

# Análisis de datos: Proyecto de Diseño Experimental

Arenas Tamara, Medina Nuria, Noriega Berenice, Picasso David, Ruiz Braulio, Vázquez Mariana

2022-12-05

## Cargando las librerías a usar

Cargando tidyverse:

```
install.packages("tidyverse")
library(tidyverse)
```

Cargando pacman:

```
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, nordest, ggplot2, car)
```

## Cargando datos en RStudio, la base de datos se llama data

```
BD <- read.csv(file = "../Data/data.csv")
str(BD)
```

```
## 'data.frame':    24 obs. of  15 variables:
## $ ID           : chr  "E_AM_S1_M" "E_AM_S2_NA" "E_AM_S3_H" "E_AM_S4_H" ...
## $ Sexo_Aplicador: chr  "M" "M" "M" "M" ...
## $ Sexo         : chr  "M" NA "H" "H" ...
## $ Chiste_1     : int   0 NA 0 0 NA NA NA 1 0 0 ...
## $ Chiste_2     : int   1 NA 0 0 NA NA NA 1 0 0 ...
## $ Chiste_3     : int   0 NA 0 0 NA NA NA 1 0 0 ...
## $ Chiste_4     : int   1 NA 1 1 NA NA NA 1 0 1 ...
## $ Chiste_5     : int   0 NA 0 0 NA NA NA 1 0 0 ...
## $ Veces_Risa   : int   2 NA 1 1 NA NA NA 5 0 1 ...
## $ ChG1         : int   3 NA 4 2 NA NA NA 3 3 2 1 ...
## $ ChG2         : int   4 NA 1 3 NA NA NA 2 2 3 1 ...
## $ ChG3         : int   4 NA 2 2 NA NA NA 3 2 2 2 ...
## $ ChG4         : int   4 NA 3 3 NA NA NA 2 2 1 2 ...
## $ ChG5         : int   3 NA 3 4 NA NA NA 3 3 3 3 ...
## $ ChGtot       : int  18 NA 13 14 NA NA NA 13 12 11 9 ...
```

Eliminamos participantes con datos faltantes (NA):

```
BD <- na.omit(BD)
```

###DATOS ANALIZADOS EN FUNCIÓN DEL SEXO DEL APLICADOR

## Obteniendo descriptivos

Esto nos permitirá calcular las medidas de estadística descriptiva.

```
BD %>% group_by(Sexo_Aplicador) %>% summarise(
  media= mean(Veces_Risa),
  mediana= median(Veces_Risa),
  ds= sd(Veces_Risa),
  varianza= var(Veces_Risa),
  minimo= min(Veces_Risa),
  maximo= max(Veces_Risa),
  muestra= n(),
  error_estandar=ds/sqrt(muestra),
  i_confianza_low= media-2*error_estandar,
  i_confianza_up= media+2*error_estandar
)

## # A tibble: 2 x 11
##   Sexo_Aplic~1 media mediana    ds varia~2 minimo maximo muestra error~3 i_con~4
##   <chr>          <dbl>   <dbl> <dbl>   <dbl> <int> <int>   <int>   <dbl>   <dbl>
## 1 H              2.62     3  1.30   1.70     1     5     8  0.460   1.70
## 2 M              1.38     1  1.60   2.55     0     5     8  0.565   0.245
## # ... with 1 more variable: i_confianza_up <dbl>, and abbreviated variable
## #   names 1: Sexo_Aplicador, 2: varianza, 3: error_estandar, 4: i_confianza_low
```

Se guarda la base generada:

```
write.csv(BD, file = "../Output/BD.csv")
```

## Normalidad y homocedasticidad de los datos

Empleamos el test de Shapiro-Wilk ya que el tamaño de la muestra de cada grupo es menor a 50.

```
# Muestra menor a 50 participantes
shapiro.test(BD$Veces_Risa[BD$Sexo_Aplicador=="M"])
```

```
##
## Shapiro-Wilk normality test
##
## data:  BD$Veces_Risa[BD$Sexo_Aplicador == "M"]
## W = 0.74855, p-value = 0.007872
shapiro.test(BD$Veces_Risa[BD$Sexo_Aplicador=="H"])
```

```
##
## Shapiro-Wilk normality test
##
## data:  BD$Veces_Risa[BD$Sexo_Aplicador == "H"]
## W = 0.87745, p-value = 0.178
```

