

Inferencia — Examen parcial segundo

Problema 1

13 de mayo de 2024

```
load ("~/Downloads/eficacia.rda")
summary (eficacia)

##           Grupo      Estado      Marcador
## Control      :10   Bajo : 9   Min.      :  1.0
## Tratamiento:15   Medio:14  1st Qu.: 42.0
##              Alto  : 2   Median   : 80.0
##              Mean   :110.9
##              3rd Qu.:132.0
##              Max.   :509.0

attach (eficacia)
```

a)

Distribuciones univariantes de estado en ambos grupos:

```
(fre <- table (Grupo, Estado))

##           Estado
## Grupo      Bajo Medio Alto
## Control      6    3    1
## Tratamiento  3   11    1

round (100 * prop.table (fre, 1)) # porcentajes

##           Estado
## Grupo      Bajo Medio Alto
## Control      60   30   10
## Tratamiento  20   73    7
```

Se aprecian diferencias notables en los porcentajes de las dos primeras categorías, pero el tamaño muestral es pequeño. Para contrastar

H_0 : El estado se distribuye igual en controles y tratamientos

H_1 : El estado se distribuye distinto en controles y tratamientos

se usará el contraste χ^2 de homogeneidad por ser “estado” una variable categórica:

```
(ji2 <- chisq.test (Grupo, Estado))

## Warning in chisq.test(Grupo, Estado): Chi-squared approximation may be incorrect

##
## Pearson's Chi-squared test
##
## data: Grupo and Estado
## X-squared = 4.7619, df = 2, p-value = 0.09246
```

Hay un aviso por aproximación asintótica no fiable, porque alguna frecuencia esperada será menor que 5:

```
ji2$expected

##
## Estado
## Grupo      Bajo Medio Alto
## Control    3.6   5.6  0.8
## Tratamiento 5.4   8.4  1.2
```

Efectivamente, hay 3 frecuencias absolutas esperadas (el 50%) menores que 5. Usaremos simulación:

```
replicate (10, chisq.test (Grupo, Estado, simulate.p.value=TRUE, B=1e6) $ p.value)

## [1] 0.1015449 0.1015279 0.1009919 0.1010589 0.1008259 0.1022359 0.1009119
## [8] 0.1011829 0.1014169 0.1006939
```

El P -valor es estable y siempre algo mayor que $\alpha = 0.1$, por lo que por los pelos se decide que no hay evidencias para rechazar que el estado se distribuye de igual modo en el grupo de control y en el de tratamiento.

Otra prueba usable en este contexto es la de Fisher, tanto exacta (ya que el tamaño muestral es pequeño) como simulada:

```
fisher.test (Grupo, Estado)

##
## Fisher's Exact Test for Count Data
##
## data: Grupo and Estado
## p-value = 0.07553
## alternative hypothesis: two.sided

fisher.test (Grupo, Estado, simulate.p.value=TRUE, B=1e6)

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 1e+06 replicates)
##
## data: Grupo and Estado
## p-value = 0.07597
## alternative hypothesis: two.sided
```

Esta prueba sí es capaz de detectar diferencias significativas en la distribución del estado entre control y tratamiento a nivel $\alpha = 0.1$.

Si se considera que “Estado” es una variable ordinal y que Kruskal-Wallis permite corregir los empates, se podría aplicar transformando las categorías en números:

```
levels (Estado) # para comprobar si los niveles están en el orden correcto

## [1] "Bajo" "Medio" "Alto"

kruskal.test (as.numeric(Estado) ~ Grupo)

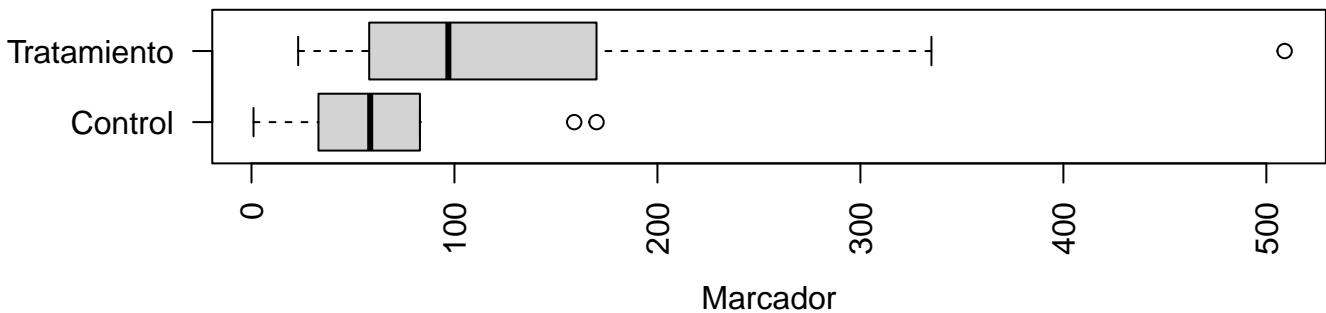
##
## Kruskal-Wallis rank sum test
##
## data: as.numeric(Estado) by Grupo
## Kruskal-Wallis chi-squared = 2.6719, df = 1, p-value = 0.1021
```

