Treball en R. Intervals de confiança i contrast d'hipòtesi

Noah Márquez
Alejandro Guzman

1 juny 2022

ÍNDEX

1 PROBLEMES PRÀCTICA 7

1 Problema 3

3. **Després d'un tractament contra l'obsesitat, els pesos en Kg de vuit dones eren:**

   58, 50, 60, 65, 64, 62, 56, 57

   Si suposem normalitat, calculeu un interval de confiança amb coeficient de confiança 0.95 **per la mitjana teòrica si la variància és desconeguda. Què passa si el calculeu amb un coeficient de confiança del** 0.9**?**
   Donat el nivell de confiança γ, calculem a, per a T_n, un *t-Student* de paràmetre n (mida de la mostra),

   $$\text{Prob}\ (|T_n| < a) = \gamma.$$

   De l'enunciat tenim que $\gamma = 0.95$ i $n = 8$, amb la distribució $t(n-1)$ i aïllem μ:

   $$0.95 = Prob(\overline{X_n} - a\frac{\tilde{S_n}}{\sqrt{n}} < \mu < \overline{X_n} + a\frac{\tilde{S_n}}{\sqrt{n}}).$$

   Amb la instrucció de **R** corresponent (totes estaran detallades a la secció Scripts en R, obtenim que l'interval de confiança per la mostra anterior és:

   $$[\mathbf{54.9288}, \mathbf{63.0712}]$$

   Si ara fem el càlcul amb un coeficient de confiança de 0.9 obtenim el següent interval de confiança:

   $$[\mathbf{55.73808}, \mathbf{62.26192}]$$

   Veiem llavors que reduir el coeficient de confiança implica acotar més l'interval que acceptem per la mostra donada.

2 Problema 4

4. **S'ha mesurat el pH del cordó umbilical de 22 nadons de dones no fumadores i de dones fumadores, obtenint:**

   Nadons de dones no fumadores:
   7.28 7.31 7.34 7.32 7.23 7.31 7.32 7.29 7.35 7.32 7.34
   7.35 7.26 7.18 7.34 7.27 7.34 7.32 7.29 7.26 7.26
   Nadons de dones fumadores:
   7.26 7.27 7.27 7.35 7.29 7.28 7.31 7.34 7.29 7.39 7.21 7.28
   7.30 7.24 7.20 7.28 7.30 7.35 7.32 7.31 7.37 7.26

   a) **Calculeu un interval de confiança del** 90% **per la mitjana de cada una de les mostres.**
      Donat el nivell de confiança γ, calculem a, per a T_n, un *t-Student* de paràmetre n (mida de la mostra),

      $$\text{Prob}\ (|T_n| < a) = \gamma.$$

      De l'enunciat tenim que $\gamma = 0.9$ i $n = 22$, amb la distribució $t(n-1)$ i aïllem μ:

      $$0.9 = Prob(\overline{X_n} - a\frac{\tilde{S_n}}{\sqrt{n}} < \mu < \overline{X_n} + a\frac{\tilde{S_n}}{\sqrt{n}}).$$

Amb la instrucció de **R** corresponent (totes estaran detallades a la secció Scripts en R, obtenim que l'interval de confiança per la mostra de dones no fumadores és:

$$[\mathbf{7.284822}, \mathbf{7.316997}]$$

I l'interval de confiança per la mostra de dones fumadores és el següent:

$$[\mathbf{7.276637}, \mathbf{7.311545}]$$

b) **Calculeu l'interval de confiança del** 95% **per a la diferència de les mitjanes del pH de les dones fumadores i de les no fumadores.**
L'interval que volem calcular amb nivell de confiança $\gamma = 0.95$ és:

$$\left(\overline{X} - \overline{Y} - qnorm(\gamma)(\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y})^{\frac{1}{2}}, \overline{X} - \overline{Y} + qnorm(\gamma)(\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y})^{\frac{1}{2}}\right).$$

