

Tarea 06 Prueba de Hipótesis

Heidy Valdelamar Gonzalez

2022-11-16

Llamamos las librerías que usaremos

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(Rmisc)
```

```
## Loading required package: lattice
## Loading required package: plyr
## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
##
## The following object is masked from 'package:purrr':
##
##   compact
```

```
library(rcompanion)
```

1. Vitacca et al. Realizaron un estudio para determinar si la posición supina o la posición sentada empeoran los flujos espiratorios forzados estáticos y las mediciones de la mecánica pulmonar. Los sujetos eran personas de edad que vivían en una residencia de ancianos, clínicamente estables y sin evidencia clínica de enfermedades cardiorrespiratorias. Entre los datos recogidos se encuentran los siguientes valores porcentuales de FEV para los sujetos en posición sentada y supina en REV_C07_46.csv.

- Realice un análisis estadístico de los datos (que incluya la comprobación de hipótesis y la construcción de intervalos de confianza) que, en su opinión, pueda aportar información útil a los investigadores.
- Indique todos los supuestos necesarios para validar su análisis.
- Encuentre los valores p para todos los estadísticos de prueba calculados.

Leer base de datos

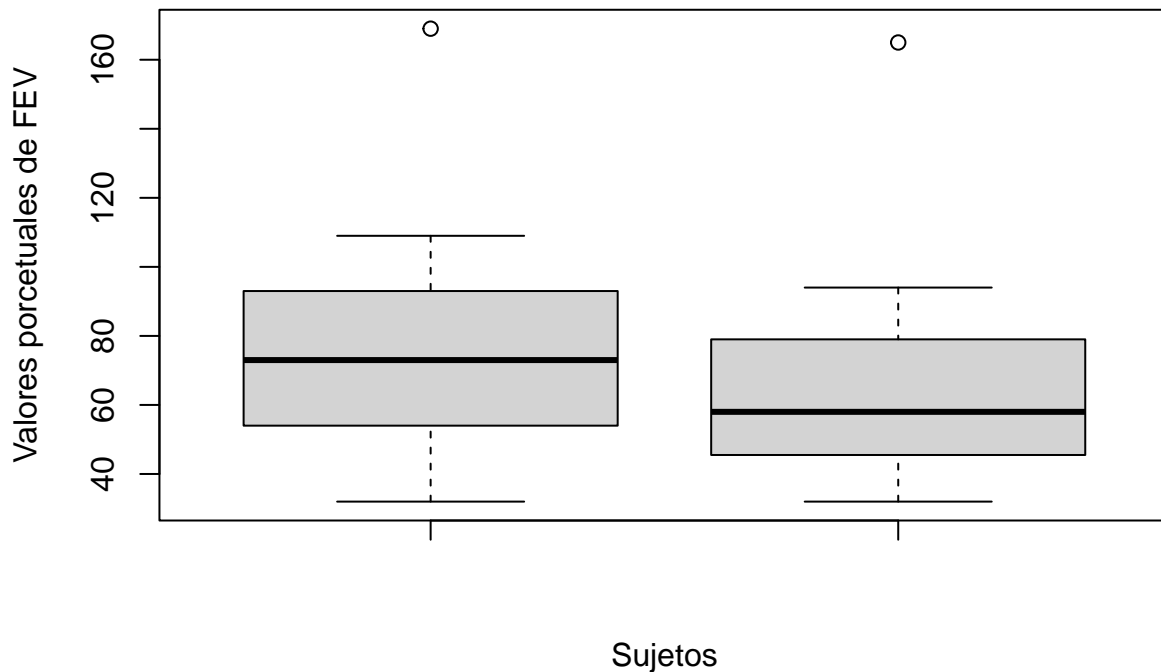
```
bd <- read_csv("REV_C07_46.csv", show_col_types = FALSE)
bd
```

```
## # A tibble: 17 x 2
##       SIT    SUP
##   <dbl> <dbl>
## 1     64    56
## 2     44    37
## 3     44    39
## 4     40    43
## 5     32    32
## 6     70    61
## 7     82    58
## 8     74    48
## 9     91    63
## 10    103    94
## 11    109    92
## 12     NA     NA
## 13    169   165
## 14     73    66
## 15     95    94
## 16     NA     NA
## 17     73    58
```

