

# Document de Treball 7

## Objectiu

L'objectiu d'aquesta sessió és:

1. Realitzar problema 3 de l'examen de Juny 2019
2. Realitzar problema 1 de l'examen de Juny de 2019
3. Realitzar problema 2 de l'examen de Juny de 2019

## 1. Problema 3 de l'examen Juny 2019

S'ha realitzat un estudi al llarg de 10 anys en el qual s'investiga la relació entre el consum de cafè i el risc de depressió en les persones. Per a això es van recollir dades de 50.739 individus sense símptomes de depressió a l'inici de l'estudi en l'any 1996, i es va fer un seguiment fins al 2006.

Els investigadors van utilitzar uns qüestionaris per recopilar dades sobre el consum de cafè (amb cafeïna) sobre el diagnòstic clínic de la malaltia de la depressió. També es van obtenir dades sobre l'ús de tractaments antidepressius. La taula següent mostra la distribució de la incidència de la malaltia de la depressió classificada segons la quantitat de cafè que solen prendre:

|                          |    | Consumo de tazas de café |                     |          |                  |                |
|--------------------------|----|--------------------------|---------------------|----------|------------------|----------------|
|                          |    | Hasta una por semana     | De 2 a 6 por semana | 1 al día | De 2 a 3 por día | 4 o más al día |
| Diagnóstico de depresión | Sí | 670                      | 373                 | 905      | 564              | 95             |
|                          | No | 11545                    | 6244                | 19329    | 11726            | 2288           |

Contesteu a les següents preguntes, justificant totes les respostes:

- a) Calculeu un interval de confiança a l'95% per a la proporció de persones que pateixen depressió. També per a la proporció de persones que no la pateix. Expliqueu el significat de cada un d'ells. Construeixi els mateixos intervals però aquesta vegada per a una confiança el 99%. ¿Observa diferències pel que fa als primers intervals calculats?
- b) Quin tipus de test és l'apropiat per avaluar l'associació estadística entre el consum de cafè i la malaltia de la depressió?
- c) Escriu les hipòtesis del test identificat en l'apartat anterior. Es compleixen les condicions d'aplicabilitat del test que has plantejat? A quina conclusió arribes?
- d) Els investigadors volen també comparar si hi ha diferències entre prendre només una tassa al dia o prendre'n més d'una. Escriu les hipòtesis, verifica les condicions d'aplicabilitat i obté una conclusió per a aquesta comparació.

```
> .Table # Counts
      Consumo tazas de cafe
Enfermedad Hasta_1_sem De_2_a_6_sem 1_dia De_2_a_3_dia 4_o_mas_dia
Si          670         373   905         564         95
No          11545        6244 19329        11726        2288

> .Test <- chisq.test(.Table, correct=FALSE)

> .Test

      Pearson's Chi-squared test

data: .Table
X-squared = 31.454, df = 4, p-value = 0.000002474
```

Tabla 3.1

```
> .Table # Counts
      Consumo tazas de cafe
Enfermedad Una_al_dia Mas_de_una_al_dia
Si          905         659
No          16329        14014

> .Test <- chisq.test(.Table, correct=FALSE)

> .Test

      Pearson's Chi-squared test

data: .Table
X-squared = 9.8204, df = 1, p-value = 0.001726
```

Tabla 3.2

```
> binom.test(2607, 53739, conf.level=.95)

      Exact binomial test

data: 2607 and 53739
number of successes = 2607, number of trials = 53739, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.04671089 0.05036257
sample estimates:
probability of success
 0.04851225
```

Tabla 3.3

```
> binom.test(2607, 53739, conf.level=.99)

      Exact binomial test

data: 2607 and 53739
number of successes = 2607, number of trials = 53739, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
99 percent confidence interval:
 0.04615576 0.05094910
sample estimates:
probability of success
 0.04851225
```

Tabla 3.4

```
> binom.test(51132, 53739, conf.level=.95)

      Exact binomial test

data: 51132 and 53739
number of successes = 51132, number of trials = 53739, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.9496374 0.9532891
sample estimates:
probability of success
 0.9514877
```

Tabla 3.5

```
> binom.test(51132, 53739, conf.level=.99)

      Exact binomial test

data: 51132 and 53739
number of successes = 51132, number of trials = 53739, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
99 percent confidence interval:
 0.9490509 0.9538442
sample estimates:
probability of success
 0.9514877
```

Tabla 3.6

## 2. Problema 1 de l'examen Juny 2019

El consum insuficient de calci és un dels principals factors de l'osteoporosi. En un estudi realitzat al País Valencià per avaluar la prevalença de la ingesta de calci en adults, es van prendre mesures a 1206 subjectes, dels quals el 51% eren dones.

Segons l'esmentat estudi, publicat per Jiménez et al (2017), la ingesta diària de calci en adults segueix una distribució aproximadament normal de mitjana 741,4 mg / persona i desviació típica 334,9 mg / persona.

Segons l'IOM (Institute Of Medicine), es considera un consum insuficient de calci si la ingesta diària no supera la quantitat mínima recomanada, fixada en els 800 mg / persona.

