Ejercicio 4A

Armando Jessica Luz

Para empezar, importamos nuestra base a R Studio

```
base <- read.csv("./Ex4A.csv")
str(base) # Ambas variables de tipo char
```

```
## 'data.frame':    120 obs. of  3 variables:
##  $ Sexo   : chr  "Hombre" "Hombre" "Hombre" "Hombre" ...
##  $ Trat   : chr  "Control" "Control" "Control" "Control" ...
##  $ Puntaje: num  11.17 11.42 9.78 9.09 11.21 ...
```

```
base$Sexo <- factor(base$Sexo) # Transformamos a factor variable Sexo
base$Trat <- factor(base$Trat) #  Transformamos a factor variable Trat
str(base)
```

```
## 'data.frame':    120 obs. of  3 variables:
##  $ Sexo   : Factor w/ 2 levels "Hombre","Mujer": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Trat   : Factor w/ 3 levels "Control","Trat1",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Puntaje: num  11.17 11.42 9.78 9.09 11.21 ...
```

```
levels(base$Sexo)
```

```
## [1] "Hombre" "Mujer"
```

```
levels(base$Trat)
```

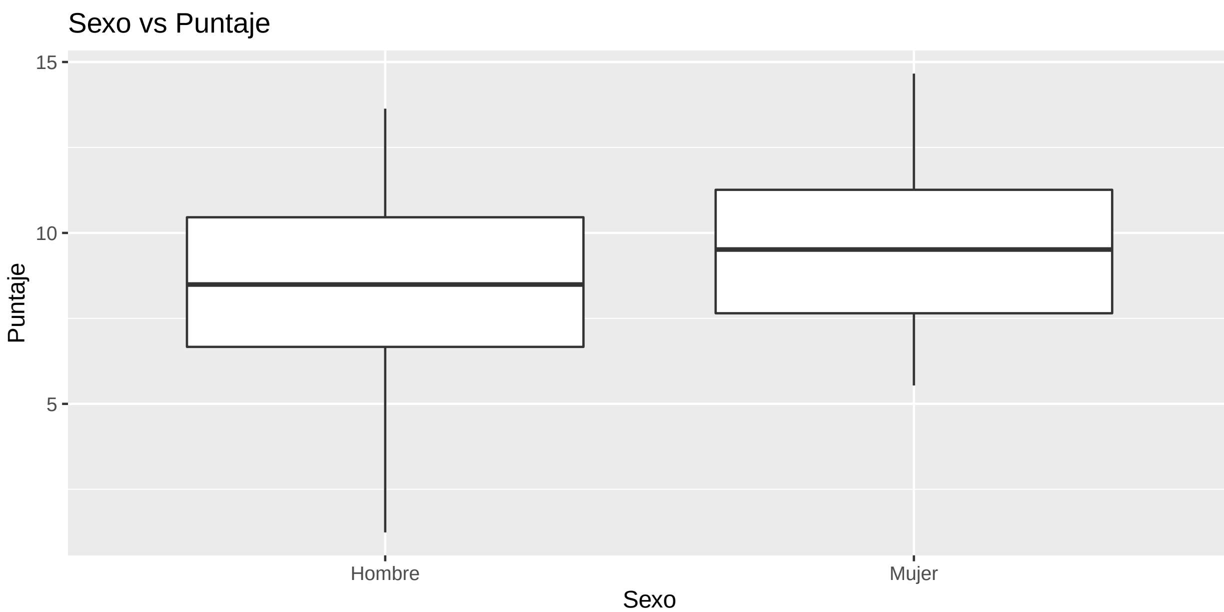
```
## [1] "Control" "Trat1"   "Trat2"
```

Obtenemos dos valores en la variable Sexo y tres en la vraible Trat # i. Realice un análisis descriptivo

