Mínimos cuadrados

J. Abellán

06/11/2015

Mínimos cuadrados.

- Hemos adquirido N pares de datos $x_i, y_i, \; i = 1, 2, ..., N$.
- Conocemos la ley que liga la variable dependiente o respuesta del sistema y de la variable independiente o de control x. Supondremos en este primer ejercicio que la variable x no presenta ningún error; dicho de otro modo: no es una variable aleatoria.
- Para empezar consideremos el caso más sencillo, una relación lineal entre causa x y efecto y:

$$y = y(\; x; \; a, b \;) = a + b \; x$$

donde a, b son los parámetros de la teoría a inferir de los datos experimentales.

- **¿Cómo se hace dicha inferencia?**
- La respuesta *mínimos cuadrados* es muy sencilla e intuitiva: se construye la función de los parámetros a, b

$$Q = Q(\; a, b \;) = \Sigma_1^N (\; y_i - (\; a + b \; x_i \;) \;)^2$$

y se buscan los valores de a, b que hacen mínima dicha función. Es decir, la recta que mejor se ajusta a los datos.

- Supongamos que no sabemos derivar: **¿cómo encontramos dicho mínimo?**

La respuesta es sencilla: a base de potencia de cálculo.

Veamos para ello el ejemplo $9.2a$ del *Ross*. Se trata de ver el contenido en agua y de cierto material en función de la humedad relativa x del lugar de almacenaje. Se ha hecho $n = 15$ medidas:

```
library("latex2exp", lib.loc="~/R/i686-pc-linux-gnu-library/3.2")

x <- c( 46, 53, 29, 61, 36, 39, 47, 49, 52, 38, 55, 32, 57, 54, 44 )

y <- c( 12, 15, 7, 17, 10, 11, 11, 12, 14, 9, 16, 8, 18, 14, 12 )

cbind( x, y )
```

```
##        x  y
##  [1,] 46 12
##  [2,] 53 15
##  [3,] 29  7
##  [4,] 61 17
##  [5,] 36 10
##  [6,] 39 11
##  [7,] 47 11
##  [8,] 49 12
##  [9,] 52 14
```

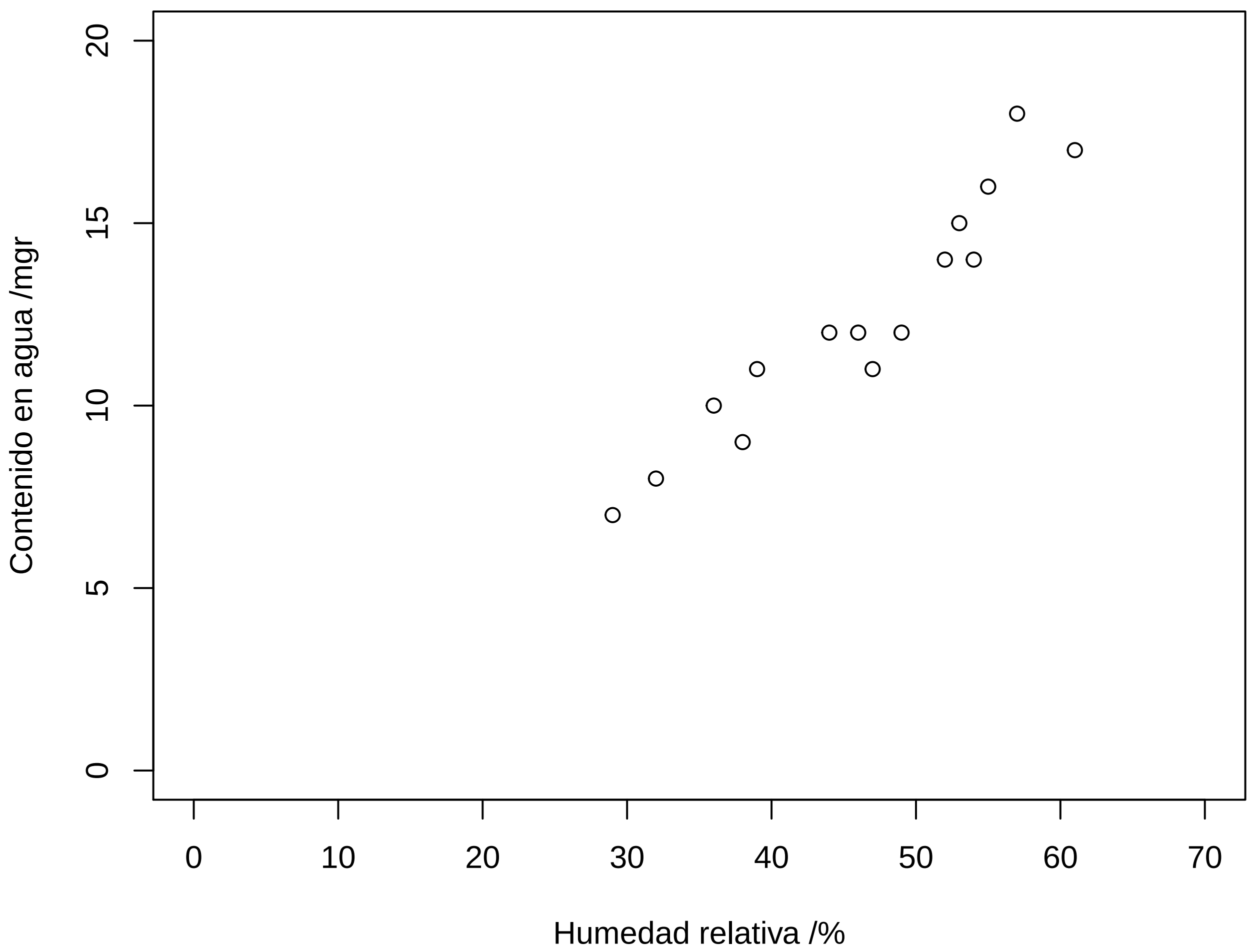
```
## [10,] 38  9
## [11,] 55 16
## [12,] 32  8
## [13,] 57 18
## [14,] 54 14
## [15,] 44 12
```

