

# Modelado estadístico marble burying

2023-11-10

```
knitr::opts_chunk$set(echo = TRUE, eval=TRUE, size=10, tidy=TRUE, tidy.opts=list(width.cutoff=65), warning=FALSE)
options(emmeans= list(emmeans = list(infer = c(TRUE, TRUE)),
                        contrast = list(infer = c(TRUE, TRUE))))
```

```
rm(list=ls()) # Borramos todo lo que podría estar guardado en el environment
ls()
```

```
## character(0)
```

```
# Cargamos todas las librerías que puedan sernos de utilidad a lo largo del análisis
```

```
library(ggplot2) # Graficos
library(car) # Supuestos
```

```
## Loading required package: carData
```

```
library(nlme) # Modelos gls para modelar var
library(ggcorrplot) # Correlacion
library(corrplot) # Correlacion
```

```
## corrplot 0.92 loaded
```

```
library(performance) # Para chequear modelos
library(lme4) # Modelos mixtos
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'lme4'
```

```

## The following object is masked from 'package:nlme':
##
##      lmList

library(pastecs) # Para armar tabla (ver biomasa.rmd)
library(ggeffects) # ggpredict
library(dplyr)     # La vamos a utilizar para armar tablas resumen

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:pastecs':
##
##      first, last

## The following object is masked from 'package:nlme':
##
##      collapse

## The following object is masked from 'package:car':
##
##      recode

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

library(emmeans) # Para comparaciones a posteriori
library(lmerTest) # para obtener un valor p en glmm debe incluir antes esta libreria

##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##      lmer

## The following object is masked from 'package:stats':
##
##      step

library(glmmTMB) # Para modelos mixtos generalizados
library(geepack) # Para modelos marginales generalizados
library(MuMIn)  # Para correr QIC en modelor marginales generalizados

##
## Attaching package: 'MuMIn'

```

```
## The following object is masked from 'package:geepack':  
##  
##      QIC
```

```
library(DHARMa) # Chequear supuestos con DHARMa
```

```
## This is DHARMa 0.4.6. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa')
```

```
library(reshape2) # Para graficar
```

## EFFECTOS DE LA ADMINISTRACIÓN DE ESTRADIOL PREPUBERAL EN COMPORTAMIENTOS RELACIONADOS CON DESÓRDENES PSIQUIÁTRICOS

### 1. Marco teórico

Los desórdenes psiquiátricos se caracterizan por diferencias en la prevalencia entre sexos, es decir, algunos son más prevalentes en hombres que en mujeres, tal como el autismo y el déficit de atención e hiperactividad, y otros más prevalentes en mujeres que en hombres, tales como la ansiedad y la depresión. La hipótesis organizacional-activacional de las diferencias sexuales del cerebro plantea que, durante el desarrollo hay una secreción perinatal de hormonas gonadales que es diferente en hembras y machos, la cual sería responsable de “masculinizar” (en el caso de los machos) el cerebro. Lo que sucedería luego es que la producción de hormonas que se da en las gónadas durante la pubertad, la cual también difiere entre sexos, activa ese cerebro “femenino” o “masculino” según sea el caso, dando así lugar a comportamientos “típicamente femeninos” o “típicamente masculinos”.

### 2. Hipótesis

La hipótesis bajo la cual se llevó a cabo el experimento que se presentará en este trabajo, es que los procesos de masculinización y feminización que ocurren durante la pubertad como consecuencia de los perfiles hormonales gonadales, son necesarios para que se observen las diferencias sexuales que se reportan en conductas relevantes para desórdenes psiquiátricos.

### 3. Diseño experimental

Para testear esta hipótesis y así estudiar un posible mecanismo que subyazca a las diferencias entre sexos que se observan en conductas relacionadas a diversos desórdenes psiquiátricos, se llevó a cabo el siguiente experimento. Se partió de 6 camadas de ratones de la cepa CF1 Crl:FCEN. Tanto los machos como las hembras de esta camada fueron separados al azar en dos grupos y a cada grupo se le asignó el tratamiento control o tratamiento con estradiol. El tratamiento consistió en la administración diaria de manera peroral de estradiol (o vehículo para el grupo control) durante todo el período de la pubertad. Cuando los animales alcanzaron la adultez, se evaluó una amplia batería de ensayos conductuales para evaluar diferentes comportamientos relacionados con desórdenes psiquiátricos. En este trabajo en particular, analizaremos los resultados del experimento de Marble Burying Test.

El test de Marble Burying es un ensayo comportamental que se utiliza para evaluar comportamientos compulsivos y ansiosos en roedores. Consiste en colocar al ratón en una jaula con un colchón de viruta y con 20 canicas dispuestas en formas de grilla. El test dura 30 minutos y durante este, debido a su comportamiento natural de enterrar objetos novedos y extraños, los ratones irán enterrando las canicas. Cuanto más rápido y cuanta mayor sea la cantidad de canicas enterradas, esto se interpreta como un comportamiento de tipo compulsivo. Por lo tanto, cada cinco minutos se evaluó la cantidad de canicas visibles en la jaula y por diferencia, se obtuvo cuantas se encontraban enterradas.

## 4. Objetivo

El objetivo de este trabajo es estudiar si la dinámica del entierro de canicas, entendida como medición del comportamiento obsesivo-compulsivo en ratones, difiere entre sexos y si el tratamiento con estradiol durante la pubertad la modifica. Definimos a la dinámica de entierro como el tiempo que cada ratón tarda en esconder un valor de canicas en particular.

## 5. Predicciones

1. Dado que los comportamientos compulsivos son más prevalentes en machos que en hembras, las hembras adultas control presentaran una dinámica más lenta de entierro de canicas, es decir, tardarán más tiempo en enterrar una determinada cantidad de canicas, que los machos adultos control.
2. Dado que al tratar a machos con estradiol simulamos el perfil hormonal en la pubertad de las hembras, los machos que fueron tratados con estradiol presentarán en la adultez la misma dinámica de entierro de canicas que las hembras control, es decir, tardarán el mismo tiempo en enterrar una determinada cantidad de canicas.

## 6. Base de datos y diccionario de variables.

La base de datos es canicas.csv, la cual contiene 59 filas y 10 columnas.

```
setwd("C:/Users/usuario/Desktop/ÚLTIMA VERSIÓN/ÚLTIMA VERSIÓN")
datos_marble <- read.csv("canicas.csv", header = TRUE, sep = ";", dec = ",")

summary(datos_marble)
```

```
##      Raton              Sexo      Tratamiento      Canada
## Length:59      Length:59      Length:59      Length:59
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##      X5      X10      X15      X20
## Min.   : 0.000 Min.   : 0.00 Min.   : 1.00 Min.   : 4.00
## 1st Qu.: 4.000 1st Qu.: 8.00 1st Qu.:10.00 1st Qu.:10.00
## Median : 7.000 Median :12.00 Median :13.00 Median :14.00
## Mean   : 7.373 Mean   :10.88 Mean   :12.51 Mean   :13.49
## 3rd Qu.:10.000 3rd Qu.:14.00 3rd Qu.:16.00 3rd Qu.:16.00
## Max.   :18.000 Max.   :18.00 Max.   :20.00 Max.   :20.00
##      X25      X30
## Min.   : 4.00 Min.   : 7.00
## 1st Qu.:12.50 1st Qu.:13.00
## Median :16.00 Median :16.00
## Mean   :14.81 Mean   :15.19
## 3rd Qu.:18.00 3rd Qu.:18.00
## Max.   :20.00 Max.   :20.00
```

