HT 5 - Estadística 2

Rudik Roberto Rompich

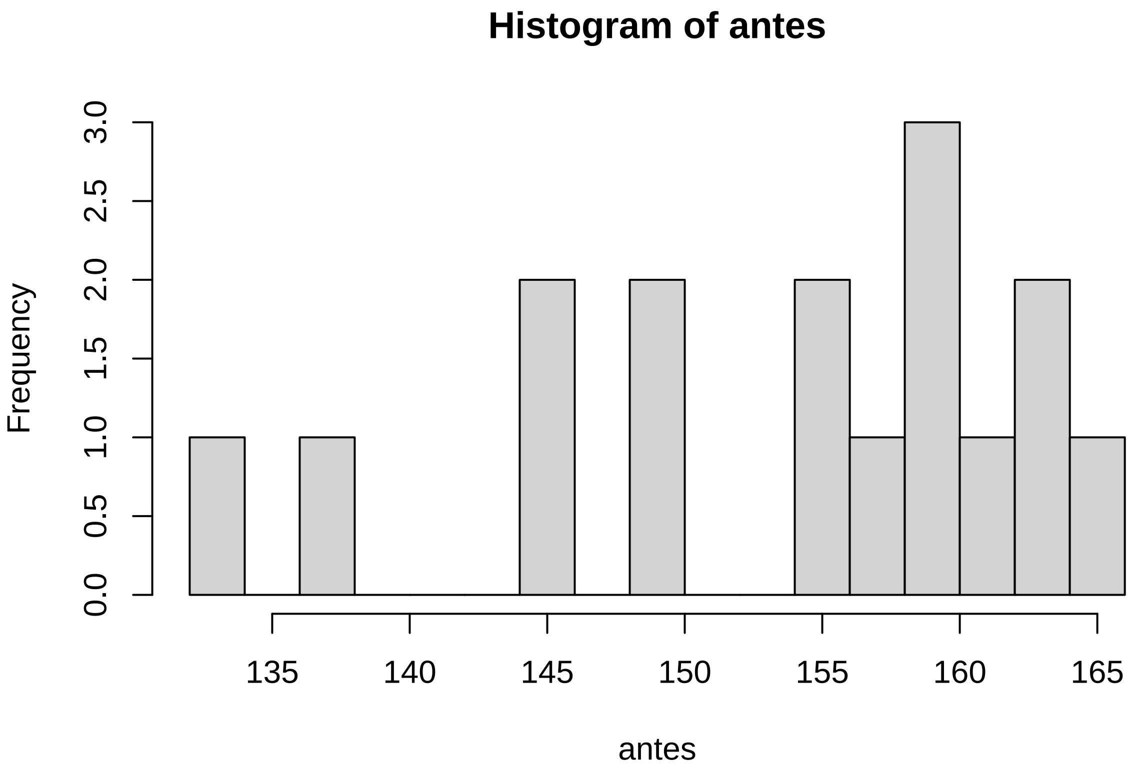
Problema 1

Los datos muestran la presión sanguínea sistólica de 16 corredores antes y después de una carrera de 8 kilómetros.