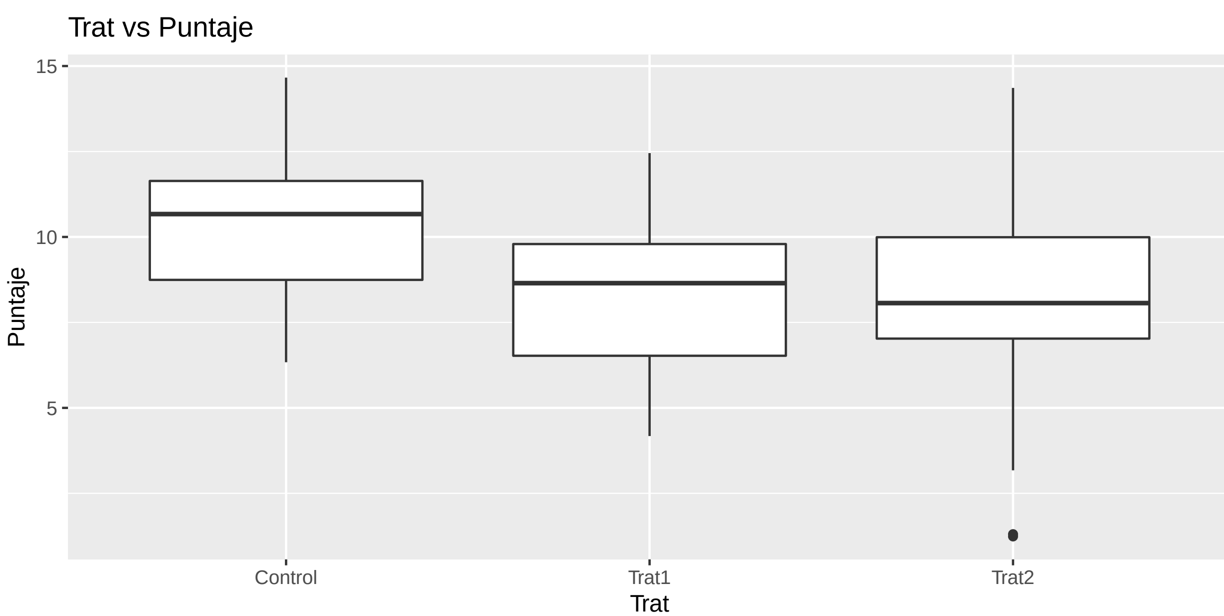
Cuadro 1: Principales estadísticas de la Efectividad por Tratamiento

Trat	Sexo	Observaciones	Media	Mediana	Varianza
Control	Hombre	20	10.2	10.2	2.8
Control	Mujer	20	10.5	11.0	5.3
Trat1	Hombre	20	8.3	8.6	5.8
Trat1	Mujer	20	8.3	8.7	4.8
Trat2	Hombre	20	6.2	6.9	6.7
Trat2	Mujer	20	10.1	9.9	3.9

