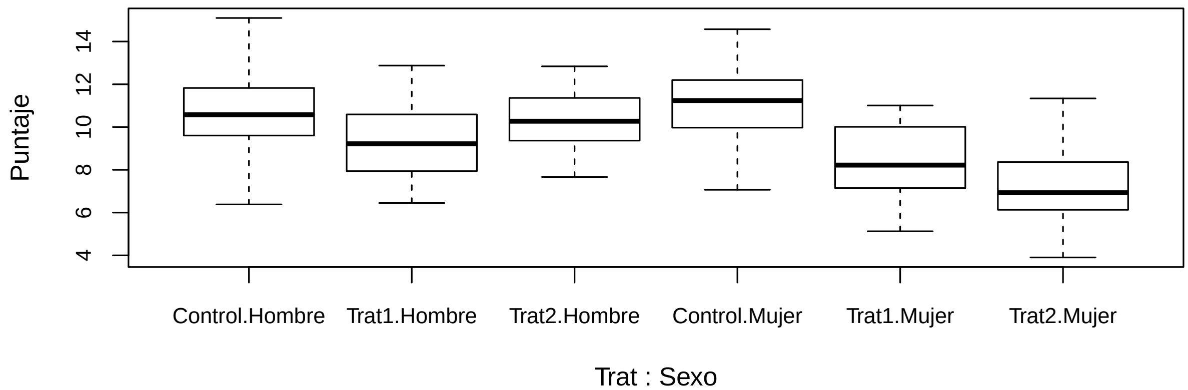
Ejercicio 4

I)



En estos boxplots podemos observar que para el caso Control se observa que de puntaje que presentan ansiedad, tamto en hombres como mujeres, se encuentra aproximadamente entre 10 y 11.

En el caso de aplicar el tratamiento actual (Trat1) podemos notar que hay una disminucion en el nivel de ansiedad que presentan tanto hombres como mujeres, aunque parece ser que el tratamiento actual reduce más el nivel de ansiedad en mujeres que en hombres.

Por otro lado, al aplicar el nuevo tratamiento (Trat2) podemos notar una ligera disminucion en el nivel de ansiedad de los hombres respecto al caso Control, pero los niveles de ansiedad son mayores en comparacion al aplicar el tratamiento actual. Mientras que en el caso de las mujeres el nuevo tratamiento reduce la ansiedad a un nivel muchisimo mas bajo en comparación al caso control e incluso a un nivel mas bajo comparandolo al aplicar el tratamiento actual.

II)

El modelo general es el siguiente:

$$E(Puntaje; Trat, Sexo) = \beta_0 + \beta_1 * Trat1 + \beta_2 * Trat2 + \beta_3 * Mujer + \beta_4(Trat1 * Mujer) + \beta_5(Trat2 * Mujer)$$

A partir del modelo general podemos obtener los modelos individuales:

$E(Puntaje; Control, Hombre) = \beta_0$

$E(Puntaje; Trat1, Hombre) = \beta_0 + \beta_1$

$E(Puntaje; Trat2, Hombre) = \beta_0 + \beta_2$

$E(Puntaje; Control, Mujer) = \beta_0 + \beta_3$

$E(Puntaje; Trat1, Mujer) = \beta_0 + \beta_1 + \beta_3 + \beta_4$

$E(Puntaje; Trat2, Mujer) = \beta_0 + \beta_2 + \beta_3 + \beta_5$

Ajunstamos nuestro modelo:

```
##
## Call:
## lm(formula = Puntaje ~ Trat * Sexo, data = datos)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3785 -1.1800 -0.0518  1.2159  4.3400
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          10.7602     0.3948  27.252  < 2e-16 ***
## TratTrat1            -1.5100     0.5584  -2.704   0.0079 **
## TratTrat2            -0.4789     0.5584  -0.858   0.3929
## SexoMujer             0.5231     0.5584   0.937   0.3509
## TratTrat1:SexoMujer  -1.3758     0.7897  -1.742   0.0842 .
## TratTrat2:SexoMujer  -3.5914     0.7897  -4.548 1.36e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.766 on 114 degrees of freedom
## Multiple R-squared:  0.4007, Adjusted R-squared:  0.3744
## F-statistic: 15.24 on 5 and 114 DF,  p-value: 1.873e-11
```