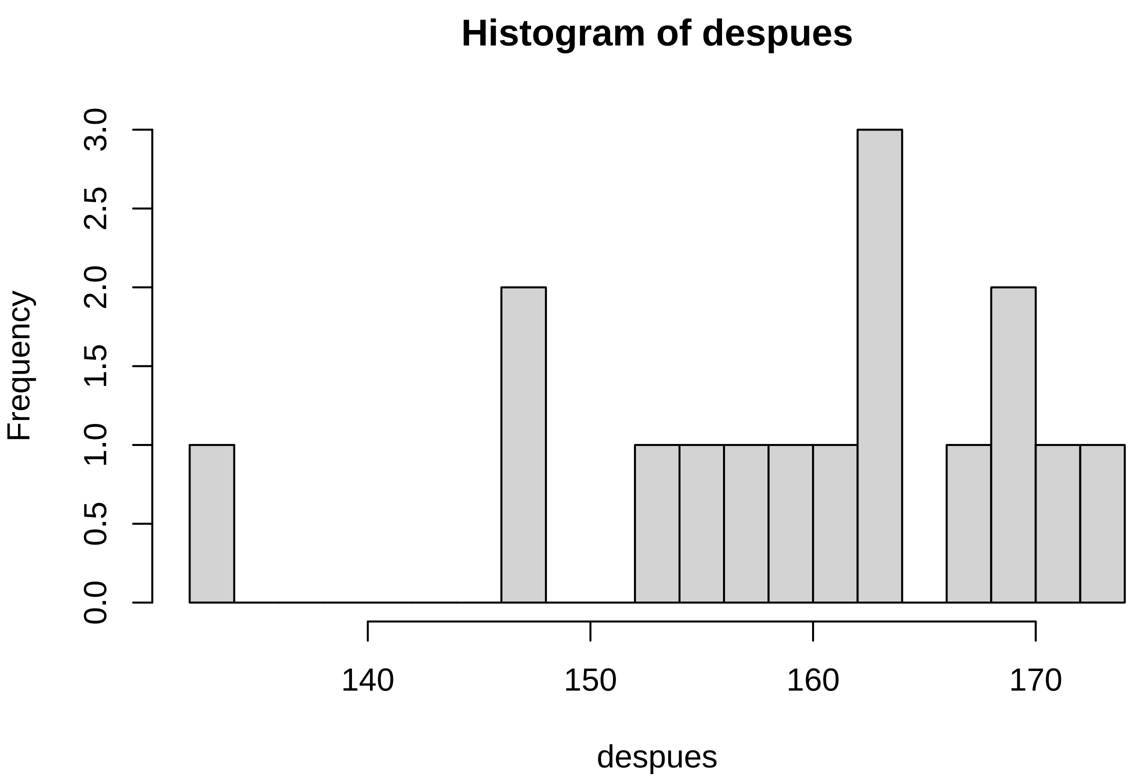
```
DB1 <- read.csv("/Users/rudiks/Desktop/p1.csv")
DB1
```

```
##    Corredor Antes Despues
## 1         1   158     164
## 2         2   149     158
## 3         3   160     163
## 4         4   155     160
## 5         5   164     172
## 6         6   138     147
## 7         7   163     167
## 8         8   159     169
## 9         9   165     173
## 10       10   145     147
## 11       11   150     156
## 12       12   161     164
## 13       13   132     133
## 14       14   155     161
## 15       15   146     154
## 16       16   159     170
```

```
#Gráficamente
antes<- DB1$Antes
despues <- DB1$Despues
hist(antes, breaks=16)
```




```r
hist(despues, breaks=16)
```



Con una significancia de 0.05, ¿será posible afirmar que correr 8 kilómetros aumenta la mediana de la presión sanguínea sistólica en menos de 8 puntos?

Solución.

Comenzamos determinando las medianas.

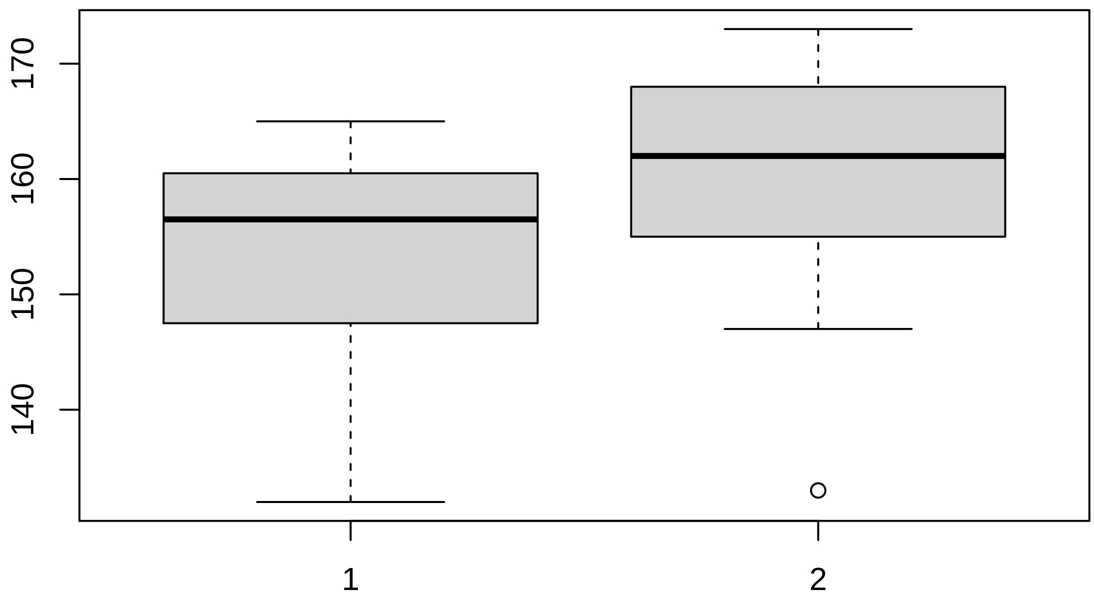
```r
#Mediana antes
median(antes)
```

```
## [1] 156.5
```

```r
#Mediana después
median(despues)
```

```
## [1] 162
```

```r
#Gráficamente con una caja de bigotes
boxplot(antes, despues)
```



