

TASCA 2. Inferència no paramètrica: Rangs.

Laura Julià Melis i Víctor Navarro Garcés

26 de Maig de 2019

Exercici 1

1. La inflamabilitat dels pijames infantils és un tema de seguretat important. Les dades de la taula adjunta corresponen a unes proves realitzades per cinc laboratoris diferents; els valors indicats són els d'un índex estàndard d'inflamabilitat. Cada laboratori va rebre 11 pijames d'un mateix model (en total 55 pijames) i va realitzar les corresponents proves. Com veurem, el problema central en aquest problema serà determinar si hi ha diferències en la manera de mesurar la inflamabilitat per part dels laboratoris.

Introducció de les dades:

```
inflamabilitat=c(
  3.91,4.15,4.11,4.78,4.1,5.25,4.7,4.9,3.8,4.09,2.99,
  2.7,3.4,3.6,3.22,4.08,4.13,3.8,3.87,4.3,3.45,3.3,
  3.3,3.35,3.5,3.51,2.89,2.83,3.21,2.84,3.8,3.55,3.8,
  3.37,3.26,3.46,2.78,2.71,3.39,2.93,3.27,2.96,2.6,2.82,
  5.17,5.19,4.8,5.29,4.19,4.53,4.88,3.58,4.72,6.55,6.96
)
laboratori = factor(rep(c("L1","L2","L3","L4","L5"), rep(11,5)))
```

a) Realitza una prova de Kruskal-Wallis per determinar si hi ha alguna diferència entre laboratoris quant a la mediana de l'índex d'inflamabilitat. Indica les hipòtesis, el model associat, l'estadístic de test i la conclusió final d'aquesta prova sense utilitzar l'ordre `kruskal.test` de l'R.

1. Definim el contrast d'hipòtesis a realitzar

Les hipòtesis nul·la (H_0) i alternativa (H_1) d'aquesta prova de Kruskal-Wallis són:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1: \mu_i \neq \mu_j \quad i = [1,2,3,4,5], j = [1,2,3,4,5],$$

per alguna parella de nivells i, j . És a dir, que algun parell sigui diferent entre si

Aquí μ fa referència a la mediana poblacional de l'índex d'inflamabilitat per a cada un dels laboratoris $i, i = 1:5$.

2. Calculem els rangs

```

# Longitud total.
N = length(inflamabilitat)
N

## [1] 55

# Grandària dels grups.
n_i = tapply(inflamabilitat, laboratori, length)
n_i

## L1 L2 L3 L4 L5
## 11 11 11 11 11

# Càlcul dels rangs com si extractes d'una sola mostra.
inflamabilitat.rank = rank(inflamabilitat)
inflamabilitat.rank

## [1] 34.0 40.0 38.0 46.0 37.0 52.0 44.0 49.0 30.5 36.0 11.0  2.0 21.0 28.0
## [15] 13.0 35.0 39.0 30.5 33.0 42.0 22.0 16.5 16.5 18.0 24.0 25.0  8.0  6.0
## [29] 12.0  7.0 30.5 26.0 30.5 19.0 14.0 23.0  4.0  3.0 20.0  9.0 15.0 10.0
## [43]  1.0  5.0 50.0 51.0 47.0 53.0 41.0 43.0 48.0 27.0 45.0 54.0 55.0

# Suma dels rangs per cada grup (Laboratori)
sum.ranks = tapply(inflamabilitat.rank, laboratori, sum)
sum.ranks

##      L1      L2      L3      L4      L5
## 417.5 282.0 203.5 123.0 514.0

```

3. Calculem l'estadístic del test (mesura de discrepància)

L'estadístic del test és:

$$H = \frac{12}{N(N+1)} \cdot \sum_{i=1}^5 \frac{R_i^2}{n_i} - 3(N+1)$$

```

H = (12 / (N * (N + 1))) * sum((sum.ranks^2) / n_i) - 3 * (N + 1)
H

## [1] 35.50679

```

Però alguns autors recomanen l'estadístic “corregit pels empats” $H' = \frac{H}{C}$, on:

$$C = 1 - \frac{\sum_{i=1}^s (t_i^3 - t_i)}{N^3 - N}$$

amb s=nombre de sèries de valors empatats i t_i = llargada de la sèrie i de valors empatat.

