

# MR - Trabajo

Alicia Losada | alicia.losada.sanchez@udc.es      María Cardoso | m.cardoso@udc.es  
Nicolás Muñiz | nicolas.muniz@udc.es

11/12/2024

## Regresión Lineal Múltiple

- Antes de empezar, cargamos los datos *OzonoLA.rda*

```
load("Datos/OzonoLA.rda")  
attach(OzonoLA)
```

### 1. Análisis descriptivo

Para el análisis descriptivo de las variables podemos comenzar con una visión general de las variables mediante las funciones `str()` y `summary()`.

```
str(OzonoLA)
```

```
## 'data.frame':   203 obs. of  13 variables:  
## $ Mes          : int  1 1 1 1 1 1 1 1 1 1 ...  
## $ DiaMes       : int  5 6 7 8 9 12 13 14 15 16 ...  
## $ DiaSemana    : int  1 2 3 4 5 1 2 3 4 5 ...  
## $ Ozono        : num  5.34 5.77 3.69 3.89 5.76 6.39 4.73 4.35 3.94 7 ...  
## $ Pres_Alt     : int  5760 5720 5790 5790 5700 5720 5760 5780 5830 5870 ...  
## $ Vel_Viento   : int  3 4 6 3 3 3 6 6 3 2 ...  
## $ Humedad      : int  51 69 19 25 73 44 33 19 19 19 ...  
## $ T_Sandburg   : int  54 35 45 55 41 51 51 54 58 61 ...  
## $ T_ElMonte    : num  45.3 49.6 46.4 52.7 48 ...  
## $ Inv_Alt_b    : int  1450 1568 2631 554 2083 111 492 5000 1249 5000 ...  
## $ Grad_Pres    : int  25 15 -33 -28 23 9 -44 -44 -53 -67 ...  
## $ Inv_T_b      : num  57 53.8 54.1 64.8 52.5 ...  
## $ Visibilidad  : int  60 60 100 250 120 150 40 200 250 200 ...
```

La salida de `str()` nos dice que los datos constan de 203 observaciones de 13 variables:

- **Mes**: Número del mes en el que se hicieron las observaciones (Entero)
- **DiaMes**: Número del día del mes en el que se hicieron las observaciones (Entero)
- **DíaSemana**: Número del día de la semana en el que se hicieron las observaciones (Entero)
- **Ozono**: Nivel de Ozono medido (Numérica)
- **Pres\_Alt**: Altura en metros a la que se alcanza una presión de 500 milibares (Entero)
- **Vel\_Viento**: Velocidad del viento en millas por hora en el Aeropuerto Internacional de Los Angeles (Entero)
- **Humedad**: Humedad en porcentaje en LAX (Entero)
- **T\_Sandburg**: Temperatura (F) en Sandburg, CA (Entero)

- T\_ElMonte: Temperatura (F) en El Monte, CA (Numérica)
- Inv\_Alt\_b: Inversion de la altura base (en pies) en LAX (Entero)
- Grand\_Pres: Gradiente de presion de LAX a Daggett, CA (Entero)
- Inv\_T\_b: Inversion de la temperatura base (F) en LAX (Numérica)
- Visibilidad: Visibilidad (millas) evaluada en LAX (Entero)

```
summary(OzonoLA)
```

```
##      Mes      DiaMes      DiaSemana      Ozono      Pres_Alt
## Min.   : 1.000   Min.    : 1.0   Min.   :1.000   Min.    : 0.72   Min.    :5320
## 1st Qu.: 3.000   1st Qu.: 9.0   1st Qu.:2.000   1st Qu.: 4.77   1st Qu.:5690
## Median : 6.000   Median :15.0   Median :3.000   Median : 8.90   Median :5760
## Mean   : 6.522   Mean    :15.7   Mean    :3.005   Mean    :11.37   Mean    :5746
## 3rd Qu.:10.000   3rd Qu.:23.0   3rd Qu.:4.000   3rd Qu.:16.07   3rd Qu.:5830
## Max.    :12.000   Max.     :31.0   Max.     :5.000   Max.     :37.98   Max.     :5950
## Vel_Viento      Humedad      T_Sandburg      T_ElMonte
## Min.   : 0.000   Min.    :19.00   Min.     :25.00   Min.     :27.68
## 1st Qu.: 3.000   1st Qu.:46.00   1st Qu.:51.50   1st Qu.:49.64
## Median : 5.000   Median :64.00   Median :61.00   Median :56.48
## Mean   : 4.867   Mean    :57.61   Mean     :61.11   Mean     :56.54
## 3rd Qu.: 6.000   3rd Qu.:73.00   3rd Qu.:71.00   3rd Qu.:66.20
## Max.    :11.000   Max.     :93.00   Max.     :93.00   Max.     :82.58
## Inv_Alt_b      Grad_Pres      Inv_T_b      Visibilidad
## Min.   : 111   Min.    :-69.00   Min.     :27.50   Min.     : 0.0
## 1st Qu.: 869   1st Qu.: -14.00   1st Qu.:51.26   1st Qu.: 60.0
## Median :2083   Median : 18.00   Median :60.98   Median :100.0
## Mean   :2602   Mean     :14.43   Mean     :60.69   Mean    :122.2
## 3rd Qu.:5000   3rd Qu.: 43.00   3rd Qu.:70.88   3rd Qu.:150.0
## Max.    :5000   Max.     :107.00   Max.     :90.68   Max.     :350.0
```

Ahora realizaremos un análisis descriptivo de cada variable:

### Análisis descriptivo de la variable Mes :

```
summary(Mes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000  3.000   6.000   6.522 10.000   12.000
```

Desviación típica y rango intercuartílico:

```
sd(Mes)
```

```
## [1] 3.594998
```

```
IQR(Mes)
```

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

Evaluamos la asimetría y kurtois

```
library(moments)
```

```
skewness(Mes, na.rm = FALSE)
```

```
## [1] 0.03220505
```

```
kurtosis(Mes, na.rm = FALSE)
```

```
## [1] 1.671129
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis menor que tres, las colas de la variable comparadas con una normal son más ligeras.

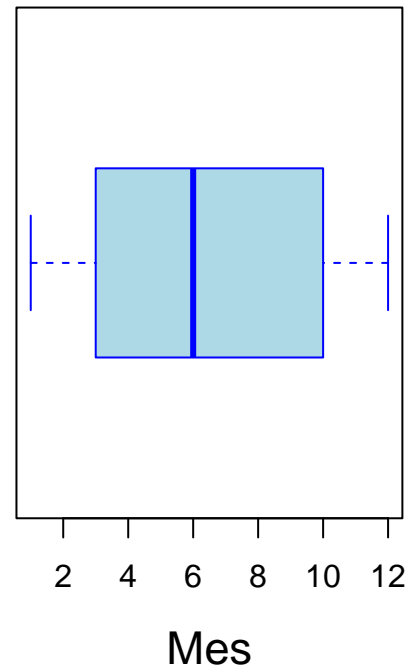
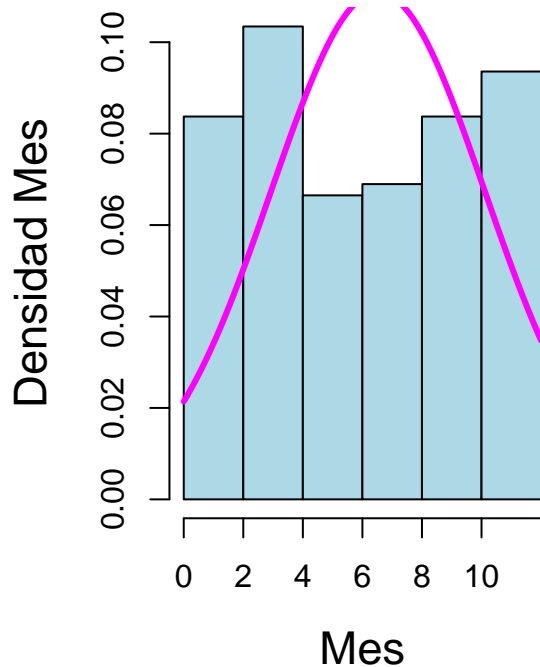
Vemos si hay registros atípicos

```
boxplot.stats(Mes)$out
```

```
## integer(0)
```

Como podemos ver no existe ningún registro atípico

```
par(mfrow=c(1,2))
hist(Mes, breaks=5, freq=FALSE, main = "", xlab="Mes",
     cex.lab=1.4, ylab = "Densidad Mes", col = "lightblue")
curve( dnorm(x, mean=mean(Mes), sd=sd(Mes)),
      col="magenta", lwd=3, add=TRUE)
boxplot(Mes, main = "", xlab="Mes",
      cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
      horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable DiaMes :

```
summary(Mes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   6.000   6.522  10.000  12.000
```

Desviación típica y rango intercuartílico:

```
sd(DiaMes)
```

```
## [1] 8.569537
```

```
IQR(DiaMes)
```

```
## [1] 14
```