- Quin percentatge, en aquesta població, ingereix com a mínim la quantitat de calci recomanada?
- Quin és el percentatge de la població que ingereix diàriament una quantitat superior a 1.000 mg / persona?
- Calcula el percentil 95 i interpreta el resultat. Es visita una empresa de 10 treballadors.

Denotem per Y el nombre d'empleats que tenen una ingesta de calci insuficient.

- Quina és la distribució de Y?
- Calcula la probabilitat que com a màxim tres empleats tinguin una ingesta de calci insuficient.

```
> pnorm(c(400,600,800,1000,1200), mean=741.4, sd=334.9, lower.tail=TRUE)
[1] 0.1540045 0.3364338 0.5694514 0.7799934 0.9145572
```

Tabla 1.1

```
> pnorm(c(400,600,800,1000,1200), mean=741.4, sd=334.9, lower.tail=FALSE)
[1] 0.84599553 0.66356617 0.43054862 0.22000664 0.08544276
```

Tabla 1.2

```
> qnorm(c(.01,.05,.50,.95,.99), mean=741.4, sd=334.9, lower.tail=TRUE)
[1] -37.6939 190.5385 741.4000 1292.2615 1520.4939
```

Tabla 1.3

```
> qnorm(c(.01,.05,.50,.95,.99), mean=741.4, sd=334.9, lower.tail=FALSE)
[1] 1520.4939 1292.2615 741.4000 190.5385 -37.6939
```

Tabla 1.4

```
> .Table <- data.frame(Probability=dbinom(0:10, size=10, prob=0.43))

  Probability
0 0.003620333
1 0.0273112864
2 0.0927146302
3 0.1865136420
4 0.2462307291
5 0.2229036074
6 0.1401294608
7 0.0604066849
8 0.0170887332
9 0.0028647779
10 0.0002161148
```

Tabla 1.5

```
> .Table <- data.frame(Probability=dbinom(0:10, size=10, prob=0.57))
      Probability
0  0.0002161148
1  0.0028647779
2  0.0170887332
3  0.0604066849
4  0.1401294608
5  0.2229036074
6  0.2462307291
7  0.1865136420
8  0.0927146302
9  0.0273112864
10 0.0036203333
```

Tabla 1.6

```
> .Table <- data.frame(Probability=dbinom(0:10, size=10, prob=0.51))
      Probability
0  0.0007979227
1  0.0083049093
2  0.0388974836
3  0.1079603626
4  0.1966420890
5  0.2456019561
6  0.2130221048
7  0.1266953626
8  0.0494499757
9  0.0114374093
10 0.0011904242
```

Tabla 1.7

### 3. Problema 2 de l'examen Juny 2019

El peix constitueix una font important d'aliment per la seva aportació nutricional de proteïnes, greixos saludables, vitamines i minerals. No obstant això, sovint els peixos concentren mercuri en els seus organismes provinent de la contaminació de l'aigua. La presència de mercuri en els peixos pot tenir repercussions sobre la salut dels éssers humans, especialment en dones embarassades i nadons.

En l'estudi titulat Fish intake, contaminants, and human health: Evaluating the risks and the benefits i signat per Mozaffarian et al (2006) s'analitza la concentració de mercuri en els peixos de dos rius de Carolina de Nord (EUA): el riu Lumber i el riu Waccamaw.

Les dades es van recollir en el conjunt de dades "mercuri". Les variables que es van mesurar en els dos grups, definits pels rius, van ser:

Conc: nivells de concentració de mercuri en els peixos (mg / kg)

Long: longituds dels peixos (cm)

- Dóna una estimació puntual de la concentració mitjana de mercuri dels peixos al riu Lumber. Seria correcte calcular un interval de confiança? En cas afirmatiu, calcula un interval de confiança a l'95%.
- Hi ha evidència estadística per afirmar a nivell  $\alpha = 0.05$  que el nivell mitjà de concentració al riu Waccamaw és superior a el nivell mitjà en el riu Lumber? Justifica el procediment emprat.

Una part de l'estudi pretén determinar la relació entre la mida dels peixos i el seu nivell de contaminació per mercuri.

- Calcula la recta de regressió adequada per predir la concentració de mercuri a partir de la longitud dels peixos. Quin és el percentatge de la variabilitat de la variable dependent que s'explica per la regressió?
- Quina concentració de mercuri s'esperaria per un peix de 50 cm de longitud?

```
> numSummary(mecurio[, "Conc"], groups=mecurio$Rio, statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

|          | mean     | sd        | 0%   | 25%    | 50%  | 75% | 100% | data:n |
|----------|----------|-----------|------|--------|------|-----|------|--------|
| Waccamaw | 1.276429 | 0.8291484 | 0.11 | 0.6175 | 1.05 | 1.8 | 3.6  | 98     |
| Lumber   | 1.078082 | 0.6486110 | 0.27 | 0.5700 | 0.91 | 1.5 | 3.5  | 73     |

Tabla 2.1