```
tail(bd)
```

```
## # A tibble: 6 x 2
##       SIT    SUP
##   <dbl> <dbl>
## 1     NA     NA
## 2    169   165
## 3     73    66
## 4     95    94
## 5     NA     NA
## 6     73    58
```

```
boxplot(bd$SIT, bd$SUP, xlab = "Sujetos", ylab = "Valores porcentuales de FEV")
```



```
# Comparamos la igualdad
t.test(bd$SIT, bd$SUP, alternative = "two.sided", mu = 0, var.equal = FALSE, conf.level = .95)
```

```
##
## Welch Two Sample t-test
##
## data: bd$SIT and bd$SUP
## t = 0.84259, df = 27.988, p-value = 0.4066
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -14.97928 35.91262
## sample estimates:
## mean of x mean of y
## 77.53333 67.06667
```

2. Montner et al. realizaron estudios para comprobar los efectos de la hiperhidratación con glicerol (GEH) en la resistencia del rendimiento ciclista. Los 11 sujetos, de edades comprendidas entre los 22 y los 40 años, practicaban regularmente el ciclismo al menos 75 millas por semana. Los datos en REV_C07_41.csv son los volúmenes de salida de orina antes del ejercicio (ml) tras la ingestión de glicerol y agua.

- Realice un análisis estadístico de los datos (que incluya la comprobación de hipótesis y la construcción de intervalos de confianza) que, en su opinión, pueda aportar información útil a los investigadores.
- Indique todos los supuestos necesarios para validar su análisis.
- Encuentre los valores p para todos los estadísticos de prueba calculados.

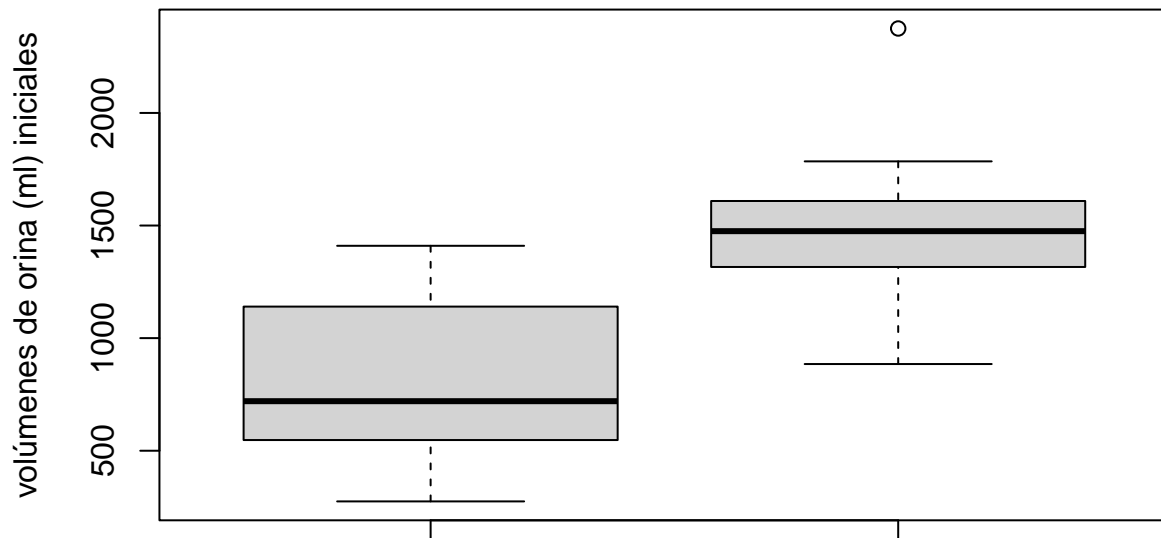
Leemos el csv

```
GEH <- read_csv("REV_C07_41.csv", show_col_types = FALSE)
GEH
```

```
## # A tibble: 11 x 3
##   Subj Glycerol Control
##   <dbl>   <dbl>   <dbl>
## 1     1     1410   2375
## 2     2     610   1610
## 3     3    1170   1608
## 4     4    1140   1490
## 5     5     515   1475
## 6     6     580   1445
## 7     7     430    885
## 8     8    1140   1187
## 9     9     720   1445
## 10    10     275    890
## 11    11     875   1785
```

Gráfica Boxplot