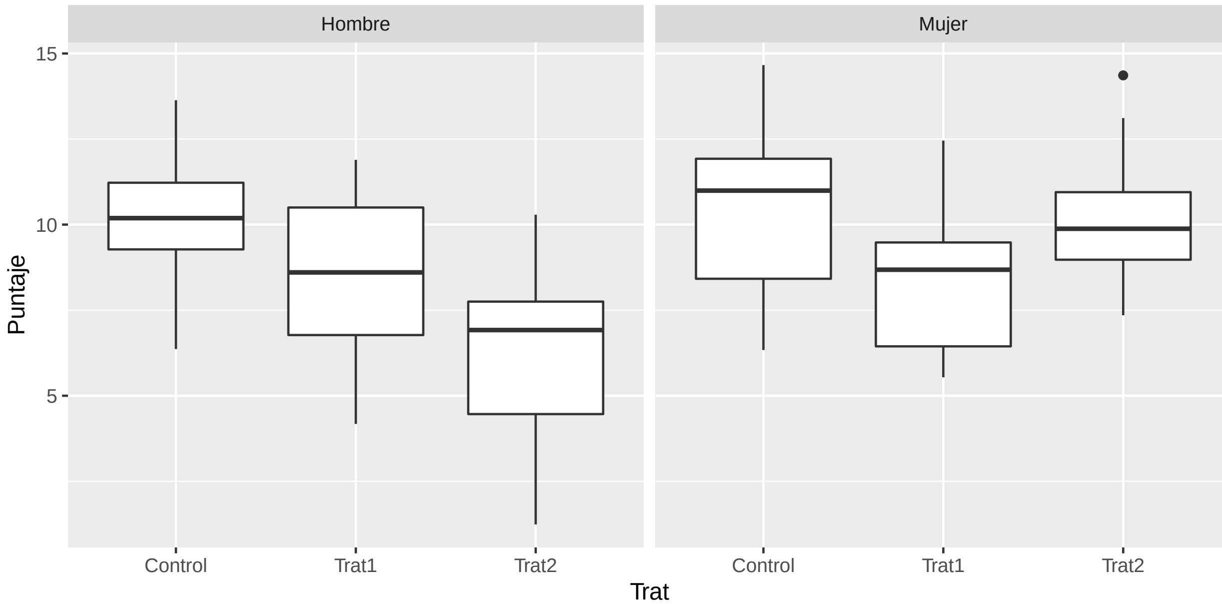
Observamos gracias a esta tabla que la media de puntaje de las mujeres es mayor que la de los hombres con los tres tratamientos, aunque en el Trat2 es donde la diferencia es mayor. Lo mismo se observa en el caso de la mediana



Mediana de las mujeres por encima de los hombres, es más probable encontrar mayor puntaje de ansiedad en una mujer que en un hombre



En cuanto a los tratameintos, la gráfica indica que el emjro tratamiento sería el Tratamiento 2 ya que las medias tanto de los que recibieron el Tratamiento 1 como de los que no recibieron ninguno está por encima de esta, e incluso recalcar que la media de los que no recibieron nada está por encima de ambos



Ahora, esta gráfica separando la eficacia de los tratamientos por sexo nos deja ver que en los hombres como en mujeres se percibe un mayor puntaje de ansiedad en los casos que no reciben medicamento. A la vez podemos observar que en los hombres tiene mayor eficacia el Trat2 mientras que en las muejeres funciona mejor el Trat1

ii.

Considerando un modelo de regresión que incluye las dos variables categóricas de forma individual y también su interacción, dé la expresión del puntaje promedio para cada valor de las variables categóricas.

```r
ajuste <- lm(Puntaje ~ Trat * Sexo, data = base)
summary(ajuste) # Toma las categorías Control y Hombre de referncia
```

```
##
## Call:
## lm(formula = Puntaje ~ Trat * Sexo, data = base)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.971 -1.642  0.346  1.339  4.258
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           10.153      0.495   20.53  < 2e-16 ***
## TratTrat1             -1.845      0.699   -2.64  0.00950 **
## TratTrat2             -3.943      0.699   -5.64  1.3e-07 ***
## SexoMujer              0.311      0.699    0.44  0.65720
## TratTrat1:SexoMujer   -0.302      0.989   -0.31  0.76046
## TratTrat2:SexoMujer    3.580      0.989    3.62  0.00044 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.2 on 114 degrees of freedom
## Multiple R-squared:  0.324,  Adjusted R-squared:  0.295
```

```
## F-statistic: 10.9 on 5 and 114 DF,  p-value: 1.29e-08
```

$$\hat{E}(y; Trat, Sexo) = \hat{\beta}_0 + \hat{\beta}_1 Trat1 + \hat{\beta}_2 Trat2 + \hat{\beta}_3 Mujer + \hat{\beta}_4 (Trat1 * Mujer) + \hat{\beta}_5 (Trat2 * Mujer)$$