Al ver los datos representados gráficamente se sospecha rápidamente una dependencia lineal entre x e y.

Así pues, proponemos $y = y(x) = a + bx$ y buscamos los valores de a y b que determinan la recta que mejor se ajusta a los datos.

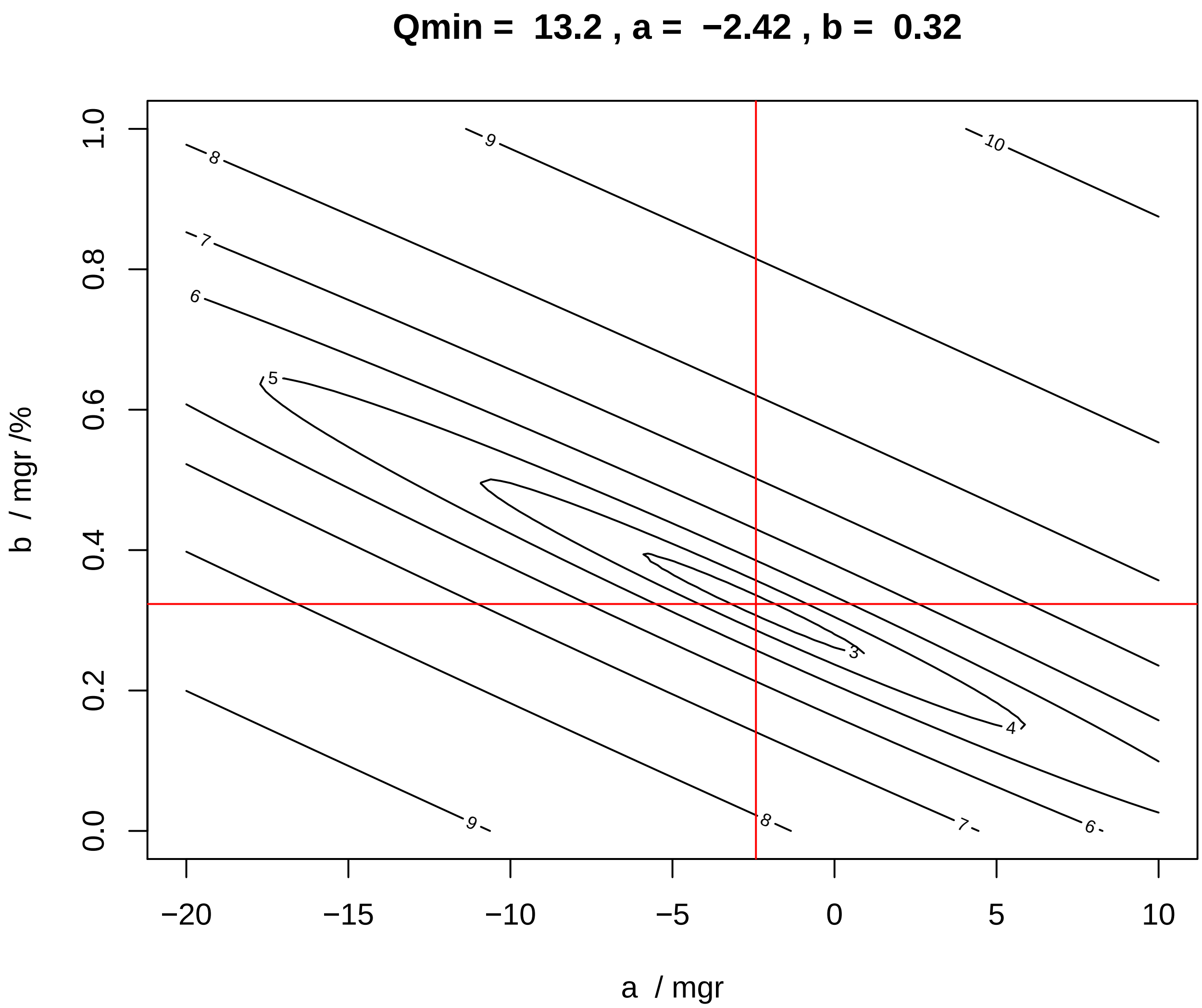
A simple vista se ve que $b \sim 1/3$ y $a \sim -5$

```
plot( x, y,

      xlab = " Humedad relativa /% ",

      ylab = " Contenido en agua /mgr ",

      xlim = c( 0, 70 ),

      ylim = c( 0, 20 )

    )
```



