10

Regresión robusta

Cuando detectamos la presencia de puntos influyentes en el método de los mínimos cuadrados o cuando la distribución de los errores no es normal y sus colas son más pesadas podemos proceder de dos formas: eliminar dichos puntos y calcular las estimaciones con el mismo método o modificar el criterio de ajuste.

Este problema ya se ha tratado en el capítulo 7 para la regresión simple. Sin embargo, los métodos explicados no son extensibles a la regresión múltiple.

En este capítulo se presentan algunos métodos de ajuste alternativos a los mínimos-cuadrados ordinarios y que evitan la vulnerabilidad de dicho método a los puntos influyentes. Son los llamados *métodos robustos* (ver el apéndice de J. Fox [30]).

10.1. Minimizar una función objetivo

Para un modelo lineal del tipo

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \qquad i = 1, \ldots, n$$

podemos estimar los parámetros $\boldsymbol{\beta}$ con un estimador $\mathbf{b}$ que minimiza una función objetivo del tipo

$$\sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i'\mathbf{b})$$

donde la función ρ precisa la contribución de cada residuo $e_i = y_i - \mathbf{x}_i'\mathbf{b}$ a la suma.

Una función ρ apropiada debe tener las siguientes propiedades:

1. $\rho(e) \geq 0$
2. $\rho(0) = 0$
3. $\rho(e) = \rho(-e)$
4. Si $|e_i| < |e_j|$, entonces $\rho(e_i) \leq \rho(e_j)$

Un ejemplo es la función $\rho(e_i) = e_i^2$ que proporciona las estimaciones mínimo-cuadráticas.

Como se ha comentado en la sección 4.3.2, los estimadores $\mathbf{b}$ que se obtengan para una función ρ determinada se llaman estimadores M o M-estimadores, ya que se pueden considerar como generalización de la estimación de máxima verosimilitud.

Si derivamos la función objetivo respecto a los coeficientes $\mathbf{b}$ e igualamos a cero las derivadas parciales, obtenemos el sistema de ecuaciones

$$\sum_{i=1}^{n} \rho'(y_i - \mathbf{x}_i'\mathbf{b})\mathbf{x}_i' = \mathbf{0}$$

Este sistema se puede reescribir en la forma

$$\sum_{i=1}^{n} \frac{\rho'(e_i)}{e_i} e_i \mathbf{x}_i' = \sum_{i=1}^{n} \omega_i \cdot (y_i - \mathbf{x}_i'\mathbf{b})\mathbf{x}_i' = \mathbf{0}$$

donde $\omega_i = \omega(e_i) = \rho'(e_i)/e_i$ es el valor para una determinada función de ponderación ω.

Esta última expresión muestra que la solución de las ecuaciones equivale al problema de estimación mínimo-cuadrada ponderada que minimiza $\sum \omega_i e_i^2$. Sin embargo, los pesos dependen de los residuos, los residuos dependen de los coeficientes estimados y los coeficientes estimados, a su vez, dependen de los pesos. Este círculo impone una solución iterativa conocida como IRLS (*iteratively reweighted least-squares*). Esta solución se consigue con los siguientes pasos:

1. Tomamos unos estimadores iniciales $\mathbf{b}^{(0)}$, por ejemplo los estimadores mínimo-cuadráticos ordinarios.

2. Para cada iteración $j = 1, 2, \ldots$, se calculan los residuos $e_i^{(j-1)}$ y sus pesos asociados $\omega_i^{(j-1)}$.

3. Se calcula la nueva estimación ponderada

$$\mathbf{b}^{(j)} = (\mathbf{X}'\mathbf{W}^{(j-1)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(j-1)}\mathbf{Y}$$

   donde $\mathbf{X}$ es la matriz del modelo y $\mathbf{W}^{(j-1)} = \mathrm{diag}(\omega_i^{(j-1)})$.

4. Se repiten los pasos 2 y 3 hasta que los coeficientes estimados converjan.

10.1.1. Funciones objetivo

La tabla 10.1 muestra las funciones objetivo y las funciones de ponderación para tres métodos: el clásico de los mínimos-cuadrados, el método de Huber y el método de Tukey de los estimadores *bisquare*.

