

## Autoevaluazioa

### 5. ARIKETA

(10 PUNTU)

**carData** paketea kontsideratuko dugu.

- a) Instalatu eta kargatu **carData** paketea. (0.5 puntu)

**carData** paketearen barnean dagoen **Salaries** datu-markoa kontsideratuko dugu.

- b) Laguntza erabiliz, azaldu **Salaries** datu-markoak ematen duen informazioa. (0.5 puntu)
- c) Bistaratu datu-markoaren azken 6 behaketak. (0.5 puntu)
- d) Zenbat irakasle dira katedradunak? Eta horietatik zenbatek irabazten dute 150000 \$ baino gehiago? Eta horietatik zenbat dira emakumezkoak? (puntu 1)
- e) Bistaratu 115000 \$ baino gehiago irabazten duten irakasle elkartuen informazioa. (puntu 1)
- f) %1 adierazgarritasun mailaz, unibertsitateko irakasleen soldatak banaketa normala jarraitzen duen aztertu. (0.5 puntu)
- g) Soldata aldagaiaren kuartilak kalkulatu, eta balio hauek erabiliz soldatak hurrengo eran sailkatu: "Baxua", "Ohikoa" eta "Altua" eta kontingentzia taula bat egin. (1.5 puntu)
- h) %1 adierazgarritasun mailaz, soldaten sailkapena eta sexua aldagaia independenteak diren aztertu. (puntu 1)
- i) %1eko adierazgarritasun-mailaz, esan al daiteke emakumeen arteko ohiko soldata dutenen proportzioa eta gizonen artean ohiko soldata dutenen proportzioa berdina al dela? (puntu 1)
- j) %1eko adierazgarritasun-mailaz, esan al daiteke soldata altua duten irakasleen artean emakumeen proportzioa eta gizonen proportzioa berdina al dela?(puntu 1)
- k) %1 adierazgarritasun mailaz, kategoria eta sexua aldagaiak independenteak diren aztertu. Zein izango litzateke ondorio, adierazgarritasun-maila %5 izango balitz? (1.5 puntu)

---

Ebazpena :

a)

**carData** paketea instalatu eta kargatuko dugu:

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

b)

**Salaries** datu-markoari buruzko informazioa lortzeko `help()` funtzioa erabiliko dugu:

```
> help(Salaries) # edo ?Salaries
```

R Documentation

## Salaries for Professors

**Description**

The 2008-09 nine-month academic salary for Assistant Professors, Associate Professors and Professors in a college in the U.S. The data were collected as part of the on-going effort of the college's administration to monitor salary differences between male and female faculty members.

**Usage**

```
Salaries
```

**Format**

A data frame with 397 observations on the following 6 variables.

`rank`

a factor with levels `AssocProf` `AsstProf` `Prof`

`discipline`

a factor with levels A ("theoretical" departments) or B ("applied" departments).

`yrs.since.phd`

years since PhD.

`yrs.service`

years of service.

`sex`

a factor with levels `Female` `Male`

`salary`

nine-month salary, in dollars.

**References**

Fox J. and Weisberg, S. (2019) *An R Companion to Applied Regression*, Third Edition, Sage.

Datu-marko honetan 2008-2009 ikasturtean A.E.B.-etako unibertsitate bateko irakasleen soldata erakusten da. Irakasleak 3 kategoriatan banatuta daude: Katedratikoak (Prof), irakasle elkartuak (AssocProf) eta irakasle laguntzaileak (AsstProf). Irakasleen jakintzagaia (jakintza teorikoa edo jakintza aplikatua), sexua, zerbitzu-urteak eta doktoregoa lortu zutenetik igarotako urteak ere ageri dira.

c)

Lehenengo 6 bekaketak bistaratzeko tail() funtzioa erabiliko dugu:

```
> tail(Salaries) # edo tail(Salaries, n=6)
  rank discipline yrs. since. phd yrs. service sex salary
392 Prof           A           30           19 Male 151292
393 Prof           A           33           30 Male 103106
394 Prof           A           31           19 Male 150564
395 Prof           A           42           25 Male 101738
396 Prof           A           25           15 Male  95329
397 AsstProf       A            8            4 Male  81035
```

d)

```
> length(which(rank=="Prof"))
[1] 266
> length(which(rank=="Prof" & salary>150000))
[1] 54
> length(which(rank=="Prof" & salary>150000 &
sex=="Female"))
[1] 2
```

Guztira 2006 irakasle katedradun daude. Horietatik 54 irakaslek 150000 \$ baino gehiago irabazten dute, baina bakarrik 2 dira emakumezkoak.

e)

```
> Salaries[rank=="AssocProf" & salary>115000, ]
  rank discipline yrs. since. phd yrs. service sex salary
 11 AssocProf     B           12            8 Male 119800
215 AssocProf     B           11            1 Male 118700
323 AssocProf     B           13           11 Male 126431
```

Irakasle elkartuen artean bakarrik hiruk 115000 \$ baino gehiago irabazten dute. Hirurak gizonetzkoak dira, jakintza aplikatuak irakasten dituzte, eta doktoregoa lortu zutenetik 10 urte baino gehiago igaro dira.

f)

Soldatan normaltasuna aztertzeko ondorengo hipotesi-contrastea planteatzen da:

$H_0$ : Unibertsitateko irakasleen soldatak banaketa normala jarraitzen du.

$H_a$ : Unibertsitateko irakasleen soldatak ez du banaketa normala jarraitzen.