```
> shapiro.test(Conc$Waccamaw)
```

Shapiro-Wilk normality test

data: Conc\$Waccamaw

W = 0.91544, p-value = 0.000009826

Tabla 2.2

```
> shapiro.test(Conc$Lumber)
```

Shapiro-Wilk normality test

data: Conc\$Lumber

W = 0.86906, p-value = 0.000001888

Tabla 2.3

```
> with(Lumber, (t.test(Conc, alternative='two.sided', mu=0.0, conf.level=.95)))
```

One Sample t-test

data: Conc

t = 14.201, df = 72, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

0.926750 1.229414

sample estimates:

mean of x

1.078082

Tabla 2.4

```
> with(Lumber, (t.test(Conc, alternative='less', mu=0.0, conf.level=.95)))
```

One Sample t-test

data: Conc

t = 14.201, df = 72, p-value = 1

alternative hypothesis: true mean is less than 0

95 percent confidence interval:

-Inf 1.204578

sample estimates:

mean of x

1.078082

Tabla 2.5

```
> leveneTest(Conc ~ Rio, data=mecurio, center="median")
```

Levene's Test for Homogeneity of Variance (center = "median")

| group | Df  | F value | Pr(>F)    |
|-------|-----|---------|-----------|
| 1     | 169 | 5.0658  | 0.02569 * |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Tabla 2.6

```
> with(Dataset, (t.test(ConcLumber, ConcWaccamaw, alternative='two.sided', conf.level=.95, paired=TRUE)))
```

Paired t-test

data: ConcLumber and ConcWaccamaw

t = 0.42572, df = 72, p-value = 0.6716

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1634435 0.2522106

sample estimates:

mean of the differences

0.04438356

Tabla 2.7

```
> with(Dataset, (t.test(ConcLumber, ConcWaccamaw, alternative='greater', conf.level=.95, paired=TRUE)))

Paired t-test

data: ConcLumber and ConcWaccamaw
t = 0.42572, df = 72, p-value = 0.3358
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-0.1293346      Inf
sample estimates:
mean of the differences
0.04438356
```

Tabla 2.8

```
> t.test(Conc~Rio, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=mecurio)

Two Sample t-test

data: Conc by Rio
t = 1.6936, df = 169, p-value = 0.09218
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.03285077 0.42954353
sample estimates:
mean in group Waccamaw mean in group Lumber
1.276429 1.078082
```

Tabla 2.9

```
> t.test(Conc~Rio, alternative='greater', conf.level=.95, var.equal=TRUE, data=mecurio)

Two Sample t-test

data: Conc by Rio
t = 1.6936, df = 169, p-value = 0.04609
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
0.004647234      Inf
sample estimates:
mean in group Waccamaw mean in group Lumber
1.276429 1.078082
```

Tabla 2.10

```
> t.test(Conc~Rio, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=mecurio)

Welch Two Sample t-test

data: Conc by Rio
t = 1.7547, df = 168.57, p-value = 0.08114
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.02481087 0.42150363
sample estimates:
mean in group Waccamaw mean in group Lumber
1.276429 1.078082
```

Tabla 2.11

```
> t.test(Conc~Rio, alternative='greater', conf.level=.95, var.equal=FALSE, data=mecurio)

Welch Two Sample t-test

data: Conc by Rio
t = 1.7547, df = 168.57, p-value = 0.04057
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.01138394      Inf
sample estimates:
mean in group Waccamaw    mean in group Lumber
      1.276429              1.078082
```

Tabla 2.12

```
> wilcox.test(Conc ~ Rio, alternative="two.sided", data=mecurio)

Wilcoxon rank sum test with continuity correction

data: Conc by Rio
W = 4006, p-value = 0.1807
alternative hypothesis: true location shift is not equal to 0
```

Tabla 2.13

```
> wilcox.test(Conc ~ Rio, alternative="greater", data=mecurio)

Wilcoxon rank sum test with continuity correction

data: Conc by Rio
W = 4006, p-value = 0.09033
alternative hypothesis: true location shift is greater than 0
```

Tabla 2.14

```
Call:
lm(formula = Conc ~ Long, data = mecurio)

Residuals:
    Min       1Q   Median       3Q      Max
-1.1499 -0.3436 -0.1022  0.3123  1.9100

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.131645   0.213615  -5.298 0.000000362 ***
Long         0.058127   0.005228  11.119   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5805 on 169 degrees of freedom
Multiple R-squared:  0.4225, Adjusted R-squared:  0.4191
F-statistic: 123.6 on 1 and 169 DF, p-value: < 2.2e-16
```

Tabla 2.15

```

Call:
lm(formula = Long ~ Conc, data = mecurio)

Residuals:
    Min       1Q   Median       3Q      Max
-15.8871  -4.6865  -0.7918   3.8823  18.4277

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  31.3085     0.9238   33.89  <2e-16 ***
Conc         7.2685     0.6537   11.12  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.492 on 169 degrees of freedom
Multiple R-squared:  0.4225, Adjusted R-squared:  0.4191
F-statistic: 123.6 on 1 and 169 DF,  p-value: < 2.2e-16

```

Tabla 2.16