En aquest cas, no tenim cap funció de l'**R** que ens el calculi. Per fer el càlcul tenim les dades de l'enunciat, el coeficient de confiança (0.95) i podem conèixer les variàncies. Fent ús d'un seguit d'instruccions d'**R** arribem a que l'interval de confiança per a la diferència de les mitjanes és el següent:

$$[\mathbf{-0.01587181}, \mathbf{0.02950818}]$$

3 Problema 5

5. **Les notes d'11 alumnes d'una classe en dos exàmens consecutius van ser:**

Alumnes	1	2	3	4	5	6	7	8	9	10	11
Examen 1	7.9	8.3	6.2	8.2	8	7.8	4.9	6.2	8.9	7.8	9.4
Examen 2	8.2	7.1	4.8	8.4	7.9	7.4	5.2	5.6	9.2	6.5	8.5

a) **Calculeu l'interval de confiança del** 95% **per a la diferència de mitjanes de les notes del primer i del segon examen.**
Si volem calcular l'interval de confiança per la diferència de les mitjanes, en el cas de dades aparellades amb X_i independents entre elles i Y_i independents entre elles, tenim una funció de **R** que ens ho calcula automàticament (descrita a Scripts en R). Ens retorna el següent interval de confiança:

$$[\mathbf{-0.02152516}, \mathbf{0.89425243}]$$

b) **Calculeu l'interval de confiança del** 80% **per a la diferència de mitjanes de les notes del primer i del segon examen. Compara els resultats amb l'apartat anterior.**
Fent servir la mateixa instrucció que l'apartat anterior, amb el respectiu coeficient de confiança, obtenim el següent interval:

$$[\mathbf{0.1543760}, \mathbf{0.7183512}]$$

Veiem llavors que reduir el coeficient de confiança implica acotar més l'interval que acceptem per la mostra donada.

2 POBLEMES PRÀCTICA 8

1 Problema 1

1. **Després d'un tractament contra l'obesitat, els pesos en Kg de vuit dons eren:**

   58, 50, 60, 65, 64, 62, 56, 57.

   Suposeu normalitat i $\alpha = 0.1$.

   a) **Podem afirmar que la mitjana teòrica μ és 61, sabent que $\sigma = 3$?**
   Per al vector de dades dels pesos calculem la mitjana empírica (amb ajuda de l'**R**) i la mitjana empírica estandaritzada:

   $$\overline{X} = 59$$
   $$Z = \frac{\overline{X}-61}{3} \cdot \sqrt{8} = -1.885618.$$

   Estem davant d'un contrast bilateral:

   $$\begin{cases} H_0 : \mu = 61 \\ H_1 : \mu \neq 61 \end{cases}$$

   amb nivell de significació $\alpha = 0.1$. Hem de calcular ara el p-valor, que serà Prob$[|Z| > z]$ que es calcula amb l'**R** i s'obté:

   $$p\text{-valor} = \mathbf{0.05934644}.$$

   Aquest valor representa la probabilitat que en una mostra de mida $n = 8$ d'una $N(\mu = 61, \sigma = 3)$ obtinguem un valor de la mitjana mostral $\overline{x}$ igual o diferent al valor obtingut a partir de la mostra.

   Com que $p < \alpha$ (nivell de significació) hem de **rebutjar la hipòtesi nul·la**, per tant **no podem afirmar que la mitjana teòrica μ sigui 61 sabent que** $\sigma = 3$.

   b) **Podem afirmar que la mitjana teòrica μ és 61, si σ és desconeguda?**
   Amb una simple instrucció d'**R** obtenim que per aquest cas,

   $$p\text{-valor} = \mathbf{0.2835}.$$

   I com que $p > \alpha$ no podem refutar la hipòtesi nul·la. És a dir**, podem afirmar que la mitjana teòrica μ és 61 amb σ desconeguda.**

   c) **Podem afirmar que la variància teòrica σ^2 és superior a 10?**
   Estem davant d'un contrast sobre variància σ^2. Suposem que tenim una mostra normal i volem fer una comparació de la variància mostral amb un σ_0^2 donat.