El grupo A (expositor mujer) tiene un p-value menor de 0.05, mientras que el grupo B (expositor hombre) tiene un p-value mayor de 0.05, por lo cual se concluye que los datos del número de eventos de risa no cumplen el supuesto de normalidad.

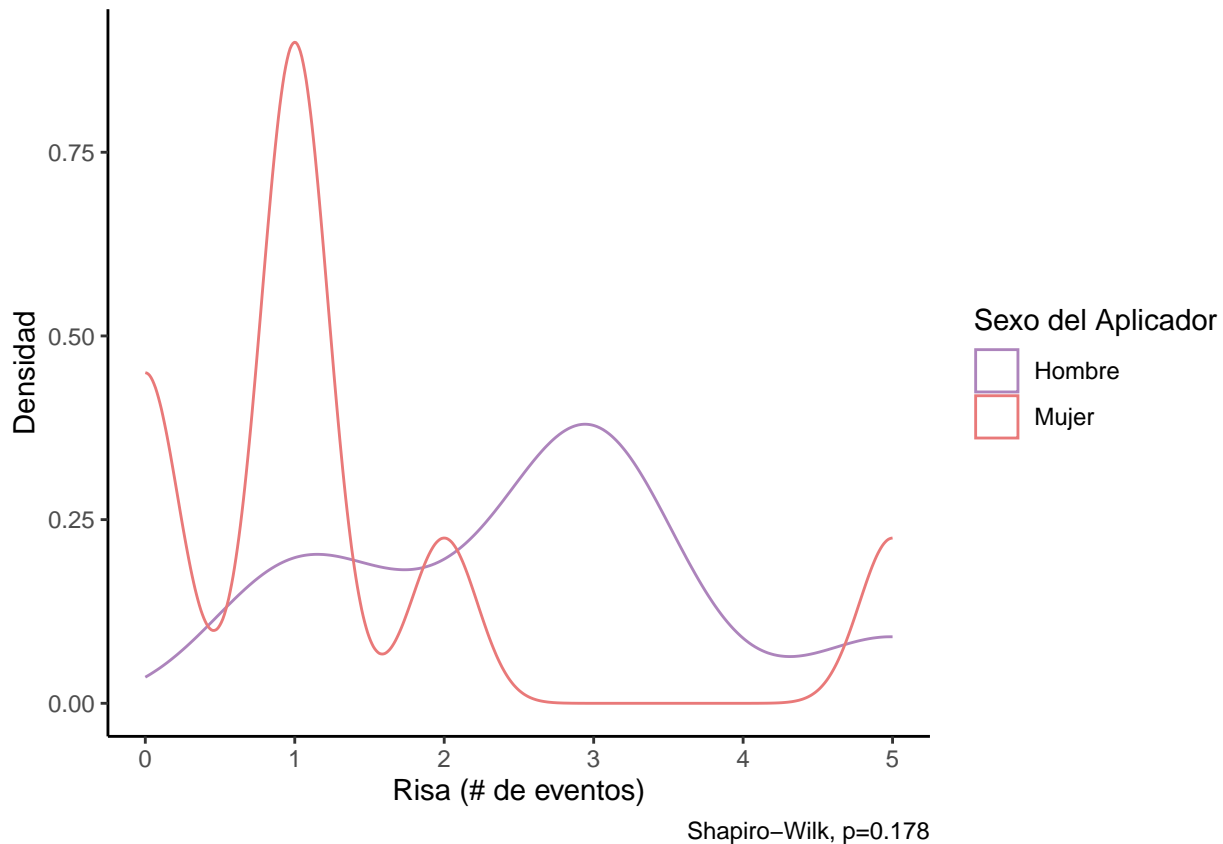
#Se cambian las abreviaciones por el nombre completo

```
sexo_apl_nom <- factor(BD$Sexo_Aplicador, levels=c("H", "M"), labels=c("Hombre", "Mujer"))
```