Per tant, procedim a calcular-ho:

```

ties<-summary(as.factor(inflamabilitat.rank), maxsum=1000)
ties<-ties[ties>1]

```

```

t_i<-ties
t_i
## 16.5 30.5
##      2      4

C = 1 - sum(t_i^3 - t_i) / (N^3 - N)
H_cor<-H/C
H_cor
## [1] 35.52088

```

Així doncs, tenim que el valor de l'estadístic és $H = 35.50679$ i que l'estadístic corregit és $H' = 35.52088$. Encara que és força semblant, per als posteriors càlculs es farà servir l'estadístic amb la correcció ja que en les nostres dades hi ha diversos empats.

4. Calculem el p-valor i la regió de rebug

Tant l'estadístic com l'estadístic corregit segueixen asimptòticament una distribució chi-quadrat amb $a - 1$ graus de llibertat, on a és el nombre de condicions diferents (en aquest cas, $a=5$ laboratoris):

$$H' \sim \chi_{a-1}^2$$

```

pvalor <- 1-pchisq(H_cor, length(n_i)-1)
pvalor
## [1] 3.630622e-07

ChiT<-qchisq(0.95, (length(n_i)-1))
ChiT
## [1] 9.487729

```

El p-valor per a la prova que estem realitzant és 3.630622e-07.

5. Conclusió.

El criteri per prendre decisions en aquest test és que, si $H \geq \chi_{\alpha}^2(a - 1)$, o bé si el p-valor és inferior al nivell de significació α (és a dir, que $P\{\chi^2 \geq H|H_0\} \leq \alpha$), llavors es rebutjarà H_0 .

En el nostre cas, com que $p\text{-valor} = 3.63e-07$ (< 0.05 , amb un nivell de significació del 5%) i $H'=35.52088 > \chi_{\alpha}^2(a - 1)=9.487729$, rebutjarem la hipòtesi nul·la d'igualtat de medianes entre les mesures dels diferents grups. Per tant, es conclou que hi ha alguna diferència entre laboratoris pel que fa a la mediana de l'índex d'inflamabilitat.

b) Utilitza l'ordre de R `kruskal.test` i compara els resultats que has obtingut amb els de l'R.

Fem servir la funció `kruskal.test` indicant com a arguments d'aquesta els vectors amb les mesures d'inflamabilitat i els diferents laboratoris:

```
kruskal.test(inflamabilitat ~ laboratori)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: inflamabilitat by laboratori
## Kruskal-Wallis chi-squared = 35.521, df = 4, p-value = 3.631e-07
```

L'output ens retorna l'estadístic del test (Kruskal-Wallis chi-squared = 35.521), els graus de llibertat (que com hem vist abans aquests es mesuraven fent $a - 1$ i en el nostre cas $a = 5$ i d'aquí que $df = 5 - 1 = 4$) i el p-valor (p-value = 3.631e-07).

Veiem que els tres valors són pràcticament iguals als obtinguts en l'apartat 1 (només canvien alguns decimals a causa de l'arrodoniment que fa la funció `kruskal.test`), per la qual cosa, la conclusió a la que s'arriba és la mateixa.

c) Fent servir altres proves addicionals si s'escau, descriu les hipòtesis de validesa del test utilitzat i indica si en aquest cas concret es compleixen o si el creus apropiat per les dades presentades.

Tres condicions de validesa del test de Kruskal-Wallis:

(1) Les dades no provenen d'una normal.

