

# Práctica 13. Ejercicios

Jesús Martín Fernández

En primer lugar, vamos a preparar una base de datos de nuestro directorio de trabajo, `df_iam3_r`, que puedes obtener de la Carpeta de la Práctica 12 en el Aula Virtual:

```
#setwd()  
#getwd ()  
  
df_iam3<-read.csv ("df_iam3_r.csv")
```

Recurda que recoge una serie de características de 984 sujetos. De todos se incluyeron características sociodemográficas ( `fecha_nac`, `sex` y `clas_soc` ), clínicas (`hta` , `DM`, `colesterol` , `salud`) y el hábito tabáquico (`fum`). En un momento en el tiempo se recogió qué sujetos habían tenido un evento tipo infarto de miocardio `iam`. Posteriormente se siguió a los sujetos hasta otro punto en el tiempo. Tras ese punto `t`, sólo se recogió si el sujeto seguía fumando tras el primer infarto (`fum_p`) , la cifra de colesterol (`colesterol_p`), la percepción de salud posterior al infarto (`salud_p`) y la ocurrencia de un nuevo reinfarto (`iam2`). Se asume que quien no tuvo `iam1` antes del tiempo `t`, no lo tuvo después.

Vamos a repetir, los análisis realizados en clase (algunos de ellos, excluyendo a los sujetos que pertenecen a la clase social alta. Esto hará cambiar los resultados, que no el código, y tendrás que reinterpretar los resultados

Primero importa la base `df_iam_r.csv` y elimina los sujetos con `clas_soc= alta`

```
df_iam3 <- read.csv("df_iam3_r.csv")  
  
# Eliminar los sujetos con clas_soc = 'Alta'  
  
#(mantengo el nombre, cambiará la "n")  
df_iam3 <- subset(df_iam3, clas_soc != "Alta")  
  
# Verificar el nuevo dataframe filtrado  
str(df_iam3)
```

## 1. Muestras independientes

Compara los niveles de colesterol en aquellos sujetos que tuvieron infarto y los que no.

Lo primero será valorar los supuestos de normalidad y homogeneidad de la varianza en ambos grupos

```
# Filtrar los datos por grupo
col_infarto <- df_iam3$colesterol[df_iam3$iam1 == "Sí"]
col_no_infarto <- df_iam3$colesterol[df_iam3$iam1 == "No"]

# Prueba de normalidad (Shapiro-Wilk) para el grupo con infarto
shapiro.test(col_infarto)

# Prueba de normalidad (Shapiro-Wilk) para el grupo sin infarto
shapiro.test(col_no_infarto)
```

Parece que en ninguno de los dos grupos la distribución es muy ajustada a la normal.

Ahora vamos a testar la hipótesis de igualdad de las varianzas, solo con la prueba F de Snedecor.

```
# Prueba F para comparar las varianzas entre hombres y mujeres
var_test <- var.test(colesterol ~ iam1, data = df_iam3)

# Mostrar resultados
print(var_test)
```

Este resultado no indica diferencias en las varianzas

Vamos a comparar las medias del colesterol en pacientes con y sin infarto con una t de student para datos independientes.

```
t_test_result <- t.test(colesterol ~ iam1, data = df_iam3, var.equal = TRUE)
t_test_result
```

Interpreta las diferencias de medias

Ahora vas a usar la prueba de Mann-Whitney para comparar la distribución del `colesterol_pen` los que tuvieron un segundo infarto y los que no. Primero comprobaremos la asunción de normalidad para esta variable

```

grupo_infarto <- df_iam3$colesterol_p[df_iam3$iam2 == "Sí"]
grupo_no_infarto <- df_iam3$colesterol_p[df_iam3$iam2 == "No"]

shapiro_infarto <- shapiro.test(grupo_infarto)
shapiro_no_infarto <- shapiro.test(grupo_no_infarto)

shapiro_infarto
shapiro_no_infarto

var_test2 <- var.test(colesterol_p ~ iam2, data = df_iam3)

print(var_test2)

```

Vemos que la distribución no es normal , aunque sí son homogéneas las varianzas. Aún así usaremos la prueba de Mann-Whitney

```

MW_test<- wilcox.test(colesterol_p ~ iam2, data = df_iam3)
MW_test

```

La salida de esta función nos dice que las distribuciones son diferentes

## 2. Datos apareados

Vamos a realizar una prueba T de Student para datos apareados comparando el colesterol medido antes y después del punto t (cuando se evidenció si había habido o no un infarto, `colesterol` y `colesterol_p`)

```

#Evaluamos normalidad de las distribución de las diferencias

diferencias <- df_iam3$colesterol - df_iam3$colesterol_p

# Test de normalidad (Shapiro-Wilk)
shapiro.test(diferencias)

```

Las distribuciones no son normales , pero sabemos que cuando la n es suficientemente grande, la t de student admite desviaciones de la normalidad , por lo que nos permitimos utilizar la t de student para datos apareados

```
t.test(df_iam3$colesterol, df_iam3$colesterol_p, paired = TRUE)
```

Nos dice que el valor de la p es muy pequeño ( $1.206e-12$ ), pero además nos ofrece cuál es la diferencia de medias y su IC del 95% (2.71236, IC95%: 1.973906 a 3.450813 mg/dl).

Vamos a utilizar la prueba de Wilcoxon para datos apareados para determinar si las medias del colesterol son iguales cuando se midieron antes y después del punto t

```
wilcox.test(df_iam3$colesterol, df_iam3$colesterol_p, paired = TRUE)
```

Nos dice que las diferencias existen y que es muy improbable que se deba al azar, pero no analiza cómo de grande es esa diferencia.