Primera aproximación gráfica a los datos:

```
# Primero tenemos que cambiar la orientacion de los datos para que los tiempos
# pasen a estar todos agrupados bajo una misma columna
```

```
#datos_copia<-datos_marble[,1:11]
datos_copia<-datos_marble # Armamos una copia del data frame original
names(datos_copia) # Nombres de las columnas
```

```
## [1] "Raton"      "Sexo"        "Tratamiento" "Camada"      "X5"
## [6] "X10"        "X15"         "X20"          "X25"         "X30"
```

```
datos_copia<-melt(datos_copia, id.vars=c("Raton","Sexo","Tratamiento","Camada"))
# Agrupamos todos los tiempos y las canicas enterradas en dos columnas nuevas
colnames(datos_copia)[5] <- "Tiempo" # Nombre de la columna nueva
colnames(datos_copia)[6] <- "N_Canicas"
names(datos_copia)
```

```
## [1] "Raton"      "Sexo"        "Tratamiento" "Camada"      "Tiempo"
## [6] "N_Canicas"
```

```
# Grafico de perfiles
ggplot(data = datos_copia, aes(x = Tiempo, y = N_Canicas,
                               colour=interaction(Tratamiento,Sexo),
                               group=interaction(Tratamiento,Sexo))) +
  geom_smooth(se = FALSE) +
  theme_grey(base_size = 16)+
  #facet_grid (. ~ Tratamiento)+
  labs(x="Tiempo (minutos)", y="Cantidad de canicas enterradas",colour="Grupo")+
  ggtitle ("Cantidad de canicas enterradas en función del Tiempo (minutos)")
```

```
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

## Cantidad de canicas enterradas en función del T

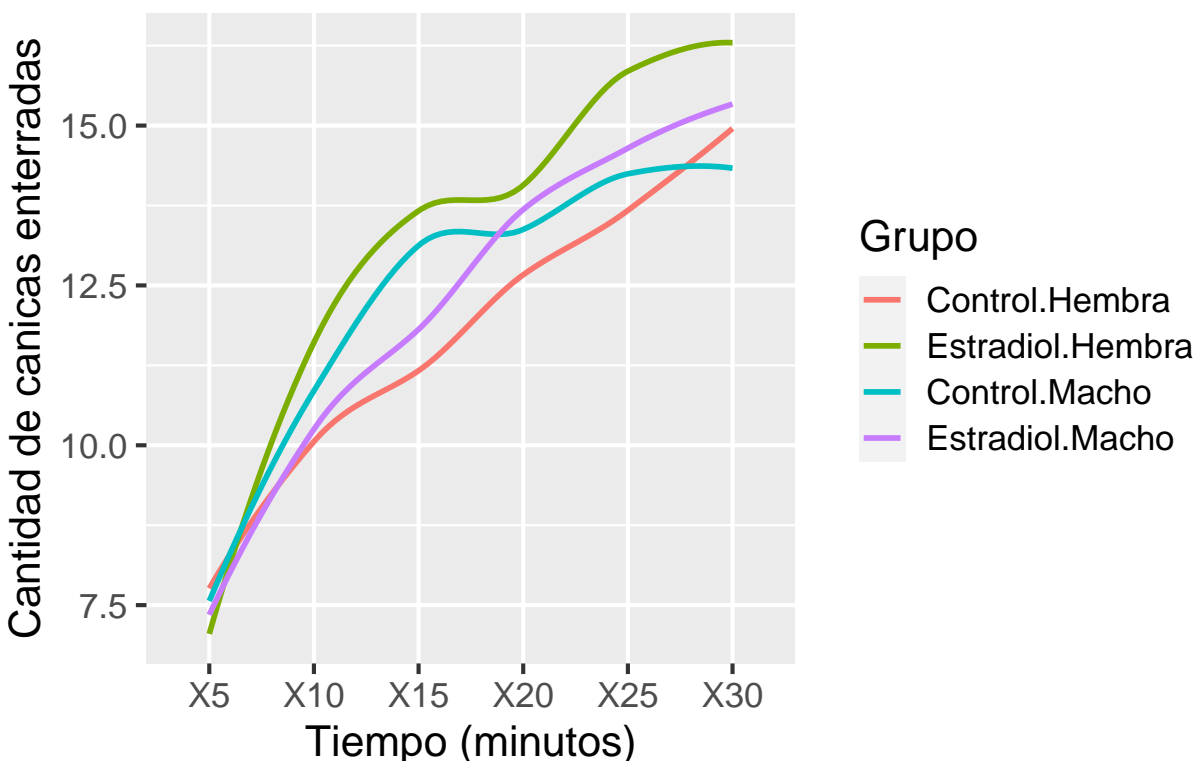


Figura 1. Cantidad de canicas enterradas en función del tiempo. Se grafica la cantidad de canicas acumuladas bajo el aserrín en función del tiempo para cada tratamiento dentro de cada sexo.

La Figura 1 fue construida con todos los datos de la base de datos. A priori se puede observar curvas diferentes para cada tratamiento, sobre todo entre grupos de hembras. Además, entre tratamientos, se puede observar que para el grupo control los machos parecen enterrar sus canicas más rápido que las hembras. Para los tratados con estradiol se observa la tendencia opuesta, donde las hembras parecen ser las que entierran sus canicas más rápido. Para confirmar esto hace falta un análisis estadístico en el cual se defina la variable respuesta.

Como cantidad de canicas enterradas acumuladas no es una variable que fácilmente podamos ajustar a una binomial de  $n=20$  (ya que  $n$ , entendido como cantidad de canicas disponibles para ser enterradas, no es verdaderamente 20, sino que se redefine tiempo a tiempo para cada ratón), se optó por tomar un determinado tiempo en el cual todos los ratones hayan enterrado cierta cantidad de canicas como variable respuesta. Esta variable nos permitirá estudiar la dinámica de entierro de canicas. Se propuso que esta nueva variable ajusta a una distribución gamma.

En primer lugar, se determinó la cantidad de canicas que se utilizarían para definir a la variable respuesta. Para determinar este valor, primero se definió cuál fue el número máximo de canicas que enterró cada ratón, y de esa lista de valores se extrajo el mínimo. De esta forma, se aseguró que todos los ratones hayan enterrado durante el test, al menos, esa cantidad de canicas que utilizaremos para definir la variable respuesta (valor de corte).

```
tiempos <- c(1:length(datos_marble$X5))*0 # Generamos un vector con tantos
# ceros como valores tenga cada columna de la base de datos.
tiempos # Nos queda un vector llamado tiempo con 59 ceros
```

```
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
## [39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
datos_marble$Tiempo <- tiempos # Armamos la columna Tiempo que por ahora toma
#valor 0 para todos los ratones. El objetivo es ir reemplazando los ceros por el
#tiempo en el cual el raton enterró tantas canicas como determine el valor de corte
datos_marble$Maximo <- apply(datos_marble[,6:11], 1, max) # Armamos una columna
# llamada Maximo que va a contener el número máximo de canicas que tuvo
# enterradas cada ratón
```

```
head(datos_marble[,11:12])
```

```
##      Tiempo Maximo
## 1         0      20
## 2         0      16
## 3         0      18
## 4         0      17
## 5         0      15
## 6         0      20
```

```
vc <- min(datos_marble$Maximo) # Como se explicó arriba, el valor de corte será
# el mínimo de esta columna Maximo. De esta manera nos aseguramos no excluir a
# ningun ratón del análisis.
vc # el valor de corte queda determinado como 7
```

```
## [1] 7
```

```
# Armamos un loop para completar la columna Tiempo. Como dijimos, esta tiene que
# contener el tiempo en el cual el ratón tenía enterradas al menos 7 canicas
```

```
for (i in 1:nrow(datos_marble)) { # recorre fila por fila
  if (datos_marble$X5[i] >= vc) { # SI a los 5 min ese raton tenía escondidas "vc" canicas o mas...
    datos_marble$Tiempo[i] <- "5" # ENTONCES en columna Tiempo escribir "5"
  }
  else if (datos_marble$X10[i] >= vc) {
    datos_marble$Tiempo[i] <- "10"
  }
  else if (datos_marble$X15[i] >= vc) {
    datos_marble$Tiempo[i] <- "15"
  }
  else if (datos_marble$X20[i] >= vc) {
    datos_marble$Tiempo[i] <- "20"
  }
  else if (datos_marble$X25[i] >= vc) {
    datos_marble$Tiempo[i] <- "25"
  }
  else if (datos_marble$X30[i] >= vc) {
    datos_marble$Tiempo[i] <- "30"
  }
}
```

```
# El loop devuelve la base de datos original con la columna Tiempo modificada.
# Esta columna es la que contiene la información que usaremos como variable
```

```
# respuesta.
```

```
head(datos_marble$Tiempo)
```

```
## [1] "5" "5" "5" "5" "10" "5"
```

```
class(datos_marble$Tiempo)
```

**Diccionario de variables:**

```
## [1] "character"
```

```
class(datos_marble$Tratamiento)
```

```
## [1] "character"
```

```
class(datos_marble$Sexo)
```

```
## [1] "character"
```

```
class(datos_marble$Camada)
```

```
## [1] "character"
```

*Variable respuesta:* tiempo en el que tarda el ratón en enterrar el 35% de las canicas (7 de 20 canicas). Cuantitativa continua. Unidades: minutos.

*Variables explicativas:*

- Tratamiento: cualitativa con 2 niveles (Estradiol/Control). Fija. Cruzada con Sexo y Camada. Clase **character**. Sin unidades.
- Sexo: cualitativa con 2 niveles (macho/hembra). Fija. Cruzada con Tratamiento y Camada. Clase **character**. Sin unidades
- Camada: cualitativa con 6 niveles (H38, H39, H47, H50, H52 y H55). Aleatoria. Cruzada con Sexo y Tratamiento. Clase **character**. Sin unidades.

*Unidad experimental:* Cada ratón

*Diseño:* Diseño de bloques al azar (DBA). Siendo la camada el bloque. Se establece este tipo de diseño experimental debido a la posibilidad de que exista variabilidad entre las camadas (por ejemplo variabilidad genética, entre otros). Bloqueando por camadas podríamos llegar a ver de mejor manera el efecto de las otras variables.

Dado que la variable respuesta es una medición del tiempo, que es una variable cuantitativa y que tiene una cota inferior (no puedo tomar valores negativos) se propone que sigue una distribución **gamma**.



```
# Antes de continuar con el análisis, pasamos a factor todas las variables
# cualitativas y a numeric las cuantitativas
```

```
datos_marble$Sexo <- as.factor(datos_marble$Sexo)
class(datos_marble$Sexo)
```

```
## [1] "factor"
```

```
datos_marble$Tratamiento <- as.factor(datos_marble$Tratamiento)
class(datos_marble$Tratamiento)
```

```
## [1] "factor"
```

```
datos_marble$Camada <- as.factor(datos_marble$Camada)
class(datos_marble$Camada)
```

```
## [1] "factor"
```

```
datos_marble$Tiempo <- as.numeric(datos_marble$Tiempo)
class(datos_marble$Tiempo)
```

```
## [1] "numeric"
```

## 7. Estadística descriptiva y gráfica

```
# Armamos tabla con las medias para cada grupo (tratamiento x sexo). Veremos la
# media del tiempo en el que los ratones de cada grupo enterraron al menos 7 canicas.
```

```
resumen <- datos_marble %>%
  group_by(Tratamiento, Sexo) %>%
  summarize(Tiempo = mean(Tiempo))
```

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

```
knitr :: kable(resumen,
  caption = "Resumen por grupo (tratamiento x sexo)")
```

Table 1: Resumen por grupo (tratamiento x sexo)

Tratamiento	Sexo	Tiempo
Control	Hembra	7.916667
Control	Macho	8.750000
Estradiol	Hembra	8.333333
Estradiol	Macho	10.000000

```

# Grafico para las medias de cada grupo
g1 <- ggplot(data = resumen, aes(x = interaction(Tratamiento,Sexo),
                                         y = Tiempo,
                                         colour=interaction(Tratamiento,Sexo),
                                         group = interaction(Tratamiento,Sexo))) +
  geom_point(data = resumen, size = 3, colour="red") +
  labs(y="Tiempo", x="Grupo",
       colour="Grupo")