En primer lloc s'estandaritzen les dades (restant a cada valor la mediana del grup al que pertany) i després, s'executa el test de normalitat.

```
inflamabilitat11 <- inflamabilitat[1:11] -
rep(median(inflamabilitat[1:11]),11)
inflamabilitat12 <- inflamabilitat[12:22] -
rep(median(inflamabilitat[12:22]),11)
inflamabilitat13 <- inflamabilitat[23:33] -
rep(median(inflamabilitat[23:33]),11)
inflamabilitat14 <- inflamabilitat[34:44] -
rep(median(inflamabilitat[34:44]),11)
inflamabilitat15 <- inflamabilitat[45:55] -
rep(median(inflamabilitat[45:55]),11)

dades.est <-
c(inflamabilitat11,inflamabilitat12,inflamabilitat13,inflamabilitat14,inflamabilitat15)

shapiro.test(dades.est)

##
## Shapiro-Wilk normality test
##
## data: dades.est
## W = 0.94104, p-value = 0.009412
```

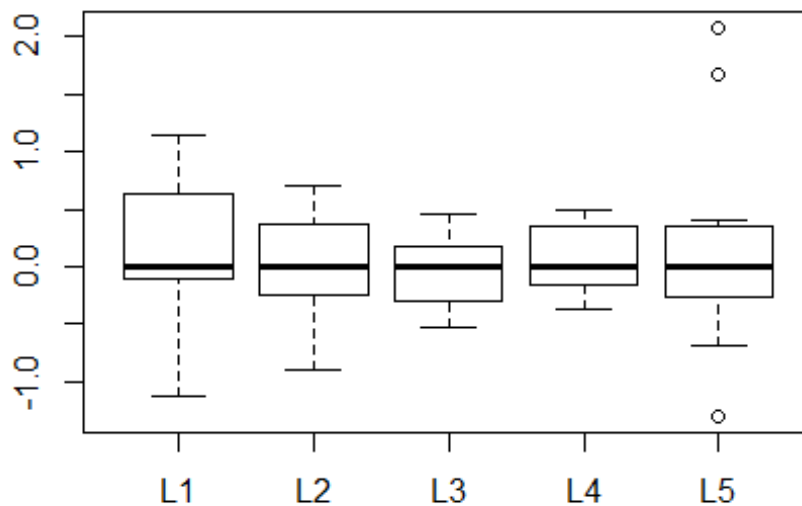
Com que el p valor del test de normalitat Shapiro-Wilk és inferior a 0.05 es conclou que les dades no provenen d'una normal.

(2) Les desviacions típiques dels a diferents grups són iguals:

$$\sigma_1 = \dots = \sigma_a$$

Ho comprovem visualment mitjançant un gràfic de boxplot:

```
boxplot(dades.est~laboratori)
```



Sembla que les dades tenen una variància similiar i per tant es pot considerar que hi ha homocedasticitat entre els grups.

(3) Les distribucions de les submostres de cadascuna de les a condicions són iguals:

$$F_1\left(\frac{Y - \mu_1}{\sigma_1}\right) = \dots = F_a\left(\frac{Y - \mu_a}{\sigma_a}\right),$$

on μ_i i σ_i són la mitjana i la desviació típica per cada gru $i = 1, 2, \dots, a$.

El test de Kolmogorov-Smirnov ens permet estudiar si les mostres provenen d'una mateixa distribució, per tant el realitzem a continuació.

1. Hipòtesis

H_0 : les mostres provenen d'una mateixa distribució.

H_1 : les mostres NO provenen d'una mateixa distribució.

2. Càlcul de l'estadístic i del p-valor mitjançant la funció `ks.test`.

```
ks.test(dades.est[1:11], dades.est[12:22])

## Warning in ks.test(dades.est[1:11], dades.est[12:22]): cannot compute
exact
## p-value with ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: dades.est[1:11] and dades.est[12:22]
## D = 0.27273, p-value = 0.8079
## alternative hypothesis: two-sided

ks.test(dades.est[1:11], dades.est[23:33])

## Warning in ks.test(dades.est[1:11], dades.est[23:33]): cannot compute
exact
## p-value with ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: dades.est[1:11] and dades.est[23:33]
## D = 0.36364, p-value = 0.4611
## alternative hypothesis: two-sided

ks.test(dades.est[1:11], dades.est[34:44])

## Warning in ks.test(dades.est[1:11], dades.est[34:44]): cannot compute
exact
## p-value with ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: dades.est[1:11] and dades.est[34:44]
## D = 0.36364, p-value = 0.4611
## alternative hypothesis: two-sided

ks.test(dades.est[1:11], dades.est[45:55])

## Warning in ks.test(dades.est[1:11], dades.est[45:55]): cannot compute
exact
## p-value with ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: dades.est[1:11] and dades.est[45:55]
## D = 0.18182, p-value = 0.9934
## alternative hypothesis: two-sided
```

3. Conclusió. Com que els p-valors són superiors a 0.05, no tenim evidències suficients per rebutjar la hipòtesis nul·la i es conclou que totes les dades provenen d'una mateixa distribució.