Hipotesi-contrastea egiteko lehenik eta behin laginaren tamaina zein den ikusiko da:

```
> length(salary)
[1] 397
```

Laginaren tamaina  $n = 397 > 50$  denez, Kolmogórov-Smirnov-en testa aplikatuko da. Horretarako, R-ko nortest paketea instalatu eta kargatu behar da:

Hipotesi-contrastea egiteko lehenik eta behin laginaren tamaina zein den ikusiko da:

```
> library(nortest)
> install.packages("nortest")
> lilliefors.test(salary)

Lilliefors (Kolmogorov-Smirnov) normality test

data: salary
D = 0.09086, p-value = 2.354e-08
```

p-balioa =  $2.354e-08 < 0.01$ , denez, %1eko adierazgarritasun mailaz hipotesi nulua errefusa daiteke, hau da, soldatak banaketa normala jarraitzen ez duela onar daiteke.

g)

```
> q1<-quantile(salary, 0.25)
> q3<-quantile(salary, 0.75)
> Salaries$kat<-as.factor(ifelse(salary<q1, "Baxua",
ifelse(salary<q3, "Ohi koa", "Al tua")))
> konting_taula<-table(sex, Salaries$kat)
> addmargins(konting_taula)

sex      Al tua Baxua Ohi koa Sum
Female   5      15      19   39
Male     95     83     180  358
Sum      100    98     199  397
```

h)

 $H_0$ : Soldaten sailkapena eta sexua aldagai independenteak dira. $H_a$ : Soldaten sailkapena eta sexua ez dira aldagai independenteak.

```
> chi sq. test(konti g_ taul a)
```

```
Pearson' s Chi -squared test
```

```
data: konti g_ taul a
```

```
X-squared = 5.9688, df = 2, p-value = 0.05057
```

p-balioa = 0.05057 > 0.01, denez, %1eko adierazgarritasun mailaz hipotesi nulua onar daiteke, hau da, soldaten sailkapena eta sexua aldagai independenteak direla onar daiteke.

i)

Kasu honetan, planteatu beharreko hipotesi-contrastea honako hau da:

 $H_0: p_E^{Ohikoa} = p_G^{Ohikoa}$  $H_a: p_E^{Ohikoa} \neq p_G^{Ohikoa}$ 

```
> prop. test(c(19, 180), c(39, 358), conf. level =0.99)
```

```
2-sample test for equality of proportions with  
continuity  
correction
```

```
data: c(19, 180) out of c(39, 358)
```

```
X-squared = 0.00027441, df = 1, p-value = 0.9868
```

```
alternative hypothesis: two.sided
```

```
99 percent confidence interval:
```

```
-0.2469406 0.2157130
```

```
sample estimates:
```

```
prop 1 prop 2
```

```
0.4871795 0.5027933
```

p-balioa = 0.9868 > 0.01, denez, %1eko adierazgarritasun mailaz hipotesi nulua onar daiteke, hortaz, %1eko adierazgarritasun-mailaz, ohiko soldata duten emakume eta gizonen proportzioa berdina dela onar daiteke.

j)

Kasu honetan, planteatutako galdera erantzuteko, nahikoa da soldata altua duten langileen artean emakumeen (edo gizonen) proportzioa 0.5 den edo ez aztertzea.

$$H_0: p_{Altua}^E = 0.5 \quad (p_{Altua}^G = 0.5)$$

$$H_a: p_{Altua}^E \neq 0.5 \quad (p_{Altua}^G \neq 0.5)$$

```
> prop.test(5, 100, p=0.5, conf.level=0.99)
```

```
1-sample proportions test with continuity correction
```

```
data: 5 out of 100, null probability 0.5
```

```
X-squared = 79.21, df = 1, p-value < 2.2e-16
```

```
alternative hypothesis: true p is not equal to 0.5
```

```
99 percent confidence interval:
```

```
0.01435777 0.14593816
```

```
sample estimates:
```

```
p
```

```
0.05
```

p-balioa =  $2.2e-16 < 0.01$ ,enez, %1eko adierazgarritasun mailaz hipotesi nulua bazter daiteke. Beraz, %1eko adierazgarritasun-mailaz, soldata altua duten irakasleen artean emakumeen proportzioa eta gizonen proportzioa desberdina dela onar daiteke.

k)

$H_0$ : Irakasleen kategoria eta sexua aldagai independenteak dira.

$H_a$ : Irakasleen kategoria eta sexua ez dira aldagai independenteak.

Kasu honetako kontingentzia-taula hurrengoa da:

```
> konting_taula2 <- table(rank, sex)
```

```
> konting_taula2
```

rank	sex	
	Female	Male
AsstProf	11	56
AssocProf	10	54
Prof	18	248

Eta hipotesi-kontrastea egitean:

```
> chi sq. test(konti g_taul a2)
```

```
Pearson' s Chi -squared test
```

```
data: konti g_taul a2
```

```
X-squared = 8.5259, df = 2, p-value = 0.01408
```

p-balioa =  $0.01408 > 0.01$ , denez, %1eko adierazgarritasun mailaz hipotesi nulua onar daiteke. Beraz, %1eko adierazgarritasun-mailaz, irakasleen kategoria eta sexua aldagai independenteak dira.

Adierazgarritasun-maila %5 bada berriz, p-balioa =  $0.01408 < 0.05$ , betetzen denez, irakasleen kategoria eta sexua aldagai independenteak ez direla onar daiteke.