   En el nostre cas hem de fer un test unilateral a la dreta:

   $$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$$

   Fent el càlcul de la p amb una instrucció d'**R** obtenim:

   $$p = \mathbf{0.02016574}.$$

   Com que $p < \alpha$, cal rebutjar H_0. Per tant, **podem afirmar que la variància teòrica σ^2 és superior a 10**.

2 Problema 3

3. **A partir de les dades de l'arxiu** *ttereny.txt***, contrasteu les hipòtesis següents amb nivell de significació** $\alpha = 0.05$**:**

   a) **El consum mitjà a 120 Km/h és de 12 litres.**
   Tenim un contrast sobre la mitjana μ amb σ desconeguda. A més, és un contrast bilateral:
   $$\begin{cases} H_0 : \mu = 12 \\ H_1 : \mu \neq 12 \end{cases}$$
   amb nivell de significació $\alpha = 0.05$.

   Amb **R** trobem que:
   $$p\text{-valor} = \mathbf{0.1575}.$$
   Per tant, com $p > \alpha$, **considerem** H_0 **com a certa**. És a dir, el **consum mitjà a 120 Km/h és de 12 litres**.

   b) **La mitjana de la velocitat màxima és de 155 Km/h.**
   De la mateixa forma que l'apartat anterior tenim un contrast sobre la mitjana μ amb σ desconeguda. A més, és un contrast bilateral:
   $$\begin{cases} H_0 : \mu = 155 \\ H_1 : \mu \neq 155 \end{cases}$$
   Amb **R** trobem que:
   $$p\text{-valor} = \mathbf{0.1664}.$$
   Per tant, com $p > \alpha$, **considerem** H_0 **com a certa**. És a dir, la **mitjana de la velocitat màxima és de 155 Km/h**.

   c) **La mitjana del consum urbà és inferior de 12.2 litres.**
   Tenim un contrast sobre la mitjana μ amb σ desconeguda. A més, és un contrast unilateral a l'esquerra:
   $$\begin{cases} H_0 : \mu = 12.2 \\ H_1 : \mu < 12.2 \end{cases}$$
   Amb **R** trobem que:
   $$p\text{-valor} = \mathbf{0.9695}.$$
   Per tant, com $p > \alpha$, **considerem** H_0 **com a certa**. És a dir, la **mitjana del consum urbà NO és inferior a 12.2 litres**.

   d) **A partir de les dades de les variables** *Consum90* **i** *Consum120***, podem acceptar com a vàlida l'afirmació que** $\mu_{120} = \mu_{90}$**?**
   Tenim un contrast d'una mostra de dades aparellades amb un test bilateral:
   $$\begin{cases} H_0 : \mu_{120} = \mu_{90} \\ H_1 : \mu_{120} \neq \mu_{90} \end{cases}$$
   Amb **R** trobem que:
   $$p\text{-valor} = \mathbf{2.2e-16}.$$
   Per tant, com $p < \alpha$, **rebutjem** H_0. És a dir, **no podem acceptar com a vàlida l'afirmació que** $\mu_{120} = \mu_{90}$.

e) **A partir de les dades de les variables** *Consum120* **i** *ConsumUrba***, podem acceptar com a vàlida l'afirmació que** $\mu_{120} = \mu_U$**?**
Tenim un contrast d'una mostra de dades aparellades amb un test bilateral:

$$\begin{cases} H_0 : \mu_{120} = \mu_U \\ H_1 : \mu_{120} \neq \mu_U \end{cases}$$

Amb **R** trobem que:

$$p\text{-valor} = \mathbf{0.06603}.$$

Per tant, com $p > \alpha$, **considerem** H_0 **com a certa**. És a dir, **podem acceptar com a vàlida l'afirmació que** $\mu_{120} = \mu_U$.