Evaluamos la asimetría y kurtois

```
library(moments)
skewness(DiaMes, na.rm = FALSE)
```

```
## [1] 0.0395616
```

```
kurtosis(DiaMes, na.rm = FALSE)
```

```
## [1] 1.868548
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis menor que tres, las colas de la variable comparadas con una normal son más ligeras.

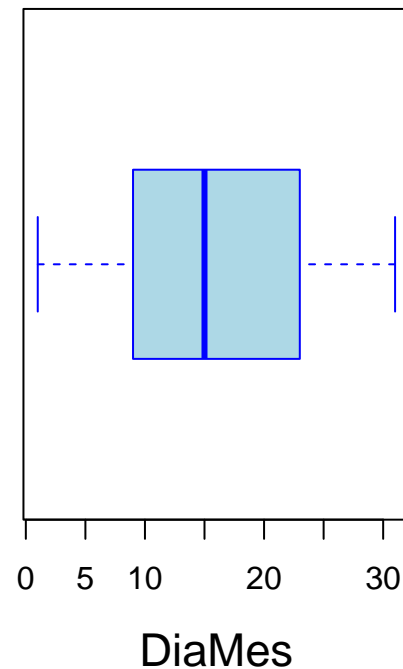
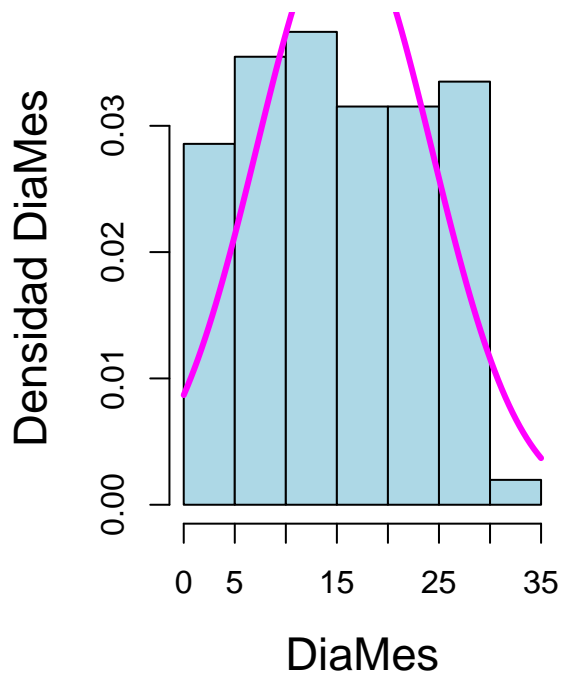
Vemos si hay registros atípicos

```
boxplot.stats(DiaMes)$out
```

```
## integer(0)
```

Como podemos ver no existe ningún registro atípico

```
par(mfrow=c(1,2))
hist(DiaMes, breaks=5, freq=FALSE, main = "", xlab="DiaMes",
     cex.lab=1.4, ylab = "Densidad DiaMes", col = "lightblue")
curve( dnorm(x, mean=mean(DiaMes), sd=sd(DiaMes)),
      col="magenta", lwd=3, add=TRUE)
boxplot(DiaMes, main = "", xlab="DiaMes",
      cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
      horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable DiaSemana :

```
summary(DiaSemana)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000  2.000   3.000   3.005  4.000   5.000
```

Desviación típica y rango intercuartílico:

```
sd(DiaSemana)
```

```
## [1] 1.401899
```

```
IQR(DiaSemana)
```

```
## [1] 2
```

Evaluamos la asimetría y kurtoisis

```
library(moments)
```

```
skewness(DiaSemana, na.rm = FALSE)
```

```
## [1] 0.04527053
```

```
kurtosis(DiaSemana, na.rm = FALSE)
```

```
## [1] 1.731687
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis menor que tres, las colas de la variable comparadas con una normal son más ligeras.

Vemos si hay registros atípicos

```
boxplot.stats(DiaSemana)$out
```

```
## integer(0)
```

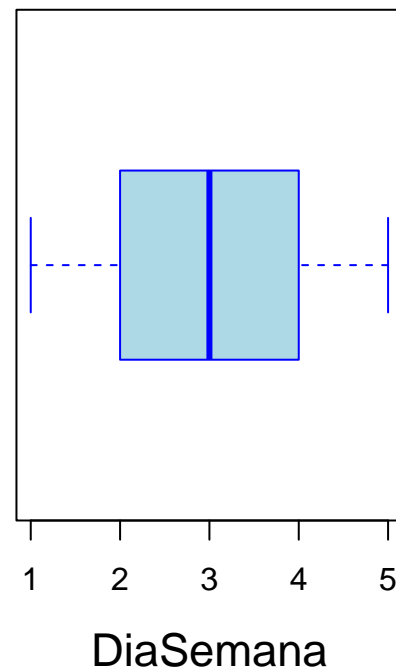
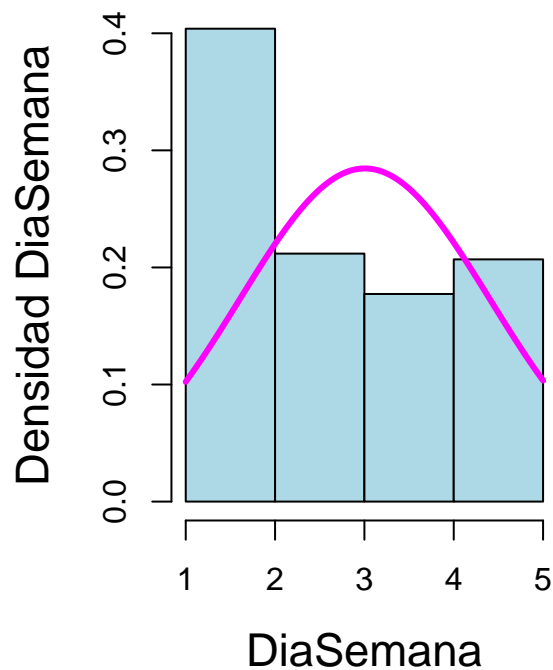
Como podemos ver no existe ningún registro atípico

```
par(mfrow=c(1,2))
```

```
hist(DiaSemana, breaks=5, freq=FALSE, main = "", xlab="DiaSemana",  
     cex.lab=1.4, ylab = "Densidad DiaSemana", col = "lightblue")
```

```
curve( dnorm(x, mean=mean(DiaSemana), sd=sd(DiaSemana)),  
      col="magenta", lwd=3, add=TRUE)
```

```
boxplot(DiaSemana, main = "", xlab="DiaSemana",  
      cex.lab=1.4, border = "blue", col= "lightblue", pch="+",  
      horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable Ozono :

```
summary(Ozono)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.72   4.77   8.90   11.37  16.07   37.98
```

Desviación típica y rango intercuartílico:

```
sd(Ozono)
```

```
## [1] 8.192652
```

```
IQR(Ozono)
```

```
## [1] 11.305
```

Evaluamos la asimetría y kurtosis

```
library(moments)
```

```
skewness(Ozono, na.rm = FALSE)
```

```
## [1] 0.9652702
```

```
kurtosis(Ozono, na.rm = FALSE)
```

```
## [1] 3.089498
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es próximo a tres, las colas de la variable son similares a las de una normal

Vemos si hay registros atípicos

```
boxplot.stats(Ozono)$out
```

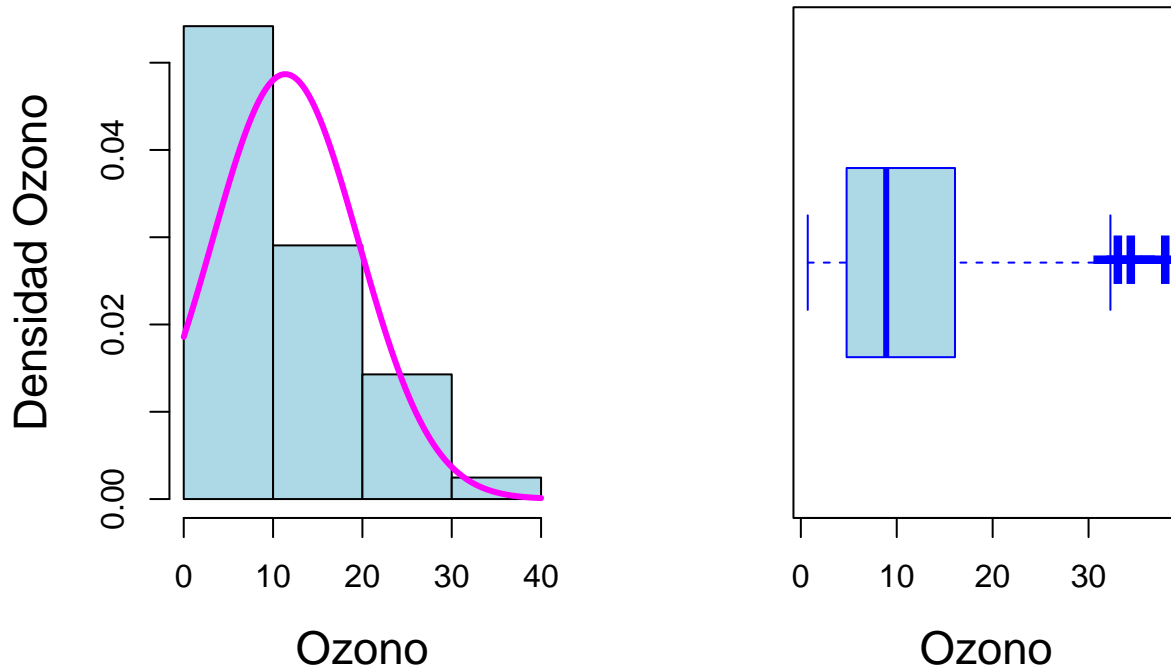
```
## [1] 33.04 34.39 37.98
```

Como podemos ver existen 4 registros atípicos

```

par(mfrow=c(1,2))
hist(Ozono, breaks=5, freq=FALSE, main = "", xlab="Ozono",
     cex.lab=1.4, ylab = "Densidad Ozono", col = "lightblue")
curve( dnorm(x, mean=mean(Ozono), sd=sd(Ozono)),
      col="magenta", lwd=3, add=TRUE)
boxplot(Ozono, main = "", xlab="Ozono",
       cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
       horizontal = TRUE, cex=3)

```



Análisis descriptivo de la variable Pres\_Alt :

```
summary(Pres_Alt)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      5320   5690   5760   5746   5830   5950
```

Desviación típica y rango intercuartílico:

```
sd(Pres_Alt)
```

```
## [1] 113.0277
```

```
IQR(Pres_Alt)
```

```
## [1] 140
```

Evaluamos la asimetría y kurtois

```
library(moments)
```

```
skewness(Pres_Alt, na.rm = FALSE)
```

```
## [1] -0.9499496
```

```
kurtosis(Pres_Alt, na.rm = FALSE)
```

```
## [1] 4.198772
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es mayor a tres, las colas de la variable son más grandes que las de una normal.