```
boxplot(GEH$Glycerol, GEH$Control, ylab = "volúmenes de orina (ml) iniciales")
```



El grupo control tuvo niveles mayores de volúmenes de orina, sin embargo, el grupo que consumió agua con glicerol es más grande, pero en promedio el volumen de orina es menor. tendríamos una diferencia negativa de ambos grupos.

Prueba T_Student pareada doble por la derecha, nivel de confianza del 99%

```
t.test(GEH$Glycerol, GEH$Control, alternative = "less", paired = TRUE, conf.level = 0.99)
```

```
##
## Paired t-test
##
## data: GEH$Glycerol and GEH$Control
## t = -7.0883, df = 10, p-value = 1.671e-05
## alternative hypothesis: true mean difference is less than 0
## 99 percent confidence interval:
##      -Inf -406.5462
## sample estimates:
```

```
## mean difference
##      -666.3636
```

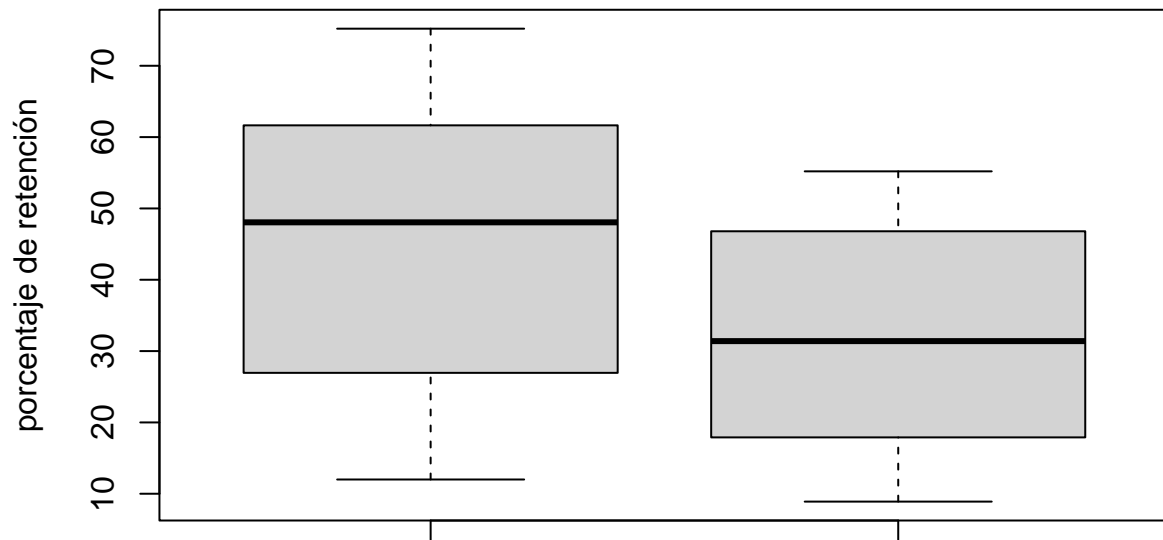
Tenemos un valor p de 1.671e-05 menor a el alpha que definimos = 0.005, por lo que descartamos la hipótesis nula. El intervalo de confianza es de -406.5462, la diferencia de los promedios es negativa: -666.3636. Concluimos que la diferencia entre los grupos es que la media del grupo control es mayor a la media del grupo que consumió agua con glycerol, por lo tanto el agua con glycerol no induce la hiperhidratación en los 11 sujetos.

3. Algunos investigadores han observado una mayor resistencia de las vías respiratorias en los fumadores que en los no fumadores. Supongamos que un estudio, realizado para comparar el porcentaje de retención braquiobronquial de partículas en gemelos monocigóticos fumadores-discordantes, arrojó los resultados en REV_C07_29.csv. ¿Apoyan estos datos la hipótesis de que el aclaramiento traqueo-bronquial es más lento en los fumadores? Sea $\alpha = 0,05$ y determine el valor p para esta prueba.

