puntu)

Ebazpena :

a)

FSA paketea instalatu eta kargatuko dugu:

```
> install.packages("FSA")  
> library(FSA)
```

b)

ChinookArg datu-marcoaren egitura bistaratzeko str() funtzioa erabiliko dugu:

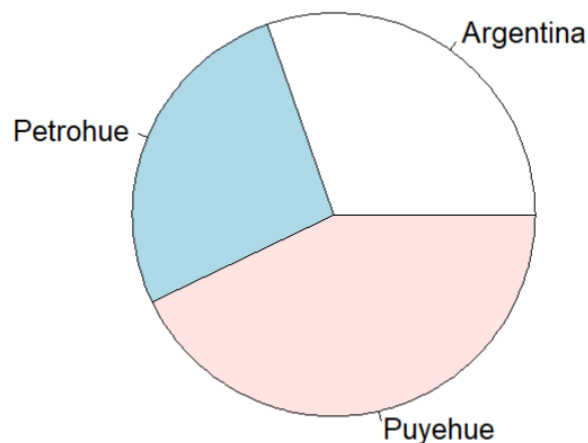
```
> str(ChinookArg)  
'data.frame': 112 obs. of 3 variables:  
 $ tl : num 120 115 111 110 110 ...  
 $ w : num 17.9 17.2 16.8 15.8 14.3 13.8 12.8 11.7 ...  
 $ loc: Factor w/ 3 levels "Argentina", "Petrohue", ...: 1 1  
 1 1 1 1 1 ...
```

Datu-marcoak 112 behaketa eta 3 aldagai ditu.

c)

```
> pie(table(loc), main="Arrantzatutako izokinen  
sektore-diagrama")
```

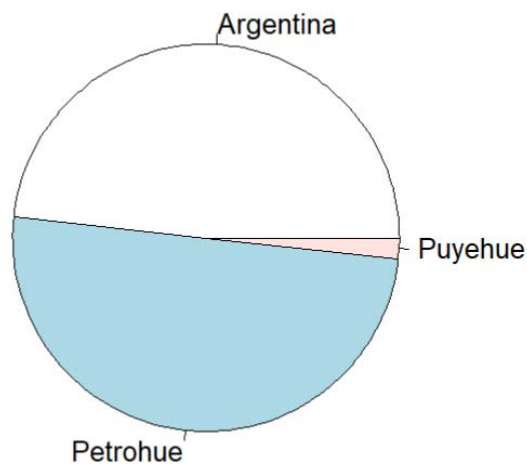
Arrantzatutako izokinen sektore-diagrama



Orain, histograma errepikatuko dugu baina bakarrik 80 cm-ko gutxienezko luzera duten izokinekin:

```
> Chi.nookArg.tl.80 <- Chi.nookArg[tl >=80, ]
> pie(table(Chi.nookArg.tl.80$loc), main="Gutxienez 80
cm-ko luzera duten izokinen sektore-diagrama")
```

Gutxienez 80 cm-ko luzera duten izokinen sektore-diagrama



Argi ikus daiteke Puyehue kokalekuan, gehienbat, luzera laburreko izokinak daudela.

d)

Argentinan arrantzatutako izokinen artean gehienez 60 cm-ko luzera edo gehienez 6kg-ko pisua duten izokinak:

```
> Chi.nookArg[(tl <=60 | w<=6) & loc=="Argenti na", ]
  tl  w  loc
20 92.0 5.9 Argenti na
21 87.7 5.9 Argenti na
26 83.8 5.7 Argenti na
33 66.8 5.2 Argenti na
34 59.9 3.9 Argenti na
```

e)

Medianaren arabera izokinak berain pisuaren arabera sailkatuko ditugu:

```
> medi ana<-medi an(w)
> Chi nookArg$kat<-as. factor(i fel se(w<medi ana, "Baxua",
"Al tua"))
> Vi ew(Chi nookArg)
```

	tl	w	loc	kat		tl	w	loc	kat
1	120.1	17.9	Argentina	Altua	14	92.2	9.1	Argentina	Altua
2	115.0	17.2	Argentina	Altua	15	94.9	11.8	Argentina	Altua
3	111.2	16.8	Argentina	Altua	16	92.7	11.3	Argentina	Altua
4	110.2	15.8	Argentina	Altua	17	92.9	11.9	Argentina	Altua
5	110.0	14.3	Argentina	Altua	18	89.9	11.8	Argentina	Altua
6	109.7	13.8	Argentina	Altua	19	88.7	10.8	Argentina	Altua
7	105.0	12.8	Argentina	Altua	20	92.0	5.9	Argentina	Baxua
8	100.1	11.7	Argentina	Altua	21	87.7	5.9	Argentina	Baxua
9	98.0	12.8	Argentina	Altua	22	85.1	9.0	Argentina	Altua
10	92.1	14.8	Argentina	Altua	23	85.1	9.8	Argentina	Altua
11	99.0	9.7	Argentina	Altua	24	82.9	8.7	Argentina	Altua
12	97.9	7.3	Argentina	Altua	25	82.9	7.8	Argentina	Altua
13	94.9	7.8	Argentina	Altua	26	83.8	5.7	Argentina	Baxua

f)