- E(y;Control, Hombre) $= \hat{\beta}_0$
- E(y;Control, Mujer) $= \hat{\beta}_0 + \hat{\beta}_3$
- E(y;Trat1, Hombre) $= \hat{\beta}_0 + \hat{\beta}_1$
- E(y;Trat1, Mujer) $= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_4$
- E(y;Trat2, Hombre) $= \hat{\beta}_0 + \hat{\beta}_2$
- E(y;Trat2, Mujer) $= \hat{\beta}_0 + +\hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_5$

Estimadores puntuales:

- E(y;Control, Hombre) $= 10.15$
- E(y;Control, Mujer) $= 10.15 + 0.31 = 10.46$
- E(y;Trat1, Hombre) $= 10.15 - 1.84 = 8.31$
- E(y;Trat1, Mujer) $= 10.15 - 1.84 + 0.31 - 0.3 = 8.32$
- E(y;Trat2, Hombre) $= 10.15 - 3.94 = 6.21$
- E(y;Trat2, Mujer) $= 10.15 - 3.94 + 0.31 + 3.58 = 10.1$

iii.

Escriba las hipótesis que se contrastan con la tabla ANOVA, calcule ésta e interprete. Use alpha = 0.05

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_\alpha : \beta_1 \neq 0 \quad \text{o} \quad \beta_2 \neq 0 \quad ... \quad \beta_5 \neq 0$$

```
mat <- matrix(c(0,1,0,0,0,0,
           0,0,1,0,0,0,
           0,0,0,1,0,0,
           0,0,0,0,1,0,
           0,0,0,0,0,1), ncol=6, nrow=5, byrow=TRUE)
res <- c(0,0,0,0,0)
summary(glht(ajuste, linfct=mat, rhs=res),test=Ftest())
```

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##        Estimate
## 1 == 0   -1.845
## 2 == 0   -3.943
## 3 == 0    0.311
## 4 == 0   -0.302
## 5 == 0    3.580
##
## Global Test:
##      F DF1 DF2   Pr(>F)
## 1 10.9   5 114 1.29e-08
```

Con un p-value = 1.29e-08 < 0.05. Con una confianza del 95 % rechazamos H_0, entonces el modelo tiene sentido, por tanto nos ayuda a modelar E(Y)

iv.

¿Se puede considerar que el sexo tiene un efecto en el puntaje, es decir, al menos para un tratamiento existe un efecto diferenciado en el puntaje derivado del sexo de los individos? Use una prueba F con = .025. Interprete.

- E(y;Control,Hombre) = E(y;Control,Mujer)
- E(y;Trat1,Hombre) = E(y;Trat1,Mujer)
- E(y;Trat2,Hombre) = E(y;Trat2,Mujer)
- E(y;Control,Hombre) = E(y;Control,Mujer) = $\beta_3 = 0$
- E(y;Trat1,Hombre) = E(y;Trat1,Mujer) = $\beta_3 + \beta_4 = 0$
- E(y;Trat2,Hombre) = E(y;Trat2,Mujer) = $\beta_3 + \beta_5 = 0$

$$H_0 : \hat{\beta}_3 = 0, \quad \hat{\beta}_3 + \hat{\beta}_4 = 0, \quad \hat{\beta}_3 + \hat{\beta}_5 = 0 \quad \text{vs} \quad H_\alpha : \hat{\beta}_3 \neq 0, \quad \hat{\beta}_3 + \hat{\beta}_4 \neq 0, \quad \hat{\beta}_3 + \hat{\beta}_5 \neq 0$$

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##        Estimate
## 1 == 0  0.31122
## 2 == 0  0.00892
## 3 == 0  3.89126
##
## Global Test:
##      F DF1 DF2   Pr(>F)
## 1 10.4   3 114 4.29e-06
```

Con un p-value = 4.29e-06 < 0.05 y una confianza del 95 % se rechaza H0, al menos un nivel de tratamiento muestra un diferencia entre hombres y mujeres.