#g1

# Grafico de medias + todos los datos de cada grupo
g2 <- g1 + geom_jitter(data=datos_marble, aes(x=interaction(Tratamiento,Sexo),
                                         y= Tiempo)) +
  ggtitle ("Tiempo (minutos) empleado por cada grupo para enterrar al
          menos el 35% de las canicas")

g2

```

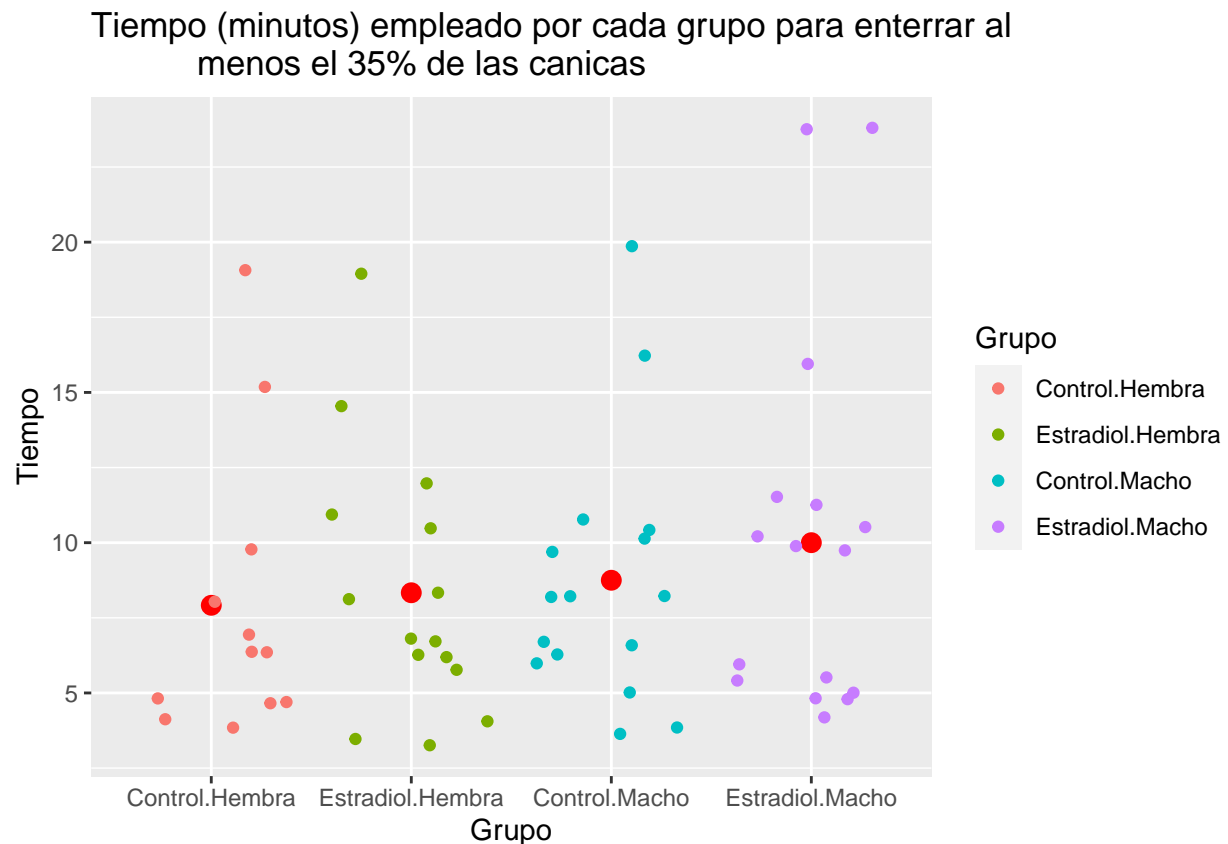


Figura 2. Tiempo (minutos) empleado por cada grupo para enterrar al menos el 35% de las canicas. Se grafica la media (punto rojo) y los valores de cada ratón por cada grupo.

Debido al diseño experimental utilizado (DBA), es de interés estudiar la variabilidad que existe entre ellas para poder explicar los componentes de la varianza del modelo. También nos permitirá elaborar sugerencias para futuros experimentos a realizar, con este paradigma y esta cepa de ratones. En primera instancia, se estudiará la variabilidad entre las camadas de manera gráfica.

```
# como hicimos antes, primero elaboramos una tabla resumen que agrupo por
# tratamiento, sexo y camada
```

```
resumen_camadas <- datos_marble %>%
  group_by(Camada, Tratamiento, Sexo) %>%
  summarize(Tiempo = mean(Tiempo))
```

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

```
knitr :: kable(resumen_camadas,
  caption = "Resumen por camada")
```

Table 2: Resumen por camada

Camada	Tratamiento	Sexo	Tiempo
H38	Control	Hembra	5.000000
H38	Control	Macho	10.000000
H38	Estradiol	Hembra	5.000000
H38	Estradiol	Macho	5.000000
H39	Control	Hembra	10.000000
H39	Control	Macho	10.000000
H39	Estradiol	Hembra	7.500000
H39	Estradiol	Macho	13.333333
H47	Control	Hembra	5.000000
H47	Control	Macho	10.000000
H47	Estradiol	Hembra	8.333333
H47	Estradiol	Macho	8.333333
H50	Control	Hembra	10.000000
H50	Control	Macho	7.500000
H50	Estradiol	Hembra	13.333333
H50	Estradiol	Macho	5.000000
H52	Control	Hembra	10.000000
H52	Control	Macho	7.500000
H52	Estradiol	Hembra	8.333333
H52	Estradiol	Macho	16.666667
H55	Control	Hembra	5.000000
H55	Control	Macho	6.666667
H55	Estradiol	Hembra	5.000000
H55	Estradiol	Macho	8.333333

```
# Grafico de perfiles para camadas
```

```
ggplot(data = resumen_camadas, aes(x = interaction(Tratamiento,Sexo),
  y = Tiempo,
  colour=Camada,
  group = Camada)) +
  #geom_point(data = resumen, size = 3,colour="red") +
  geom_point(size=3)+
  #geom_smooth(se=F)+
  geom_line(size = 1)+
  ggtitle ("Tiempo (minutos) medio empleado por cada grupo dentro de cada camada")
```

```
para enterrar al menos el 35% de las canicas.")+
# facet_grid (. ~ Sexo)+
labs(y="Tiempo", x="Grupo",
colour="Grupo")
```

Tiempo (minutos) medio empleado por cada grupo dentro de cada camada para enterrar al menos el 35% de las canicas.

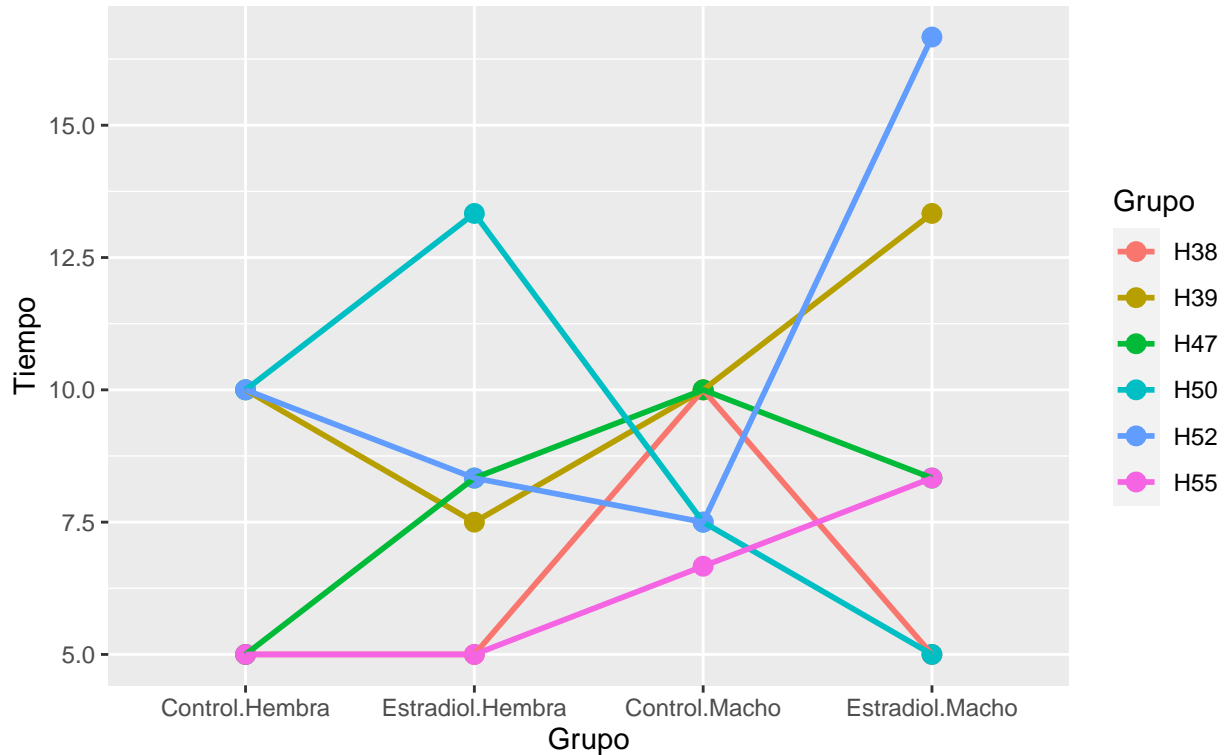


Figura 3. Tiempo (minutos) medio empleado por cada grupo dentro de cada camada para enterrar al menos el 35% de las canicas. Se grafican las medias como puntos y se trazan las líneas que unen a los grupos experimentales dentro de la misma camada.

A partir del gráfico de la *Figura 3* podemos observar que las camadas no se comportan de manera muy similar entre ellas ya que no se observa paralelismo entre las rectas que unen los tiempos medios de cada tratamiento. En otras palabras, no se ven las mismas “bajadas” y “subidas” entre camadas. Se planteará un modelo que incluya la variable camada como variable aleatoria, consistente con el diseño experimental utilizado. De esta manera, se podrán estimar los efectos aleatorios de cada una y estimar un ICC que nos hable de la variabilidad que estas aportan a la variabilidad total. Es importante mencionar que no todos los grupos (tratamiento x sexo) tienen la misma cantidad de ratones en todas las camadas y que algunos puntos puede que no sean medias sino valores puntuales para un único ratón.

## 8. Modelos

### Modelo 1: valor de corte en 7 canicas

Se propone un primer modelo lineal generalizado mixto con variable respuesta de distribución *gamma* y se llevará a cabo una comparación de medias ya que todas las variables explicativas son cualitativas.

### Componente aleatorio

$$Y_i \sim \text{Gamma}(\alpha, \beta)$$

$$E(Y_i) = \alpha \cdot \beta$$

$$\text{Var}(Y_i) = \alpha \cdot \beta^2$$

**NOTA:** Los  $\alpha$  y  $\beta$  anteriores son parámetros de la distribución *gamma* y son distintos al  $\alpha_i$  y  $\beta_j$  que se escriben a continuación para el modelo.

### Predictor lineal

$$\eta_i = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + C_k$$

$\alpha_i$  = efecto del tratamiento

$\beta_j$  = efecto del sexo

$\alpha\beta_{ij}$  = efecto de la interacción entre tratamiento y sexo

$C_k$  = efecto aleatorio de las camadas

$$i = 1 : 2; \quad j = 1 : 2; \quad k = 1 : 6$$

$$C_k \sim N(0, \sigma_{camada}^2)$$

### Función de enlace

$$\log E(Y_i) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + C_k$$

### Escala de la variable respuesta

$$E(Y_i) = \mu_i = e^{\eta_i} = e^{\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + C_k}$$

### Pruebas de hipótesis

$$H_{o \ 1} : \alpha_i = 0$$

No existe efecto del tratamiento con estradiol sobre el tiempo que tardan los ratones en enterrar una determinada cantidad de canicas.

$$H_{o \ 2} : \beta_j = 0$$

No existe efecto del sexo sobre el tiempo que tardan los ratones en enterrar una determinada cantidad de canicas.

$$H_{o \ 3} : \alpha\beta_{ij} = 0$$

El efecto del tratamiento sobre el tiempo que tardan los ratones en enterrar una determinada cantidad de canicas es independiente del efecto del sexo, y viceversa.

$$H_o : \sigma_{camada}^2 = 0$$

La varianza entre los niveles de camada de un mismo tratamiento es igual a cero.

