

Práctica 10. Ejercicios

Jesús Martín Fernández

Selecciona el directorio de trabajo y carga la base de datos `df_iam3`, que puedes obtener de la Carpeta de la Práctica 10 en el Aula Virtual:

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

Recuerda que, en esta base se recogen una serie de características de 984 sujetos. 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`) y la ocurrencia de un nuevo reinfarcto (`iam2`)

Obten la variable edad (en años cumplidos), suponiendo que la fecha final de seguimiento es el 30/6/2024 a partir de la variable `fecha_nac`

```
#fecha_nac viene definida como character, cambiamos a formato fecha  
df_iam3$fecha_nac <- as.Date(df_iam3$fecha_nac)  
fecha_fin <- as.Date("2024-06-30")  
df_iam3$edad <- (fecha_fin-df_iam3$fecha_nac)/365.25  
df_iam3$edad <- as.numeric (round (df_iam3$edad,0))
```

Seguidamente etiquetas, como se hizo en la práctica correspondiente, las variables `sex` (etiquetas: 0="mujer; 1="Varón"), `hta` , `fum` , `DM` , `fum_p` , `iam1` e `iam2` , con las etiquetas (0="No", 1="Sí") y la variable `clas_soc` con las estiquetas (0= "baja"; 1=" Media", 2= Alta")

```
df_iam3$sex <- factor(df_iam3$sex, levels = c(0, 1),  
                      labels = c("Mujer", "Varón"))  
df_iam3$hta <- factor(df_iam3$hta, levels = c(0, 1), labels = c("No", "Sí"))  
df_iam3$fum <- factor(df_iam3$fum, levels = c(0, 1), labels = c("No", "Sí"))
```

```
df_iam3$DM <- factor(df_iam3$DM, levels = c(0, 1), labels = c("No", "Sí"))
df_iam3$fum_p <- factor(df_iam3$fum_p,
                        levels = c(0, 1), labels = c("No", "Sí"))
df_iam3$iam1 <- factor(df_iam3$iam1, levels = c(0, 1), labels = c("No", "Sí"))
df_iam3$iam2 <- factor(df_iam3$iam2, levels = c(0, 1), labels = c("No", "Sí"))
df_iam3$clas_soc <- factor(df_iam3$clas_soc,
                           levels = c(0, 1, 2),
                           labels = c("Baja", "Media", "Alta"))
```

Comprueba que ha funcionado todo bien

```
head (df_iam3)
```

1. Comparación de proporciones muestrales.

Compara las proporciones de varones y diabéticos en pacientes con el primer infarto. Comprueba que todas las frecuencias esperadas sean mayores que 5. Si alguna fuese inferior utiliza la corrección de Yates y el test exacto de Fisher. Interpreta los resultados

```
# Para la variable fum
tabla_sex_iam1 <- table(df_iam3$sex, df_iam3$iam1)

(tabla_sex_iam1)

prop_sex_iam1 <- prop.table(tabla_sex_iam1, margin = 1)

prop_sex_iam1

chi_test_1 <- chisq.test(tabla_sex_iam1, correct = FALSE)

chi_test_1$expected

# Ver el resultado del test
chi_test_1
```

El p-value es muy grande, no nos permite afirmar que las proporciones sean diferente de las esperadas por azar- Como no hay ninguna frecuencia esperada menor que 5, mantenemos el test sin corrección de Yates. Si quisiésemos aplicar esta corrección usaríamos la opción `correct=TRUE`.

Repetimos el proceso para la variable DM

```

# Para la variable DM
tabla_DM_iam1 <- table(df_iam3$DM, df_iam3$iam1)

tabla_DM_iam1

prop_DM_iam1 <- prop.table(tabla_DM_iam1, margin = 1)

prop_DM_iam1

chi_test_2 <- chisq.test(tabla_DM_iam1, correct = FALSE)

chi_test_2$expected

# Ver el resultado del test
chi_test_2

```

En este caso tampoco podemos afirmar que haya diferencias entre ambas proporciones (aunque la p no es muy grande, aceptamos el convenio del punto de corte de 0,05). Como no hay ninguna frecuencia esperada menor que 5, mantenemos el test sin corrección de Yates.

