

# Lección 6: Modelado de relaciones

## Bioestadística con R

### Ejercicios

**Ejercicio 1:** La beta endorfina humana (HBE) es una hormona secretada por la glándula pituitaria bajo condiciones de estrés. Un fisiólogo mide la concentración de HBE en estado de reposo de tres grupos de personas: 15 que acaban de inscribirse en un programa de ejercicio, 11 que corre de manera regular y 10 personas sedentarias. Los niveles de HBE (pg/ml) se muestran en la siguiente tabla.

Table 1: Concentración de HBE (pg/ml)

	Programa de ejercicio	Corredores	Sedentarios
Media	38.7	35.7	42.5
SD	16.1	13.4	12.8
n	15.0	11.0	10.0

A partir de esta información:

- Escribir una hipótesis nula dado el contexto de los datos.
- Representar la hipótesis nula como una fórmula matemática.
- Realizar un ANOVA y probar la hipótesis nula con un  $\alpha = 0.05$ .

**Ejercicio 2:** El volumen espiratorio forzado (FEV) es una medida del ritmo de flujo de aire (L/min) durante una exhalación profunda. A 75 estudiantes se les midió el FEV tres veces y esos resultados fueron promediados y mostrados en la siguiente tabla. Algunos de los estudiantes pertenecían al club de fútbol de la Universidad.

Table 2: Media de FEV en L/min

	Mujer, no fútbol	Mujer, fútbol	Hombre, no fútbol	Hombre, fútbol
Media	396	472	560	653
SD	14	17	20	24

- Realizar un ANOVA de dos vías de estos datos.
- *Actividad extra:* Realizar un gráfico de interacciones con `ggplot2`.

### Soluciones

**Ejercicio 1:** Lo primero sería establecer nuestra hipótesis nula, que podría quedar de la siguiente manera  $H_0$ : Los valores medios de los niveles de la concentración de HBE de los tres grupos de personas es igual.

Para representar esto de manera matemática, lo que decimos es que las medias de los tres grupos son iguales, por ende,  $H_0 : \mu_{programa} = \mu_{corredor} = \mu_{sedentario}$ . De igual manera, la hipótesis alternativa sería que realmente sí existen diferencias en al menos uno de los grupos, por ende,  $H_A : \mu_{programa} \neq \mu_{corredor} \neq \mu_{sedentario}$ .

Con esto tenemos los primeros dos puntos resueltos, ahora tenemos que realizar el ANOVA. Primero, como se nos dieron valores medios y desviaciones estándares, debemos generar números aleatorios que correspondan a dichos valores. Para eso utilizamos la función `rnorm()` con los argumentos `mean`, `sd` y `n`.

```
set.seed(2023) # Creamos una semilla para trabajar siempre con los mismos números aleatorios.
programa <- rnorm(n = 15, mean = 38.7, sd = 16.1)
corredor <- rnorm(n = 11, mean = 35.7, sd = 13.4)
sedentario <- rnorm(n = 10, mean = 42.5, sd = 12.8)
ej1 <- data.frame(HBE = c(programa, corredor, sedentario),
                  Grupo = factor(rep(c("Programa", "Corredor", "Sedentario"),
                                     times = c(length(programa),
                                               length(corredor),
                                               length(sedentario))))))
```

Ahora sí, podemos calcular nuestro ANOVA con la función `aov()`.

```
ANOVA.ej1 <- aov(HBE ~ Grupo, ej1)
ANOVA.ej1
```

```
## Call:
##   aov(formula = HBE ~ Grupo, data = ej1)
##
## Terms:
##              Grupo Residuals
## Sum of Squares   533.290 5911.419
## Deg. of Freedom      2      33
##
## Residual standard error: 13.38409
## Estimated effects may be unbalanced
```

Para ver un resumen de nuestros datos utilizamos la función `summary()`.

```
summary(ANOVA.ej1)

##              Df Sum Sq Mean Sq F value Pr(>F)
## Grupo         2    533   266.6    1.489   0.24
## Residuals    33   5911   179.1
```

**Ejercicio 2:** En este caso, lo primero es generar los datos.

```
mujer.nof <- rnorm(n = 14, mean = 396)
mujer.f <- rnorm(n = 17, mean = 472)
hombre.nof <- rnorm(n = 20, mean = 560)
hombre.f <- rnorm(n = 24, mean = 653)
ej2 <- data.frame(FEV = c(mujer.nof, mujer.f, hombre.nof, hombre.f),
                  Sexo = factor(c(rep("Mujer", length(mujer.nof)),
                                   rep("Mujer", length(mujer.f)),
                                   rep("Mujer", length(mujer.f)),
                                   rep("Mujer", length(mujer.f)))))
```

```

      rep("Hombre", length(hombre.nof)),
      rep(("Hombre"), length(hombre.f))),
  Condicion = factor(c(rep("No futbol", length(mujer.nof)),
      rep(("Futbol"), length(mujer.f)),
      rep(("No futbol"), length(hombre.nof)),
      rep(("Futbol"), length(hombre.f))))))

```

Aunque el código de arriba se ve terrorífico, en realidad son cosas muy sencillas, ¡Pero cuidado con los paréntesis! Ahora podemos realizar nuestro ANOVA de dos vías.

```

ANOVA.ej2 <- aov(FEV ~ Sexo + Condicion, ej2)
ANOVA.ej2