Método	*Función objetivo*	*Función de ponderación*
MC	$\rho_{MC}(e) = e^2$	$\omega_{MC}(e) = 1$
Huber	$\rho_H(e) = \begin{cases} \frac{1}{2}e^2 & \text{para } \lvert e\rvert \le k \\ k\lvert e\rvert - \frac{1}{2}k^2 & \text{para } \lvert e\rvert > k \end{cases}$	$\omega_H(e) = \begin{cases} 1 & \text{para } \lvert e\rvert \le k \\ k/\lvert e\rvert & \text{para } \lvert e\rvert > k \end{cases}$
Bisquare	$\rho_B(e) = \begin{cases} \frac{k^2}{6}\left[1 - \left(1 - \left(\frac{e}{k}\right)^2\right)^3\right] & \text{para } \lvert e\rvert \le k \\ k^2/6 & \text{para } \lvert e\rvert > k \end{cases}$	$\omega_B(e) = \begin{cases} \left[1 - \left(\frac{e}{k}\right)^2\right]^2 & \text{para } \lvert e\rvert \le k \\ 0 & \text{para } \lvert e\rvert > k \end{cases}$

Tabla 10.1: Funciones objetivo y funciones de ponderación para los métodos MC, Huber y Bisquare.

En los gráficos de la figura 10.1 se pueden comparar las curvas de los tres métodos. En las funciones objetivo del método de los mínimos-cuadrados y del método de Huber destaca el crecimiento no acotado cuando el residuo se aparta notablemente del cero, cuadráticamente o linealmente según el método. Mientras que la función objetivo bicuadrada se estabiliza para los residuos tales que $|e| > k$. En cuanto a las funciones de ponderación, podemos ver que el método MC asigna el mismo peso a todas las observaciones; el método de Huber decrece para los residuos tales que $|e| > k$; y para el método de la función bicuadrada los pesos decrecen en cuanto nos apartamos del cero y son 0 para $|e| > k$.

El valor de k para el método de Huber y *bisquare* se llama la constante de afinado (*tunning constant*). Un valor de k pequeño produce mayor resistencia a los valores atípicos, pero a costa de una menor