```
# Modelo 1
m1 <-glmmTMB(Tiempo~Tratamiento*Sexo+(1|Camada),
             data=datos_marble,
             family=Gamma(link="log"))
```

## Supuestos

- Muestra aleatoria y observaciones independientes
- Ausencia de sobre o subdispersión (la varianza de la variable es mayor o menor, respectivamente, a la que se espera según la distribución Gamma)
- Normalidad y varianza constante para los efectos aleatorios.

```
# Calculo del estadístico de dispersión
GL.resid <- df.residual(m1)
(dispersion<-sum(resid(m1, type="pearson")^2/df.residual(m1)))
```

```
## [1] 0.3089953
```

Como el estadístico es menor a 0.5 podríamos llegar a observar subdispersión. Para confirmarlo, se correra el modelo con una función del paquete DHARMA que nos permitirá detectar el incumplimiento de los supuestos, en caso de que esto sea lo que esté sucediendo con el modelo.

```
# Para terminar de confirmar si existe o no subdispersión y además chequear
# outliers, corremos el modelo con el paquete DHARMA
simulationOutput <- simulateResiduals(fittedModel = m1, refit=T, plot = T)
```

```
## Model family was recognized or set as continuous, but duplicate values were detected in the response
```

## DHARMA residual

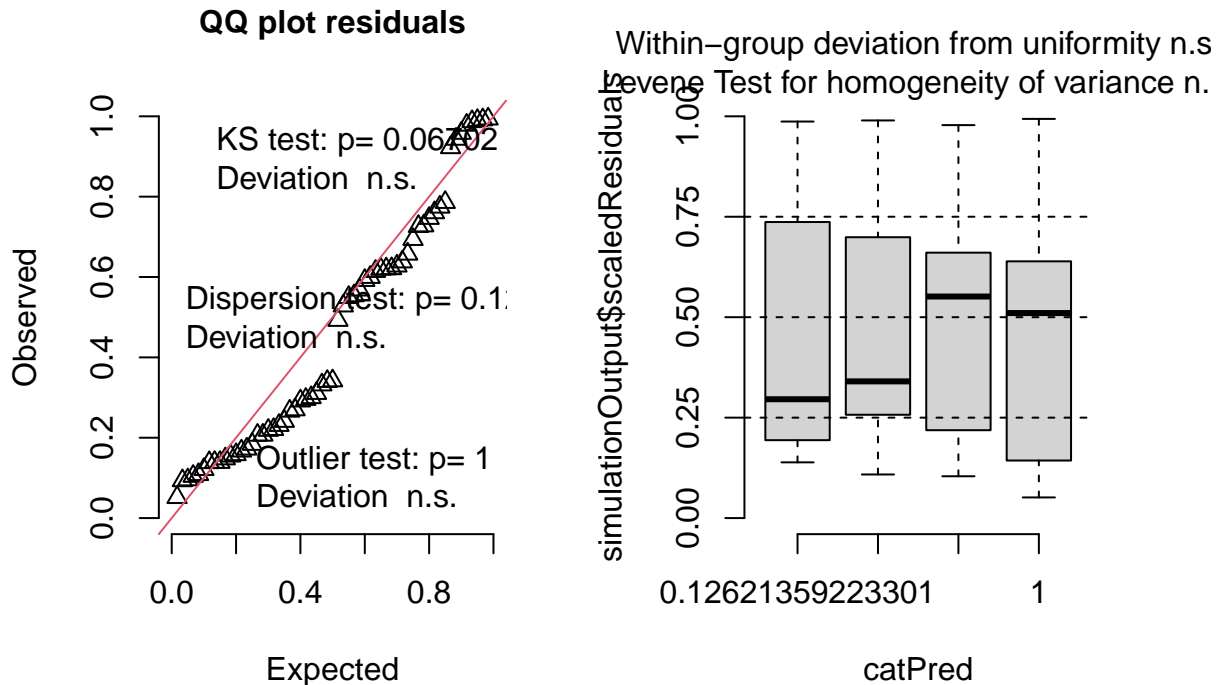


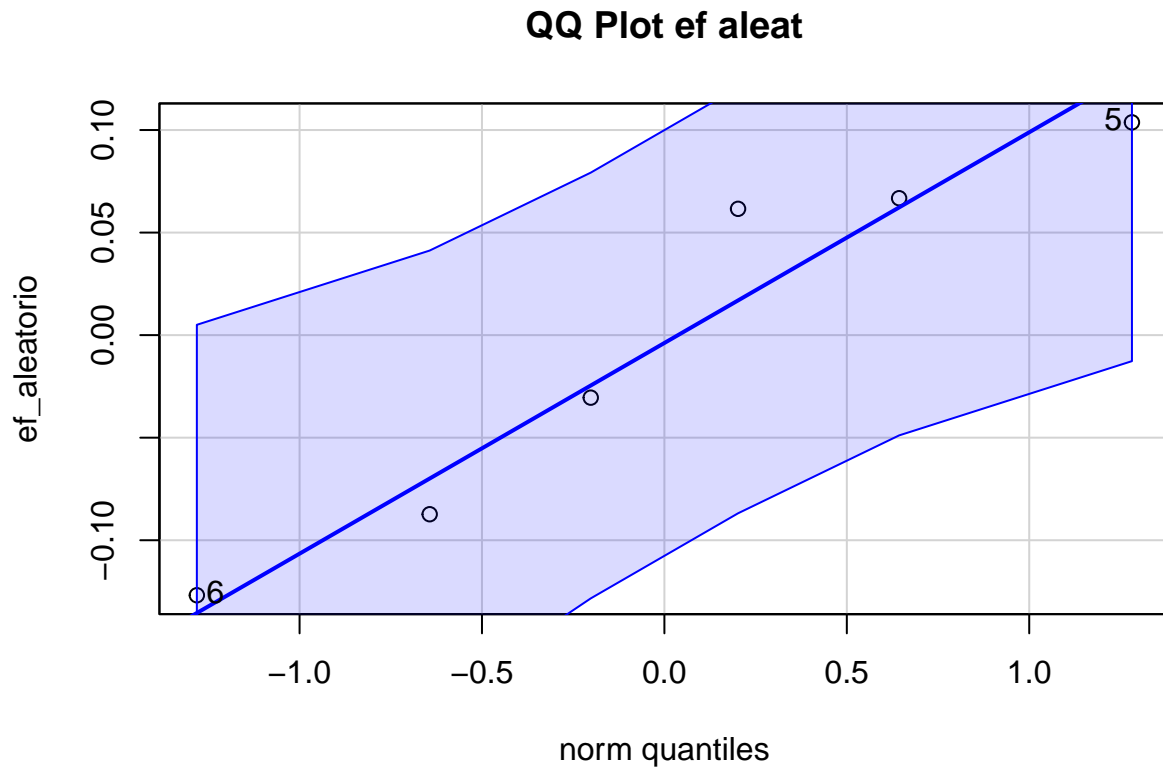
Figura 4. Puesta a prueba del modelo  $m1$ . Se puso a prueba el modelo a partir de las funciones del paquete DHARMA. Izquierda: qqPlot de los residuos simulados por DHARMA. Derecha: gráfico de residuos vs predichos. Se indican sobre el gráfico las distintas pruebas realizadas por la función y sus respectivos p-valores.

No hay evidencias del mal ajuste al modelo: De acuerdo al gráfico de la izquierda, no se rechaza la distribución uniforme para los residuos escalados de DHARMA ( $p$ -valor = 0.06), no se detecta la presencia de outliers y tampoco se detecta sobre o subdispersion ( $p$ -valor = 0.12). Además, no se observan patrones en el gráfico de residuos escalados vs predichos (gráfico de la derecha).

```
# Prueba de normalidad para la varianza aportada por una variable aleatoria

#Ho Normalidad: La distribución de cada subpoblacion es normal
#Esta hipótesis se pone a prueba con el qqPlot y con la prueba de Shapiro Wilks.
ef_aleatorio<-as.data.frame(ranef(m1))$condval # Extraemos los valores de los
# efectos aleatorios de las camadas y los guardamos en un elemento llamado
# ef_aleatorio

qqPlot(ef_aleatorio, main = "QQ Plot ef aleat")
```



```
## [1] 6 5
```

```
shapiro.test(ef_aleatorio)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ef_aleatorio
## W = 0.91349, p-value = 0.4597
```

*Figura 5. qqPlot para los efectos aleatorios de las camadas.* Se grafican los efectos aleatorios en función de los cuantiles de una distribución normal.

No hay evidencias para rechazar que los efectos aleatorios siguen una distribución normal (p-valor = 0.4597). Además, en el qqPlot se puede observar que los efectos aleatorios se encuentran dentro de los valores esperados para una distribución normal.

```
#Parte aleatoria
model_performance(m1, metrics = "ICC") #variabilidad
```

**Análisis del modelo**



```
## # Indices of model performance
##
## ICC
## -----
## 0.070
```

*Parte aleatoria* Al evaluar la parte aleatoria vimos que el ICC dió 0,070. Esto quiere decir que el 7% de la variabilidad total estaría explicada por la variabilidad entre camadas.

```
#Parte fija
Anova(m1)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: Tiempo
##              Chisq Df Pr(>Chisq)
## Tratamiento    0.4284 1    0.5128
## Sexo           1.6668 1    0.1967
## Tratamiento:Sexo 0.0189 1    0.8907
```

*Parte fija* A partir del modelo con valor de corte igual a 7 (es decir, en el 35% de las canicas), la prueba de hipótesis para la interacción es no significativa (p-valor = 0.8907). Esto quiere decir que no hay evidencia para rechazar que el efecto del tratamiento es independiente del sexo. Además, observamos que no hay efecto significativo del sexo (p-valor = 0.1967) ni del tratamiento con estradiol (p-valor = 0.5128) en el tiempo en que los ratones tardan en enterrar 7 canicas.

## Redefinición de la variable respuesta

Debido a que en la estadística descriptiva vimos que podía llegar a haber una diferencia entre machos y hembras, decidimos evaluar el modelo con un valor de corte mayor para ver si detectábamos diferencias.

Para esto se definió por un valor de corte de 10 (50% de canicas enterradas) de manera arbitraria.