Se puede comprobar visualmente este hecho:

```
ggplot(data= BD, aes(x=Veces_Risa, color=sexo_apl_nom)) +
  geom_density() +
```

```
scale_color_manual(values = c("#AD84BC", "#E97979"))+
  theme_classic()+
  labs(color = 'Sexo del Aplicador',
        x = 'Risa (# de eventos)',
        y = 'Densidad',
        caption = "Shapiro-Wilk, p=0.178")
```



Dado que los datos no cumplen el supuesto de normalidad, se pone a prueba el supuesto de homocedasticidad con la prueba de Levene

```
car::leveneTest(y=BD$Veces_Risa, group = as.factor(BD$Sexo_Aplicador), center = median)
```

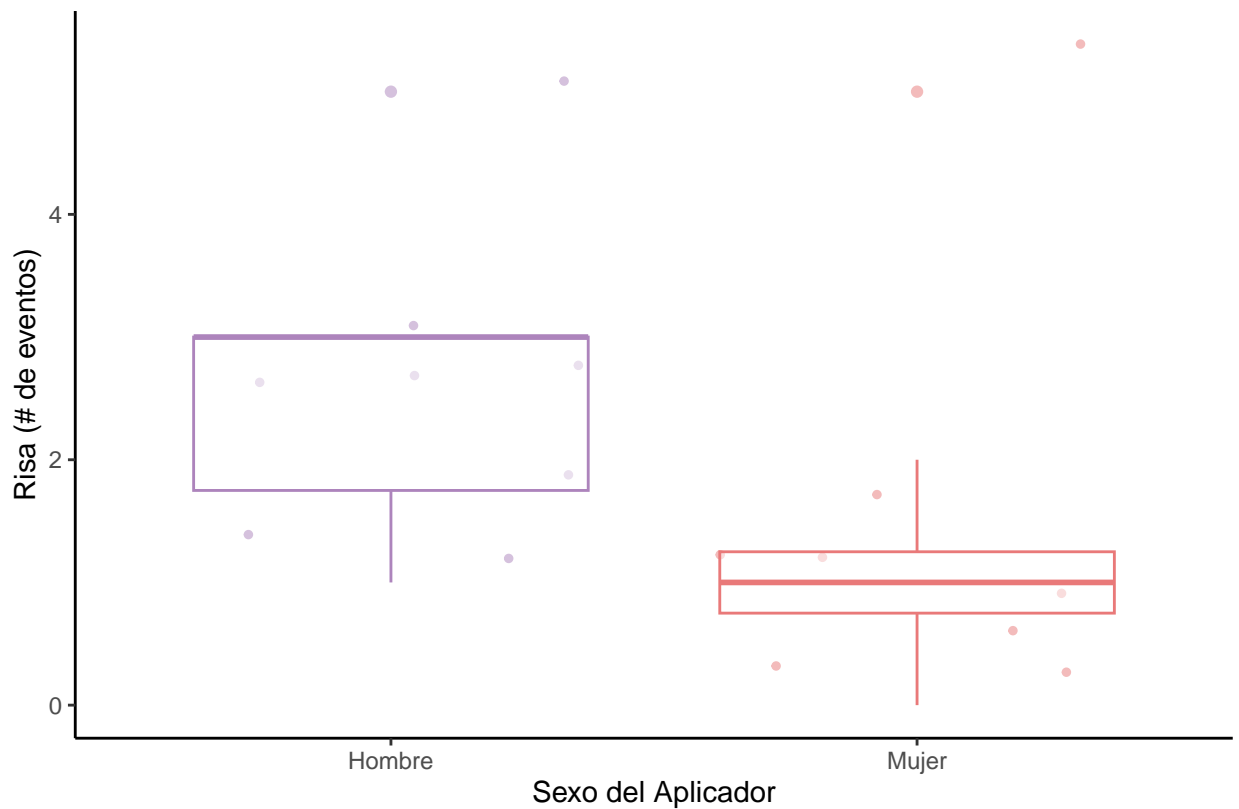
```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 1      0      1
##      14
```

Como el p-value es un valor menor de 0.05, podemos rechazar la hipótesis nula (no rechazamos la hipótesis alternativa).