f) **A partir de les dades de les variables** *Consum120* **i** *Consum90***, podem acceptar com a vàlida l'afirmació que** $\mu_{120} = \mu_{90} + 2$**?**
Tenim un contrast d'una mostra de dades aparellades amb un test bilateral:

$$\begin{cases} H_0 : \mu_{120} = \mu_{90} + 2 \\ H_1 : \mu_{120} \neq \mu_U + 2 \end{cases}$$

Amb **R** trobem que:

$$p\text{-valor} = \mathbf{1.36e-15}.$$

Per tant, com $p < \alpha$, **rebutjem** H_0. És a dir, **no podem acceptar com a vàlida l'afirmació que** $\mu_{120} = \mu_{90} + 2$.

3 Problema 7

7. **Les dades de la taula següent representen el sou actual i el sou inicial, en euros, de 10 empleats d'una empresa:**

Homes		Dones	
Sou actual	Sou inicial	Sou actual	Sou inicial
57000	27000	21450	12000
40200	18750	21900	13200
45000	21000	21900	9750
32100	13500	27900	12750
36000	18750	24000	13500

a) **Podem afirmar que la mitjana teòrica del sou actual dels homes és de 42000 euros?** (Nivell de significació $\alpha = 0.05$)
Tenim un contrast sobre la mitjana μ amb σ desconeguda. A més, és un contrast bilateral:

$$\begin{cases} H_0 : \mu = 42000 \\ H_1 : \mu \neq 42000 \end{cases}$$

Amb **R** trobem que:

$$p\text{-valor} = \mathbf{0.9896}.$$

Per tant, com $p > \alpha$, **considerem** H_0 **com a certa**. És a dir, **la mitjana teòrica del sou actual dels homes és de 42000 euros**.

b) **Les dones, han tingut un augment de sou significatiu?** (Nivell de significació $\alpha = 0.05$)
Estem davant un contrast d'una mostra de dades aparellades amb un test unilateral a la dreta:

$$\begin{cases} H_0 : \mu_{\text{dones\_actual}} = \mu_{\text{dones\_inicial}} \\ H_1 : \mu_{\text{dones\_actual}} > \mu_{\text{dones\_inicial}} \end{cases}$$

Amb **R** trobem que:

$$p\text{-valor} = \mathbf{0.0003091}.$$

Per tant, com $p < \alpha$, **rebutgem** H_0. És a dir, **les dones han tingut un augment de sou significatiu**.

c) **Podem afirmar que el sou mitjà inicial dels homes és més alt que el de les dones?**
Suposem que tenim normalitat. Volem comparar les mitjanes de les homes i les dones. Observem en primer lloc que les dades no són aparellades.

Per resoldre aquest problema hem de fer dos contrastos. Farem el primer contrast per veure si podem suposar que les variàncies són iguals.

Plantegem la hipòtesis nul·la i alternativa.

$$\begin{cases} H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1 \\ H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1 \end{cases}$$

Fixem el nivell de significació, que com en aquest apartat no ens diuen res, suposem que $\alpha = 0.01$.

Amb una instrucció d'**R** trobem que

$$p\text{-valor} = \mathbf{0.04243}.$$

Per tant, com $p > \alpha$, **acceptem** H_0. És a dir, **podem afirmar que les variàncies són iguals**. En el cas d'haver agafat $\alpha = 0.05$ com en els anteriors apartats, no podríem haver acceptat la hipòtesi nul·la.

Ara volem comparar les mitjanes dels dos grups. Estem per tant en un cas de comparació de mitjanes amb variàncies iguals (ho acabem de comprovar) però desconegudes:

$$\begin{cases} H_0 : \mu_{\text{homes\_inicial}} = \mu_{\text{dones\_inicial}} \\ H_1 : \mu_{\text{homes\_inicial}} > \mu_{\text{dones\_inicial}} \end{cases}$$

Fixem el nivell de significació, que com en aquest apartat no ens diuen res, suposem que $\alpha = 0.01$.