Així doncs, les hipòtesis de validesa del test de Kruskal-Wallis es compleixen i és apropiat haver-lo utilitzat amb aquestes dades.

Exercici 2

2. Independentment del que hagi contestat a l'exercici anterior, es demana determinar si hi ha diferències entre els laboratoris 1 i 4, i entre 1 i 5 fent servir la prova basada en rangs de les que hem vist a classe que consideris més adient.

a) Indica les hipòtesis, l'estadístic de test i la conclusió final d'aquesta prova sense utilitzar les ordres ad hoc de l'R.

Per comparar els paràmetres de dos grups independents, es recomanable l'utilització del Test de Mann-Whitney-Wilcoxon. Com no s'especifica quin tipus de desigualtat es vol estudiar, es farà el test bilateral.

- ENTRE ELS GRUPS 1 I 4:

1. Definim el contrast d'hipòtesis a realitzar

$$H_0: \mu_1 = \mu_4$$

$$H_1: \mu_1 \neq \mu_4$$

2. Calculem els rangs

```
# Longitud total.
inf1_4 <- inflamabilitat[c(1:11,34:44)]
lab1_4 <- laboratori[c(1:11,34:44)]
lab1_4 <- droplevels(lab1_4)
N1_4 = length(inf1_4)
N1_4

## [1] 22

# Grandària dels grups.
n_1_4 = tapply(inf1_4, lab1_4, length)
n_1_4

## L1 L4
## 11 11

# Càlcul dels rangs com si extractes d'una sola mostra.
inflamabilitat.rank1_4 = rank(inf1_4)
inflamabilitat.rank1_4

## [1] 14 18 17 20 16 22 19 21 13 15 7 10 8 12 3 2 11 5 9 6 1 4
```

```
# Suma dels rangs per cada grup (Laboratori)
sum.ranks1_4 = tapply(inflamabilitat.rank1_4, lab1_4, sum)
sum.ranks1_4

##   L1   L4
## 182   71
```

3. Calculem l'estadístic del test (mesura de discrepància)

- Estadístic U de Mann-Whitney

$$U = \min(R_{1\cdot} - \frac{n1 * (n1 + 1)}{2}, R_{2\cdot} - \frac{n2 * (n2 + 1)}{2})$$

- Estadístic W de Wilcoxon

$$W = R_{1\cdot} - \frac{n1 * (n1 + 1)}{2}$$

```
u1_4 =min(sum.ranks1_4[1]-((n_1_4[1]*(n_1_4[1]+1))/2),sum.ranks1_4[2]-
((n_1_4[2]*(n_1_4[2]+1))/2))
w1_4 <- sum.ranks1_4[1]-((n_1_4[1]*(n_1_4[1]+1))/2)
u1_4

## [1] 5

w1_4

##   L1
## 116
```

Hem obtingut :

$$W = 116$$

$$U = 5$$

S'assumira un nivell de significació α del 0.05.

Es rebutjara H_0 si:

$$U \leq U_{\alpha/2}(n, m)$$

$$U_{\alpha/2}(n, m) = 30$$

$$U = 5$$

Com l'estadístic U es igual a 5, es pot considerar que hi ha suficient evidència per rebutjar H_0 i per tant per considerar que els grups 1 i 4 no són iguals, ja que es compleix que $U \leq U_{0.05/2}(11, 11)$