Se observa este hallazgo visualmente:

```
ggplot(data = BD, aes(x = sexo_apl_nom, y = Veces_Risa)) +
  geom_jitter(aes(color = sexo_apl_nom), size = 1, alpha = 0.5) +
  geom_boxplot(aes(color = sexo_apl_nom), alpha = 0.5) +
  scale_color_manual(values = c("#AD84BC", "#E97979"))+
  theme_classic()+
  labs(x = "Sexo del Aplicador",
        y = 'Risa (# de eventos)',
```

```
caption="Levene, p=0"
) +
theme(legend.position = "none")
```



Levene, p=0

###DATOS ANALIZADOS EN FUNCION DEL SEXO DEL PARTICIPANTE

## Obteniendo descriptivos

Se obtiene el análisis de los datos en función del sexo del participante.

```
BD %>% group_by(Sexo) %>% summarise(
  media= mean(Veces_Risa),
  mediana= median(Veces_Risa),
  ds= sd(Veces_Risa),
  varianza= var(Veces_Risa),
  minimo= min(Veces_Risa),
  maximo= max(Veces_Risa),
  muestra= n(),
  error_estandar=ds/sqrt(muestra),
  i_confianza_low= media-2*error_estandar,
  i_confianza_up= media+2*error_estandar
)
```

```
## # A tibble: 2 x 11
##   Sexo  media mediana    ds varianza minimo maximo muestra error_estan-1 i_con-2
##   <chr> <dbl>   <dbl> <dbl>   <dbl>   <int>   <int>   <int>         <dbl>   <dbl>
## 1 H      2.25     1.5  1.91    3.64     0     5     8         0.675   0.900
```

```
## 2 M      1.75      1.5 1.16      1.36      0      3      8      0.412 0.926
## # ... with 1 more variable: i_confianza_up <dbl>, and abbreviated variable
## #   names 1: error_estandar, 2: i_confianza_low
```

## Normalidad y homocedasticidad de los datos

Empleamos el test de Shapiro-Wilk ya que el tamaño de la muestra de cada grupo es menor a 50.

```
# Muestra menor a 50 participantes
shapiro.test(BD$Veces_Risa[BD$Sexo=="M"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  BD$Veces_Risa[BD$Sexo == "M"]
## W = 0.8577, p-value = 0.1139
shapiro.test(BD$Veces_Risa[BD$Sexo=="H"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  BD$Veces_Risa[BD$Sexo == "H"]
## W = 0.86525, p-value = 0.1354
```

