

Ejercicios Sesión 7

2025-06-05

Se cuenta con un conjunto de datos que representa información clínica de 300 pacientes, con variables demográficas, de tratamiento y de laboratorio.

Las variables incluidas son:

- paciente_id: identificador del paciente
- edad: edad en años
- sexo: “Hombre” o “Mujer”
- grupo_tratamiento: grupo asignado (“A”, “B” o “C”)
- presion_sistolica: presión arterial sistólica (mmHg)
- presion_diastolica: presión arterial diastólica (mmHg)
- colesterol: nivel de colesterol total (mg/dL)
- glucosa_baseline: glucosa basal (mg/dL)
- glucosa_6meses: glucosa a los 6 meses de seguimiento (mg/dL)

Pregunta inicial: ¿Qué tipo de variables hay? ¿Cuáles son categóricas y cuáles cuantitativas? ¿Qué tipo de análisis podrías hacer?

```
set.seed(123)

n <- 500

library(tibble)
set.seed(123) # Para reproducibilidad
n <- 300

# Generación de variables base
paciente_id <- 1:n
edad <- round(rnorm(n, mean = 65, sd = 15))
sexo <- sample(c("Hombre", "Mujer"), n, replace = TRUE)
grupo_tratamiento <- sample(c("A", "B", "C"), n, replace = TRUE)

# Presión sistólica depende del grupo de tratamiento
presion_sistolica <- round(rnorm(n, mean = 130 +
                                ifelse(grupo_tratamiento == "A", 0,
                                ifelse(grupo_tratamiento == "B", 5, 10)),
                                sd = 10))

# Colesterol depende del sexo
```

```

colesterol <- round(rnorm(n, mean = 200 +
                        ifelse(sexo == "Hombre", 4, -4),
                        sd = 20))

# Resto de variables sin modificar
presion_diastolica <- round(rnorm(n, mean = 80, sd = 10))
glucosa_baseline <- round(rnorm(n, mean = 100, sd = 15))
glucosa_6meses <- round(rnorm(n, mean = 97, sd = 15))

# Crear el tibble
data <- tibble(
  paciente_id,
  edad,
  sexo,
  grupo_tratamiento,
  presion_sistolica,
  presion_diastolica,
  colesterol,
  glucosa_baseline,
  glucosa_6meses
)

head(data)

```

```

## # A tibble: 6 x 9
##   paciente_id edad sexo grupo_tratamiento presion_sistolica presion_diastolica
##         <int> <dbl> <chr> <chr>                <dbl>                <dbl>
## 1           1    57 Mujer A                   111                   91
## 2           2    62 Mujer A                   131                   86
## 3           3    88 Mujer B                   141                   84
## 4           4    66 Homb~ C                   125                   84
## 5           5    67 Homb~ C                   145                   93
## 6           6    91 Homb~ B                   127                   70
## # i 3 more variables: colesterol <dbl>, glucosa_baseline <dbl>,
## #   glucosa_6meses <dbl>

save(data, file = "datasets/data.Rda" )

```

Análisis Exploratorio

Antes de realizar pruebas estadísticas, es esencial explorar los datos.

Reflexiona: ¿Hay diferencias visibles en las distribuciones por sexo o tratamiento? ¿Existen posibles valores atípicos o distribuciones sesgadas?

```
summary(data)
```

```

##   paciente_id      edad      sexo      grupo_tratamiento
## Min.   : 1.00   Min.   : 30.00   Length:300   Length:300
## 1st Qu.: 75.75   1st Qu.: 56.00   Class :character   Class :character