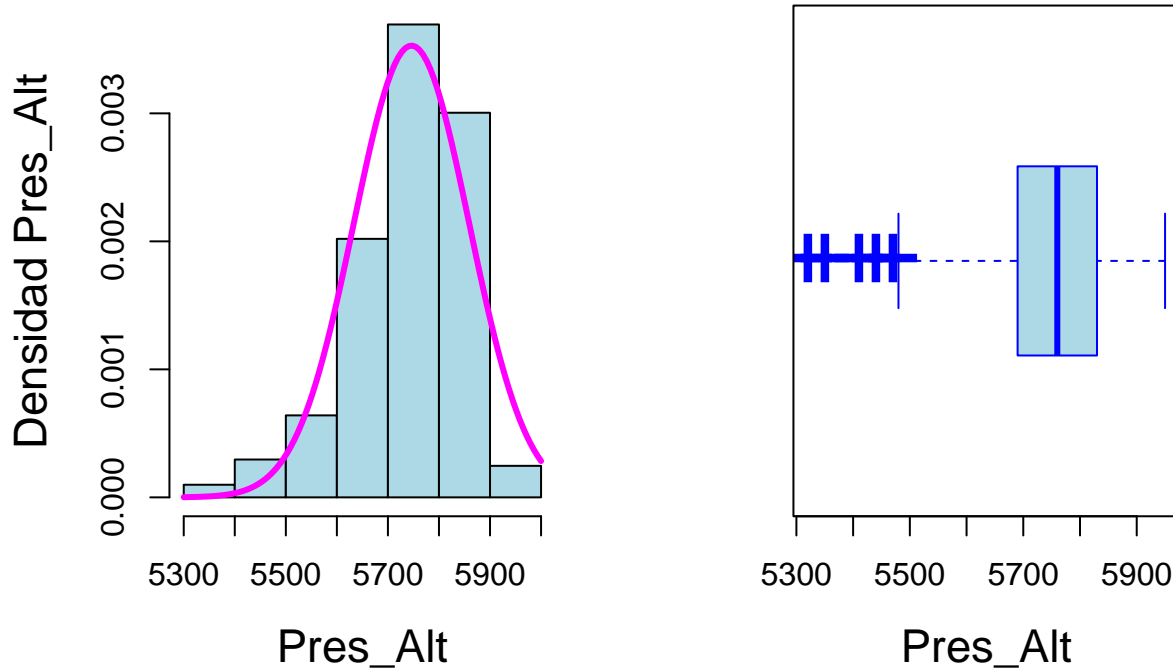
Vemos si hay registros atípicos

```
boxplot.stats(Pres_Alt)$out
```

```
## [1] 5410 5350 5470 5320 5440
```

Como podemos ver existen 5 registros atípicos

```
par(mfrow=c(1,2))
hist(Pres_Alt, breaks=5, freq=FALSE, main = "", xlab="Pres_Alt",
     cex.lab=1.4, ylab = "Densidad Pres_Alt", col = "lightblue")
curve( dnorm(x, mean=mean(Pres_Alt), sd=sd(Pres_Alt)),
      col="magenta", lwd=3, add=TRUE)
boxplot(Pres_Alt, main = "", xlab="Pres_Alt",
      cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
      horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable Vel\_Viento :

```
summary(Vel_Viento)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   3.000   5.000   4.867   6.000  11.000
```

Desviación típica y rango intercuartílico:

```
sd(Vel_Viento)
```

```
## [1] 2.105402
```

```
IQR(Vel_Viento)
```

```
## [1] 3
```



Evaluamos la asimetría y kurtosis

```
library(moments)
skewness(Vel_Viento, na.rm = FALSE)
```

```
## [1] 0.09612047
```

```
kurtosis(Vel_Viento, na.rm = FALSE)
```

```
## [1] 3.378636
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es próximo a tres, las colas de la variable son similares a las de una normal.

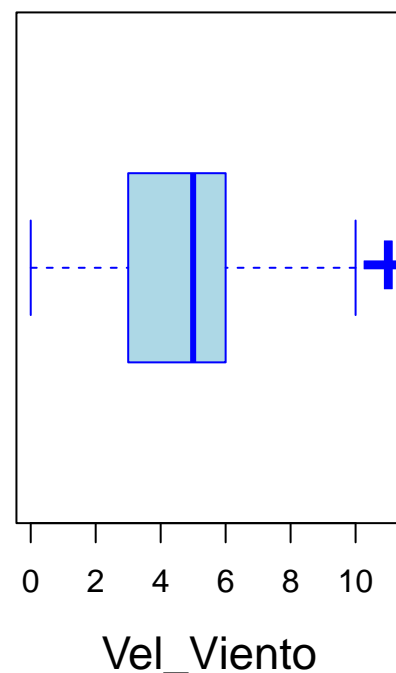
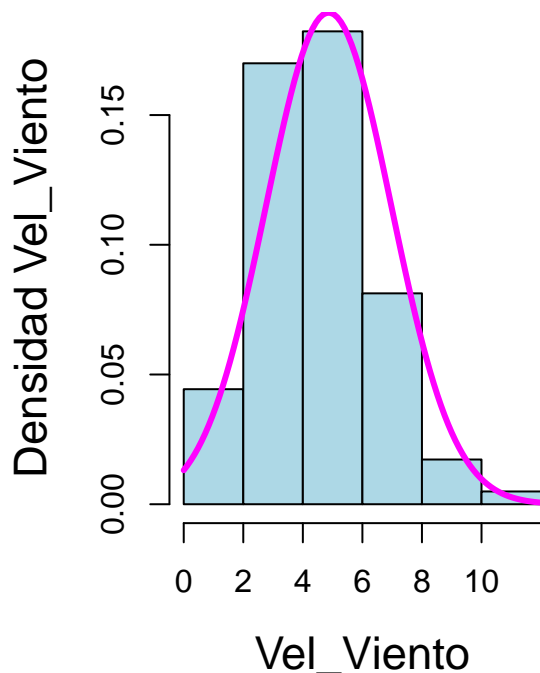
Vemos si hay registros atípicos

```
boxplot.stats(Vel_Viento)$out
```

```
## [1] 11 11
```

Como podemos ver existen 2 registros atípicos

```
par(mfrow=c(1,2))
hist(Vel_Viento, breaks=5, freq=FALSE, main = "", xlab="Vel_Viento",
     cex.lab=1.4, ylab = "Densidad Vel_Viento", col = "lightblue")
curve( dnorm(x, mean=mean(Vel_Viento), sd=sd(Vel_Viento)),
      col="magenta", lwd=3, add=TRUE)
boxplot(Vel_Viento, main = "", xlab="Vel_Viento",
      cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
      horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable Humedad :

```
summary(Humedad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  19.00   46.00   64.00   57.61   73.00   93.00
```

Desviación típica y rango intercuartílico:

```
sd(Humedad)
```

```
## [1] 20.84766
```

```
IQR(Humedad)
```

```
## [1] 27
```

Evaluamos la asimetría y kurtois

```
library(moments)
```

```
skewness(Humedad, na.rm = FALSE)
```

```
## [1] -0.6935066
```

```
kurtosis(Humedad, na.rm = FALSE)
```

```
## [1] 2.307891
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es próximo a tres, las colas de la variable son similares a las de una normal.

Vemos si hay registros atípicos

```
boxplot.stats(Humedad)$out
```

```
## integer(0)
```

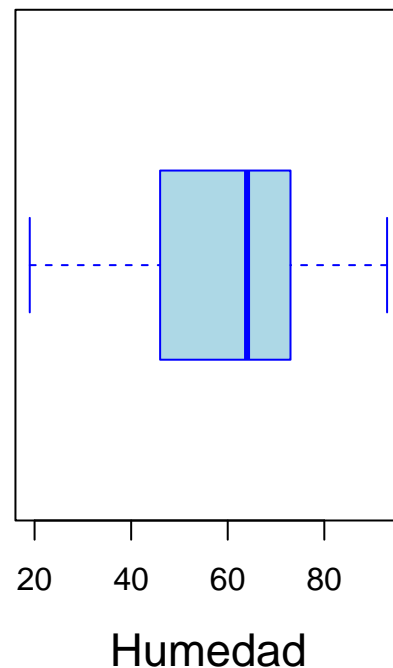
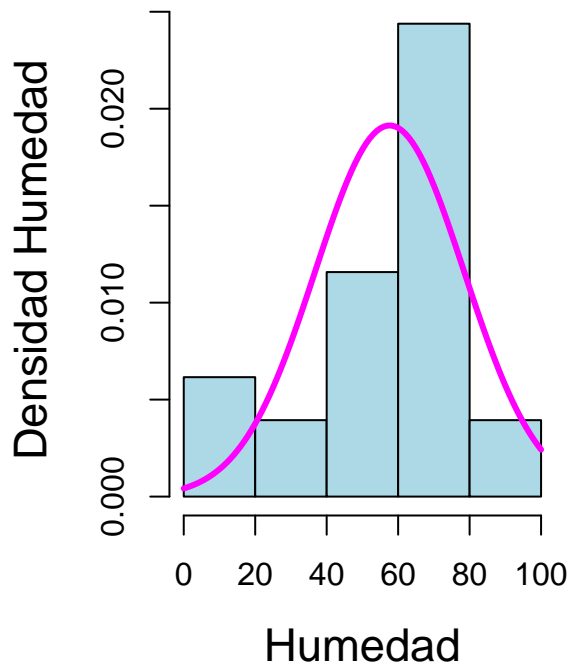
Como podemos ver no existen registros atípicos

```
par(mfrow=c(1,2))
```

```
hist(Humedad, breaks=5, freq=FALSE, main = "", xlab="Humedad",  
      cex.lab=1.4, ylab = "Densidad Humedad", col = "lightblue")
```

```
curve( dnorm(x, mean=mean(Humedad), sd=sd(Humedad)),  
       col="magenta", lwd=3, add=TRUE)
```

```
boxplot(Humedad, main = "", xlab="Humedad",  
        cex.lab=1.4, border = "blue", col= "lightblue", pch="+",  
        horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable T\_Sandburg :

```
summary(T_Sandburg)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  25.00  51.50   61.00   61.11  71.00   93.00
```

Desviación típica y rango intercuartílico:

```
sd(T_Sandburg)
```

```
## [1] 14.20647
```

```
IQR(T_Sandburg)
```

```
## [1] 19.5
```

Evaluamos la asimetría y kurtois

```
library(moments)
```

```
skewness(T_Sandburg, na.rm = FALSE)
```

```
## [1] 0.006212875
```

```
kurtosis(T_Sandburg, na.rm = FALSE)
```

```
## [1] 2.510297
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es próximo a tres, las colas de la variable son similares a las de una normal.

