

## Enunciat

6. En un estudi per a determinar l'efectivitat de la vitamina B1 per estimular el creixement dels fongs *Trichophyton album*, d'una mostra de 24 plaques de cultiu que contienien el fong, a 13 seleccionades a l'atzar els fou aplicada vitamina B1 mentre que les restants serviren de control. El pes final dels fongs a cada placa va ser:

Control 18 14.5 13.5 12.5 23 24 21 17 18.5 9.5 14

Vitamina B1 27 34 20.5 29.5 20 28 20 26.5 22 24.5 34 35.5 19

A un nivell 0.01, es pot afirmar que la vitamina B1 incrementa significativament el creixement d'aquests fongs? Utilitza l'aproximació normal a partir dels següents càlculs R (els que siguin adequats):

```
> control = c(18, 14.5, 13.5, 12.5, 23, 24, 21, 17, 18.5, 9.5, 14)
> vitamina = c(27, 34, 20.5, 29.5, 20, 28, 20, 26.5, 22, 24.5, 34, 35.5, 19)
> mostra = c(control, vitamina)
> n1 = length(control)
> n2 = length(vitamina)
> N = n1 + n2
> n1 * (n1 + 1) / 2
[1] 66
> n2 * (n2 + 1) / 2
[1] 91
> N * (N + 1) / 2
[1] 300
> n1 * n2
[1] 143
> n1 * n2 * (n1 + n2 + 1)
[1] 3575
> n1 * n2 * (n1 + n2 - 1)
[1] 3289
> rangs = rank(mostra)
> rangs
[1] 7.0 5.0 3.0 2.0 15.0 16.0 13.0 6.0 8.0 1.0 4.0 19.0 22.5
12.0 21.0 10.5 20.0 10.5 18.0 14.0 17.0 22.5 24.0 9.0
> sum(rangs[1:n1])
[1] 80
>
```

## Solució

Aquest problema està pensat per a realitzar-lo sense l'ajut d'R ni de cap programari estadístic, solament cal fer càlculs senzills pels quals una calculadora és suficient. Els càlculs una mica més complexos ja estan fets als llistats R, que hem de saber interpretar.

Planteja una prova d'hipòtesis unilateral:  $H_0 : \mu_{control} = \mu_{vitamina}$  vs  $H_1 : \mu_{control} < \mu_{vitamina}$  on  $\mu$  aquí correspondria a la mediana poblacional de la variable “creixement”, contínua. També es podria plantejar com  $H_0 : \delta = 0$  vs  $\delta < 0$  si  $\delta = \mu_{control} - \mu_{vitamina}$ .

Intuïtivament, sembla que la informació que ens proporcionen els rangs suggereix que és veritat que la producció sota vitamina B1 és superior: la suma de rangs “control” és 80 mentre que la suma de rangs “vitamina” és  $N(N+1)/2 - 80 = 300 - 80 = 220$ . També és cert que el segon grup és una mica més gran (mides mostrals 11 i 13), però això no justificaria tanta diferència, només cal observar que les respectives mitjanes de rangs són 7.27 i 16.92. La darrera paraula la tindria una prova d'hipòtesis com la de Mann-Whitney-Wilcoxon. Calcularem l'estadístic

$$U = \min\{R_1 - \frac{n_1(n_1+1)}{2}, R_2 - \frac{n_2(n_2+1)}{2}\} = \min\{80 - 66, 220 - 91\} = 14$$

on  $R_1$  i  $R_2$  són les sumes dels rangs de les observacions del primer i del segon grup, respectivament.

L'enuniat demana que utilitzem l'aproximació normal. Si no hi haguessin empats, la variància a utilitzar seria

$$\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = \frac{3575}{12} = 297.9167$$

però a aquest valor cal restar-li

$$\frac{n_1 n_2 \left( \sum_{i=1}^s (t_i^3 - t_i) \right)}{12 (n_1 + n_2) (n_1 + n_2 - 1)}$$

on  $s$  vol dir quantes sèries d'empats hi ha (aquí  $s = 2$ ) i  $t_i$  és la llargada de la sèrie  $i = 1, 2$ . Per tant, la variància a utilitzar serà:

$$\sigma^2 = 297.9167 - \frac{11 \times 13 ((2^3 - 2) + (2^3 - 2))}{12 (11 + 13) (11 + 13 - 1)} = 297.6576.$$

Per tant, l'estadístic  $Z$  serà:

$$Z = \frac{14 - \frac{11 \times 13}{2}}{\sqrt{297.6576}} = -3.332801.$$

Atès que cal fer la prova a nivell 0.01, buscarem a la taula de la normal aquell valor que deixa a una cua una probabilitat  $\alpha = 0.01$ , és a dir, a la taula de 2 cues una probabilitat de  $2\alpha = 0.02$ . Aquest valor és  $z_{2\alpha} = 2.326347$ . Com que  $|-3.332801| > 2.326347$  i la mitjana dels rangs del grup control és inferior a la mitjana dels rangs del grup amb vitamina B1, podem rebutjar la hipòtesi nul·la, podem concloure que la mediana del creixement és superior en el grup tractat amb vitamina.

Encara que no es demana a l'enunciat, podríem afegir que l'interval de confiança associat al test anterior seria de la forma  $(-\infty, d_{(\nu)}]$  on  $\nu$  és el valor arrodonit a zero decimals de:

$$\nu^* = 1 + \frac{n_1 n_2}{2} + z_{2\alpha} \sigma$$

i  $d_{()}$  correspon al vector ordenat de totes les diferències entre els valors del grup “control” i el grup “vitamina”:

```
> control = c(18, 14.5, 13.5, 12.5, 23, 24, 21, 17, 18.5, 9.5, 14)
> vitamina = c(27, 34, 20.5, 29.5, 20, 28, 20, 26.5, 22, 24.5, 34, 35.5, 19)
> n1 = length(control)
> n2 = length(vitamina)
> d = sort(outer(control, vitamina, "-"))
> nu = round(1 + n1*n2/2 + 2.326347 * sqrt(297.6576))
> # Extrem superior de l'interval:
> d[nu]

[1] -3.5
```