Por tanto las estimaciones puntuales son:

$E(Puntaje; Control, Hombre) = 10.7602$

$E(Puntaje; Trat1, Hombre) = 10.7602 + (-1.51) = 9.2502$

$E(Puntaje; Trat2, Hombre) = 10.7602 + (-0.4798) = 10.2804$

$E(Puntaje; Control, Mujer) = 10.7602 + 0.5231 = 11.2833$

$E(Puntaje; Trat1, Mujer) = 10.7602 + (-1.5100) + 0.5231 + (-1.3758) = 8.3975$

$E(Puntaje; Trat2, Mujer) = 10.7602 + (-0.4789) + 0.5231 + (-3.5914) = 7.213$

III)

Las hipótesis que se contrastan con la tabla ANOVA son:

$H_0 : \beta_0 = 0,\ \beta_1 = 0,\ \beta_2 = 0,\ \beta_3 = 0,\ \beta_4 = 0,\ \beta_5 = 0$ vs $H_\alpha : \beta_0 \neq 0$ ó $\beta_1 \neq 0\ \beta_2 \neq 0,\ \beta_3 \neq 0,\ \beta_4 \neq 0,\ \beta_5 \neq 0$

Con el comando $summary(fit)$ vemos que el p-value asociado a la tabla ANOVA es de 1.873e-11 que es menor a nuestra significancia de .05, por lo que rechazamos H_0.

IV)

Para determinar si el sexo tiene un efecto en el puntaje realizaremos una prueba de hipotesis, la cual tiene como hipotesis nula:

$$H_0 = \begin{cases} E(Puntaje; Control, Hombre) &=& E(Puntaje; Control, Mujer) \\ E(Puntaje; Trat1, Hombre) &=& E(Puntaje; Trat1, Mujer) \\ E(Puntaje; Trat2, Hombre) &=& E(Puntaje; Trat2, Mujer) \end{cases} \iff$$

$$\iff H_0 = \begin{cases} \beta_0 &=& \beta_0 + \beta_3 \\ \beta_0 + \beta_1 &=& \beta_0 + \beta_1 + \beta_3 + \beta_4 \\ \beta_0 + \beta_2 &=& \beta_0 + \beta_2 + \beta_3 + \beta_5 \end{cases} \iff H_0 = \begin{cases} 0 &=& \beta_3 \\ 0 &=& \beta_3 + \beta_4 \\ 0 &=& \beta_3 + \beta_5 \end{cases}$$

Notemos que al comparar dos a dos las igualdades de H_0 podemos obtener una hipotesis nula que es equivalente, la cual es:

$$H_0 = \begin{cases} 0 &=& \beta_3 \\ 0 &=& \beta_4 \\ 0 &=& \beta_5 \end{cases}$$

Por lo tanto, la prueba de hipotesis a realizar es:

$H_0 : \beta_3 = 0,\ \beta_4 = 0,\ \beta_5 = 0$ vs $H_\alpha : \beta_3 \neq 0,\ \beta_4 \neq 0,\ \beta_5 \neq 0$

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##        Estimate
## 1 == 0   0.5231
## 2 == 0  -1.3758
## 3 == 0  -3.5914
##
## Global Test:
##       F DF1 DF2    Pr(>F)
## 1 11.13   3 114 1.828e-06
```

Vemos que el p-valor es de 1.827792e-06, por lo que rechazamos H_0. Ahora realizaremos una prueba simultanea para ver si podemos decir que el sexo tiene algun efecto en el tratamiento

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ Trat * Sexo, data = datos)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   0.5231     0.5584   0.937    0.623
## 2 == 0  -1.3758     0.7897  -1.742    0.182
## 3 == 0  -3.5914     0.7897  -4.548   <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Con la prueba simultanea podemos ver que es plausible quitar a β_3 y β_4 del modelo, pues el p-value en ambos es mayor a .025

Por tanto nuestro modelo reducido seria:

$$E(Puntaje; Trat, Sexo) = \beta_0 + \beta_1 * Trat1 + \beta_2 * Trat2 + \beta_3(Trat2 * Mujer)$$

V)

Ajustamos nuestro modelo reducido:

```
##
## Call:
## lm(formula = Puntaje ~ I(Trat == "Trat1") + I(Trat == "Trat2") +
##     I((Sexo == "Mujer") * (Trat == "Trat2")), data = datos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6400 -1.0701  0.0033  1.0982  4.1249
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               11.0217     0.2806  39.272  < 2e-16
## I(Trat == "Trat1")TRUE                    -2.1979     0.3969  -5.538 1.94e-07
## I(Trat == "Trat2")TRUE                    -0.7405     0.4861  -1.523     0.13
```

```
## I((Sexo == "Mujer") * (Trat == "Trat2"))  -3.0683     0.5613  -5.466 2.67e-07
##
## (Intercept)                               ***
## I(Trat == "Trat1")TRUE                    ***
## I(Trat == "Trat2")TRUE
## I((Sexo == "Mujer") * (Trat == "Trat2")) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.775 on 116 degrees of freedom
## Multiple R-squared:  0.3838, Adjusted R-squared:  0.3678
## F-statistic: 24.08 on 3 and 116 DF,  p-value: 3.461e-12
```

Entonces las expresiones del puntaje promedio de cada uno de los valores en las variables categoricas, junton con su estimacion puntual, son:

$E(Puntaje; Control, Hombre) = \beta_0 = 11.0217$

$E(Puntaje; Trat1, Hombre) = \beta_0 + \beta_1 = 11.0217 + (-2.1979) = 8.8238$

$E(Puntaje; Trat2, Hombre) = \beta_0 + \beta_2 = 11.0217 + (-0.7405) = 10.2812$

$E(Puntaje; Control, Mujer) = \beta_0 = 11.0217$

$E(Puntaje; Trat1, Mujer) = \beta_0 + \beta_1 = 11.0217 + (-2.1979) = 8.8238$

$E(Puntaje; Trat2, Mujer) = \beta_0 + \beta_2 + \beta_3 = 11.0217 + (-0.7405) + (-3.0683) = 7.2129$

VI)

Queremos realizar una prueba de hipotesis para ver si el nuevo tratamiento tiene mejor desempeño, por lo que nuestra hipotesis alternativa seria:

$$H_\alpha = \begin{cases} E(puntaje; Trat2, Hombre) & < & E(puntaje; Control, Hombre) \\ E(puntaje; Trat2, Hombre) & < & E(puntaje; Trat1, Hombre) \\ E(puntaje; Trat2, Mujer) & < & E(puntaje; Control, Mujer) \\ E(puntaje; Trat2, Mujer) & < & E(puntaje; Trat1, Mujer) \end{cases} \iff$$

$$\iff H_\alpha = \begin{cases} \beta_0 + \beta_2 & < & \beta_0 \\ \beta_0 + \beta_2 & < & \beta_0 + \beta_1 \\ \beta_0 + \beta_2 + \beta_3 & < & \beta_0 \\ \beta_0 + \beta_2 + \beta_3 & < & \beta_0 + \beta_1 \end{cases} \iff H_\alpha = \begin{cases} 0 & < & -\beta_2 \\ 0 & < & \beta_1 - \beta_2 \\ 0 & < & -\beta_2 - \beta_3 \\ 0 & < & \beta_1 - \beta_2 - \beta_3 \end{cases}$$

Entonces nuestra prueba de hipotesis a realizar es:

$$H_0 = \begin{cases} 0 & \geq & -\beta_2 \\ 0 & \geq & \beta_1 - \beta_2 \\ 0 & \geq & -\beta_2 - \beta_3 \\ 0 & \geq & \beta_1 - \beta_2 - \beta_3 \end{cases} \quad \text{vs} \quad H_\alpha = \begin{cases} 0 & < & -\beta_2 \\ 0 & < & \beta_1 - \beta_2 \\ 0 & < & -\beta_2 - \beta_3 \\ 0 & < & \beta_1 - \beta_2 - \beta_3 \end{cases}$$