De nuevo, el P -valor está en torno a $\alpha = 0.1$. Esta vez, por encima, por lo que no se rechazaría H_0 .

b)

Se trata de realizar un contraste de homogeneidad de una variable continua. Descriptivos:

```
par (mar=c(5,6,1,0)) # para que se vea "Tratamiento" completo
boxplot (Marcador ~ Grupo, horizontal=TRUE, las=2, ylab="")
```



El marcador toma valores algo más altos en la muestra de “Tratamiento”. Vamos a comprobar a nivel $\alpha = 0.05$ si hay evidencias de que en la población ocurre lo mismo. Hipótesis:

H_0 : El marcador se distribuye igual en controles y tratamientos

H_1 : El marcador se distribuye distinto en controles y tratamientos

Comprobaremos en primer lugar si podemos usar métodos paramétricos mediante un contraste de bondad de ajuste de gaussianidad:

H_0 : La distribución es gaussiana

H_1 : La distribución no es gaussiana

```
(pval <- by (Marcador, Grupo, function (x) shapiro.test(x)$p.value))

## Grupo: Control
## [1] 0.1592133
## -----
## Grupo: Tratamiento
## [1] 0.001751438

p.adjust (pval) # corrección de Holm

##      Control Tratamiento
## 0.159213258 0.003502876
```

A nivel $\alpha = 0.05$ se rechaza la gaussianidad en el grupo “Tratamiento” incluso con la corrección de Holm para contrastes múltiples. Por tanto, recurriremos a métodos no paramétricos para comparar igualdad de distribuciones en muestras independientes (Mann-Whitney-Wilcoxon y Kruskal-Wallis):

```
## P-valor exacto
wilcox.test (Marcador ~ Grupo)

##
## Wilcoxon rank sum exact test
##
## data:  Marcador by Grupo
## W = 43, p-value = 0.08046
## alternative hypothesis: true location shift is not equal to 0

## P-valor asintótico
kruskal.test (Marcador ~ Grupo)

##
## Kruskal-Wallis rank sum test
##
## data:  Marcador by Grupo
## Kruskal-Wallis chi-squared = 3.1508, df = 1, p-value = 0.07589
```

En ambos casos, a nivel $\alpha = 0.05$ no se rechaza H_0 y, por tanto, no hay evidencias para afirmar que existen diferencias entre las distribuciones de “Marcador” en ambos grupos (lo que, por otro lado, parece contradecir que la distribución en un grupo no es gaussiana y, en el otro, puede serlo). Por tanto, los valores mayores observados en el diagrama de cajas para la muestra de “Tratamiento” pueden deberse al azar.

Otra prueba que podríamos haber aplicado es la de Kolmogórov y Smirnov para dos muestras:

```
control <- Marcador [Grupo == "Control"]
tratamiento <- Marcador [Grupo == "Tratamiento"]
ks.test (control, tratamiento)

##
## Exact two-sample Kolmogorov-Smirnov test
##
## data:  control and tratamiento
```

```
## D = 0.4, p-value = 0.2521
## alternative hypothesis: two-sided
```

Vemos que el P -valor obtenido es aun mayor que el de Wilcoxon, lo que no es sorprendente ya que esta prueba es habitualmente menos potente. En consecuencia, la conclusión es la misma: no se detectan diferencias significativas.

Otra prueba aun menos potente pero aplicable sería la prueba de la mediana¹ para dos muestras, que contrasta $H_0 : \text{Mediana}X = \text{Mediana}Y$, $H_1 : \text{Mediana}X \neq \text{Mediana}Y$:

```
median (control)
## [1] 58.5

median (tratamiento)
## [1] 97

(M <- median (c (control, tratamiento)))
## [1] 80

(U <- sum (control < M))
## [1] 7

(t <- U + sum (tratamiento <= M))
## [1] 13

val <- 0 : min (t, length(control))
round (dis <- setNames (dhyper (val, length(control), length(tratamiento), t),
                           val),
        3)
##      0      1      2      3      4      5      6      7      8      9     10
## 0.000 0.001 0.012 0.069 0.202 0.312 0.260 0.115 0.026 0.003 0.000

(R.C. <- sort (val [cumsum (sort (dis)) <= 0.05]))
## [1] 0 1 2 3 4 5

U %in% R.C.
## [1] FALSE
```

El estadístico U no está en la región crítica, por lo que no hay evidencias para rechazar la hipótesis nula de igualdad de medianas, lo que era esperable por la baja potencia de este método y los resultados anteriores.