```
datos_marble$Tiempo_50 <- tiempos

pc <- 10
for (i in 1:nrow(datos_marble)) { # recorre fila por fila
  if (datos_marble$X5[i] >= pc) {
    datos_marble$Tiempo_50[i] <- "5"
  }
  else if (datos_marble$X10[i] >= pc) {
    datos_marble$Tiempo_50[i] <- "10"
  }
  else if (datos_marble$X15[i] >= pc) {
    datos_marble$Tiempo_50[i] <- "15"
  }
  else if (datos_marble$X20[i] >= pc) {
    datos_marble$Tiempo_50[i] <- "20"
  }
  else if (datos_marble$X25[i] >= pc) {
    datos_marble$Tiempo_50[i] <- "25"
  }
  else if (datos_marble$X30[i] >= pc) {
```

```

    datos_marble$Tiempo_50[i] <- "30"
  }
}

head(datos_marble)

```

```

##   Raton   Sexo Tratamiento Canada X5 X10 X15 X20 X25 X30 Tiempo Maximo
## 1    w3 Hembra Estradiol   H50 14  16  17  18  20  17     5     20
## 2    w2 Hembra Estradiol   H47 10  16  16  15  15  15     5     16
## 3    w1 Hembra Estradiol   H39  9  15  15  15  17  18     5     18
## 4    w4 Hembra Estradiol   H52 11  13  12  15  17  17     5     17
## 5    w8 Hembra Estradiol   H47  5  12  14  14  15  14    10     15
## 6    w5 Hembra Estradiol   H55  7  14  17  16  20  20     5     20
##   Tiempo_50
## 1          5
## 2          5
## 3         10
## 4          5
## 5         10
## 6         10

```

```

datos_marblePC10 <- subset(datos_marble,datos_marble$Tiempo_50 >= 1) #Con este
# código eliminamos de la base de datos a aquellos ratones que no llegaron a
# eliminar 10 canicas. Creamos una nueva base de datos que contiene a los ratones
# que vamos a usar para el análisis, de manera que no perdemos la base de datos
# original.
length(datos_marble$Tiempo_50)

```

```
## [1] 59
```

```
length(datos_marblePC10$Tiempo_50) # Se eliminaron 2 ratones únicamente
```

```
## [1] 57
```

Al realizar un modelo con un valor de corte mayor se pierden dos observaciones, correspondientes a dos ratones que no tienen en ninguno de los tiempos registrados esa cantidad determinada de canicas enterradas (10 canicas).

```

# Igual que antes, pasamos a numeric las variables cuantitativas
datos_marblePC10$Tiempo_50 <- as.numeric(datos_marblePC10$Tiempo)
class(datos_marblePC10$Tiempo_50)

```

```
## [1] "numeric"
```

## Estadística descriptiva y gráfica

Al tener una nueva variable respuesta, es necesario generar los gráficos que ilustren este nuevo caso de estudio.

```
# Armamos tabla con las medias para cada grupo (tratamiento x sexo).
# Veremos la media del tiempo en el que los ratones de cada grupo enterraro
# al menos 10 canicas.
resumenPC10 <- datos_marblePC10 %>%
  group_by(Tratamiento, Sexo) %>%
  summarize(Tiempo_50 = mean(Tiempo_50))
```

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

```
knitr :: kable(resumenPC10,
  caption = "Resumen por grupo (tratamiento x sexo)")
```

Table 3: Resumen por grupo (tratamiento x sexo)

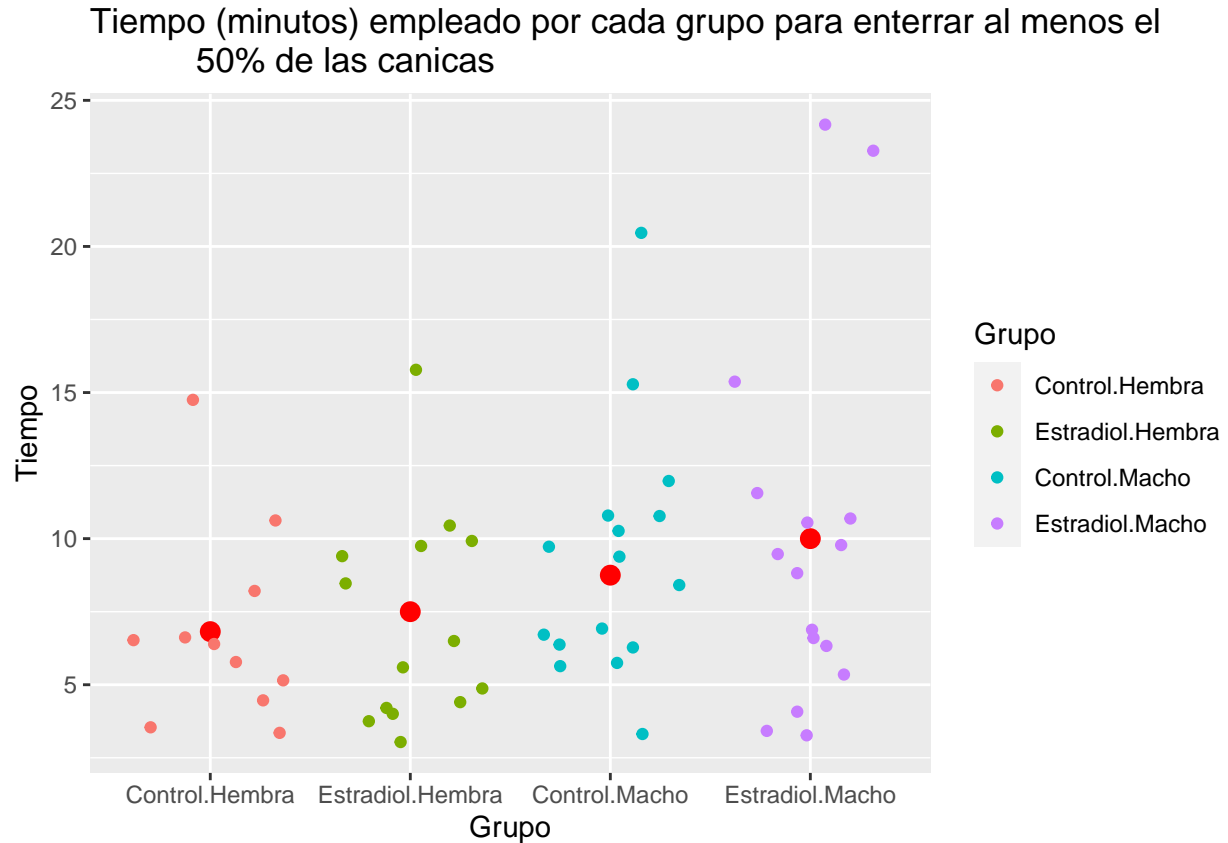
Tratamiento	Sexo	Tiempo_50
Control	Hembra	6.818182
Control	Macho	8.750000
Estradiol	Hembra	7.500000
Estradiol	Macho	10.000000

```
# Grafico para las medias de cada grupo
g3 <- ggplot(data = resumenPC10, aes(x = interaction(Tratamiento,Sexo),
  y = Tiempo_50,
  colour=interaction(Tratamiento,Sexo),
  group = interaction(Tratamiento,Sexo))) +
  geom_point(data = resumenPC10, size = 3,colour="red") +
  labs(y="Tiempo", x="Grupo",
  colour="Grupo")

#g1

# Grafico de medias + todos los datos de cada grupo
g4 <- g3 + geom_jitter(data=datos_marblePC10, aes(x=interaction(Tratamiento,Sexo), y= Tiempo_50)) +
  ggtitle ("Tiempo (minutos) empleado por cada grupo para enterrar al menos el
  50% de las canicas")

g4
```



\*\*Figura 6. Tiempo (minutos) empleado por cada grupo para enterrar al menos el 50% de las canicas.\* Se grafica la media (punto rojo) y los valores de cada ratón por cada grupo.

Igual que antes, es importante estudiar la variabilidad entre las camadas a partir de un gráfico de perfiles.

```
# como hicimos antes, primero elaboramos una tabla resumen que agrupo por
# tratamiento, sexo y camada
resumen_camadasPC10 <- datos_marblePC10 %>%
  group_by(Camada, Tratamiento, Sexo) %>%
  summarize(Tiempo_50 = mean(Tiempo_50))
```

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

```
knitr :: kable(resumen_camadasPC10,
  caption = "Resumen por camada")
```

Table 4: Resumen por camada

Camada	Tratamiento	Sexo	Tiempo_50
H38	Control	Hembra	5.000000
H38	Control	Macho	10.000000
H38	Estradiol	Hembra	5.000000
H38	Estradiol	Macho	5.000000
H39	Control	Hembra	10.000000

Camada	Tratamiento	Sexo	Tiempo_50
H39	Control	Macho	10.000000
H39	Estradiol	Hembra	7.500000
H39	Estradiol	Macho	13.333333
H47	Control	Hembra	5.000000
H47	Control	Macho	10.000000
H47	Estradiol	Hembra	8.333333
H47	Estradiol	Macho	8.333333
H50	Control	Hembra	5.000000
H50	Control	Macho	7.500000
H50	Estradiol	Hembra	10.000000
H50	Estradiol	Macho	5.000000
H52	Control	Hembra	10.000000
H52	Control	Macho	7.500000
H52	Estradiol	Hembra	8.333333
H52	Estradiol	Macho	16.666667
H55	Control	Hembra	5.000000
H55	Control	Macho	6.666667
H55	Estradiol	Hembra	5.000000
H55	Estradiol	Macho	8.333333