Vemos si hay registros atípicos

```
boxplot.stats(T_Sandburg)$out
```

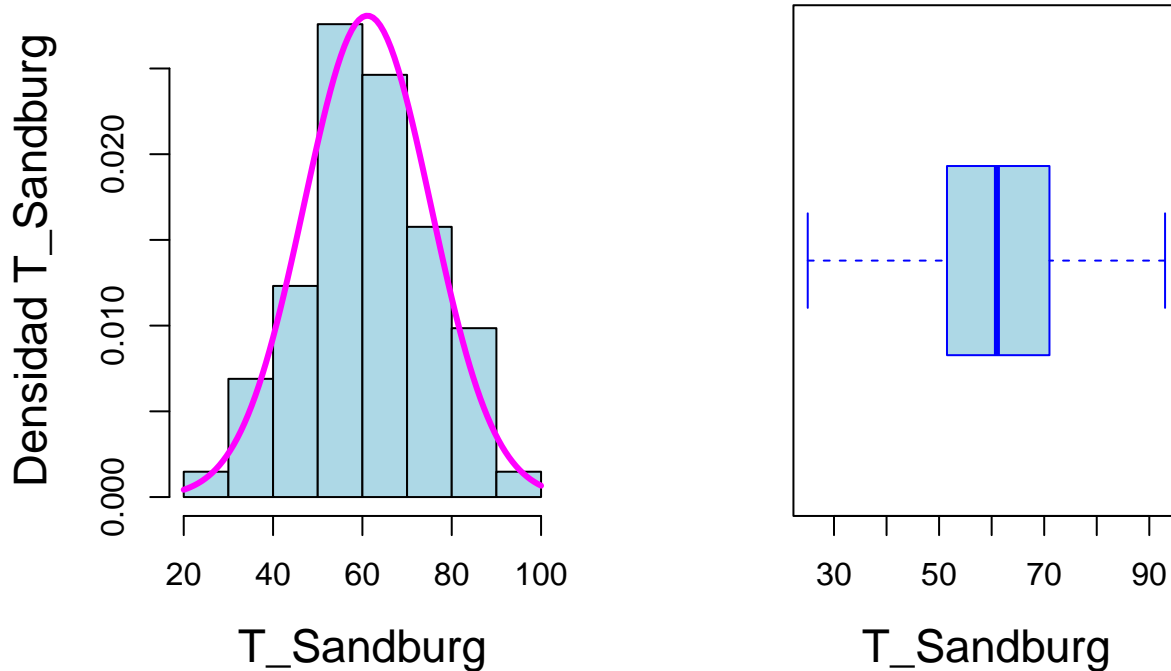
```
## integer(0)
```

Como podemos ver no existen registros atípicos

```

par(mfrow=c(1,2))
hist(T_Sandburg, breaks=5,freq=FALSE, main = "", xlab="T_Sandburg",
     cex.lab=1.4, ylab = "Densidad T_Sandburg", col = "lightblue")
curve( dnorm(x,mean=mean(T_Sandburg),sd=sd(T_Sandburg)),
      col="magenta", lwd=3, add=TRUE)
boxplot(T_Sandburg, main = "", xlab="T_Sandburg",
       cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
       horizontal = TRUE, cex=3)

```



- ANÁLISIS DESCRIPTIVO VARIABLE 'T\_ElMonte'

```
summary(T_ElMonte)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  27.68  49.64   56.48   56.54  66.20   82.58
```

Desviación típica y rango intercuartílico:

```
sd(T_ElMonte)
```

```
## [1] 11.74267
```

```
IQR(T_ElMonte)
```

```
## [1] 16.56
```

Evaluamos la asimetría y kurtois

```
library(moments)
```

```
skewness(T_ElMonte, na.rm = FALSE)
```

```
## [1] -0.1025587
```

```
kurtosis(T_ElMonte, na.rm = FALSE)
```

```
## [1] 2.486231
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es próximo a tres, las colas de la variable son similares a las de una normal.

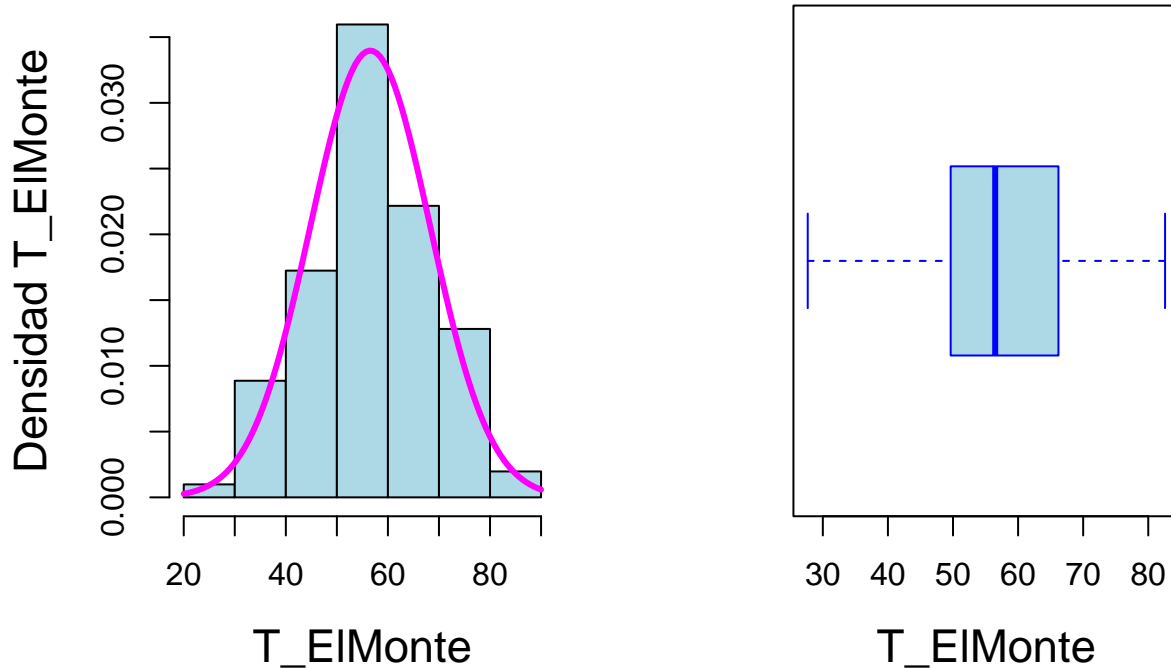
Vemos si hay registros atípicos

```
boxplot.stats(T_ElMonte)$out
```

```
## numeric(0)
```

Como podemos ver no existen registros atípicos

```
par(mfrow=c(1,2))
hist(T_ElMonte, breaks=5, freq=FALSE, main = "", xlab="T_ElMonte",
     cex.lab=1.4, ylab = "Densidad T_ElMonte", col = "lightblue")
curve( dnorm(x, mean=mean(T_ElMonte), sd=sd(T_ElMonte)),
      col="magenta", lwd=3, add=TRUE)
boxplot(T_ElMonte, main = "", xlab="T_ElMonte",
      cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
      horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable Inv\_Alt\_b :

```
summary(Inv_Alt_b)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      111    869    2083    2602    5000    5000
```

Desviación típica y rango intercuartílico:

```
sd(Inv_Alt_b)
```

```
## [1] 1859.889
```

```
IQR(Inv_Alt_b)
```

```
## [1] 4131
```

Evaluamos la asimetría y kurtosis

```
library(moments)
skewness(Inv_Alt_b, na.rm = FALSE)
```

```
## [1] 0.2355015
```

```
kurtosis(Inv_Alt_b, na.rm = FALSE)
```

```
## [1] 1.374057
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es menor a tres, las colas de la variable son más ligeras a las de una normal.

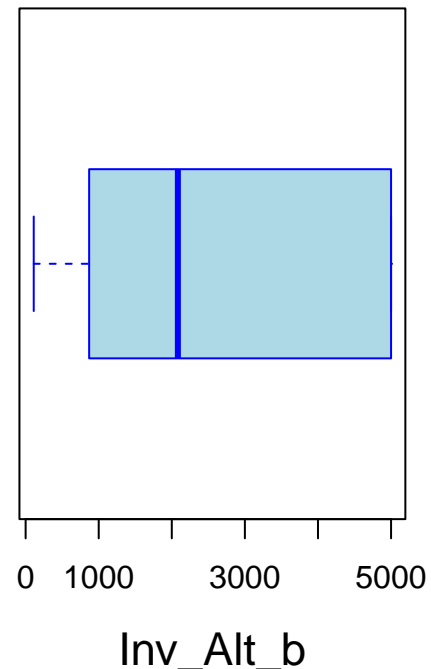
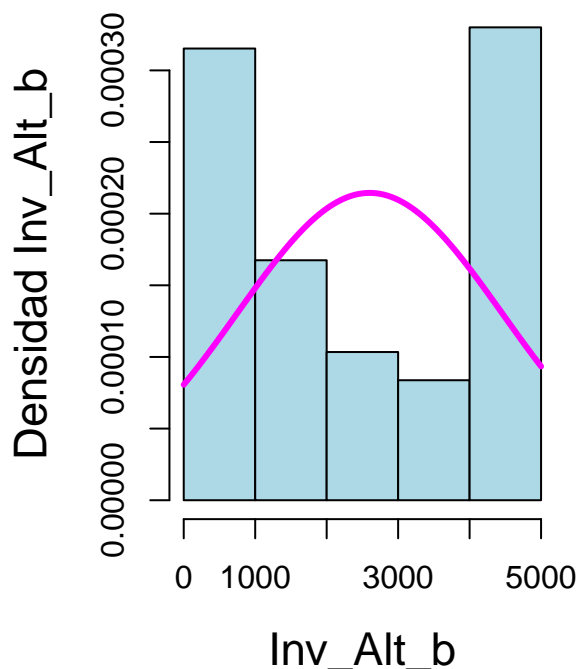
Vemos si hay registros atípicos