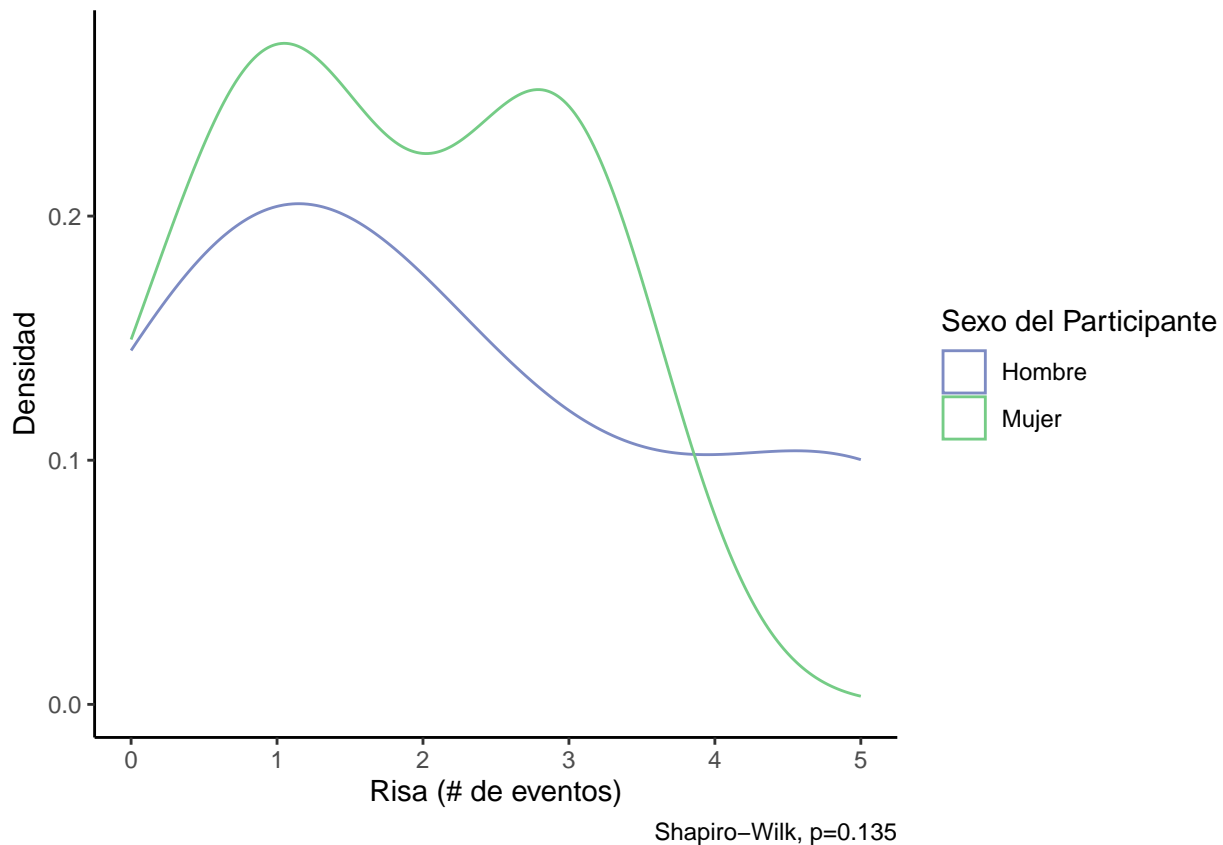
Tanto el grupo de participantes mujeres como el grupo de participantes hombres obtienen un p-value mayor de 0.05, por lo cual se concluye que los datos del número de eventos de risa cumplen el supuesto de normalidad.

#Se cambian las abreviaciones por el nombre completo

```
sexo_par_nom <- factor(BD$Sexo,levels=c("H","M"),labels=c("Hombre","Mujer"))
```

Se puede comprobar visualmente este hecho:

```
ggplot(data= BD, aes(x=Veces_Risa, color=sexo_par_nom)) +
  geom_density() +
  scale_color_manual(values = c("#7D8BC3","#75CC86"))+
  theme_classic()+
  labs(color = 'Sexo del Participante',
       x = 'Risa (# de eventos)',
       y='Densidad',
       caption="Shapiro-Wilk, p=0.135")
```



Dado que los datos cumplen el supuesto de normalidad, se pone a prueba el supuesto de homocedasticidad con la prueba de Bartlett para corroborar.