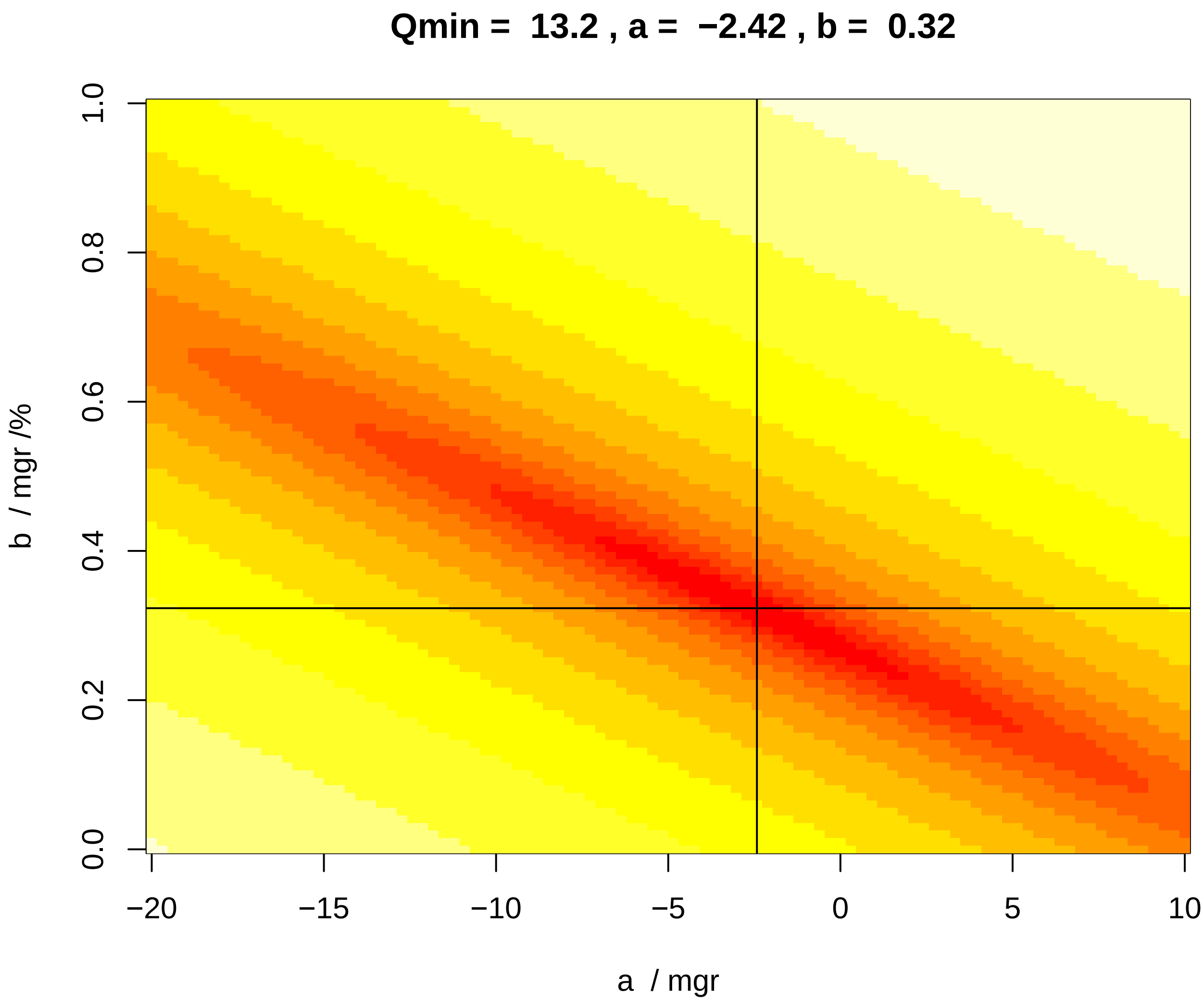
```r
#Espacio de parámetros
n <- 100

#Barrido en a
A <- seq( - 20, 10, len = n )

#Barrido en b
B <- seq( 0, 1, len = n )

#Defino la matriz Q para guardar los valores suma de los caudrados
Q <- matrix( 0, nrow = n, ncol = n )

Qmin <- 1e8

for (i in 1 : n ) {

  #Fijo un valor del parámetro a
  a <- A[ i ]

  for ( j in  1 : n ) {

    #Fijo un valor del parámetro b
    b <- B[ j ]

    #Suma de los cuadrados de los errores
    Q[ i, j ] <- sum( ( y - a - b * x )^2 )

    #Busco los índice i,j que hacen mínima Q
    if ( Q[ i, j ] < Qmin ) {
      Qmin <- Q[ i, j ]
      imin <- i ; jmin <-j
    }

  }

}

#Los parámetros buscados
amin <- A[ imin ]

bmin <- B[ jmin ]

#Veo la función ln(Q) mediante las curvas de nivel
titulo <- paste( " Qmin = ", round( Qmin, 2 ),

                   ", a = ", round( amin, 2 ),

                   ", b = ", round( bmin, 2 )

                )

contour( A, B, log( Q ),
```

```
        xlab = " a  / mgr ",
        ylab = " b  / mgr /% ",
        main = titulo
       )
abline( v = amin, col = 2 )
abline( h = bmin, col = 2 )
```