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ I(Trat == "Trat1") + I(Trat == "Trat2") +
##     I((Sexo == "Mujer") * (Trat == "Trat2")), data = datos)
##
## Linear Hypotheses:
##        Estimate Std. Error t value  Pr(>t)
## 1 <= 0   0.7405     0.4861   1.523 0.18134
```

```
## 2 <= 0  -1.4574     0.4861  -2.998 1.00000
## 3 <= 0   2.9383     0.7425   3.957 < 0.001 ***
## 4 <= 0   1.6109     0.4861   3.314 0.00229 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Con esta prueba podemos ver que hay evidencia para no rechazar H0, pues tenemos que en los hombres el nuevo tratamiento no resulta ser mejor. Por lo que podriamos decir que el nuevo tratamiento no tiene mejor desempeño

VII)

Queremos realizar una prueba de hipotesis para ver que el tratamiento nuevo tiene mejor desempeño en mujeres mientras que el tratamiento actual lo tiene en hombres, por lo que nuestra hipotesis alternativa seria:

$$H_\alpha = \begin{cases} E(puntaje; Trat1, Hombre) & < & E(puntaje; Control, Hombre) \\ E(puntaje; Trat1, Hombre) & < & E(puntaje; Trat2, Hombre) \\ E(puntaje; Trat2, Mujer) & < & E(puntaje; Control, Mujer) \\ E(puntaje; Trat2, Mujer) & < & E(puntaje; Trat1, Mujer) \end{cases}$$

$$\iff$$

$$H_\alpha = \begin{cases} \beta_0 + \beta_1 & < & \beta_0 \\ \beta_0 + \beta_1 & < & \beta_0 + \beta_2 \\ \beta_0 + \beta_2 + \beta_3 & < & \beta_0 \\ \beta_0 + \beta_2 + \beta_3 & < & \beta_0 + \beta_1 \end{cases} \iff H_\alpha = \begin{cases} 0 & < & -\beta_1 \\ 0 & < & \beta_2 - \beta_1 \\ 0 & < & -\beta_2 - \beta_3 \\ 0 & < & \beta_1 - \beta_2 - \beta_3 \end{cases}$$

Entonces nuestra prueba de hipotesis a realizar es:

$$H_0 = \begin{cases} 0 & \geq & -\beta_1 \\ 0 & \geq & \beta_2 - \beta_1 \\ 0 & \geq & -\beta_2 - \beta_3 \\ 0 & \geq & \beta_1 - \beta_2 - \beta_3 \end{cases} \quad \text{vs} \quad H_\alpha = \begin{cases} 0 & < & -\beta_1 \\ 0 & < & \beta_2 - \beta_1 \\ 0 & < & -\beta_2 - \beta_3 \\ 0 & < & \beta_1 - \beta_2 - \beta_3 \end{cases}$$

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = Puntaje ~ I(Trat == "Trat1") + I(Trat == "Trat2") +
##     I((Sexo == "Mujer") * (Trat == "Trat2")), data = datos)
##
## Linear Hypotheses:
##        Estimate Std. Error t value  Pr(>t)
## 1 <= 0   2.1979     0.3969   5.538 < 1e-04 ***
## 2 <= 0   1.4574     0.4861   2.998 0.00622 **
## 3 <= 0   3.8087     0.4861   7.835 < 1e-04 ***
## 4 <= 0   1.6109     0.4861   3.314 0.00242 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Con esta prueba podemos ver que todos los p-value son menores a .05, por lo que podemos decir que el medicamento actual es mejor en los hombres mientras que el nuevo medicamento tiene mejor desempeño en las mujeres.