Examen 2 Ejercicio 5

Armando Jessica Luz

Para empezar, importamos nuestra base a R Studio

```
base <- read.csv("./Ex5.csv", header = TRUE)
str(base) # Nuestra variable Trat es char
```

```
## 'data.frame':    300 obs. of  3 variables:
##  $ Ant : num  19.1 14.1 19.4 14.7 12.3 ...
##  $ Trat: chr  "Med" "Med" "Med" "Med" ...
##  $ Edad: int  27 39 20 56 53 35 40 49 40 25 ...
```

```
base$Trat <- factor(base$Trat) # Pasamos a factor la variable Trat
base <- base %>% rename(y = Ant,
                        x1 = Edad) %>%
                   mutate(xA = if_else(Trat == "Med",1,0),
                          xB = if_else(Trat == "Control",1,0))
levels(base$Trat) # Ya considera niveles en nuestra base
```

```
## [1] "Control" "Med"
```

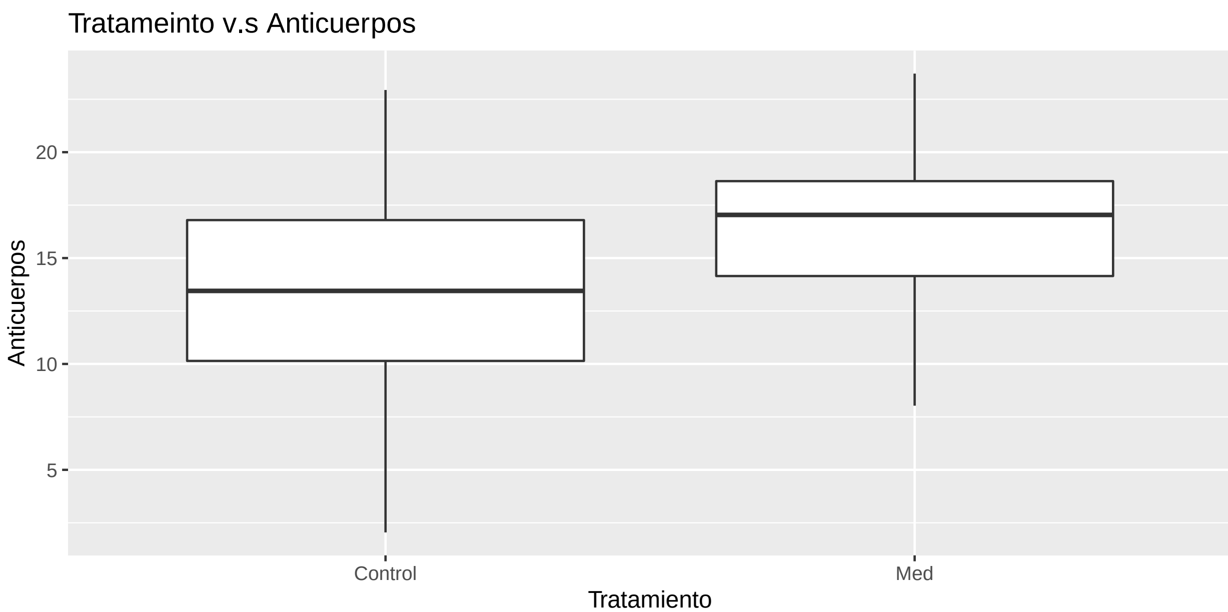
Nos encontramos con dos níveles en la variable categórica