```
boxplot.stats(Inv_Alt_b)$out
```

```
## integer(0)
```

Como podemos ver no existen registros atípicos

```
par(mfrow=c(1,2))
hist(Inv_Alt_b, breaks=5, freq=FALSE, main = "", xlab="Inv_Alt_b",
     cex.lab=1.4, ylab = "Densidad Inv_Alt_b", col = "lightblue")
curve( dnorm(x, mean=mean(Inv_Alt_b), sd=sd(Inv_Alt_b)),
       col="magenta", lwd=3, add=TRUE)
boxplot(Inv_Alt_b, main = "", xlab="Inv_Alt_b",
        cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
        horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable Grad\_Pres :

```
summary(Grad_Pres)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -69.00  -14.00   18.00   14.43  43.00  107.00
```

Desviación típica y rango intercuartílico:

```
sd(Grad_Pres)
```

```
## [1] 36.3172
```

```
IQR(Grad_Pres)
```

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

Evaluamos la asimetría y kurtoisis

```
library(moments)
```

```
skewness(Grad_Pres, na.rm = FALSE)
```

```
## [1] -0.131977
```

```
kurtosis(Grad_Pres, na.rm = FALSE)
```

```
## [1] 2.316879
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es menor a tres, las colas de la variable son más ligeras a las de una normal.

Vemos si hay registros atípicos

```
boxplot.stats(Grad_Pres)$out
```

```
## integer(0)
```

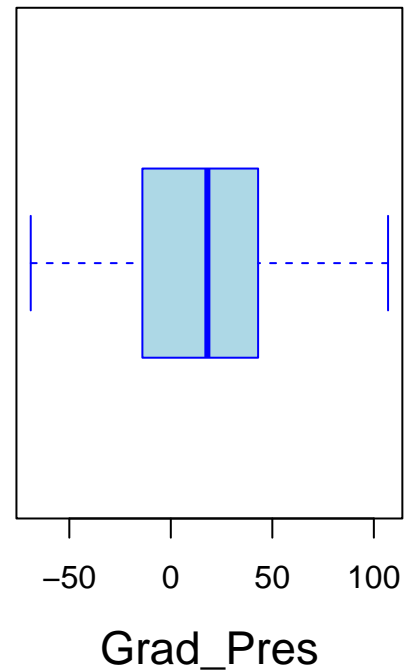
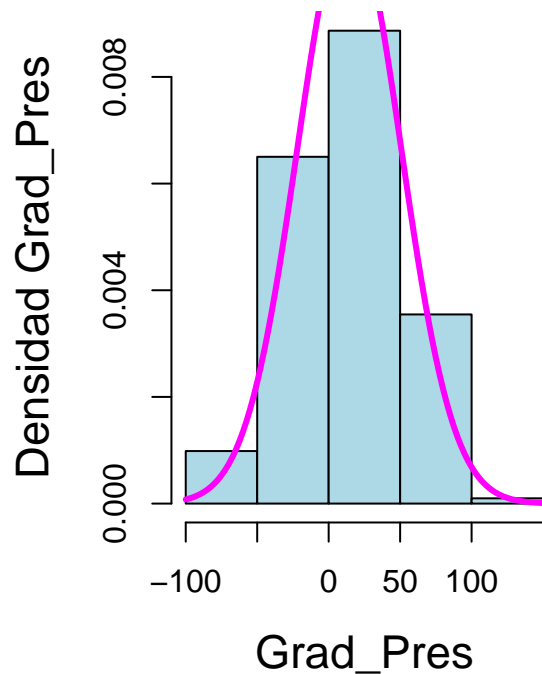
Como podemos ver no existen registros atípicos

```
par(mfrow=c(1,2))
```

```
hist(Grad_Pres, breaks=5, freq=FALSE, main = "", xlab="Grad_Pres",  
     cex.lab=1.4, ylab = "Densidad Grad_Pres", col = "lightblue")
```

```
curve( dnorm(x, mean=mean(Grad_Pres), sd=sd(Grad_Pres)),  
       col="magenta", lwd=3, add=TRUE)
```

```
boxplot(Grad_Pres, main = "", xlab="Grad_Pres",  
        cex.lab=1.4, border = "blue", col= "lightblue", pch="+",  
        horizontal = TRUE, cex=3)
```



Análisis descriptivo de la variable Inv\_T\_b :

```
summary(Inv_T_b)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  27.50  51.26   60.98   60.69  70.88   90.68
```

Desviación típica y rango intercuartílico:

```
sd(Inv_T_b)
```

```
## [1] 14.12473
```

```
IQR(Inv_T_b)
```

```
## [1] 19.62
```

Evaluamos la asimetría y kurtois

```
library(moments)
```

```
skewness(Inv_T_b, na.rm = FALSE)
```

```
## [1] -0.1886259
```

```
kurtosis(Inv_T_b, na.rm = FALSE)
```

```
## [1] 2.354789
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis es menor a tres, las colas de la variable son más ligeras a las de una normal.

Vemos si hay registros atípicos

```
boxplot.stats(Inv_T_b)$out
```

```
## numeric(0)
```

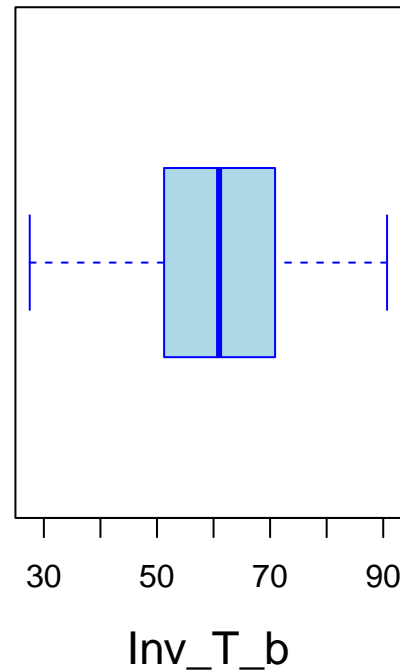
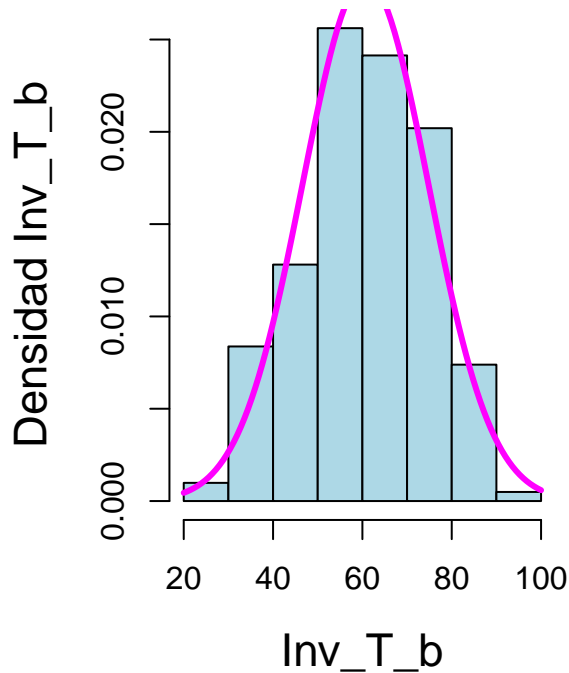
Como podemos ver no existen registros atípicos



```

par(mfrow=c(1,2))
hist(Inv_T_b, breaks=5, freq=FALSE, main = "", xlab="Inv_T_b",
     cex.lab=1.4, ylab = "Densidad Inv_T_b", col = "lightblue")
curve( dnorm(x, mean=mean(Inv_T_b), sd=sd(Inv_T_b)),
       col="magenta", lwd=3, add=TRUE)
boxplot(Inv_T_b, main = "", xlab="Inv_T_b",
        cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
        horizontal = TRUE, cex=3)

```



Análisis descriptivo de la variable Visibilidad :

```
summary(Visibilidad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   60.0   100.0   122.2   150.0   350.0
```

Desviación típica y rango intercuartílico:

```
sd(Visibilidad)
```

```
## [1] 81.17132
```

```
IQR(Visibilidad)
```

```
## [1] 90
```

Evaluamos la asimetría y kurtosis

```
library(moments)
```

```
skewness(Visibilidad, na.rm = FALSE)
```

```
## [1] 0.8067613
```

```
kurtosis(Visibilidad, na.rm = FALSE)
```

```
## [1] 2.903426
```

Podemos ver que al ser el coeficiente de asimetría cercano a 0 que puede ser una variable simétrica y al ser el coeficiente de Kurtosis próximo a tres, las colas de la variable son próximas a las de una normal.