```

```

## Call:
## aov(formula = FEV ~ Sexo + Condicion, data = ej2)
##
## Terms:
##              Sexo Condicion Residuals
## Sum of Squares 542136.1 136411.7 1426.9
## Deg. of Freedom      1         1      72
##
## Residual standard error: 4.451783
## Estimated effects may be unbalanced

```

Utilizamos la función `summary()` para ver el resumen de nuestros resultados y nuestra tabla de ANOVA.

```
summary(ANOVA.ej2)
```

```

##              Df Sum Sq Mean Sq F value Pr(>F)
## Sexo          1 542136  542136    27355 <2e-16 ***
## Condicion     1 136412  136412     6883 <2e-16 ***
## Residuals    72   1427      20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Actividad extra:** Para el gráfico de interacción del ejercicio 2 vamos a necesitar de la librería de `ggplot2` y `dplyr`.

```

library(ggplot2)
library(dplyr)

```

Ahora obtenemos la media de cada uno de nuestros grupos, por sexo y por condición.

```
ej2.grafico <- ej2 %>% group_by(Sexo, Condicion) %>% summarise(media = mean(FEV))
```

```

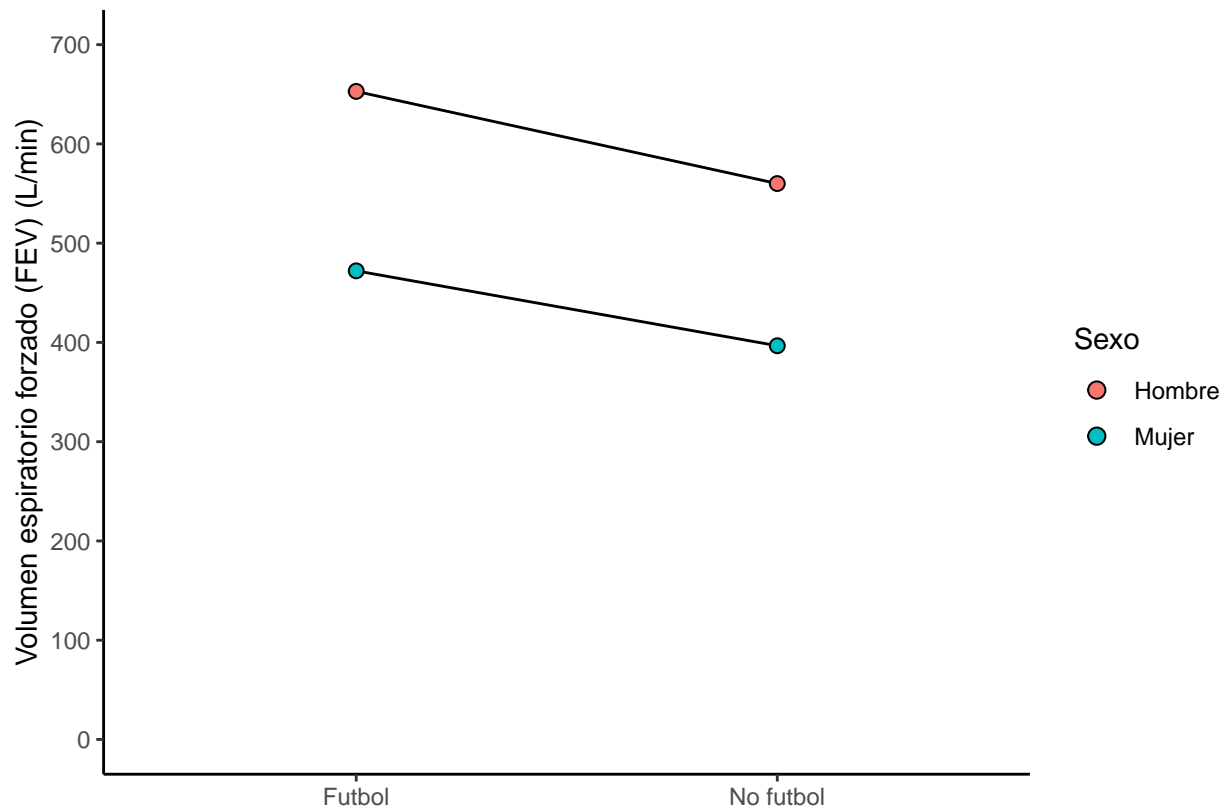
## `summarise()` has grouped output by 'Sexo'. You can override using the
## `.groups` argument.

```

```
ej2.grafico$Condicion <- factor(ej2.grafico$Condicion, levels = c("Futbol", "No futbol"))
```

Ahora podemos utilizar `ggplot2` para nuestro gráfico. Utilizaremos dos geometrías, `geom_line` y `geom_dotplot`.

```
ggplot(ej2.grafico, aes(x = Condicion, y = media, fill = Sexo)) +  
  geom_line(aes(group = Sexo)) +  
  geom_dotplot(binaxis = "y", stackdir = "center", binwidth = 15) +  
  scale_y_continuous(limits = c(0,700), breaks = seq(0, 700, by = 100)) +  
  labs(x = "", y = "Volumen espiratorio forzado (FEV) (L/min)") +  
  theme_classic()
```



En este caso vemos que el valor del FEV es independiente del sexo, pues en el gráfico de interacción estos valores no se cruzan.