i. Realice un análisis descriptivo

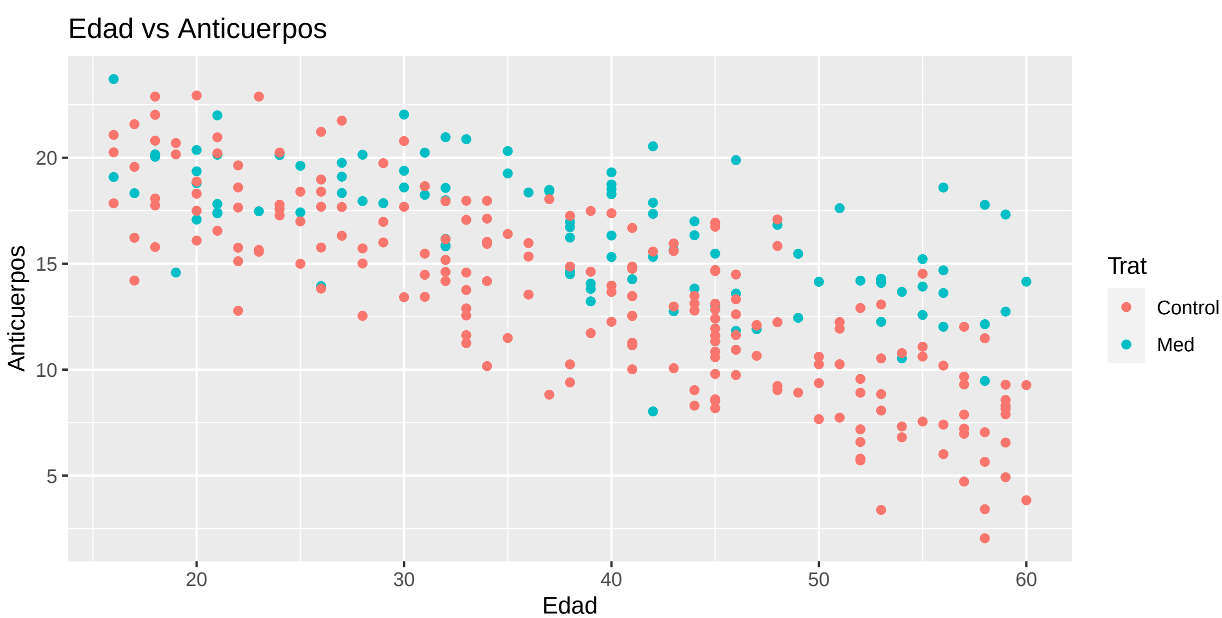
Cuadro 1: Principales estadísticas de la Efectividad por Tratamiento

Trat	Observaciones	Media	Mediana	Varianza
Control	200	13	13	19.7
Med	100	17	17	9.2

- La cantidad de observaciones es el doble en la categoría de personas que reibieron placebo(n=200) en contra de las que recibieron medicamento (n=100)
- En promedio la cantidad de anticuerpos en personas medicadas es de 16.54 mientras que en las no medicadas es de 13.40
- Hay mayor variabilidad en pacientes que que no recibieron el medicamento(19.71), casi el doble de los que sí lo hicieron(9.18)



Observamos en nuestras boxplots que la media de la categoría Med está por encima de la categoría Control, en otras palabras no está diciendo que es más probable encontrar una cantidad mayor de anticuerpos en una persona que sí recibió el medicamento



Lo que observamos en cambio con nuestra variable Edad es que al parecer, mientras mayor sea nuestro paciente, menor es la cantidad de anticuerpos presentes. Aunque también nos damos cuenta de que esto se refleja más aún en pacientes que recibieron placebo Por tanto, esta gráfica indica que la edad influye en el estudio de nuestros pacientes

Descriptor de varaibles:

- y_i = Medida de efectividad del tratamiento para el individuo i.
- x_{i1} = Edad del individuo i.
- $x_{i\mathrm{A}} = \begin{cases} 1, & \text{Si el individuo } i \text{ recibió el Medicamento} \\ 0, & \text{e.o.c} \end{cases}$