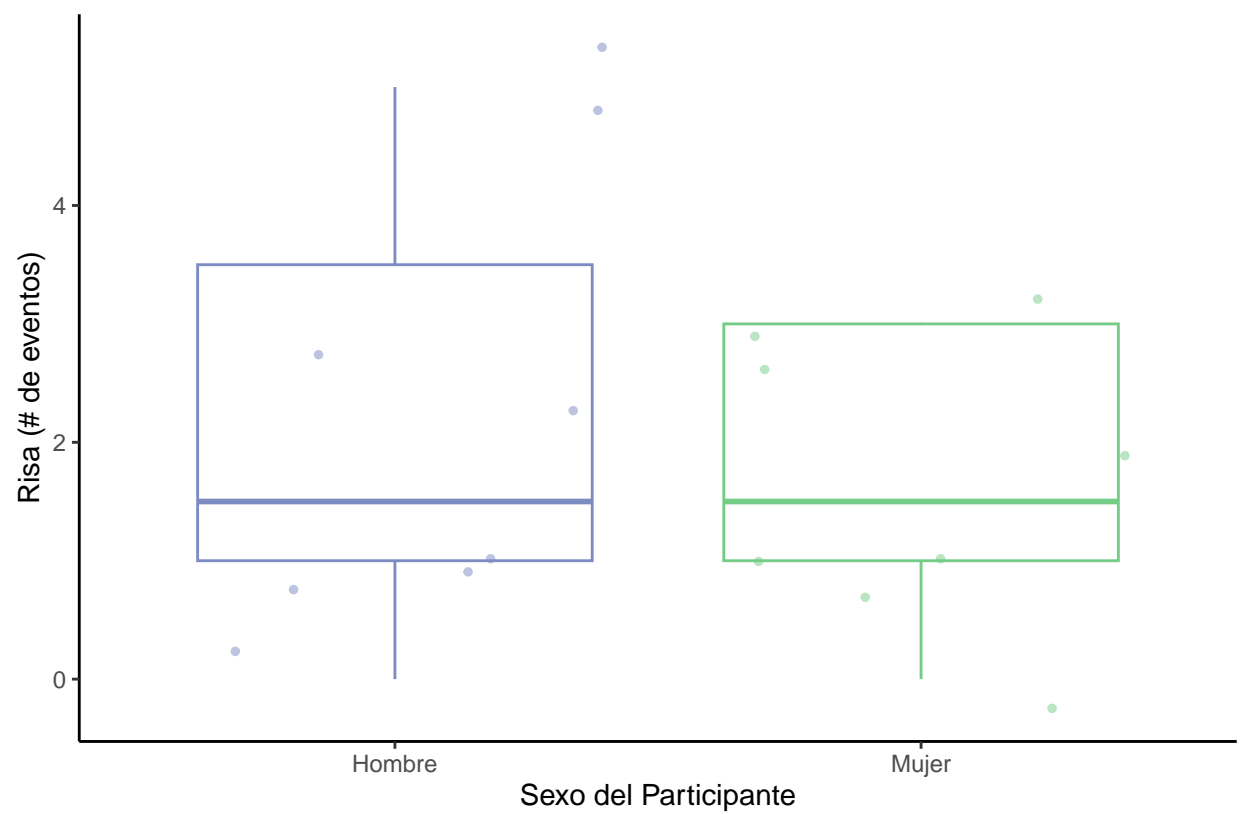
```
bartlett.test(BD$Veces_Risa ~ BD$Sexo)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: BD$Veces_Risa by BD$Sexo
## Bartlett's K-squared = 1.5316, df = 1, p-value = 0.2159
```

Como el p-value es un valor mayor de 0.05, aceptamos la hipótesis nula ( $H_0$ ). Esto nos indica que nuestras dos muestras presentan varianzas iguales. Es decir: no se encuentran diferencias significativas entre las varianzas de los dos grupos.

Podemos observar esta aseveración en el siguiente gráfico:

```
ggplot(data = BD, aes(x = sexo_par_nom, y = Veces_Risa)) +
  geom_boxplot(aes(color = sexo_par_nom), alpha = 0.5) +
  geom_jitter(aes(color = sexo_par_nom), size = 1, alpha = 0.5) +
  scale_color_manual(values = c("#7D8BC3", "#75CC86"))+
  theme_classic()+
  labs(x = 'Sexo del Participante',
       y = 'Risa (# de eventos)',
       caption="Bartlett, p=0.2159"
  ) +
  theme(legend.position = "none")
```



Bartlett, p=0.2159