Atal honetan, independentzia proba bat egin behar da, horretarako lehenik eta behin kontingentzia taula eraiki behar da:

```
> konti g_ taul a<-tabl e(l oc, Chi nookArg$kat)
> konti g_ taul a
l oc          Al tua Baxua
Argenti na    29     5
Petrohue      25     5
Puyehue       2     46
```

Kontingentzia taula erabiliz, hurrengo hipotesi-contraste ez-parametrikoa egin daiteke:

H_0 : Izokinen sailkapena eta arrantzatutako lekua independenteak dira.

H_a : Izokinen sailkapena eta arrantzatutako lekua ez dira independenteak.

```
> chi sq. test(konti g_ taul a)
      Pearson' s Chi -squared test

data:  konti g_ taul a
X-squared = 70.608, df = 2, p-value = 4.653e-16
```

$4.653e-16 > 0.01$, hau da, p-balioa adierazgarritasun maila baino txikiagoa denez hipotesi nulua ez da onartuko. %1eko adierazgarritasun mailaz, ezin daiteke onartu izokinen sailkapena eta arrantzatutako lekua aldagai independenteak direla.

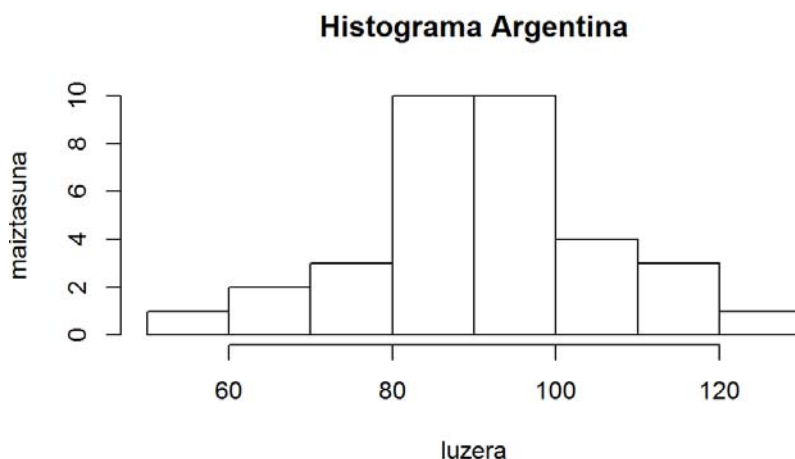
g)

Lehenengo eta behin, ChinookArg datu markotik Argentinako datuak Argentina izeneko beste datu marko batean gordetzen dira.

```
> Argenti na<-ChinookArg[iloc=="Argenti na", ]
```

Ondoren, tl aldagaiaren, hau da, izokinen pisua laburbiltzen duen aldagaiaren histograma irudikatzen da:

```
> hi st(Argenti na$tl ,mai n="Hi stograma Argenti na", xl ab =
      "Luzera", yl ab="Mai ztasuna")
```



Histograman aldagaiak banaketa normala jarrai dezakeela dirudi. Ondorengo hipotesi kontraste ez-parametrikoren bidez susmoa baieztatzen den azter daiteke:

H_0 : Argentinan arrantzatutako izokinen luzerak banaketa normala jarraitzen du.