Considere una prueba simultánea que ayude a identificar para qué tratamiento se puede considerar que el sexo tiene un efecto, con los resultados de esa prueba reduzca el modelo si es posible.

```
K2 <- matrix(c(0,0,0,1,0,0,
              0,0,0,1,1,0,
              0,0,0,1,0,1), ncol=6, nrow=3, byrow=TRUE)
m2 <- c(0,0,0)
summary(glht(ajuste, linfct=K2, rhs=m2))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ Trat * Sexo, data = base)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0  0.31122    0.69946    0.44     0.96
## 2 == 0  0.00892    0.69946    0.01     1.00
## 3 == 0  3.89126    0.69946    5.56   <1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Podemos observar que se rechaza b3 + b5 = 0 con un p-value = 1e-05 < 0.05 Por otro lado asumimos con p-value > 0.05 en ambos casos que b3 = 0 y b3 + b4 = 0, la segunda prueba es redundante una vez que tenemos b3 = 0 por tanto se traduce a que b4 = 0

Lo anterior indica que las variables relacionadas a ambas betas no suman a nuestro modelo

Ajustamos nuevo modelo

$$E(y; Trat, Sexo) = \hat{\beta}_0 + \hat{\beta}_1 Trat1 + \hat{\beta}_2 Trat2 + \hat{\beta}_3 (Trat2 * Mujer)$$

v.

En caso de que en el inciso anterior se haya reducido el modelo, ajuste de nuevo la regresión y dé la expresión del puntaje promedio para cada valor en las variables categóricas.

```
ajuste2 <- lm(Puntaje ~ Trat + I((Trat == "Trat2") *(Sexo=="Mujer")), data=base)
summary(ajuste2)
```

```
##
## Call:
## lm(formula = Puntaje ~ Trat + I((Trat == "Trat2") * (Sexo ==
##     "Mujer")), data = base)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.971 -1.641  0.317  1.367  4.352
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                10.309      0.347   29.71  < 2e-16
## TratTrat1                                  -1.996      0.491   -4.07  8.7e-05
## TratTrat2                                  -4.098      0.601   -6.82  4.4e-10
## I((Trat == "Trat2") * (Sexo == "Mujer"))    3.891      0.694    5.61  1.4e-07
##
## (Intercept)                              ***
## TratTrat1                                ***
## TratTrat2                                ***
## I((Trat == "Trat2") * (Sexo == "Mujer")) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.2 on 116 degrees of freedom
## Multiple R-squared:  0.323,  Adjusted R-squared:  0.306
## F-statistic: 18.5 on 3 and 116 DF,  p-value: 7.41e-10
```

- E(y;Control, Hombre) = $\hat{\beta}_0$
- E(y;Control, Mujer) = $\hat{\beta}_0$
- E(y;Trat1, Hombre) = $\hat{\beta}_0 + \hat{\beta}_1$
- E(y;Trat1, Mujer) = $\hat{\beta}_0 + \hat{\beta}_1$
- E(y;Trat2, Hombre) = $\hat{\beta}_0 + \hat{\beta}_2$
- E(y;Trat2, Mujer) = $\hat{\beta}_0 + +\hat{\beta}_2 + \hat{\beta}_3$

Estimadores puntuales:

- E(y;Control, Hombre) = 10.3
- E(y;Control, Mujer) = 10.3
- E(y;Trat1, Hombre) = $10.3 - 1.99 = 8.31$
- E(y;Trat1, Mujer) = $10.3 - 1.99 = 8.31$

- E(y;Trat2, Hombre) $= 10.3 - 4.1 = 6.2$
- E(y;Trat2, Mujer) $= 10.3 - 4.1 + 3.89 = 10.09$

vi.

Realice una prueba de hipótesis para argumentar en favor o en contra de la hipótesis: el nuevo tratamiento tiene el mejor desempeño. Use $\alpha = .05$

Trat 2 mejor que control y Trat 2 mejor que Trat 1

E(y; Trat2) > E(y; Control) y E(y; Trat2) > E(y; Trat1)

$$\hat{\beta}_0 + \hat{\beta}_2 > \hat{\beta}_0 \quad \text{y} \quad \hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3 > \hat{\beta}_0 + \hat{\beta}_1$$

$$H_\alpha : \quad \hat{\beta}_2 < 0 \quad \text{y} \quad \hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3 < 0$$