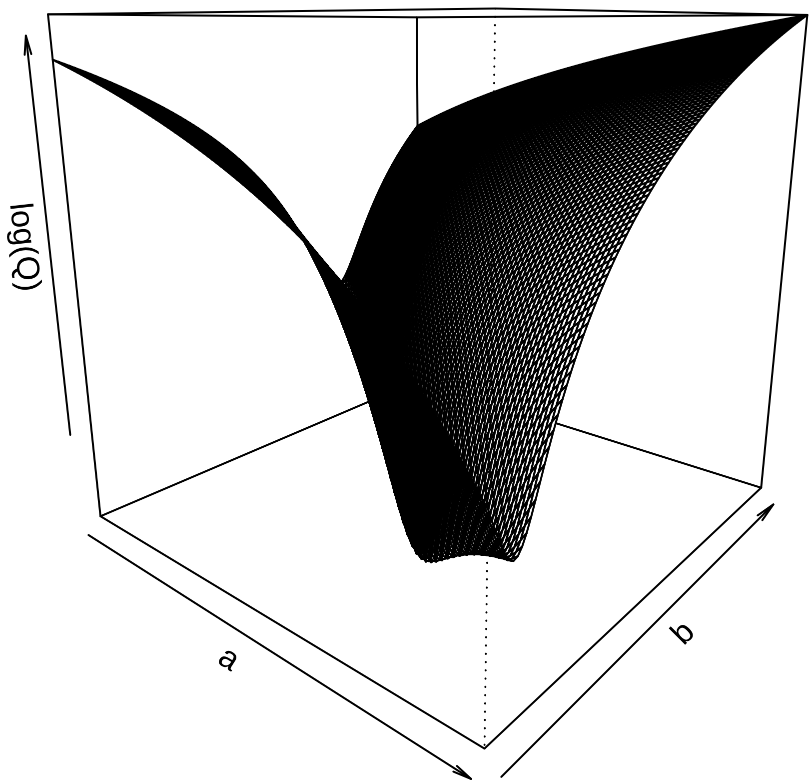



```
# En color
image( A, B, log( Q ),
      xlab = " a  / mgr ",
      ylab = " b  / mgr /% ",
      main = titulo
```