```

```
## Median :150.50   Median : 64.00   Mode  :character   Mode  :character
## Mean    :150.50   Mean    : 65.54
## 3rd Qu.:225.25   3rd Qu.: 74.25
## Max.    :300.00   Max.    :114.00
## presion_sistolica presion_diastolica   colesterol   glucosa_baseline
## Min.    :105.0    Min.    : 53.00    Min.    :145.0    Min.    : 54.0
## 1st Qu.:129.0    1st Qu.: 74.00    1st Qu.:185.0    1st Qu.: 90.0
## Median :136.0    Median : 81.00    Median :201.0    Median :100.0
## Mean    :135.9    Mean    : 80.65    Mean    :199.9    Mean    :100.1
## 3rd Qu.:144.0    3rd Qu.: 87.00    3rd Qu.:213.0    3rd Qu.:112.0
## Max.    :166.0    Max.    :114.00    Max.    :261.0    Max.    :149.0
## glucosa_6meses
## Min.    : 58.00
## 1st Qu.: 87.00
## Median : 96.00
## Mean    : 96.78
## 3rd Qu.:107.00
## Max.    :139.00
```

Análisis Estadístico. Comparación del colesterol según el sexo.

Los investigadores creen que podría haber una diferencia en los niveles de colesterol entre hombres y mujeres.

Plantea:

- ¿Cuál es la hipótesis nula y alternativa?
- ¿La variable es continua y la agrupación es binaria? ¿Qué prueba usarías?
- ¿Debes usar un test paramétrico (t-test) o no paramétrico (Wilcoxon)? Porque?

```
t.test(colesterol ~ sexo, data = data)
```

```
##
## Welch Two Sample t-test
##
## data:  colesterol by sexo
## t = 3.9439, df = 295.11, p-value = 0.0001002
## alternative hypothesis: true difference in means between group Hombre and group Mujer is not equal to 0
## 95 percent confidence interval:
##  4.638226 13.877874
## sample estimates:
## mean in group Hombre mean in group Mujer
##           204.4156           195.1575
```

```
wilcox.test(colesterol ~ sexo, data = data)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  colesterol by sexo
## W = 14186, p-value = 8.821e-05
## alternative hypothesis: true location shift is not equal to 0
```

Comparación pareada de glucosa antes y después

Los clínicos creen que tras 6 meses de tratamiento, los niveles de glucosa han disminuido significativamente.

Plantea:

- ¿Qué tipo de prueba corresponde si los datos son pareados?
- ¿Los datos parecen normales? ¿Debes usar un t-test pareado o Wilcoxon pareado?

Reflexiona: ¿Se cumple la hipótesis del equipo clínico? ¿Hay evidencia estadística de cambio?

```
t.test(data$glucosa_baseline, data$glucosa_6meses, paired = TRUE)
```

```
##
## Paired t-test
##
## data: data$glucosa_baseline and data$glucosa_6meses
## t = 2.5771, df = 299, p-value = 0.01044
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.7815769 5.8317564
## sample estimates:
## mean difference
##      3.306667
```

```
wilcox.test(data$glucosa_baseline, data$glucosa_6meses, paired = TRUE)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: data$glucosa_baseline and data$glucosa_6meses
## V = 25292, p-value = 0.01338
## alternative hypothesis: true location shift is not equal to 0
```

Comparación entre 3 grupos de tratamiento: presión sistólica

El equipo de investigación sospecha que los distintos grupos de tratamiento podrían estar asociados con diferencias en la presión sistólica.

Plantea:

- ¿Cuántos grupos tienes? ¿Qué tipo de análisis utilizarías (ANOVA o Kruskal-Wallis)?

Reflexiona: ¿Se observan diferencias significativas? ¿La prueba elegida fue la adecuada?

```
data$grupo_tratamiento <- as.factor(data$grupo_tratamiento)
anova_result <- aov(presion_sistolica ~ grupo_tratamiento, data = data)
summary(anova_result)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## grupo_tratamiento  2   4476   2238.0    20.8 3.51e-09 ***
## Residuals        297  31955    107.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(multcomp)
```

```
## Cargando paquete requerido: mvtnorm
```

```
## Cargando paquete requerido: survival
```

```
## Cargando paquete requerido: TH.data
```

```
## Cargando paquete requerido: MASS
```

```
##
```

```
## Adjuntando el paquete: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
```