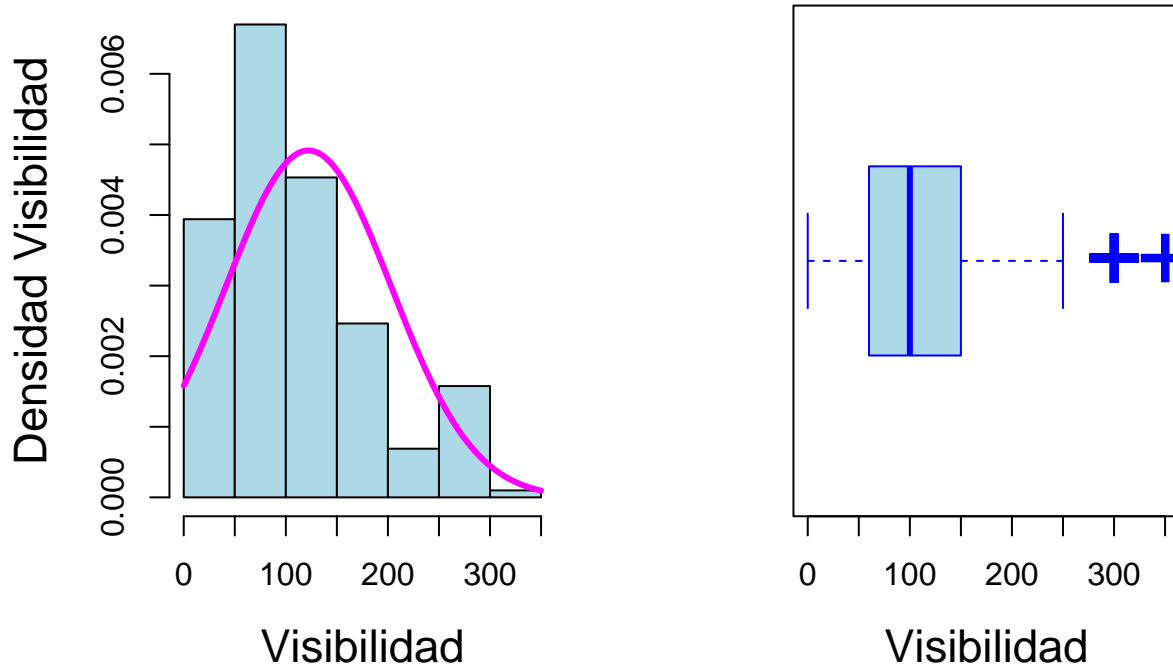
Vemos si hay registros atípicos

```
boxplot.stats(Visibilidad)$out
```

```
## [1] 350 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300
```

Como podemos ver no existen registros atípicos

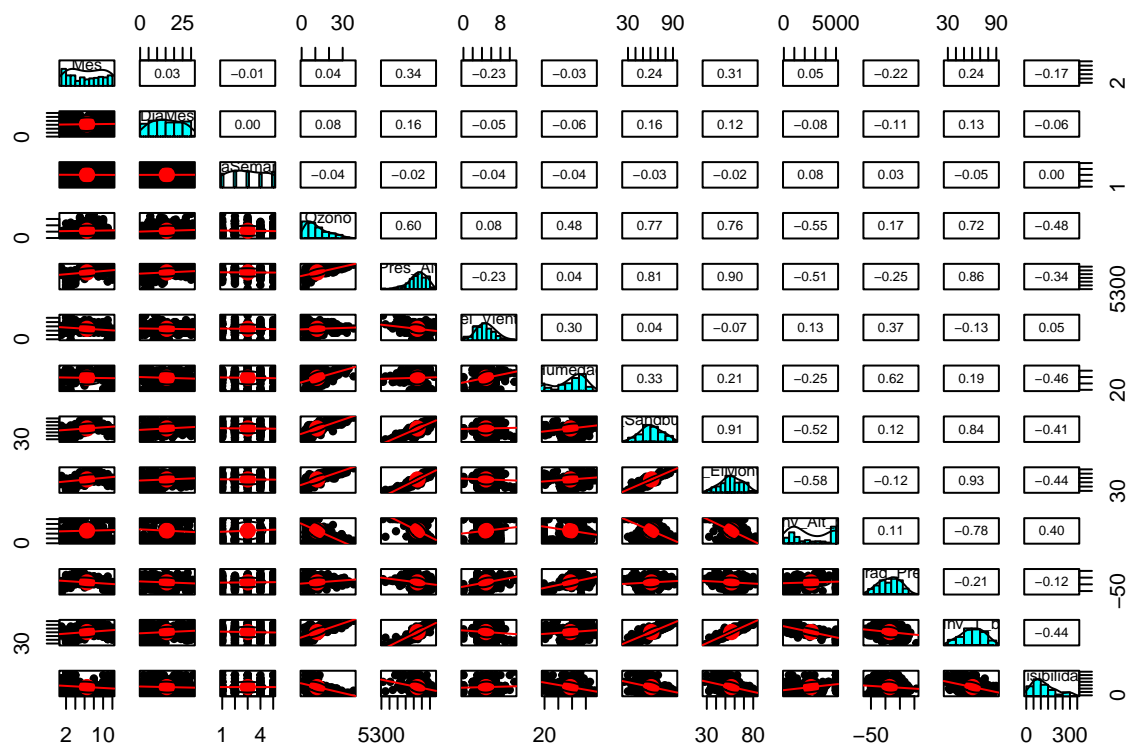
```
par(mfrow=c(1,2))
hist(Visibilidad, breaks=5, freq=FALSE, main = "", xlab="Visibilidad",
     cex.lab=1.4, ylab = "Densidad Visibilidad", col = "lightblue")
curve( dnorm(x, mean=mean(Visibilidad), sd=sd(Visibilidad)),
      col="magenta", lwd=3, add=TRUE)
boxplot(Visibilidad, main = "", xlab="Visibilidad",
        cex.lab=1.4, border = "blue", col= "lightblue", pch="+",
        horizontal = TRUE, cex=3)
```



## 2. Análisis de correlación

- Correlaciones simples bivariantes (análisis gráfico y numérico):

```
library(psych)
pairs.panels(OzonoLA, smooth = TRUE, density=TRUE, digits = 2,
            ellipses=TRUE, method="pearson", pch = 20,
            lm=TRUE, cor=TRUE)
```



`cor(OzonoLA)`

```
##          Mes      DiaMes      DiaSemana      Ozono      Pres_Alt
## Mes      1.000000000  0.029780944 -6.406562e-03  0.04417525  0.33793183
## DiaMes   0.029780944  1.000000000  3.418381e-03  0.08364060  0.15808064
## DiaSemana -0.006406562  0.003418381  1.000000e+00 -0.03750993 -0.02206218
## Ozono     0.044175248  0.083640605 -3.750993e-02  1.00000000  0.59612683
## Pres_Alt  0.337931827  0.158080640 -2.206218e-02  0.59612683  1.00000000
## Vel_Viento -0.226893006 -0.046090839 -3.667633e-02  0.08179858 -0.23161673
## Humedad   -0.034727288 -0.064739863 -3.855381e-02  0.47947091  0.03869121
## T_Sandburg  0.235445072  0.157156363 -3.035349e-02  0.77335204  0.80633038
## T_ElMonte  0.314323892  0.117127229 -2.481044e-02  0.76001956  0.89689385
## Inv_Alt_b   0.045305170 -0.082352709  7.998485e-02 -0.55196217 -0.50891157
## Grad_Pres  -0.218837079 -0.111239793  3.418479e-02  0.17391799 -0.24549047
## Inv_T_b     0.236540625  0.127530054 -5.365959e-02  0.71756186  0.85642134
## Visibilidad -0.167796386 -0.057896954 -8.572216e-06 -0.47629112 -0.34272720
##          Vel_Viento      Humedad      T_Sandburg      T_ElMonte      Inv_Alt_b
## Mes      -0.22689301 -0.03472729  0.23544507  0.31432389  0.04530517
## DiaMes   -0.04609084 -0.06473986  0.15715636  0.11712723 -0.08235271
## DiaSemana -0.03667633 -0.03855381 -0.03035349 -0.02481044  0.07998485
## Ozono     0.08179858  0.47947091  0.77335204  0.76001956 -0.55196217
## Pres_Alt  -0.23161673  0.03869121  0.80633038  0.89689385 -0.50891157
## Vel_Viento  1.00000000  0.30356343  0.04122208 -0.06983510  0.12834881
## Humedad    0.30356343  1.00000000  0.33132296  0.21158607 -0.24703914
## T_Sandburg  0.04122208  0.33132296  1.00000000  0.91396229 -0.51539621
## T_ElMonte  -0.06983510  0.21158607  0.91396229  1.00000000 -0.57965832
## Inv_Alt_b   0.12834881 -0.24703914 -0.51539621 -0.57965832  1.00000000
## Grad_Pres  0.37328762  0.62433536  0.11765666 -0.12091597  0.11350236
## Inv_T_b    -0.12959891  0.19101936  0.84310310  0.93080989 -0.78286145
## Visibilidad 0.04534341 -0.45750232 -0.41038641 -0.43897902  0.39669789
##          Grad_Pres      Inv_T_b      Visibilidad
```

```
## Mes      -0.21883708  0.23654062 -1.677964e-01
## DiaMes    -0.11123979  0.12753005 -5.789695e-02
## DiaSemana  0.03418479 -0.05365959 -8.572216e-06
## Ozono      0.17391799  0.71756186 -4.762911e-01
## Pres_Alt   -0.24549047  0.85642134 -3.427272e-01
## Vel_Viento 0.37328762 -0.12959891  4.534341e-02
## Humedad    0.62433536  0.19101936 -4.575023e-01
## T_Sandburg 0.11765666  0.84310310 -4.103864e-01
## T_ElMonte  -0.12091597  0.93080989 -4.389790e-01
## Inv_Alt_b   0.11350236 -0.78286145  3.966979e-01
## Grad_Pres   1.00000000 -0.20663872 -1.200549e-01
## Inv_T_b     -0.20663872  1.00000000 -4.377177e-01
## Visibilidad -0.12005488 -0.43771768  1.000000e+00
```