```
K3 <- matrix(c(0,0,1,0,
          0,1,-1,-1), ncol=4, nrow=2, byrow=TRUE)
m3 <- c(0,0)
summary(glht(ajuste2, linfct=K3, rhs=m3, alternative="less"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ Trat + I((Trat == "Trat2") * (Sexo ==
##     "Mujer")), data = base)
##
## Linear Hypotheses:
##        Estimate Std. Error t value  Pr(<t)
## 1 >= 0   -4.098      0.601   -6.82 4.4e-10 ***
## 2 >= 0   -1.789      0.601   -2.98  0.0035 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Con un p-value < 0.05 es posible asumir que el Trat 2 es mejor que el Control, al igual que el Trat1. Por tanto, con una confianza del 95 % argumentamos a favor de la hipótesis: el nuevo tratamiento tiene mejor desempeño.

vii.

Realice una prueba de hipótesis para argumentar en favor o en contra de la hipótesis: el nuevo tratamiento tiene el mejor desempeño en hombres, aunque el tratamiento actual lo tiene en mujeres. Use $\alpha = .05$

E(y;Trat2;Hombre) > E(y;Trat1;Hombre) y E(y;Trat2;Hombre) > E(y;Control;Hombre)

$$\hat{\beta}_0 + \hat{\beta}_2 > \hat{\beta}_0 + \hat{\beta}_1 \quad \text{y} \quad \hat{\beta}_0 + \hat{\beta}_2 > \hat{\beta}_0$$

$$\hat{\beta}_2 - \hat{\beta}_1 < 0 \quad \text{y} \quad \hat{\beta}_2 < 0$$

```
K4 <- matrix(c(0,-1,1,0,
               0,0,1,0), ncol=4, nrow=2, byrow=TRUE)
m4 <- c(0,0)
summary(glht(ajuste2, linfct=K4, rhs=m4, alternative="less"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ Trat + I((Trat == "Trat2") * (Sexo ==
##     "Mujer")), data = base)
##
## Linear Hypotheses:
##        Estimate Std. Error t value  Pr(<t)
## 1 >= 0   -2.102      0.601   -3.50 0.00063 ***
## 2 >= 0   -4.098      0.601   -6.82 4.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

E(y;Trat1;Mujer) > E(y;Trat2;Mujer) y E(y;Trat1;Mujer) > E(y;Control;Mujer)

$$\hat{\beta}_0 + \hat{\beta}_1 > \hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3 \quad \text{y} \quad \hat{\beta}_0 + \hat{\beta}_1 > \hat{\beta}_0$$

$$\hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3 < 0 \quad \text{y} \quad \hat{\beta}_1 < 0$$

```r
K5 <- matrix(c(0,1,-1,-1,
               0,1,0,0), ncol=4, nrow=2, byrow=TRUE)
m5 <- c(0,0)
summary(glht(ajuste2, linfct=K5, rhs=m5, alternative="less"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ Trat + I((Trat == "Trat2") * (Sexo ==
##     "Mujer")), data = base)
##
## Linear Hypotheses:
##        Estimate Std. Error t value  Pr(<t)
## 1 >= 0   -1.789      0.601   -2.98  0.0035 **
## 2 >= 0   -1.996      0.491   -4.07 8.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Observamos que en ambos casos los p-values son menores a nuestra significancia de 0.05, por tanto es factible concluir que el Trat2 es más efectivo en hombres y que el Trat1 es más efectivo en mujeres