- $x_{i\text{B}} = \begin{cases} 1, & \text{Si el individuo } i \text{ recibió el Placebo} \\ 0, & \text{e.o.c} \end{cases}$

ii. Ajuste un modelo adecuado

Ajuste un modelo adecuado para evaluar la efectividad del medicamento ajustando por la edad de los pacientes. Es decir, un modelo que incluya como explicativas las variables edad, la binaria asociada a la administración del medicamento y la interacción obtenida como el producto de estas dos.

```r
fit <- lm(y ~ x1 + xB + x1:xB, data = base) # Ajustamos nuestro modelo a los datos
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x1 + xB + x1:xB, data = base)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.935 -1.689 -0.013  1.623  6.261
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.5452     0.8073   27.93  < 2e-16 ***
## x1           -0.1567     0.0201   -7.81  1.0e-13 ***
## xB            2.0300     0.9834    2.06     0.04 *
## x1:xB        -0.1297     0.0243   -5.34  1.8e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.5 on 296 degrees of freedom
## Multiple R-squared:  0.672,  Adjusted R-squared:  0.669
## F-statistic:  203 on 3 and 296 DF,  p-value: <2e-16
```

El modelo queda parametrizado de la siguiente manera:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{iB} + \beta_3 x_{i1} x_{iB}$$

Nuestro modelo quedó parametrizado considerando la categoría Med(xA) como nivel de referencia

$$\hat{E}(y_i; Tratamiento; Edad) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{iB} + \hat{\beta}_3 x_{i1} x_{iB}$$

Donde:

- y_i = Medida de efectividad del tratamiento para el individuo i.
- x_{i1} = Edad del individuo i.
- $x_{i\text{B}} = \begin{cases} 1, & \text{Si el individuo } i \text{ recibió el placebo} \\ 0, & \text{e.o.c} \end{cases}$

Las expresiones quedan de la siguiente manera:

- Si el individuo recibió el medicamento, entonces $x_{iB} = 0$

$$\begin{aligned} \Rightarrow \hat{E}(y_i; Med; x_{1i}) &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{iB} + \hat{\beta}_3 x_{i1} x_{iB} \\ &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 * 0 + \hat{\beta}_3 x_{i1} * 0 \\ &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} \end{aligned}$$

- Si el individuo recibió el placebo, entonces $x_{1B} = 1$

$$\begin{aligned} \Rightarrow \hat{E}(y_i; Control; x_{1i}) &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{iB} + \hat{\beta}_3 x_{i1} x_{iB} \\ &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 * 1 + \hat{\beta}_3 x_{i1} * 1 \\ &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 + \hat{\beta}_3 x_{i1} \end{aligned}$$

Observemos lo que sucede con nuestros parámetros:

```
## (Intercept)          x1          xB       x1:xB
##       22.55       -0.16        2.03       -0.13
```

Las estimaciones de los parámetros son:

$$\hat{\beta}_0 = 22.55,\ \hat{\beta}_1 = -0.16,\ \hat{\beta}_2 = 2.03,\ \hat{\beta}_3 = -0.13$$

Las expresiones del valor esperado de la efectividad para cada tratamiento son:

- Caso Med

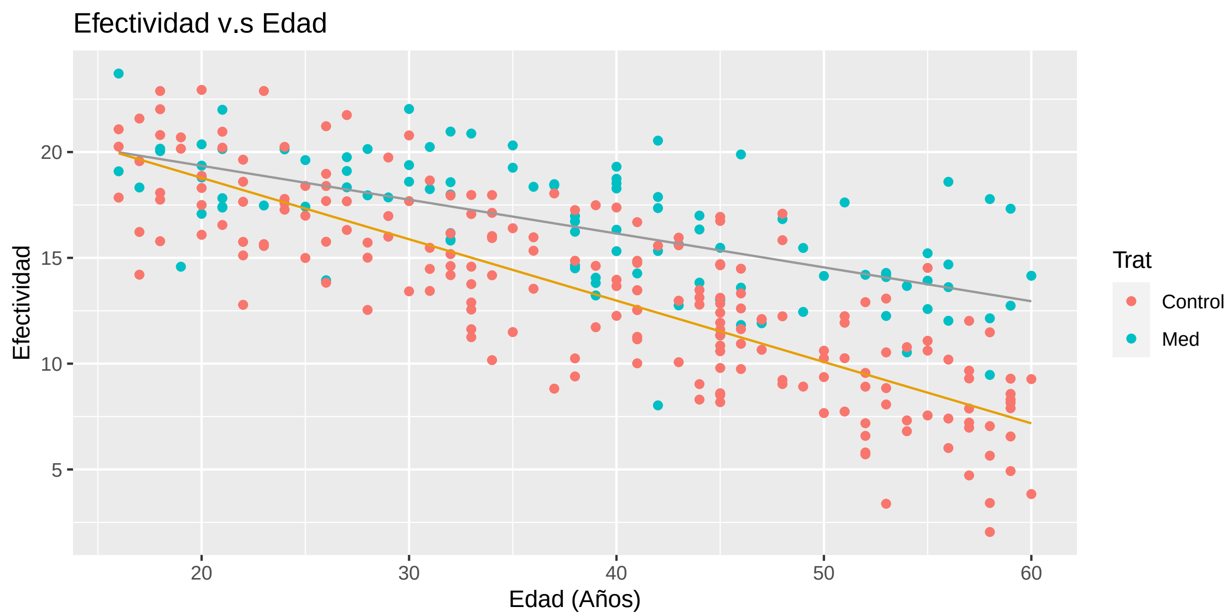
$$\begin{aligned} \Rightarrow \hat{E}(y_i; Med; x_{1i}) &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} \\ &= 22.55 - 0.16 * x_{i1} \end{aligned}$$

- Caso Control

$$\begin{aligned} \Rightarrow \hat{E}(y_i; Control; x_{1i}) &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 + \hat{\beta}_3 x_{i1} \\ &= 22.55 - 0.16 * x_{i1} + 2.03 - 0.13 x_{i1} \\ &= (22.55 + 2.03) + (-0.16 - 0.13) * x_{i1} \\ &= 24.58 - 0.29 * x_{i1} \end{aligned}$$

Con base en esto graficaremos las rectas asociadas

$$\therefore \widehat{\mathrm{E}}(y_i; \text{Tratamiento}, x_{i1} = x_1^*) = \begin{cases} 22.55 - 0.16 x_1^*, & \text{Si el individuo } i \text{ recibió el Medicamento} \\ \\ 24.58 - 0.29 x_1^*, & \text{Si el individuo } i \text{ recibió el Placebo} \end{cases}$$



iii. Indique las expresiones asociadas a la relación

De acuerdo con el modelo ajustado, indique las expresiones asociadas a la relación de la generación promedio de anticuerpos con la edad en a) el grupo control y b) en el grupo que recibe el medicamento.

$$H_0 : \hat{\beta}_1 = 0 \quad \text{y} \quad \hat{\beta}_2 = 0 \quad \text{y} \quad \hat{\beta}_3 = 0 \quad \text{vs} \quad H_\alpha : \hat{\beta}_1 \neq 0 \quad \text{o} \quad \hat{\beta}_2 \neq 0 \quad \text{o} \quad \hat{\beta}_3 \neq 0$$

```
matz <- matrix(c(0,1,0,0,
                0,0,1,0,
                0,0,0,1), ncol=4, nrow=3, byrow=TRUE)
z <- c(0,0,0)
summary(glht(fit, linfct=matz, rhs=z),test=Ftest())
```

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##        Estimate
## 1 == 0   -0.157
## 2 == 0    2.030
## 3 == 0   -0.130
##
## Global Test:
##     F DF1 DF2   Pr(>F)
## 1 203   3 296 2.12e-71
```

Veamos que con la prueba asociada a la tabla ANOVA, concluimos con una confianza de .95, se rechaza H0 en el contraste ay que el p-value < 0.05

iv.

¿Se puede decir que la edad afecta de la misma forma la generación de anticuerpos en el grupo control que en el grupo que recibe el medicamento? Realice una prueba de hipótesis apropiada e interprete.

Planteamos la prueba de Hipótesis ante la duda si nuestras rectas tienen la misma pendiente:

$$H_0 : \beta_3 = 0 \quad \text{v.s} \quad H_\alpha : \beta_3 \neq 0$$

$$H_0 : \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix} \quad \text{v.s} \quad H_a : \beta_3 \neq 0$$

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##        Estimate
## 1 == 0    -0.13
##
## Global Test:
##      F DF1 DF2   Pr(>F)
## 1 28.5   1 296 1.84e-07
```

