

Práctica 14. Ejercicios

Jesús Martín Fernández

1. ANOVA de un factor

En primer lugar, vamos a preparar una base de datos de nuestro directorio de trabajo, `df_iam4r_r`, que puedes obtener de la Carpeta de la Práctica 14 en el Aula Virtual:

```
#setwd()  
#getwd ()  
  
df_iam4<-read.csv ("df_iam4_r.csv")
```

En estos ejercicios vamos a trabajar solo con las variables `clas_soc`, `colesterol`, y `colesterol_p`. Además vamos a seleccionar a los sujetos que tuvieron un infarto (`iam=1`)

```
#setwd()  
#getwd ()  
  
df_iam4r<-df_iam4[df_iam4$iam1=="Sí",]  
  
#Nos aseguramos de que clas_soc es un factor  
df_iam4r$clas_soc<- as.factor(df_iam4r$clas_soc)
```

Deben quedar 125 sujetos

Vamos, de nuevo a comparar, el colesterol en cada grupo de clase social.

Primero evaluaremos en un gráfico las diferencias

```
boxplot(colesterol ~ clas_soc, data = df_iam4r,  
        main = "Colesterol en cada clase Social",  
        xlab = "Clase Social",  
        ylab = "Colesterol (mg/dl)",  
        col = c("red", "blue"), # Colores para las cajas  
        border = "black") # Color del borde
```

Parece que la clase “Alta” tiene valores más bajos de colesterol,, que la “media” y esta que la “Baja”.

Para estudiarlo con un ANOVA, debemos comprobar los supuestos de normalidad y homogeneidad de varianzas

```
shapiro_results <- by(df_iam4r$colesterol, df_iam4r$clas_soc, shapiro.test)

shapiro_results
```

Parece que la variable tiene una distribución normal, o que no se puede rechazar la hipótesis de normalidad

Ahora recurriremos al paquete `car` para hacer la prueba de Levene

```
library(car)
leveneTest(colesterol ~ clas_soc, data = df_iam4r)
```

Parece que las varianzas son homogéneas

Haremos también un análisis de residuales, pero primero hay que generar el modelo

```
test_anova<- aov(colesterol~clas_soc,data=df_iam4r)

qqnorm(residuals(test_anova))
qqline(residuals(test_anova))
```

Parece haber buen cumplimiento de los criterios de normalidad por lo que procedemos a hacer y analizar el ANOVA

```
# Ya teníamos el modelo ANOVA (test_anova)
anova_results <- anova(test_anova)

anova_results

# Extraer información específica
grados_libertad <- anova_results$Df
suma_cuadrados <- anova_results$`Sum Sq`
media_cuadrados <- anova_results$`Mean Sq`
estadistico_F <- anova_results$`F value`
p_valor <- anova_results$`Pr(>F)`
```

```
# Imprimir resultados específicos usando paste()
print(paste("Grados de libertad:", grados_libertad))
print(paste("Suma de cuadrados:", suma_cuadrados))
print(paste("Media de cuadrados:", media_cuadrados))
print(paste("Estadístico F:", estadistico_F))
print(paste("P-valor:", p_valor))
```

Los resultados del análisis de varianza (ANOVA) solo nos dan un p-value al borde de la significación estadística

2. Análisis post-hoc

Como las varianzas eran homogéneas usaremos las pruebas de Bonferroni y Tukey, y para tener sus intervalos de confianza usaremos el paquete **emmeans**

```
#install.packages("emmeans")#solo si no está instalado previamente
library(emmeans)
emm <- emmeans(test_anova, ~ clas_soc)

results_bonferroni<-pairs(emm, adjust = "bonferroni", infer = TRUE)

results_tukey <- pairs(emm, adjust = "tukey", infer = TRUE)
results_bonferroni
results_tukey
```

Interpreta estos resultados