4. Calculem l'interval de confiança


```

## Primer es calculen les nm diferencies
inf1 <- inf1_4[1:11]
inf4 <- inf1_4[12:22]
d1_4 <- vector()
k <- 1
for (i in 1:n_1_4[1]){
  for (j in 1:n_1_4[2]){
    d1_4[k] <- inf1[i]-inf4[j]
    k <- 1+k
  }
}

## S'ordenen de menor a major les nm diferencies
d1_4 <- sort(d1_4)

## el nostre U d'alpha mig a la taula es 30
u1_4 <- 30

alpha <- u1_4+1
v <- n_1_4[1]*n_1_4[2] - alpha+1

c(d1_4[alpha],d1_4[v])

## [1] 0.74 1.63

```

Interval de confiança de les diferencies = [0.74, 1.63]. Això vol dir que el grup 1 normalment es més gran que el 4, si fosin dependents en aquest interval tindria més rellevància el 0.

- ENTRE ELS GRUPS 1 I 5: **1. Definim el contrast d'hipòtesis a realitzar**

$$H_0: \mu_1 = \mu_5$$

$$H_1: \mu_1 \neq \mu_5$$

2. Calculem els rangs

```

# Longitud total.
inf1_5 <- inflamabilitat[c(1:11,45:55)]
lab1_5 <- laboratori[c(1:11,45:55)]
lab1_5 <- droplevels(lab1_4)
N1_5 = length(inf1_5)
N1_5

## [1] 22

# Grandària dels grups.
n_1_5 = tapply(inf1_5, lab1_5, length)
n_1_5

```

```
## L1 L4
## 11 11

# Càlcul dels rangs com si extractes d'una sola mostra.
inflamabilitat.rank1_5 = rank(inf1_5)
inflamabilitat.rank1_5

## [1] 4 8 7 13 6 19 11 16 3 5 1 17 18 14 20 9 10 15 2 12 21 22

# Suma dels rangs per cada grup (Laboratori)
sum.ranks1_5 = tapply(infamabilitat.rank1_5, lab1_5, sum)
sum.ranks1_5

## L1 L4
## 93 160
```

3. Calculem l'estadístic del test (mesura de discrepància)

- Estadístic U de Mann-Whitney

$$U = \min(R_1 - \frac{n_1 * (n_1 + 1)}{2}, R_2 - \frac{n_2 * (n_2 + 1)}{2})$$

- Estadístic W de Wilcoxon

$$W = R_1 - \frac{n_1 * (n_1 + 1)}{2}$$

```
u1_5 <- min(sum.ranks1_5[1] - ((n_1_5[1] * (n_1_5[1] + 1)) / 2), sum.ranks1_5[2] -
((n_1_5[2] * (n_1_5[2] + 1)) / 2))
w1_5 <- sum.ranks1_5[1] - ((n_1_5[1] * (n_1_5[1] + 1)) / 2)
u1_5

## [1] 27

w1_5

## L1
## 27
```

Hem obtingut :

$$W = 27$$

$$U = 27$$

S'assumira un nivell de significació α del 0.05.

Es rebutjara H_0 si:

$$U \leq U_{\alpha/2}(n, m)$$

$$U_{0.05/2}(11, 11) = 30$$

$$U = 27$$

Com l'estadístic U es igual a 27, es pot considerar que hi ha suficient evidència per rebutjar H_0 i per tant per considerar que els grups 1 i 5 són diferents, ja que es compleix que $U \leq U_{0.05/2}(11,11)$

4. Calculem l'interval de confiança

```
## Primer es calculen les nm diferències
inf1 <- inf1_5[1:11]
inf5 <- inf1_5[12:22]
d1_5 <- vector()
k <- 1
for (i in 1:n_1_5[1]){
  for (j in 1:n_1_5[2]){
    d1_5[k] <- inf1[i]-inf5[j]
    k <- 1+k
  }
}

## S'ordenen de menor a major les nm diferències
d1_5 <- sort(d1_5)

## el nostre U d'alpha mig a la taula es 30
u1_5 <- 30

alpha2 <- u1_5+1
v2 <- n_1_5[1]*n_1_5[2] - alpha2+1

c(d1_5[alpha2],d1_5[v2])

## [1] -1.38 -0.04
```

Interval de confiança de les diferències = [-1.38, -0.04]. Això vol dir que el grup 5 normalment és més gran que el 1, si fosin iguals en aquest interval tindria més rellevància el 0.