Se indentifica la hipótesis:

$$H_0 : \tilde{\mu}_A - \tilde{\mu}_D \geq 8$$
$$H_a : \tilde{\mu}_A - \tilde{\mu}_D < 8$$

```r
#Es pareada
#Por la hipótesis es de cola inferior, por eso es "less".
wilcox.test(antes, despues, mu=8, alternative = "less", paired = T, exact=F, conf.int = 0.95)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  antes and despues
## V = 0, p-value = 0.000236
## alternative hypothesis: true location shift is less than 8
## 95 percent confidence interval:
##       -Inf -4.999982
## sample estimates:
## (pseudo)median
##      -6.000087
```

∴ Por la prueba del valor-p con un significancia 0.05; 0.000236<0.05. Por lo tanto, la hipótesis nula se rechaza, es decir aumenta la mediana de la presión sanguínea sistólica en menos de 8 puntos en una carrera de 8 km.

Problema 2

Los siguientes datos representan los pesos, en kilogramos, del equipaje personal que llevan, en diferentes vuelos, un miembro de un equipo de béisbol y un jugador de un equipo de baloncesto.

```r
DB <- read.csv("/Users/rudiks/Desktop/p2.csv")
DB
```

```
##    Jugador_de_beisbol Jugador_de_baloncesto
## 1                16.3                  15.4
## 2                18.1                  17.7
## 3                15.9                  18.6
## 4                14.1                  12.7
## 5                17.7                  15.0
## 6                16.3                  15.9
## 7                13.2                  16.3
## 8                20.0                  18.1
## 9                15.0                  16.8
## 10               18.6                  14.1
## 11               14.5                  13.6
## 12               19.1                  25.0
## 13               13.6                    NA
## 14               17.2                    NA
## 15               18.6                    NA
## 16               15.4                    NA
## 17               15.6                    NA
## 18               18.3                    NA
## 19               17.4                    NA
## 20               14.8                    NA
## 21               16.5                    NA
```

Con una significancia de 0.05. ¿Podemos afirmar que el peso del equipaje para los dos tipos de atletas es diferente?

Solución.

Se tiene la hipótesis:

$$
\begin{aligned}
H_0 :& \tilde{\mu}_{beisbol} = \tilde{\mu}_{baloncesto} \\
H_a :& \tilde{\mu}_{beisbol} \neq \tilde{\mu}_{baloncesto}
\end{aligned}
$$

Como la muestra no es pareada, descartamos la prueba de signos y la única posibilidad es Wilcoxon. Es decir:

```r
baloncesto <- DB$Jugador_de_baloncesto
beisbol <- DB$Jugador_de_beisbol
wilcox.test(baloncesto, beisbol, alternative ="t", paired = F, exact= F, conf.int = 0.95)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  baloncesto and beisbol
```

```
## W = 115.5, p-value = 0.7079
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -1.999973  1.400000
## sample estimates:
## difference in location
##             -0.3999956
```

$\therefore$ Por la prueba del valor-p con un significancia 0.05; 0.7079>0.05. Por lo tanto, la hipótesis nula NO se rechaza, es decir que no podemos decir que el peso del equipaje es distinto.

Problema 3

```r
DB <- read.csv("/Users/rudiks/Desktop/p3.csv")
DB
```

```
##   Metodo_1 Metodo_2 Metodo_3 Metodo_4
## 1       65       75       59       94
## 2       87       69       78       89
## 3       73       83       67       80
## 4       79       81       62       88
## 5       81       72       83       90
## 6       69       79       76       62
## 7       NA       90       NA       71
```

Los datos mostrados se recolectaron usando un diseño completamente aleatorizado. Son las calificaciones del examen de aprovechamiento para cuatro diferentes grupos de estudiantes, donde cada grupo recibió enseñanza mediante una técnica diferente. ¿Podemos afirmar que todos los métodos producen resultados similares?

Solución.

Tenemos la hipótesis

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3 = \tilde{\mu}_4$$
$$H_a : \tilde{\mu}_1 \neq \tilde{\mu}_2 \neq \tilde{\mu}_3 \neq \tilde{\mu}_4$$

Por las características del problema, únicamente se puede utilizar Kruskal-Wallis.

```r
m1 <- DB$Metodo_1
m2 <- DB$Metodo_2
m3 <- DB$Metodo_3
m4 <- DB$Metodo_4

kruskal.test(list(m1,m2,m3,m4))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  list(m1, m2, m3, m4)
## Kruskal-Wallis chi-squared = 4.406, df = 3, p-value = 0.2208
```

$\therefore$ Por la prueba del valor-p con un significancia no específicada, pero se asumirá $\alpha = 0.5$; 0.2208>0.05. Por lo tanto, la hipótesis nula NO se rechaza, es decir que podemos decir que todos los métodos producen resultados similares.