```
fuma <- read_csv("REV_C07_29.csv", show_col_types = FALSE )
fuma
```

```
## # A tibble: 12 x 2
##   Smoking Nonsmoking
##   <dbl>      <dbl>
## 1    60.6      47.5
## 2     12      13.3
## 3     56      33
## 4    75.2     55.2
## 5    12.5     21.9
## 6    29.7     27.9
## 7    57.2     54.3
## 8    62.7     13.9
## 9    28.7      8.9
## 10    66      46.1
## 11    25.2     29.8
## 12    40.1     36.2
```

```
boxplot(fuma$Smoking, fuma$Nonsmoking, xlab = "Sujetos", ylab = "porcentaje de retención")
```



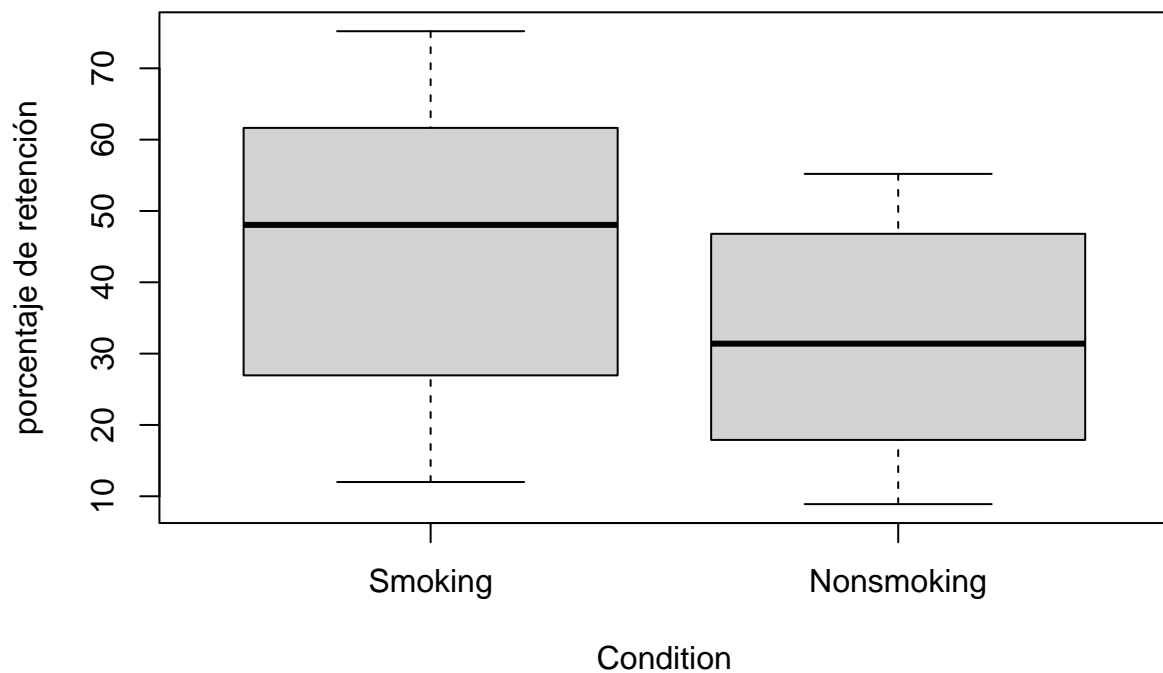
Sujetos

```
# Para hacer una prueba t usando notación de fórmula fuma ~ Condition
# Se necesitan alargar los datos

fuma_long <- fuma %>% pivot_longer(cols = everything(), names_to = "Condition", values_to = "fuma")

# y estar seguros que Condition es un factor en el orden adecuado

fuma_long <- fuma_long %>% mutate( Condition = Condition %>% fct_relevel("Smoking", "Nonsmoking"))
boxplot(fuma ~ Condition, data = fuma_long, ylab = "porcentaje de retención")
```



```
t.test(fuma ~ Condition, data = fuma_long, alternative = "less", mu=0, paired = TRUE, var.equal = TRUE,
```

```
##  
## Paired t-test  
##  
## data: fuma by Condition  
## t = 2.4847, df = 11, p-value = 0.9848  
## alternative hypothesis: true mean difference is less than 0  
## 95 percent confidence interval:  
##      -Inf 19.7975  
## sample estimates:  
## mean difference  
##      11.49167
```