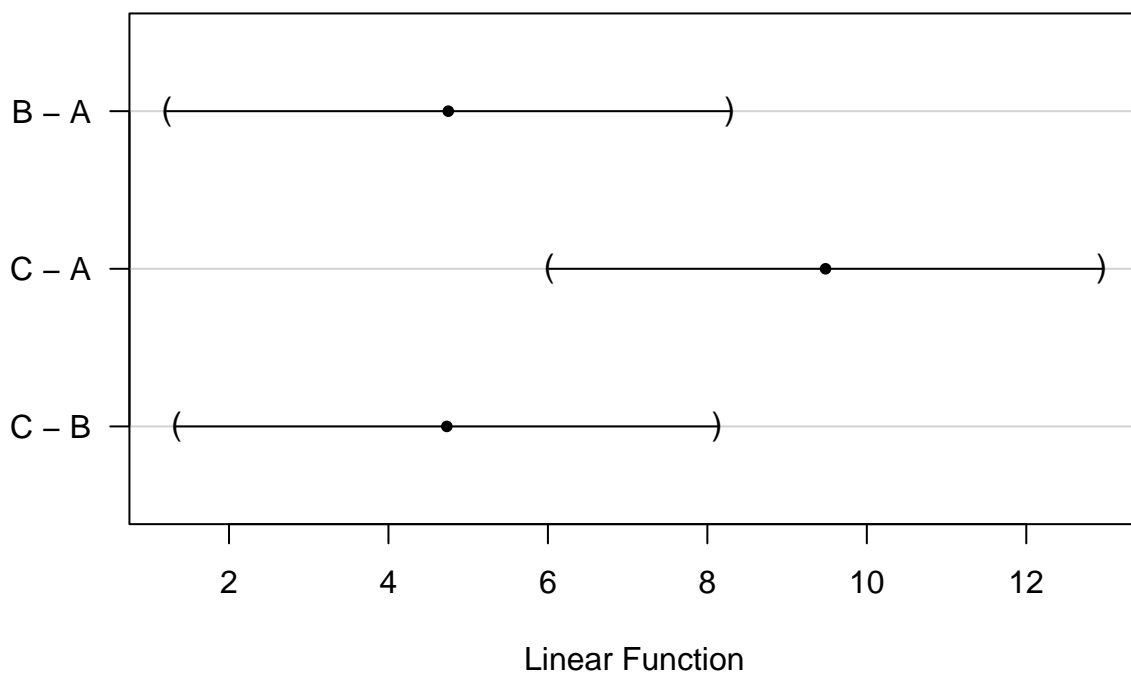
```
##
```

```
##      geyser
```

```
tuk <- glht(anova_result, linfct = mcp(grupo_tratamiento = "Tukey"))
```

```
plot(summary(tuk)) # pairwise tests
```

95% family-wise confidence level



```
kruskal.test(presion_sistolica ~ grupo_tratamiento, data = data)
```

```
##
## Kruskal-Wallis rank sum test
##
## data:  presion_sistolica by grupo_tratamiento
## Kruskal-Wallis chi-squared = 35.431, df = 2, p-value = 2.024e-08
```

Comparaciones múltiples: ¿Qué grupo es diferente?

Si se detectan diferencias globales en el análisis anterior, el equipo desea saber entre qué grupos hay diferencias.

Plantea:

- ¿Qué prueba post-hoc sería adecuada si el test fue no paramétrico?

Reflexiona: ¿Cuál(es) grupo(s) difieren? ¿Qué implicaciones clínicas podría tener?

```
library(dunn.test)
dunn.test(data$presion_sistolica, data$grupo_tratamiento, method="bonferroni")
```

```
## Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 35.4312, df = 2, p-value = 0
##
##
## Comparison of x by group
## (Bonferroni)
## Col Mean-|
## Row Mean |      A      B
## -----+-----
##      B | -2.641198
##      |  0.0124*
##      |
##      C | -5.923117 -3.305956
##      |  0.0000*  0.0014*
##
## alpha = 0.05
## Reject Ho if p <= alpha/2
```