- Correlaciones parciales:

```
partial.r(OzonoLA)
```

```
##           Mes      DiaMes      DiaSemana      Ozono      Pres_Alt
## Mes      1.000000000 -0.01473632 -0.029646884 -0.239632308 -0.008364478
## DiaMes   -0.014736319  1.000000000  0.017131467  0.023224469  0.074079502
## DiaSemana -0.029646884  0.01713147  1.000000000 -0.015463849 -0.014083279
## Ozono     -0.239632308  0.02322447 -0.015463849  1.000000000 -0.134822542
## Pres_Alt  -0.008364478  0.07407950 -0.014083279 -0.134822542  1.000000000
## Vel_Viento -0.192898039  0.01519492 -0.052672027 -0.040039195 -0.292700944
## Humedad    0.160860221 -0.03992322 -0.050358261  0.262774072 -0.095321178
## T_Sandburg 0.008578204  0.20842819 -0.037515653  0.141155532  0.108888567
## T_ElMonte  0.131026789 -0.12847809  0.050717722  0.312487718  0.344311253
## Inv_Alt_b   0.230043843 -0.02868566  0.036820690 -0.111064127  0.120880379
## Grad_Pres  -0.127208517 -0.13665426  0.068684046  0.001780773 -0.044096421
## Inv_T_b     0.048692150 -0.02999001 -0.008230412 -0.076866881  0.140848869
## Visibilidad -0.108506988 -0.06279200 -0.037003418 -0.074160846  0.014979648
##           Vel_Viento      Humedad      T_Sandburg      T_ElMonte      Inv_Alt_b
## Mes      -0.19289804  0.16086022  0.008578204  0.13102679  0.23004384
## DiaMes    0.01519492 -0.03992322  0.208428191 -0.12847809 -0.02868566
## DiaSemana -0.05267203 -0.05035826 -0.037515653  0.05071772  0.03682069
## Ozono     -0.04003920  0.26277407  0.141155532  0.31248772 -0.11106413
## Pres_Alt  -0.29270094 -0.09532118  0.108888567  0.34431125  0.12088038
## Vel_Viento 1.00000000  0.15651029  0.089387359  0.11902520  0.11170466
## Humedad    0.15651029  1.00000000 -0.044727403 -0.04353431 -0.05762633
## T_Sandburg 0.08938736 -0.04472740  1.000000000  0.35489823  0.18928541
## T_ElMonte  0.11902520 -0.04353431  0.354898232  1.00000000  0.39942102
## Inv_Alt_b   0.11170466 -0.05762633  0.189285412  0.39942102  1.00000000
## Grad_Pres  0.05542912  0.50554293  0.498084949 -0.05195235 -0.15571589
## Inv_T_b     0.01217894  0.06712657  0.229456614  0.57959707 -0.81884177
## Visibilidad 0.11148387 -0.32142715  0.085393863 -0.12200008  0.09905698
##           Grad_Pres      Inv_T_b      Visibilidad
## Mes      -0.127208517  0.048692150 -0.10850699
## DiaMes   -0.136654263 -0.029990011 -0.06279200
## DiaSemana 0.068684046 -0.008230412 -0.03700342
## Ozono     0.001780773 -0.076866881 -0.07416085
## Pres_Alt  -0.044096421  0.140848869  0.01497965
## Vel_Viento 0.055429122  0.012178940  0.11148387
## Humedad    0.505542925  0.067126570 -0.32142715
## T_Sandburg 0.498084949  0.229456614  0.08539386
```

```
## T_ElMonte -0.051952353 0.579597071 -0.12200008
## Inv_Alt_b -0.155715887 -0.818841765 0.09905698
## Grad_Pres 1.000000000 -0.326942874 0.01948577
## Inv_T_b -0.326942874 1.000000000 0.03558761
## Visibilidad 0.019485768 0.035587611 1.00000000
```

### 3. Modelo matemático

$$\mathbb{E}(\vec{Y}|\mathbf{X}) = \beta_0 + \sum_{i=1}^n \beta_i X_{ij} \quad (1)$$

```
ajuste <- lm(Ozono~., data=OzonoLA)
ajuste
```

```
##
## Call:
## lm(formula = Ozono ~ ., data = OzonoLA)
##
## Coefficients:
## (Intercept)      Mes      DiaMes      DiaSemana      Pres_Alt      Vel_Viento
## 55.4279486    -0.3431326    0.0120308   -0.0473689   -0.0133495   -0.0959961
## Humedad      T_Sandburg      T_ElMonte      Inv_Alt_b      Grad_Pres      Inv_T_b
## 0.0880372     0.1366231     0.5597690   -0.0006176    0.0003624   -0.1244500
## Visibilidad
## -0.0049469
```

## Regresión Logística

- Antes de empezar, cargamos los datos *Oro.rda*

```
load("Datos/Oro.rda")
attach(Oro)
explicativas.oro <- Oro[,1:3]      # Almacenamos las explicativas
respuesta.oro <- Proximidad        # Almacenamos la variable de respuesta
```

### 1. Análisis descriptivo

Para el análisis descriptivo de las variables podemos comenzar con una visión general de las variables mediante las funciones `str()` y `summary()`.

```
str(Oro)
```

```
## 'data.frame':   64 obs. of  4 variables:
## $ As          : num  6.77 15.03 6.43 0.1 0.1 ...
## $ Sb          : num  3.08 6.15 2.35 0.3 0.3 9.62 0.51 3.71 4.32 0.8 ...
## $ Corredor    : int   1 1 1 0 0 1 0 1 0 0 ...
## $ Proximidad  : int   1 1 1 0 0 1 0 1 0 0 ...
```

La salida de `str()` nos dice que los datos constan de 64 observaciones de 4 variables:

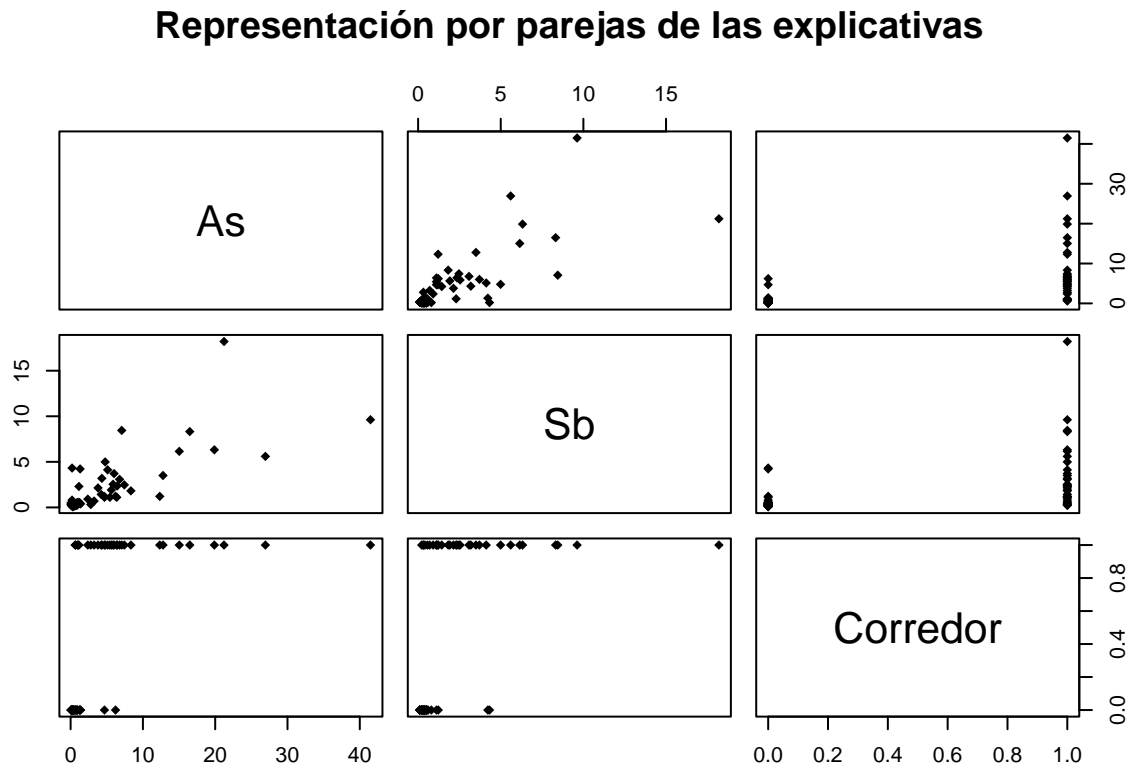
- **As**: Nivel de concentración de arsénico en la muestra de agua. (numérica)
- **Sb**: Nivel de concentración de antimonio en la muestra de agua. (numérica)
- **Corredor**: Variable binaria indicando si la zona muestreada está (1) o no está (0) en alguno de los corredores delimitados por las líneas sobre el mapa. (categórica)

Más la variable de respuesta Proximidad, que toma los valores 1 o 0 según que el depósito esté próximo o esté muy lejano al lugar.

```
summary(Oro)
```