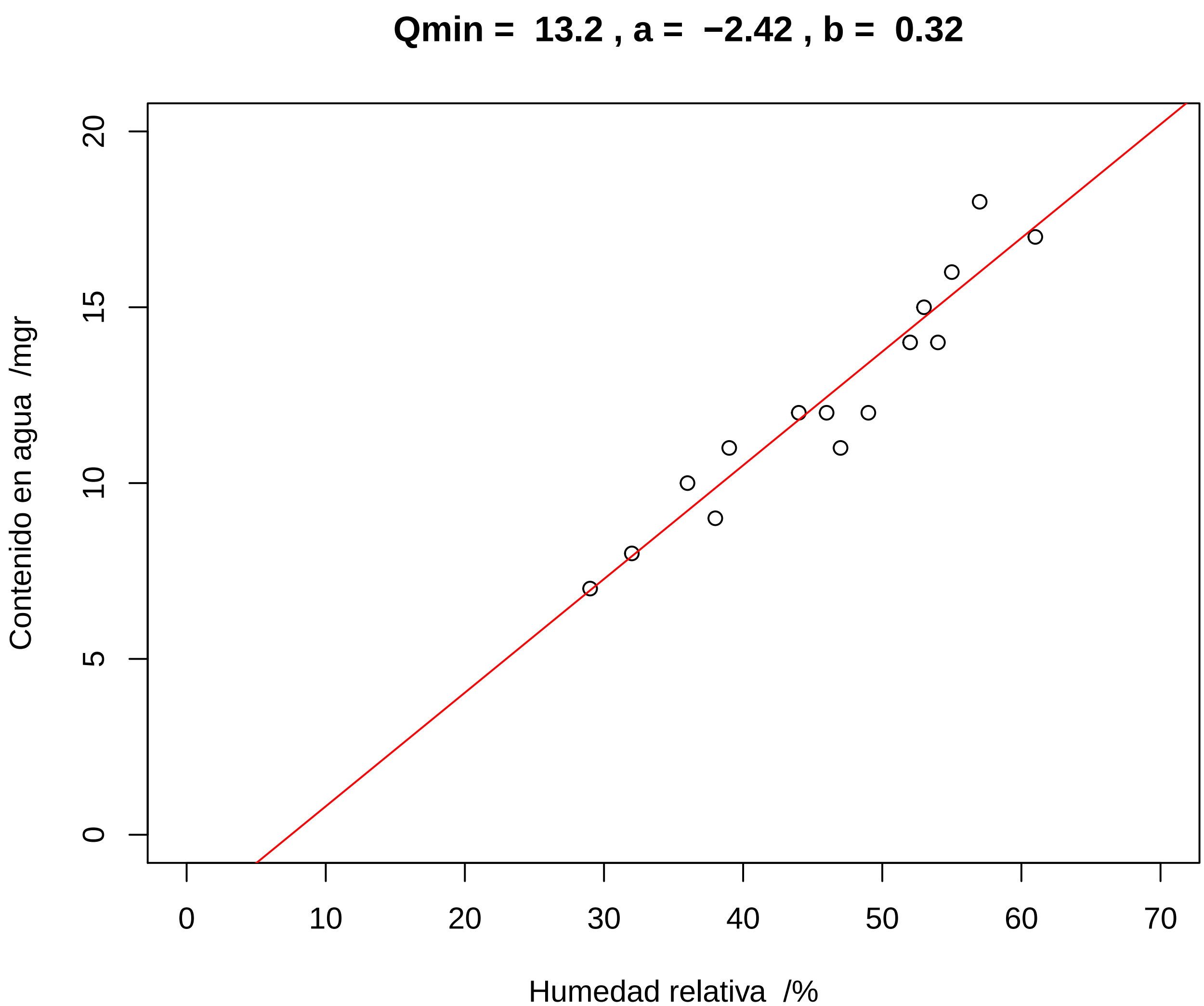
```
      )

abline( v = amin )

abline( h = bmin )
```




```
#También puedo verlo en perspectiva
persp( A, B, log( Q ), xlab = "a", ylab = "b", theta = 40 )
```