H_a : Argentinan arrantzatutako izokinen luzerak ez du banaketa normala jarraitzen.

```
> nrow(Argenti na)
[1] 34
```

Laginaren tamaina $n = 34 \leq 50$ denez, aldagaiaren normaltasuna kontrastatzeko Shapiro-Wilk testa erabiliko da:

```
> shapiro.test(Argenti na$t1)
Shapiro-Wilk normality test

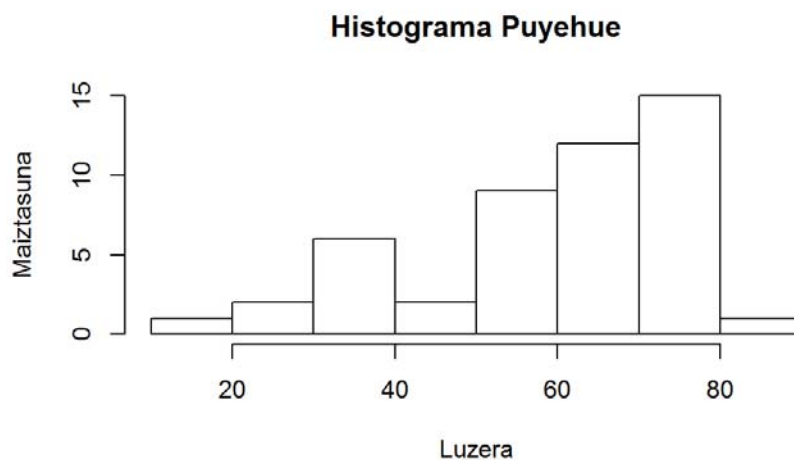
data:  Argenti na$t1
W = 0.984, p-value = 0.8875
nrow(Argenti na)
```

$0.8875 > 0.05$, hau da, p-balioa adierazgarritasun maila baino handiagoa denez hipotesi nulua onar daiteke. Hau da, %5eko adierazgarritasun mailaz, Argentinan arrantzatutako izokinen luzerak banaketa normala jarraitzen dutela onar daiteke.

h)

Aurreko atalean bezalaxe, Puyehuen arrantzatutako datuak soilik dituen datu-markoa sortu eta histograma irudikatzen da:

```
> Puyehue<-Chi nookArg[Loc=="Puyehue", ]
> hist(Puyehue$t1, mai n="Hi stograma Puyehue", xlab =
"Luzera", ylab="Mai ztasuna")
```



Histograman aldagaiak banaketa normala jarraitzen ez duela dirudi. Dagokion hipotesi kontraste ez-parametrikoren bidez susmoa baieztatzen den ikusiz:

H_0 : Puyehuen arrantzatutako izokinen luzerak banaketa normala jarraitzen du.

H_a : Puyehuen arrantzatutako izokinen luzerak ez du banaketa normala jarraitzen.

Kasu honetan ere laginaren tamaina 50 baino txikiagoa denez, Shapiro-Wilk kontrastea erabiliko da normaltasuna aztertzeko:

```
> nrow(Puyehue)
[1] 48
> shapiro.test(Puyehue$t1)
Shapiro-Wilk normality test

data:  Puyehue$t1
W = 0.91801, p-value = 0.002532
```

$0.002532 < 0.05$, hau da, p-balioa adierazgarritasun maila baino txikiagoa denez ezin da hipotesi nulua onartu. Hau da, %5eko adierazgarritasun mailaz, Puyehuen arrantzatutako izokinen luzerak banaketa normala jarraitzen duela bazter daiteke.

i)

Populazioaren batezbestekoaren puntu-estimazioa laginaren batezbestekoa denez, Argentinan arrantzatutako izokinen batezbesteko luzeraren puntu-estimazioa ondorengoa da:

```
> mean(Argentina$t1)
[1] 91.30882
```

%5eko konfiantza-mailako tarte-estimazioa berriz:

```
> t.test(Argentina$t1)$conf
[1] 86.42377 96.19388
attr(,"conf.level")
[1] 0.95
```

$$I_{\mu}^{0.95} = [86.42377, 96.19388]$$

Hortaz, $85 \notin I_{\mu}^{0.95}$, betetzen denez, 5% adierazgarritasun-mailaz, Argentinan arrantzatutako izokinen batezbestekoa 85 dela bazter daiteke.

j)

Populazioaren bariantzaren puntu-estimazioa laginaren kuasibariantza da beraz, Argentinan arrantzatutako izokinen luzeraren bariantzaren puntu-estimazioa ondorengo da:

```
> var(Argenti na$I )  
[1] 196. 0178
```

Bestalde, %5eko konfiantza-mailako konfiantza-tartea:

```
> n<-length((Argenti na$I ))  
> c((n-1)*var(Argenti na$I )/qchi sq(0. 95, n-1) ,  
  (n-1)*var(Argenti na$I )/qchi sq(0. 05, n-1))  
[1] 136. 4684 309. 9982
```

$195 \in I_{\sigma^2}^{0.95} = [136.4684, 309.9982]$ betetzen denez, %5 adierazgarritasun-mailaz, Argentinan arrantzatutako izokinen luzeraren bariantza 195 cm^2 dela onar daiteke.