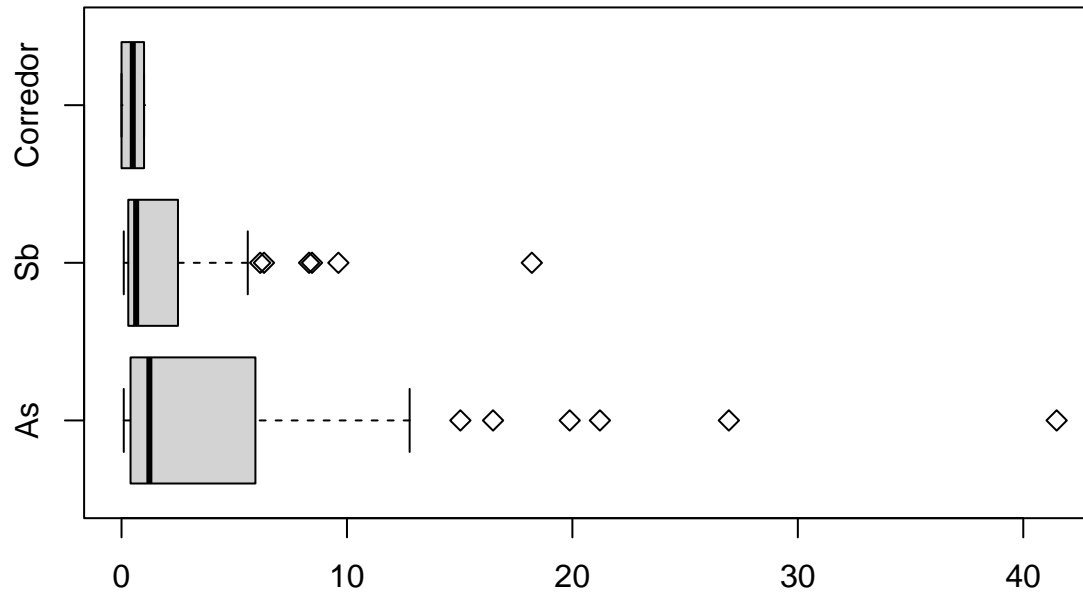
```
##           As           Sb           Corredor           Proximidad
##  Min.    : 0.100   Min.    : 0.100   Min.    :0.0   Min.    :0.0000
## 1st Qu.: 0.400   1st Qu.: 0.300   1st Qu.:0.0   1st Qu.:0.0000
## Median : 1.235   Median : 0.650   Median :0.5   Median :0.0000
## Mean   : 4.645   Mean    : 2.039   Mean    :0.5   Mean    :0.4375
## 3rd Qu.: 5.905   3rd Qu.: 2.487   3rd Qu.:1.0   3rd Qu.:1.0000
## Max.    :41.480   Max.    :18.200   Max.    :1.0   Max.    :1.0000
```

```
plot(explicativas.oro, pch=18,
     main="Representación por parejas de las explicativas")
```



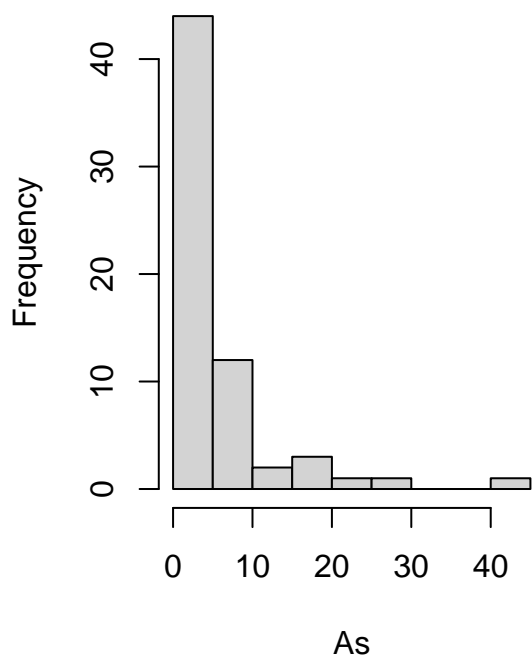
```
boxplot(explicativas.oro, horizontal=T, pch=5,
       main="Diagrama de cajas de las explicativas")
```

## Diagrama de cajas de las explicativas

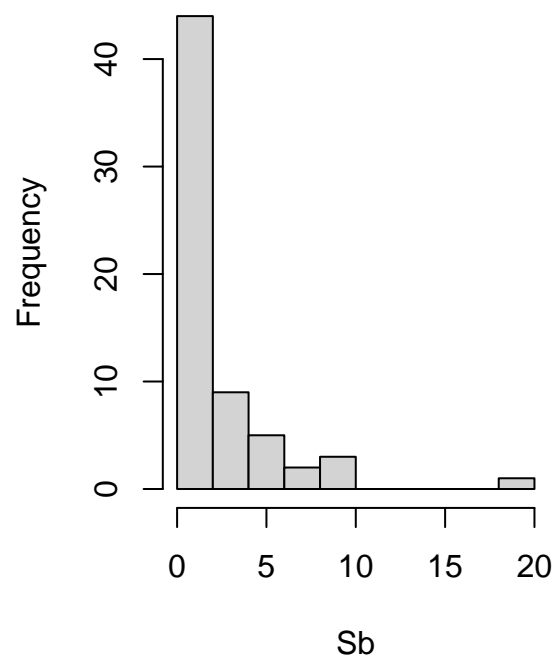


```
old.par <- par(mfrow=c(1,2))
hist(As, main="Concentración de Arsénico")
hist(Sb, main="Concentración de Antimonio")
```

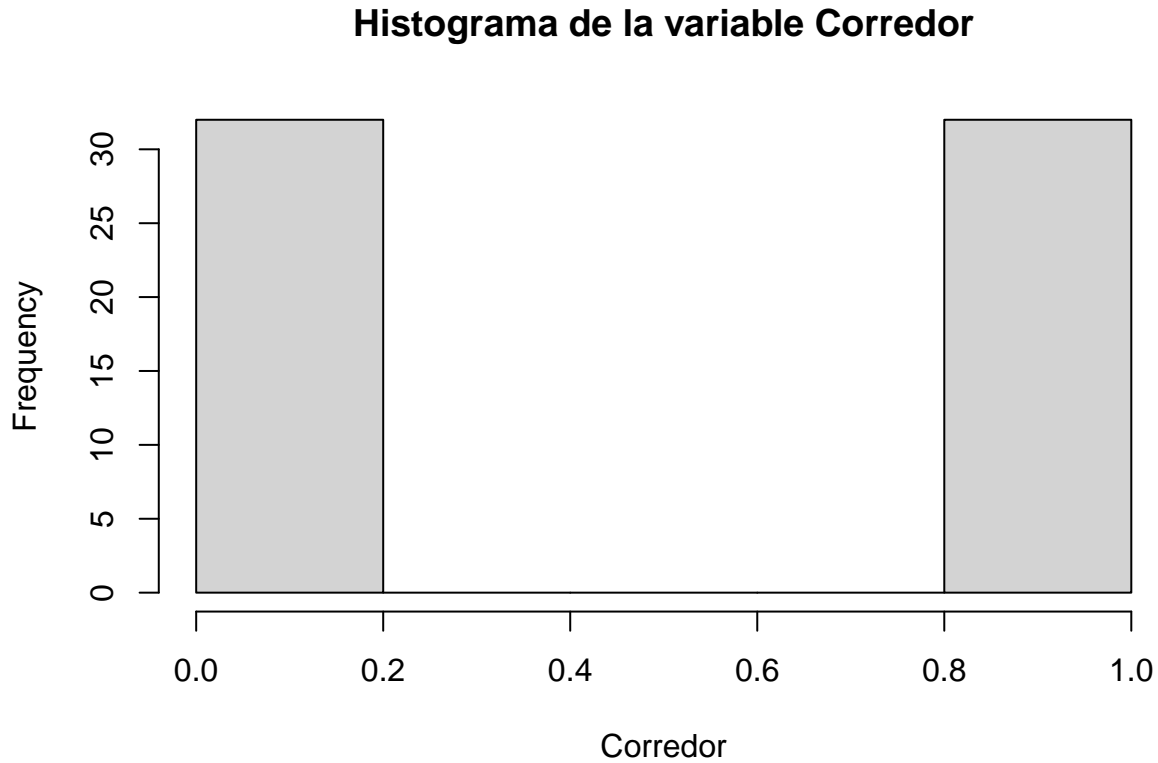
**Concentración de Arsénico**



**Concentración de Antimonio**



```
par(old.par)
hist(Corredor, main="Histograma de la variable Corredor")
```



## 2. Modelo matemático

Dado que la variable de respuesta, *Proximidad*, es binaria (0 o 1), deberemos de elegir un modelo que tenga esto en cuenta. En nuestro caso hemos elegido una transformación del modelo lineal, definida por la distribución logística de la ecuación 2

$$F(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} \quad (2)$$

Por tanto, nuestro modelo logístico quedaría de la forma

$$\mathbb{E}(Y|\vec{X}_i) = p_i = \mathbb{P}(Y = 1|\vec{X}_i) = \frac{1}{1 + e^{-\eta}} \quad (3)$$

tal que  $\eta = \vec{\beta}^t \vec{X}_i$ . Además,

$$1 - p_i = \mathbb{P}(Y = 0|\vec{X}_i) = 1 - \frac{1}{1 + e^{-\eta}} = \frac{e^{-\eta}}{1 + e^{-\eta}} \quad (4)$$



### 3. Interpretación del modelo

Para una mejor interpretación del modelo, podemos definir el **odds**<sub>*i*</sub> de manera que

$$odds_i = odds(Y|\vec{X}_i) = \frac{p_i}{1-p_i} = e^\eta = e^{\vec{\beta}^t \vec{X}_i} = e^{\beta_0} e^{\beta_1 X_{i1}} \dots e^{\beta_k X_{ik}} = e^{\beta_0} \prod_{j=1}^k e^{\beta_j X_{ij}} \quad , \quad 1 \leq i \leq n \quad (5)$$

Este es un modelo multiplicativo, en el cual  $e^{\beta_0}$  es la respuesta cuando  $\vec{X}_i = \vec{0}$ , mientras que  $e^{\beta_j}$ , para  $1 \leq j \leq k$ , es el incremento multiplicativo  $(e^{\beta_j})^l$  en el odds para algún incremento  $l$  en  $X_j$

También podemos expresar el modelo aplicando logaritmos a la ecuación 5, de manera que

$$\ln\left(\frac{p_i}{1-p_i}\right) = \eta = \vec{\beta}^t \vec{X}_i \quad (6)$$

Los cuales denominaremos como **logit**<sub>*i*</sub>. Estos logits son interpretables mucho más fácilmente, aunque debido a que

### 4. Inferencia

### 5. Bondad del ajuste