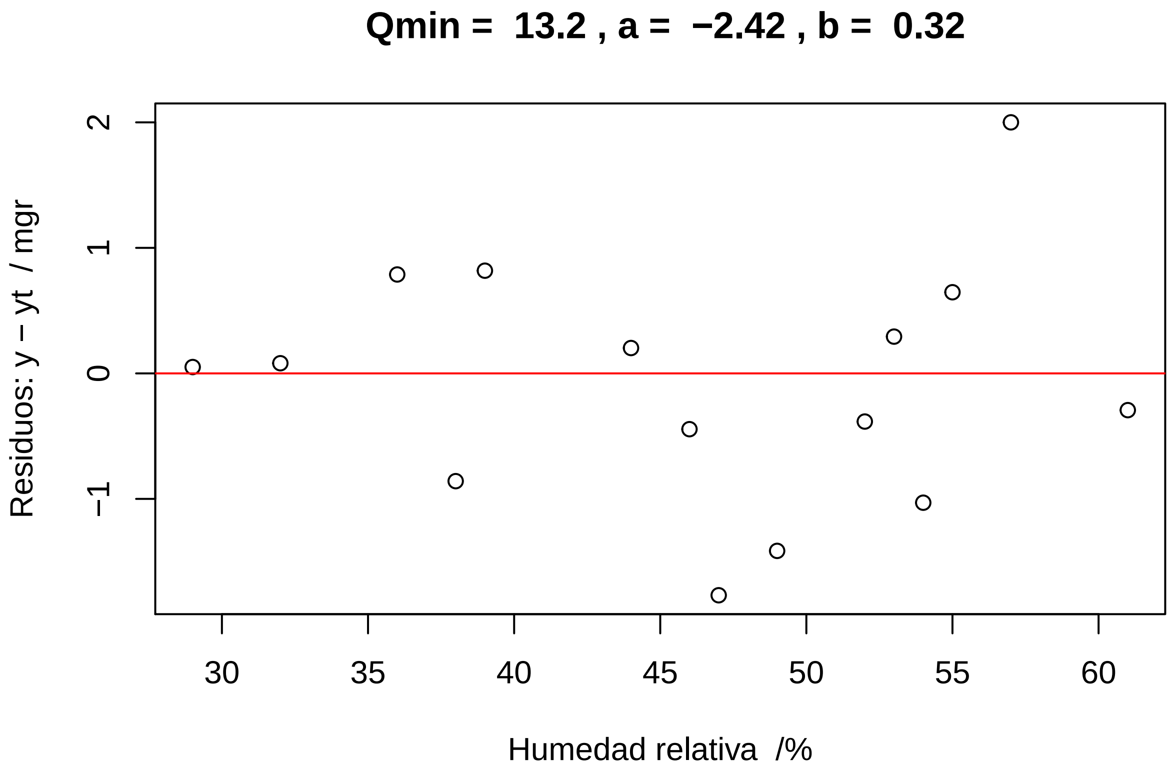



```r
#Finalmente comprobamos el ajuste de la recta teórica a los datos
plot( x, y,

      xlab = " Humedad relativa  /% ",

      ylab = " Contenido en agua  /mgr ",

      xlim = c( 0, 70 ),

      ylim = c( 0, 20),

      main = titulo

    )

#La recta que mejor se ajusta a los datos
abline( amin, bmin, col = 2 )
```



- Para comprobar que el ajuste es bueno podemos dibujar los *residuos*, esto es, las diferencias entre lo medido y lo postulado por la teoría $r_i = y_i - y(x_i, a, b)$ para comprobar que, en efecto, no hay sesgo, es decir, las desviaciones son al azar:

```
#La recta que mejor se ajusta
yt <- amin + bmin * x

#Finalmente comprobamos el ajuste de la recta teórica a los datos
plot( x, y - yt,

      xlab = " Humedad relativa  /% ",

      ylab = " Residuos: y - yt  / mgr ",

      main = titulo

    )

#el eje x
abline( h = 0, col = 2 )
```


Qmin = 13.2 , a = −2.42 , b = 0.32
Residuos: y − yt / mgr
2
1
0
−1
30
35
40
45
50
55
60
Humedad relativa /%