Amb una instrucció d'**R** trobem que

$$p\text{-valor} = \mathbf{0.005333}.$$

Per tant, com $p < \alpha$, **rebutgem** H_0. És a dir, **podem afirmar que el sou mitjà inicial dels homes és més alt que el de les dones**.

3 CONCLUSIONS

En aquest treball hem pogut treballar en l'àmbit dels intervals de confiança i el contrast d'hipòtesi, fent exercicis de pràctiques realitzades a classe. Ens ha ajudat a treballar d'una forma molt més visual i descriptiva i alhora tenir més contacte amb el funcionament del llenguatge **R**.

Ens va ser bastant fàcil començar la pràctica pel fet que alguns apartats ja els teníem fets de les hores de laboratori. És per això que hem pogut dedicar-hi més temps a repassar els conceptes i alhora fer els apartats restants en profunditat.

En general la realització d'aquest treball ens ha ajudat a tenir un bon contacte i entendre molt millor els conceptes de teoria i el potencial que té el llenguatge **R** en l'àmbit de l'estadística.

4 Scripts en R

1 Pràctica 7

1.1 Problema 3

```
# Creem el vector amb les dades
x <- c(58,50,60,65,64,62,56,57)

# Calculem l'interval de confianca amb un coeficient de
   confianca de 0.95:
t.test(x)$conf.int

# Calculem l'interval de confianca amb un coeficient de
   confianca de 0.9:
t.test(x, conf.level = 0.9)$conf.int # S'acota més l'interval
```

1.2 Problema 4

```
# (a) Calculeu un interval de confianca del 90% per la mitjana
    de cada una de les mostres:
# NO-fumadores:
no_fumadores <- c(7.28, 7.31, 7.34, 7.34, 7.32, 7.23, 7.31,
   7.32, 7.29, 7.35, 7.32, 7.34, 7.35, 7.26, 7.18, 7.34, 7.27,
    7.34, 7.32, 7.29, 7.26, 7.26)
t.test(no_fumadores, conf.level=0.9)$conf.int

# Fumadores:
fumadores <- c(7.26, 7.27, 7.27, 7.35, 7.29, 7.28, 7.31, 7.34,
    7.29, 7.39, 7.21, 7.28, 7.30, 7.24, 7.20, 7.28, 7.30,
   7.35, 7.32, 7.31, 7.37, 7.26)
t.test(fumadores, conf.level=0.9)$conf.int

# (b) Calculeu l'interval de confianca del 95% per a la
   diferencia de les mitjanes del pH de les dones fumadores i
   de les no fumadores.
mitjana_no_fumadores <- mean(no_fumadores)
mitjana_fumadores <- mean(fumadores)
var_no_fumadores <- var(no_fumadores)
var_fumadores <- var(fumadores)

a <- qnorm(0.95)
d <- a*sqrt((var_no_fumadores/length(no_fumadores)) + (var_
   fumadores/length(fumadores)))
d1 <- mitjana_no_fumadores - mitjana_fumadores - d # Part
   esquerra de l'interval
d2 <- mitjana_no_fumadores - mitjana_fumadores + d # Part
   dreta de l'interval
```

1.3 Problema 5

```
# (a) Calculeu l'interval de confianca del 95% per a la
   diferencia de mitjanes de les notes del primer i del segon
   examen:
examen_1 <- c(7.9, 8.3, 6.2, 8.2, 8, 7.8, 4.9, 6.2, 8.9, 7.8,
   9.4)
examen_2 <- c(8.2, 7.1, 4.8, 8.4, 7.9, 7.4, 5.2, 5.6, 9.2,
   6.5, 8.5)
t.test(examen_1, examen_2, paired=TRUE)$conf.int

# (b) Calculeu l'interval de confianca del 80% per a la
   diferencia de mitjanes de les notes del primer i del segon
   examen. Compara els resultats amb l'apartat anterior.

t.test(examen_1, examen_2, conf.level=0.8, paired=TRUE)$conf.
   int # S'acota l'interval de confianca
```