```
# Grafico de perfiles para camadas
ggplot(data = resumen_camadasPC10, aes(x = interaction(Tratamiento,Sexo),
y = Tiempo_50,
colour=Camada,
group = Camada)) +
  #geom_point(data = resumen, size = 3,colour="red") +
  geom_point(size=3)+
  #geom_smooth(se=F)+
  geom_line(size = 1)+
  ggtitle ("Tiempo (minutos) medio empleado por cada grupo dentro de cada camada
para enterrar al menos el 50% de las canicas.")+
# facet_grid (. ~ Sexo)+
labs(y="Tiempo", x="Grupo",
colour="Grupo")
```

Tiempo (minutos) medio empleado por cada grupo dentro de cada camada para enterrar al menos el 50% de las canicas.

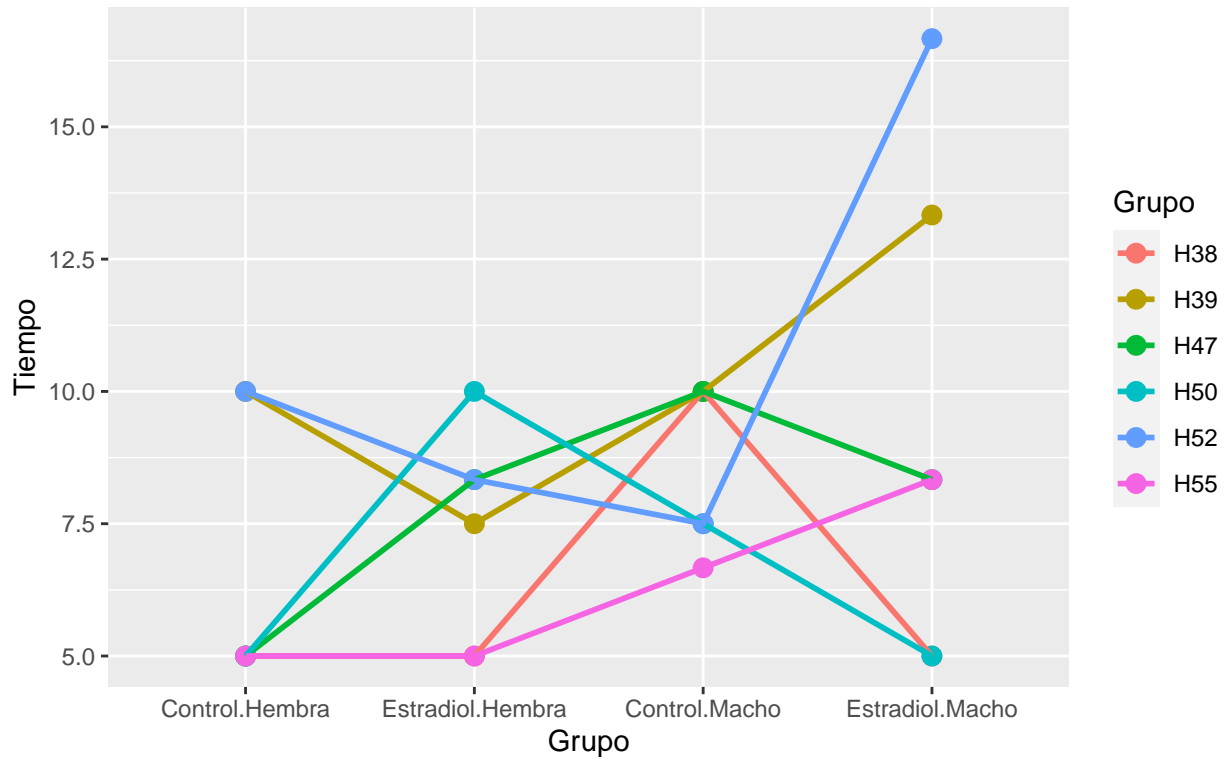


Figura 7. Tiempo (minutos) medio empleado por cada grupo dentro de cada camada para enterrar al menos el 50% de las canicas. Se grafican las medias como puntos y se trazan las líneas que unen a los grupos experimentales dentro de la misma camada.

A partir del gráfico de la Figura 7 podemos observar que las camadas no se comportan de manera muy similar entre ellas ya que no se observa paralelismo entre las rectas que unen los tiempos medios de cada tratamiento. Sin embargo, hay camadas que parecen seguir tendencias similares. Es posible que exista variabilidad entre camadas, pero que no sea mucho más grande de lo que se observó con la variable respuesta anterior.

Igual que antes, es importante mencionar que no todos los grupos (tratamiento x sexo) tienen la misma cantidad de ratones en todas las camadas y que algunos puntos puede que no sean medias sino valores puntuales para un único ratón.

## Modelo 2: valor de corte en 10 canicas

El predictor lineal, la función de enlace y el componente aleatorio son los mismos que para el modelo anterior. Lo único que cambia es la información que ingresamos como variable respuesta. Las variables explicativas también quedan igual que en el modelo anterior.

*Variable respuesta:* tiempo en el que tarda el ratón en enterrar el 50% de las canicas. Cuantitativa. Unidades: minutos.

```
m2 <- glmmTMB(Tiempo_50 ~ Tratamiento*Sexo + (1|Camada),
               data = datos_marblePC10,
               family = Gamma(link = "log"))
```

```
# Calculo del estadístico de dispersión
GL.resid <- df.residual(m2)
(dispersion<-sum(resid(m2, type="pearson")^2/df.residual(m2)))
```

## Supuestos

```
## [1] 0.2421632
```

Como el estadístico es menor a 0.5 podríamos llegar a observar subdispersión. Nuevamente, para confirmarlo, se corra el modelo con una función del paquete DHARMA que nos permitirá detectar el incumplimiento de los supuestos, en caso de que esto sea lo que esté sucediendo con el modelo.

```
# Para terminar de confirmar si existe o no subdispersión y además chequear
# outliers, corremos el modelo con el paquete DHARMA
simulationOutput <- simulateResiduals(fittedModel = m2, refit=T, plot = T)
```

```
## Model family was recognized or set as continuous, but duplicate values were detected in the response
```

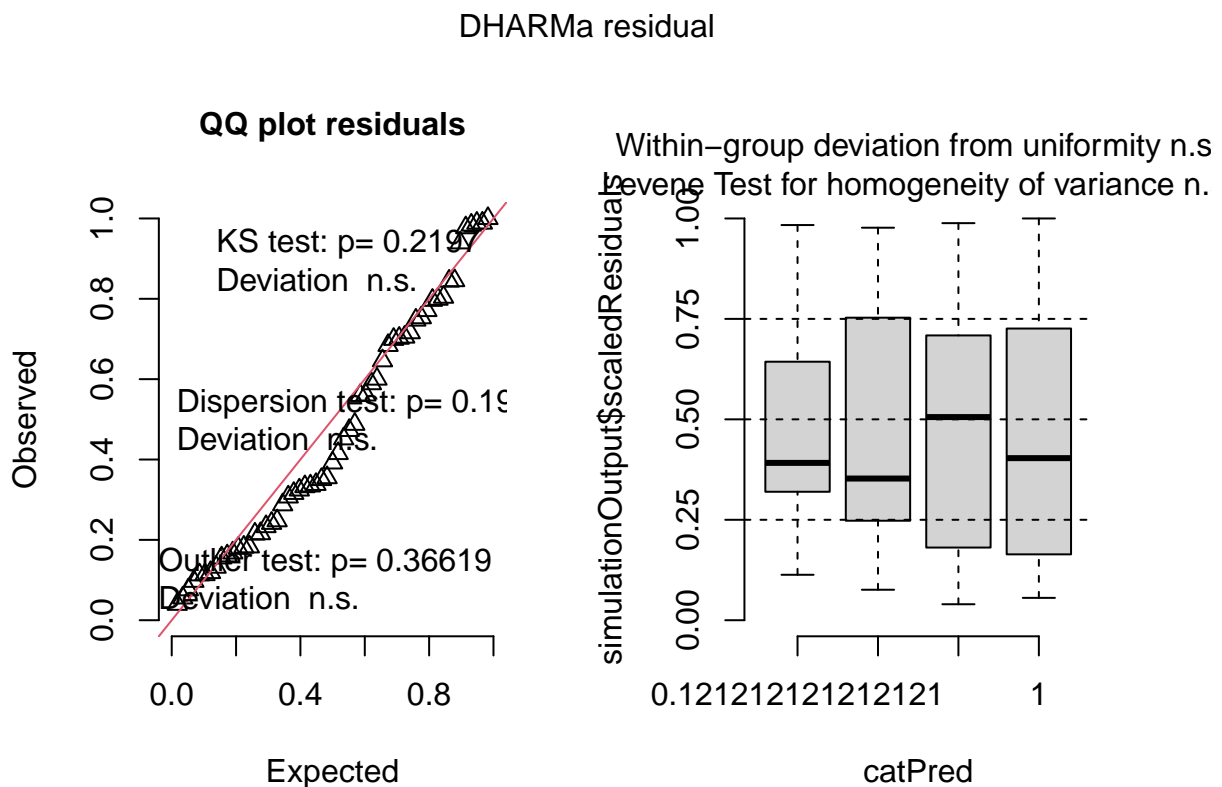


Figura 8. Puesta a prueba del modelo m2. Se puso a prueba el modelo a partir de las funciones del paquete DHARMA. Izquierda: qqPlot de los residuos simulados por DHARMA. Derecha: gráfico de residuos vs predichos. Se indican sobre el gráfico las distintas pruebas realizadas por la función y sus respectivos p-valores.

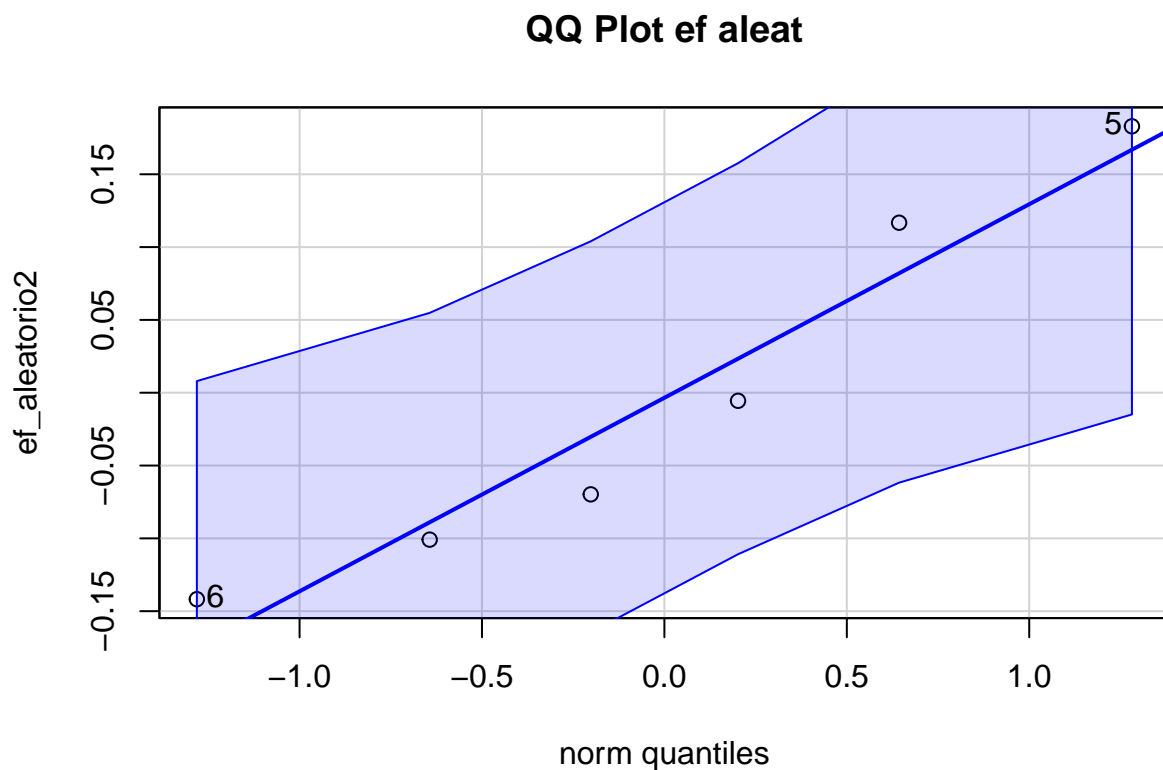
No hay evidencias del mal ajuste al modelo: De acuerdo al gráfico de la izquierda, no se rechaza la distribución uniforme para los residuos escalados de DHARMA (p-valor = 0.2), no se detecta la presencia de

outliers y tampoco se detecta sobre o sub dispersion ( $p\text{-valor} = 0.192$ ). Además, según el gráfico de la derecha no se observan patrones en el gráfico de residuos escalados vs predichos.