Conclusions Generals

Es pot veure com l'estadístic U entre els grups 1 i 4 és més baix que el dels grups 1 i 5, això es degut a que hi ha més diferència entre els grups 1 i 4. Es pot corroborar amb els intervals de confiança, l'interval de confiança entre els grups 1 i 5 està molt proper al 0 per la part superior, mentre que l'interval de la comparació entre els grups 1 i 4 és més llunyà al 0.

b) Compara els resultats que has obtingut amb les que obtindries utilitzant els tests implementats amb l'R i discuteix ambdós resultats de manera similar a com hem fet a classe.

- Entre grup 1 i 4

```

wilcox.test(inf1,inf4,conf.int = T, conf.level = .95)

##
## Wilcoxon rank sum test
##
## data:  inf1 and inf4
## W = 116, p-value = 5.387e-05
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  0.74 1.63
## sample estimates:
## difference in location
##                1.24

w1_4

## L1
## 116

c(d1_4[alpha],d1_4[v])

## [1] 0.74 1.63

```

Es pot comprovar com l'estadístic W dona igual pels dos mètodes. Segons la funció d'R es rebutjaria H_0 , per tant igual que com s'ha dit en el mètode fet a "mà".

Pel que fa a l'interval de confiança, s'han obtingut els mateixos límits superiors i inferiors fent-lo manualment que amb la funció `wilcox.test`.

- Entre grup 1 i 5

```

wilcox.test(inf1,inf5,conf.int = T, conf.level = .95)

##
## Wilcoxon rank sum test
##
## data:  inf1 and inf5
## W = 27, p-value = 0.02807
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -1.38 -0.04
## sample estimates:
## difference in location
##                -0.7

w1_5

## L1
## 27

c(d1_5[alpha2],d1_5[v2])

## [1] -1.38 -0.04

```

Es pot comprovar com l'estadístic W dona igual pels dos mètodes. Segons la funció d'R es rebutjaria H_0 , per tant igual que com s'ha dit en el mètode fet a "mà".

Pel que fa a l'interval de confiança, s'han obtingut els mateixos límits superiors i inferiors fent-lo manualment que amb la funció `wilcox.test`.

Conclusions Generals

Es pot veure que igual que s'ha concluit a l'apartat a) hi ha més evidència en la no igualtat entre el grup 1 i 4 que entre el grup 1 i 5. El p-valor de la primera comparativa és molt més petit que el de la segona. Per tant es pot concloure que s'han obtingut els mateixos resultats pels dos mètodes.

c) Creus que cal fer una correcció addicional? Quina seria? Si creus que s'hauria de fer alguna altra correcció, aplica-la.

És adient calcular els p-valors ajustats amb el mètode de Bonferroni per tal de corregir i verificar els p-valors i les conclusions anteriors. Ho farem mitjançant la funció `p.adjust` del paquet "Stat".

```
w1 <- wilcox.test(inf1,inf4,conf.int = T, conf.level = .95)
w2 <- wilcox.test(inf1,inf5,conf.int = T, conf.level = .95)

library(knitr)
kable(data.frame("pvalor" = c(w1$p.value, w2$p.value), "pvalor ajustat"
=p.adjust(c(w1$p.value,
w2$p.value),method="bonferroni"),row.names=c("Laboratoris 1 i 4","Laboratoris
1 i 5")), caption= "Taula comparativa", align='c')
```

Taula comparativa

	pvalor	pvalor.ajustat
Laboratoris 1 i 4	0.0000539	0.0001077
Laboratoris 1 i 5	0.0280651	0.0561301

Pel que fa a la comparació entre els laboratoris 1 i 4, s'observa com el p-valor ha variat una mica però segueix essent significatiu de manera que la conclusió de rebutjar H_0 i considerar que els grups 1 i 4 no són iguals segueix vigent. En canvi, quant als laboratoris 1 i 5, la correcció ha fet que el p-valor passés (per poc) a ser no significatiu, per un nivell de significació del 5%; així, es conclou finalment que hi ha diferències entre els laboratoris 1 i 5.