2 Pràctica 8

2.1 Problema 1

```
# Pesos en Kg de vuit dones:
pesos <- c(58, 50, 60, 65, 64, 62, 56, 57)

# (a) Podem afirmar que la mitjana teorica és 61, sabent que
   sigma = 3?
# Per al vector pesos calculem la mitjana empírica:
pesosm <- mean(pesos)

# I la mitjana empírica estandaritzada:
z <- (pesosm-61)/3*sqrt(8)

# Obtenim z = -1.885618, és un valor raonable? Per poder
   contestar aquesta pregunta calculem el p-valor Prob[|Z| > z
   ]:
p <- 2 * (1-pnorm(abs(z)))

# (b) Podem afirmar que la mitjana teorica és 61, si sigma és
   desconeguda?
t.test(pesos, mu=61, alternative="two.sided", conf.level
   =1-0.1)

# (c) Podem afirmar que la variancia teorica sigma^2 és
   superior a 10?
q <- (length(pesos) - 1) * var(pesos) / 10
p <- 1-pchisq(q, length(pesos)-1)
```

2.2 Problema 3

```
# Carreguem el fitxer
dades <- read.table(file.choose(), header=TRUE)

# (a) Consum mitja a 120 Km/h és de 12 litres.
t.test(dades$Consum120, mu=12, alternative="two.sided", conf.
   level = 1-0.05)

# (b) La mitjana de la velocitat maxima és de 155 Km/h.
t.test(dades$Velocitat, mu=155, alternative="two.sided", conf.
   level = 1-0.05)

# (c) La mitjana del consum urba és inferior de 12.2 litres.
t.test(dades$Consum.Urba, mu = 12.2, alternative="less", conf.
   level = 1-0.05)

# (d) A partir de les dades de les variables Consum90 i
   Consum120, podem acceptar com a valida l'afirmació que
   mu120 = mu90?
t.test(dades$Consum120, dades$Consum90, alternative="two.sided
   ", paired=TRUE)

# (e) A partir de les dades de les variables Consum120 i
   ConsumUrba, podem acceptar com a valida l'afirmació que
   mu120 = mU?
t.test(dades$Consum120, dades$Consum.Urba, alternative="two.
   sided", paired=TRUE)

# (f) A partir de les dades de les variables Consum120 i
   Consum90, podem acceptar com a valida l'afirmació que mu120
    = mu90 + 2?
t.test(dades$Consum120, dades$Consum90+2, alternative="two.
   sided", paired=TRUE)
```

2.3 Problema 7

```
# Carreguem les dades
homes_actual <- c(57000, 40200, 45000, 32100, 36000)
homes_inicial <- c(27000, 18750, 21000, 13500, 18750)

dones_actual <- c(21450, 21900, 21900, 27900, 24000)
dones_inicial <- c(12000, 13200, 9750, 12750, 13500)

# (a) Podem afirmar que la mitjana teorica del sou actual dels
    homes és de 42000 euros? (Nivell de significació = 0.05)
t.test(homes_actual, mu=42000, alternative="two.sided", conf.
   level = 1-0.05)
```

```
# (b) Les dones, han tingut un augment de sou significatiu? (
   Nivell de significació = 0.05)
t.test(dones_actual, dones_inicial, alternative="greater",
   paired=TRUE)

# (c) Podem afirmar que el sou mitja inicial dels homes és més
    alt que el de les dones?
var.test(homes_inicial, dones_inicial, alternative="two.sided"
   )
t.test(homes_inicial, dones_inicial, alternative="greater",
   var.equal = TRUE)
```