```
# Prueba de normalidad para la varianza aportada por una variable aleatoria

#Ho Normalidad: La distribución de cada camada es normal
#Esta hipótesis se pone a prueba con el qqPlot y con la prueba de Shapiro Wilks.
ef_aleatorio2<-as.data.frame(ranef(m2))$condval # Extraemos los valores de los
# efectos aleatorios de las camadas y los guardamos en un elemento llamado
# ef_aleatorio2

qqPlot(ef_aleatorio2, main = "QQ Plot ef aleat")
```



```
## [1] 5 6
```

```
shapiro.test(ef_aleatorio2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ef_aleatorio2
## W = 0.92344, p-value = 0.5305
```

Figura 8. qqPlot para los efectos aleatorios de las camadas. Se grafican los efectos aleatorios en función de los cuantiles de una distribución normal.



No hay evidencias para rechazar que los efectos aleatorios siguen una distribución normal (p-valor = 0.5305). Además, en el qqPlot se puede observar que los efectos aleatorios se encuentran dentro de los valores esperados para una distribución normal.

```
#Parte aleatoria
model_performance(m2, metrics = "ICC") #variabilidad
```

## Análisis del modelo

```
## # Indices of model performance
##
## ICC
## -----
## 0.116
```

Al evaluar la parte aleatoria se vio que el ICC dió 0,116. Esto quiere decir que el 11.6% de la variabilidad total estaría explicada por la variabilidad entre camadas. A pesar de que la variabilidad explicada por las camadas es baja, no la podemos sacar del modelo ya que el diseño empleado es de bloques al azar, siendo los bloques las camadas. Sería recomendable que para futuros experimentos se aleatorize entre todos los ratones sin considerar la camada debido a la baja variabilidad explicada por parte de las camadas.

```
#Parte fija
Anova(m2)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: Tiempo_50
##           Chisq Df Pr(>Chisq)
## Tratamiento    0.7065  1    0.40060
## Sexo           5.6827  1    0.01713 *
## Tratamiento:Sexo 0.0395  1    0.84250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A partir de analizar el modelo con el nuevo valor de corte observamos que la prueba de hipótesis para la interacción dió no significativa. Esto quiere decir que no hay evidencia para rechazar que el efecto del tratamiento es independiente del sexo (p-valor = 0.84250). Sin embargo, al evaluar los efectos principales, observamos que la variable explicativa sexo tiene un efecto significativo sobre la variable respuesta, es decir, sobre el tiempo en que tardan los ratones en enterrar al menos 10 canicas (el 50%) (p-valor = 0.01713). El efecto del tratamiento sobre la variable respuesta es no significativo (p-valor = 0.40060).

## 9. Magnitud de efecto efecto del sexo

En escala del predictor lineal:

```
comp <- emmeans(m2, pairwise ~ Sexo)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
comp
```

```
## $emmeans
## Sexo    emmean      SE df asymp.LCL asymp.UCL z.ratio p.value
## Hembra    1.94 0.110 Inf      1.72      2.16 17.592 <.0001
## Macho     2.22 0.101 Inf      2.02      2.42 22.017 <.0001
##
## Results are averaged over the levels of: Tratamiento
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
## contrast      estimate      SE df asymp.LCL asymp.UCL z.ratio p.value
## Hembra - Macho   -0.282 0.118 Inf    -0.514   -0.051 -2.392 0.0168
##
## Results are averaged over the levels of: Tratamiento
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
```

El logaritmo de la media del tiempo en el que los ratones esconden el 50% de las canicas es 0.282 unidades más grande para machos, independientemente del tratamiento, comparando con hembras.

```
p<-plot(comp$emmeans, comparisons = TRUE)
p + ggtitle("Comparaciones a posteriori entre sexos: escala log")
```

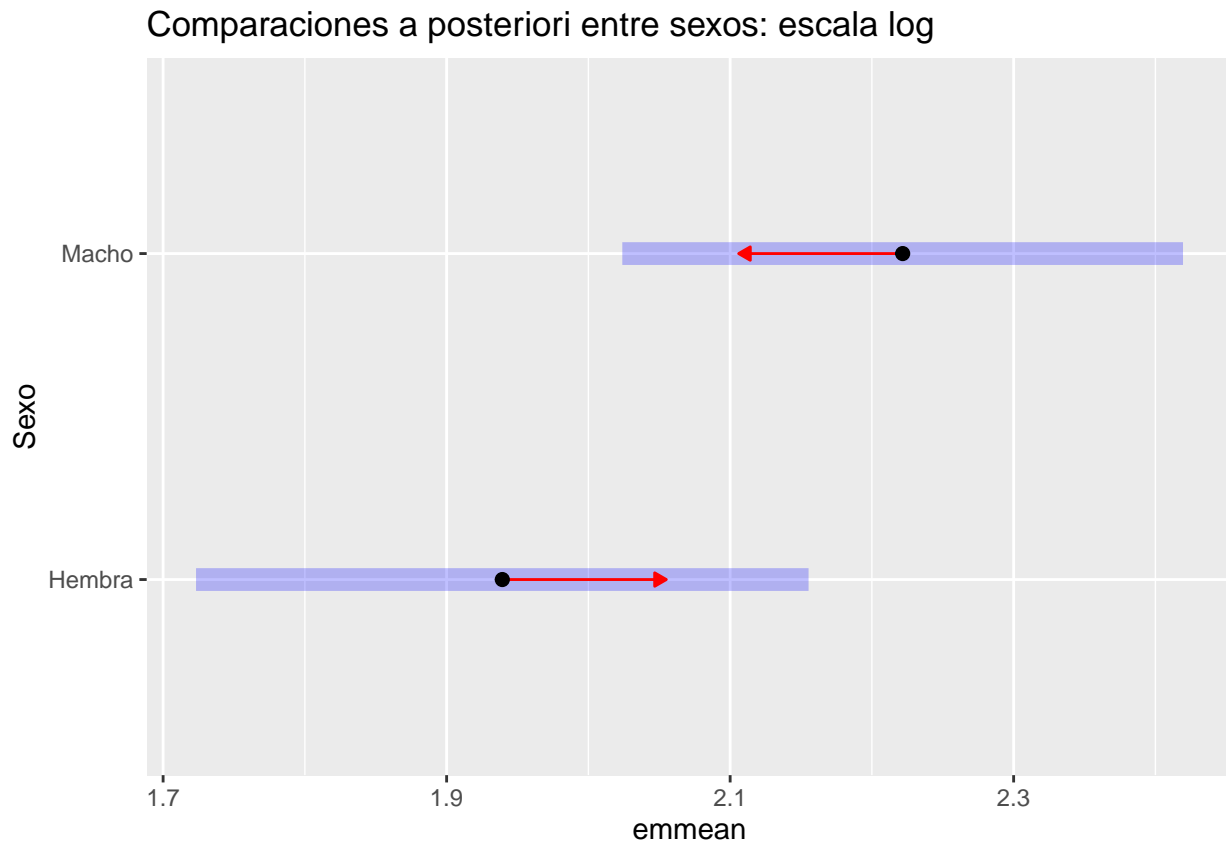


Figura 9. Comparaciones a posteriori entre sexos en escala del predictor lineal. La no superposición de las flechas rojas indica que hay diferencias significativas entre sexos.

En escala de la variable respuesta:

```
comp_vr <- emmeans(m2, pairwise ~ Sexo, type="response")
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
comp_vr
```

```
## $emmeans
##   Sexo   response      SE df asymp.LCL asymp.UCL null z.ratio p.value
##   Hembra    6.95 0.767 Inf    5.60      8.63    1 17.592 <.0001
##   Macho     9.22 0.931 Inf    7.57     11.24    1 22.017 <.0001
##
## Results are averaged over the levels of: Tratamiento
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
## Tests are performed on the log scale
##
## $contrasts
##   contrast      ratio      SE df asymp.LCL asymp.UCL null z.ratio p.value
##   Hembra / Macho 0.754 0.089 Inf    0.598      0.95    1 -2.392 0.0168
##
## Results are averaged over the levels of: Tratamiento
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
## Tests are performed on the log scale
```

El tiempo medio empleado por hembras en esconder el 50% de las canicas es 0.754 veces que el de los machos. En otras palabras, el tiempo que tardan las hembras en enterrar 10 canicas disminuye en promedio 24.6% en comparacion con los machos, independientemente del tratamiento. Se puede afirmar que el tiempo en hembras adultas de la cepa CF1 sera entre 40% y 5% menor que en los machos adultos de la misma cepa con una confianza del 95% para ratones de una camada promedio.

```
p2<-plot(comp_vr, comparisons=T)
p2 + ggtitle("Comparaciones a posteriori entre sexos: escala variable respuesta")
```