Ahora vamos a aplicar un test exacto de Fisher para ver si hay diferencias en la proporción de hombres que sufren segundo infarto, respecto a los que no lo sufren. Vamos a calcular manualmente la frecuencias esperadas:

```

tabla_sex_iam2 <- table(df_iam3$sex, df_iam3$iam2)

tabla_sex_iam2

fisher_sex_iam2 <- fisher.test(tabla_sex_iam2)
fisher_sex_iam2

```

La hipótesis nula en el test de Fisher es que la Odds Ratio es 1 y no puede descartarse en este caso.

```

# Calcular las frecuencias esperadas manualmente
n <- sum(tabla_sex_iam2) # Total general
a <- tabla_sex_iam2[1, 1] # 0, 0
b <- tabla_sex_iam2[1, 2] # 0, 1
c <- tabla_sex_iam2[2, 1] # 1, 0
d <- tabla_sex_iam2[2, 2] # 1, 1

```

```

# Calcular frecuencias esperadas
E_11 <- (a + b) * (a + c) / n # Frecuencia esperada para 0, 0
E_12 <- (a + b) * (b + d) / n # Frecuencia esperada para 0, 1
E_21 <- (c + d) * (a + c) / n # Frecuencia esperada para 1, 0
E_22 <- (c + d) * (b + d) / n # Frecuencia esperada para 1, 1

# Crear una matriz de frecuencias esperadas
frecuencias_esperadas <- matrix(c(E_11, E_12, E_21, E_22), nrow = 2, byrow = TRUE,
                                dimnames = list(c("0", "1"), c("0", "1")))

# Ver las frecuencias esperadas
print("Frecuencias esperadas:")
print(frecuencias_esperadas)

```

Al observar las frecuencias esperadas observamos que no es necesaria la prueba exacta de Fisher.

2. Datos apareados, test de McNemar

Vamos a repetir el ejercicio de la práctica, comparar la proporción de fumadores inicial (`fum`) con la proporción de fumadores tras el primer infarto (`fum_p`), pero en el subgrupo que ha sufrido un segundo infarto (`iam2`="Sí")

```

# Filtrar los datos donde iam2 == 1
subgrupo_iam2 <- df_iam3[df_iam3$iam2 == "Sí", ]

tabla_fum <- table(subgrupo_iam2$fum, subgrupo_iam2$fum_p)

tabla_fum

mcnemar_test <- mcnemar.test(tabla_fum)

mcnemar_test

```

El p-value tan pequeño hace poco probable que las dos proporciones sean iguales

Como hay un valor de una casilla =0, podría ser apropiada la corrección de Yates, pero para deberíamos calcular los valores esperados (simplemente copia el código)

```

# Se puede construir la tabla manualmente o partior de la anterior tabla_fum
#Probamos de la segunda forma
tabla_fum <- matrix(c(20, 0, 12, 21), nrow = 2, byrow = TRUE,
                    dimnames = list(c("Fumador", "No Fumador"),
                                    c("Fumador", "No Fumador")))

# Ver la tabla
print(tabla_fum)

# Frecuencias observadas
cat("Frecuencias observadas:\n")
print(tabla_fum)

# Calcular frecuencias esperadas manualmente
n <- sum(tabla_fum) # Total general
a <- tabla_fum[1, 1] # Fumador, Fumador
b <- tabla_fum[1, 2] # Fumador, No Fumador
c <- tabla_fum[2, 1] # No Fumador, Fumador
d <- tabla_fum[2, 2] # No Fumador, No Fumador

# Calcular frecuencias esperadas
E_11 <- (a + b) * (a + c) / n
E_12 <- (a + b) * (b + d) / n
E_21 <- (c + d) * (a + c) / n
E_22 <- (c + d) * (b + d) / n

# Mostrar frecuencias esperadas
cat("\nFrecuencias esperadas:\n")
cat("Frecuencia esperada (Fumador, Fumador):", E_11, "\n")
cat("Frecuencia esperada (Fumador, No Fumador):", E_12, "\n")
cat("Frecuencia esperada (No Fumador, Fumador):", E_21, "\n")
cat("Frecuencia esperada (No Fumador, No Fumador):", E_22, "\n")

```

Pues no, no hay ninguna frecuencia esperada menor que 5, así que podíamos quitar la corrección de Yates

```

mcnemar_test <- mcnemar.test(tabla_fum, correct=FALSE)

mcnemar_test

```

El resultado no es muy diferente del anterior

3. Diferencia de proporciones

Vamos a estudiar las diferencias de proporciones de fumadores tras el primer infarto (`fum_p`) en pacientes que sufren el segundo infarto (`iam2`)

```
tabla_fum_p_iam2 <- table(df_iam3$fum_p, df_iam3$iam2)

tabla_fum_p_iam2

dif_prop <- prop.test(tabla_fum_p_iam2)

dif_prop
```

Primero muestra el p-value

p-value = 7.17e-05

Y debajo muestra las proporciones de no fumadores en quienes no han tenido segundo infarto y los que sí

p0=0.9628253 p1=0.8870056

y el IC 95% de p0-p1

0.02394365 0.12769561