Realizando la prueba relacionada a nuestras pruebas de hipótesis, con una significancia del 5 %y un p-value = 1.8e-07 < 0.05, rechazando H_0, podemos concluir que el efecto de la variable Edad(x1) es diferente para las personas que recibieron el medicamento y las que no lo hicieron. Así que ambas pendientes son distintas

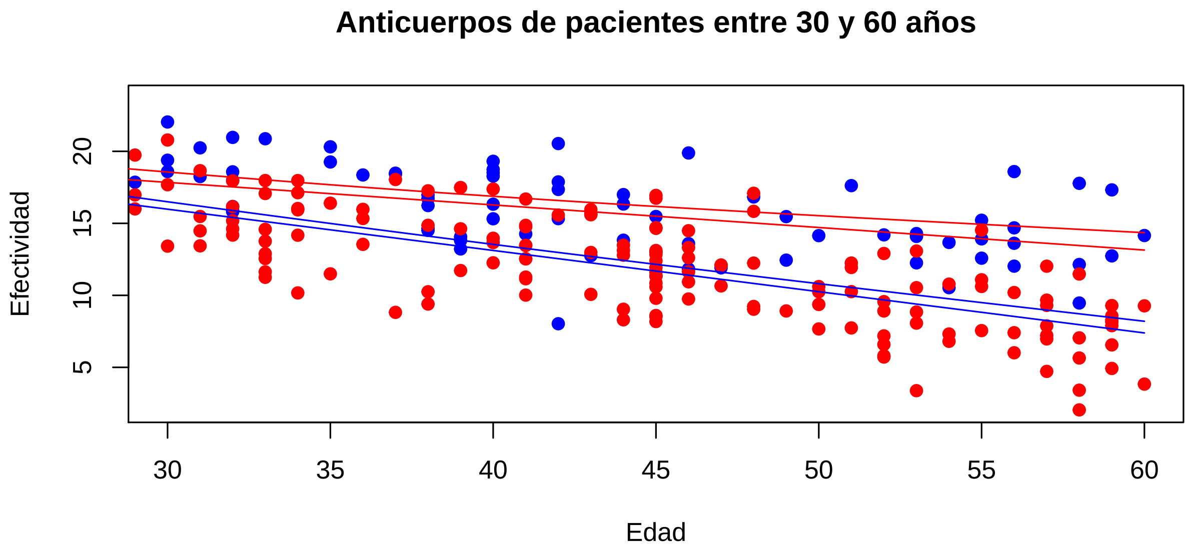
V.

Comente sobre el ajuste del modelo incluyendo la interpretación de cada uno de los coeficientes.

- El ajuste del modelo parece ser el indicado puesto que gracias a nuestras gráficas es fácil ver que los datos de ambas categorías no siguen el mismo camino.
- De igual forma nos apoyamos en el hecho de que en nuestro ajuste fit, analizando de manera individual el p-value de cada una de nuestras variables, con una confianza del 95 % no es necesario eliminar ninguna de estas variables pues estas sí nos agregan información a nuestro modelo.

VI.

Argumente en contra o a favor de la afirmación: "El medicamente funciona aumentando el número de anticuerpos para todos los pacientes entre 30 y 60 años". Se puede apoyar de pruebas de hipótesis o intervalos de confianza simultáneos.



Observamos que en todas las edades entre 30 a 60 años nuestros intervalos en un caso quedan por encima del otro en cuanto a los tratamientos. Los intervalos de las personas que recibieron el medicamento están por encima de los que no lo recibieron. Por tanto, concluimos que el medicamento aumenta el número de anticuerpos en personas entre los 30 y 60 años de edad.