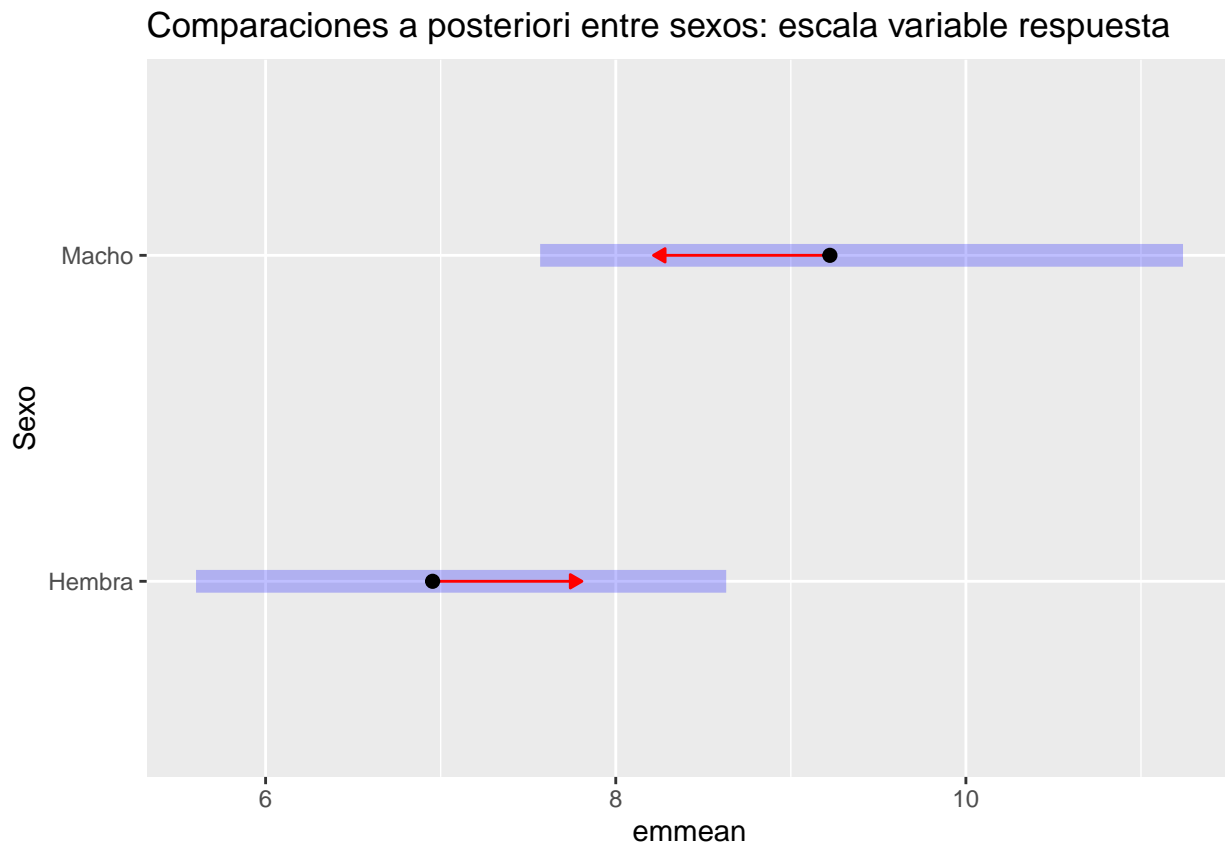


Figura 10 Comparaciones a posteriori entre sexos en escala de la variable respuesta. La no superposición de las flechas rojas indica que hay diferencias significativas entre sexos.

## 10. Metodología estadística

El tiempo que tardaron los ratones en enterrar el 50% de las canicas se analizó utilizando un modelo lineal generalizado mixto utilizando una distribución *Gamma* con función de enlace log. Como variables explicativas de efectos fijos se incluyeron el sexo y el tratamiento, mientras que como variable de efectos aleatorios se incluyó a la camada. El análisis de los supuestos de la parte fija del modelo se efectuó utilizando el paquete 'DHARMA' (Hartig, 2020). Se estudió la normalidad de los efectos aleatorios de las camadas de manera gráfica con un qqPlot y de manera analítica a partir de la prueba de Shapiro. Se puso a prueba la interacción entre sexo y tratamiento y los efectos principales de cada una de estas variables. Se estimó la magnitud efecto utilizando el método de Tukey. Todos los análisis se efectuaron con R (R Core Team).

## 11. Conclusiones y gráfico final

```
a<-ggpredict(m2, terms = c("Sexo", "Tratamiento"))

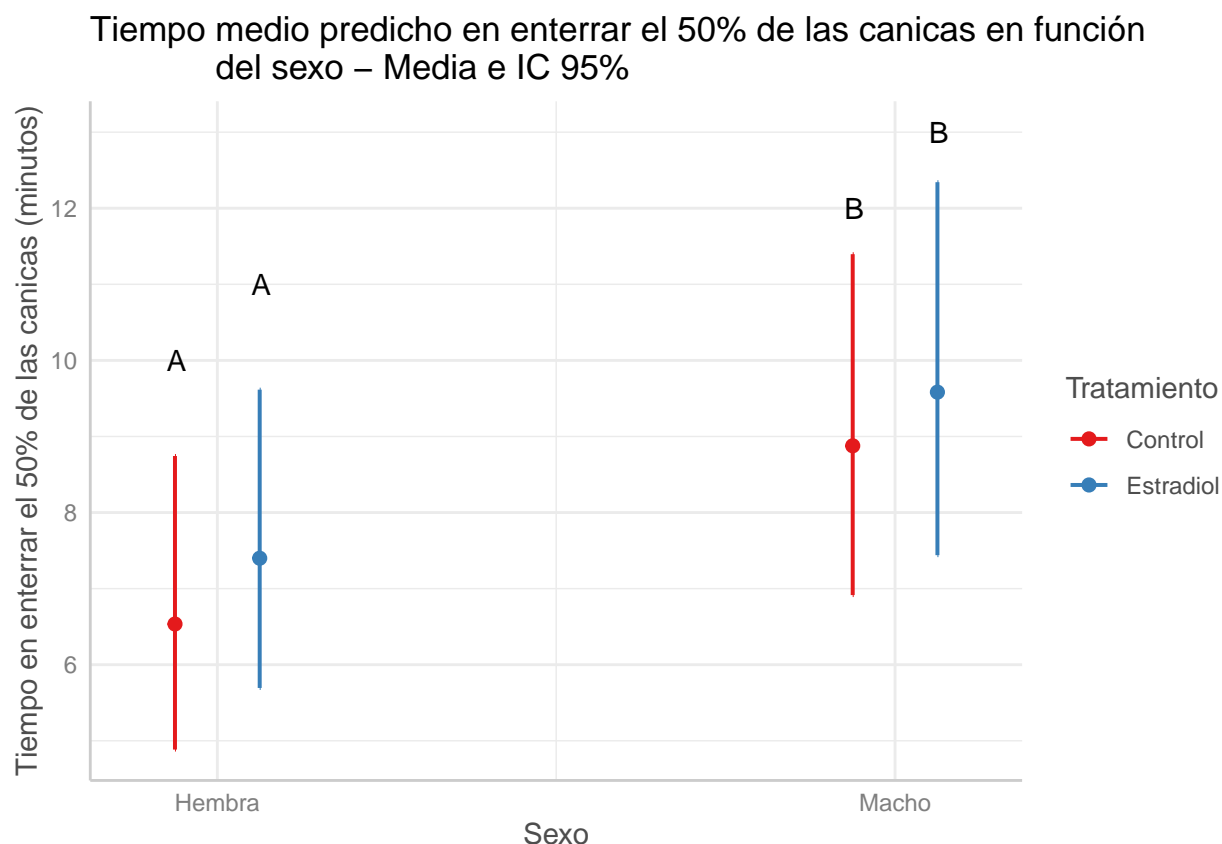
p <- plot(a, add.data = F, grid = F)

p + ggtitle("Tiempo medio predicho en enterrar el 50% de las canicas en función
del sexo - Media e IC 95%") +
  annotate("text", x=c(0.94), y=c(10), label = c("A")) +
  annotate("text", x=c(1.065), y=c(11), label = c("A")) +
```

```

annotate("text", x=c(1.94), y=c(12), label = c("B")) +
annotate("text", x=c(2.065), y=c(13), label = c("B")) +
labs(x="Sexo") + labs(y="Tiempo en enterrar el 50% de las canicas (minutos)")

```



*Figura 11. Tiempo medio predicho en enterrar el 50% de las canicas en función del sexo, para cada tratamiento (media e intervalo de confianza del 95%).* A partir de un modelo lineal generalizado mixto incluyendo como factor aleatorio a la camada, no se encontró interacción entre tratamiento y sexo ( $p=0.88845$ ). Se evidencian diferencias entre machos y hembras en el tiempo en que tardan en enterrar el 50% de las canicas ( $p=0.04224$ ), no así entre tratamientos ( $p=0.58117$ ). Las letras distintas determinan que hay diferencias significativas entre esos grupos, mientras que letras iguales indican que no hay diferencias significativas.

## 12. Discusión

Al analizar los resultados del estudio estadístico, se evidencia que ninguna de las dos predicciones formuladas se ha cumplido. En primer lugar, se anticipaba que los machos mostrarían, en promedio, un menor tiempo que las hembras en enterrar el 50% de las canicas iniciales, ya que los comportamientos obsesivo-compulsivos son más comunes en trastornos psiquiátricos que afectan predominantemente a los hombres en lugar de las mujeres. No obstante, los resultados del análisis revelaron que los machos en realidad presentaron tiempos más largos que las hembras, lo que sugiere, según este estudio, una menor manifestación de comportamiento compulsivo. Aunque estos resultados contradicen nuestra hipótesis, es importante señalar que otros investigadores han obtenido resultados similares en investigaciones previas (Burke et al, 2014). Al profundizar en la revisión de la literatura relacionada con estudios que examinan las diferencias en el test de Marble burying según el sexo, se observa que, en general, no se han obtenido resultados concluyentes. Esto podría deberse en parte a la posibilidad de que este test no sea realmente informativo ni representativo de este tipo de comportamiento, ya que existe cierta controversia en la comunidad científica al respecto. Además, en lo

que respecta al tratamiento con estradiol en machos, teníamos la expectativa de que su desempeño en el test fuera similar al de las hembras del grupo de control. No obstante, nuestros resultados no respaldaron esta suposición. Esto podría atribuirse a dos posibles razones. En primer lugar, es posible que el tratamiento con estradiol en los ratones no haya sido efectivo en alcanzar los niveles deseados de hormona en el organismo. Otra explicación plausible es que la configuración de las estructuras neuronales subyacentes a este tipo de comportamiento no se desarrolle durante la etapa puberal o que no dependa de los cambios en los perfiles hormonales gonadales que ocurren durante la pubertad.