¹Véase <https://bellman.ciencias.uniovi.es/~carleos/grado/inferencia/ce/15-no-param/nopar.html#orgb844229>

c)

Se trata de obtener un intervalo de confianza para la mediana de la población de control.

```
median (control)
## [1] 58.5
n <- length (control)
```

En la muestra de control hay $n = 2$ individuos. Un intervalo para la mediana M puede conseguirse buscando r y s tales que

$$1 - \alpha \geq \Pr[X_{(r)} < M < X_{(s)}] = \sum_{k=r}^{s-1} \Pr[B(n, 1/2) = k]$$

Como n es pequeño podemos buscar a ojo:

```
dbinom (0:n, n, 1/2)
## [1] 0.0009765625 0.0097656250 0.0439453125 0.1171875000 0.2050781250
## [6] 0.2460937500 0.2050781250 0.1171875000 0.0439453125 0.0097656250
## [11] 0.0009765625
sum (dbinom (3:7, n, 1/2))
## [1] 0.890625
sum (dbinom (3:8, n, 1/2))
## [1] 0.9345703
sum (dbinom (2:7, n, 1/2))
## [1] 0.9345703
x <- sort (control)
setNames (x, 1:n)
## 1 2 3 4 5 6 7 8 9 10
## 1 8 33 36 55 62 69 83 159 170
```

Por tanto, para superar $\alpha = 0.9$ podemos considerar los siguientes intervalos de confianza para la mediana del marcador en la población de control, cada uno de ellos con nivel de confianza 0.9345703:

- $r = 3, s = 8 + 1 \Rightarrow \text{I.C.} = (x_{(3)}, x_{(9)}) = (33, 159)$
- $r = 2, s = 7 + 1 \Rightarrow \text{I.C.} = (x_{(2)}, x_{(8)}) = (8, 83)$

Si buscamos el intervalo con un paquete de R, obtendremos un intervalo simétrico aproximado mediante interpolación lineal:

```

BSDA::SIGN.test (x)

##
## One-sample Sign-Test
##
## data:  x
## s = 10, p-value = 0.001953
## alternative hypothesis: true median is not equal to 0
## 95 percent confidence interval:
##  16.11111 134.34222
## sample estimates:
## median of x
##      58.5
##
## Achieved and Interpolated Confidence Intervals:
##
##                Conf.Level  L.E.pt  U.E.pt
## Lower Achieved CI    0.8906 33.0000  83.0000
## Interpolated CI     0.9500 16.1111 134.3422
## Upper Achieved CI    0.9785  8.0000 159.0000

```

Por ejemplo, el límite inferior del Interpolated CI lo calcula así:

```

(c1 <- sum (dbinom (3:7, n, 1/2)))

## [1] 0.890625

(c2 <- sum (dbinom (2:8, n, 1/2)))

## [1] 0.9785156

8 + (33-8) * (c2-0.95) / (c2-c1)

## [1] 16.11111

```

Otra posibilidad es considerar que la muestra se puede considerar gaussiana:

```

shapiro.test (control)

##
## Shapiro-Wilk normality test
##
## data:  control
## W = 0.88757, p-value = 0.1592

```

En consecuencia, por ser simétrica la campana de Gauß, la mediana coincidiría con la media y el intervalo de confianza puede obtener como es habitual para la media en poblaciones gaussianas:

```
t.test (control, conf.level=0.9)

##
## One Sample t-test
##
## data: control
## t = 3.7369, df = 9, p-value = 0.004648
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
## 34.43895 100.76105
## sample estimates:
## mean of x
## 67.6

## o bien, a mano:
mean (control) + c(-1,1) * qt(0.95,n-1) * sd(control) / sqrt(n)

## [1] 34.43895 100.76105
```

Otra posibilidad más sería usar *bootstrap*, por ejemplo un *bootstrap* percentil:

```
quantile (replicate (10000, median(sample(control,,TRUE))),
          c(0.05,0.95))

## 5% 95%
## 33 83
```

...o uno básico:

$$\begin{aligned}
 1 - \alpha &= \Pr[a < \text{Mediana}\vec{X} - \text{Mediana}X < b] = (*) \approx (\star) \\
 (\star) &\approx \Pr[a < \text{Mediana}\vec{X}^* - \text{Mediana}\vec{x} < b] \rightarrow \text{estimas } \hat{a}, \hat{b} \\
 (*) &= \Pr[\text{Mediana}\vec{X} - b < \text{Mediana}X < \text{Mediana}\vec{X} - a] \\
 &\approx \Pr[\text{Mediana}\vec{X} - \hat{b} < \text{Mediana}X < \text{Mediana}\vec{X} - \hat{a}] \\
 &\Rightarrow \text{I.C.} = [\text{Mediana}\vec{x} - \hat{b}, \text{Mediana}\vec{x} - \hat{a}] = \text{Mediana}\vec{x} - [\hat{b}, \hat{a}]
 \end{aligned}$$

```
median(control) - quantile (replicate (10000,
                                     median(sample(control,,TRUE)) - median(control)),
                             c(0.95,0.05))

## 95% 5%
## 34 84
```

Por otro lado, `wilcox.test` ofrece un intervalo de confianza de la pseudomediana, es decir, la mediana de la semisuma de dos copias independientes de la variable. Esto coincide con la mediana si hay simetría (pero entonces la mediana coincide también con la media) pero en el caso general no es cierto y se trataría de una aproximación burda:


```
wilcox.test (control, conf.int=TRUE, conf.level=0.9)

##
## Wilcoxon signed rank exact test
##
## data: control
## V = 55, p-value = 0.001953
## alternative hypothesis: true location is not equal to 0
## 90 percent confidence interval:
## 33 107
## sample estimates:
## (pseudo)median
## 59.5
```