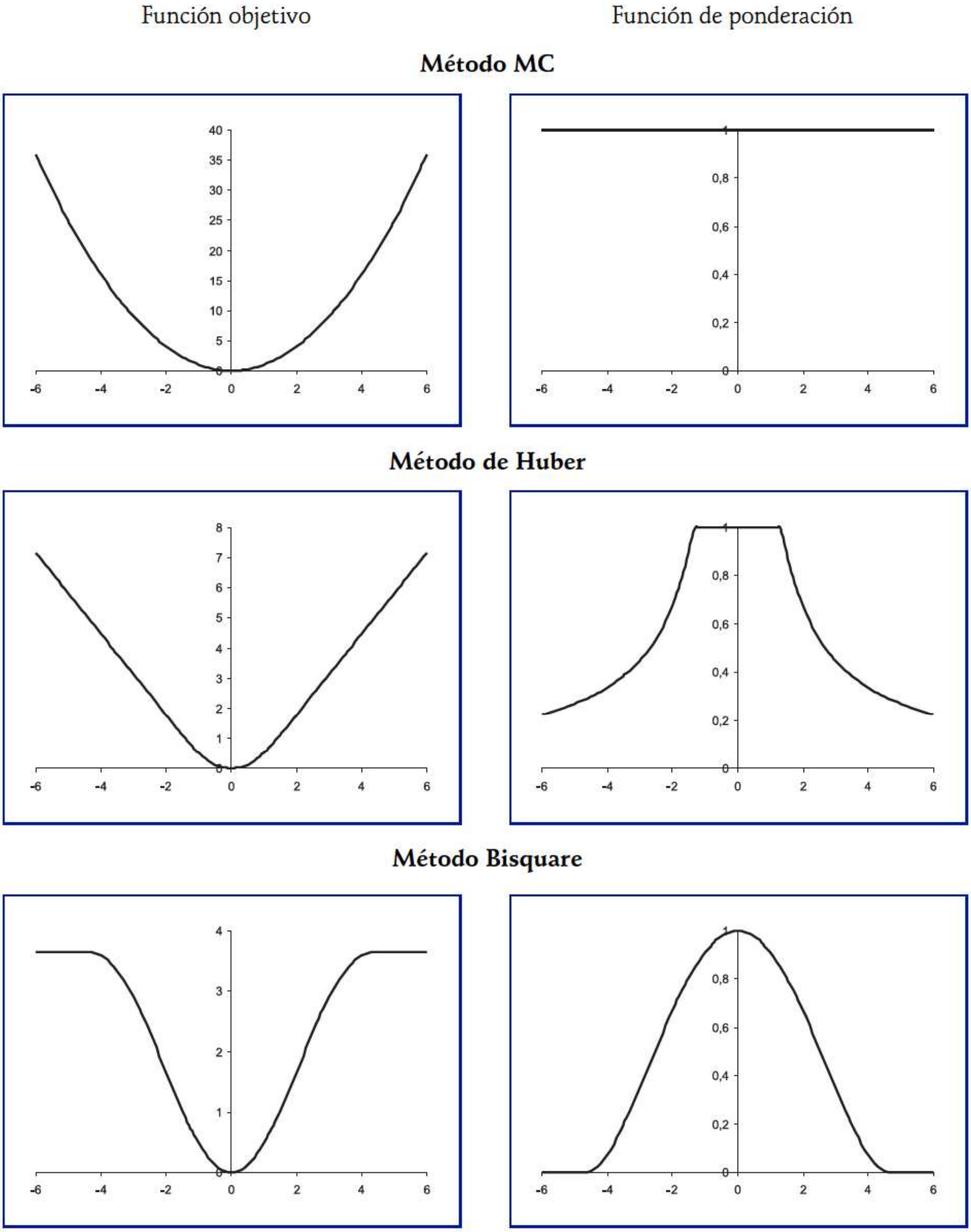


Figura 10.1: Gráficos de la función objetivo y la función de ponderación.

eficiencia cuando los errores siguen realmente la distribución normal. En general, esta constante se fija en un valor que produzca una razonable eficiencia en el caso normal. En particular, $k = 1.345\sigma$ para Huber y $k = 4.685\sigma$ para la bicuadrada (donde σ es la desviación estándar de los errores) producen una eficiencia del 95 % cuando los errores son normales y, al mismo tiempo, todavía protegen frente a valores atípicos.

En la práctica es preciso tener una estimación de la desviación estándar de los errores. Para ello se utiliza una medida de variabilidad robusta, mucho más indicada que la desviación estándar de los residuos. Por ejemplo se puede considerar $\hat{\sigma} = \text{MAR}/0.6745$, donde MAR es la mediana de los valores absolutos de los residuos.

En la tabla 10.2 se muestran todas las funciones de ponderación consideradas por el método de regresión `rreg` de S-PLUS.

Método	*Función de ponderación*	*Constante por defecto*
andrews	$\omega_A(e) = \begin{cases} \text{sen}(e/k)/(e/k) & \text{para } \lvert e\rvert \leq \pi \cdot k \\ 0 & \text{para } \lvert e\rvert > k \end{cases}$	$k = 1.339$
bicuadrada	$\omega_B(e) = \begin{cases} (1-(e/k)^2)^2 & \text{para } \lvert e\rvert \leq k \\ 0 & \text{para } \lvert e\rvert > k \end{cases}$	$k = 4.685$
cauchy	$\omega_C(e) = 1/(1+(e/k)^2)$	$k = 2.385$
fair	$\omega_F(e) = 1/(1+\lvert e/k\rvert)^2$	$k = 1.4$
hampel	$\omega_H(e) = \begin{cases} 1 & \text{para } \lvert e\rvert \leq a \\ a/\lvert e\rvert & \text{para } a < \lvert e\rvert \leq b \\ (a \cdot \frac{c-\lvert e\rvert}{c-b})/\lvert e\rvert & \text{para } b < \lvert e\rvert \leq c \\ 0 & \text{para } c < \lvert e\rvert \end{cases}$	$a = 2, b = 4, c = 8$
huber	$\omega_H(e) = \begin{cases} 1 & \text{para } \lvert e\rvert \leq k \\ k/\lvert e\rvert & \text{para } \lvert e\rvert > k \end{cases}$	$k = 1.345$
logística	$\omega_L(e) = \tanh(e/k)/(e/k)$	$k = 1.205$
talworth	$\omega_T(e) = \begin{cases} 1 & \text{para } \lvert e\rvert \leq k \\ 0 & \text{para } \lvert e\rvert > k \end{cases}$	$k = 2.795$
welsch	$\omega_W(e) = \exp(-2\lvert e/2k\rvert^2)$	$k = 2.985$

Tabla 10.2: Funciones de ponderación para los métodos propuestos en la función `rreg` de S-PLUS.

10.2. Regresión robusta mínimo-cuadrada recortada

Como ya se ha comentado, la regresión mínimo-cuadrática ordinaria es muy sensible a la suposición de normalidad de manera que las técnicas robustas son muy útiles cuando aparecen muchos valores atípicos en la respuesta. Sin embargo, los M-estimadores son especialmente delicados para puntos de alta influencia. Técnicamente, se dice que los M-estimadores no tienen la "influencia acotada" (*bounded influence*). Se trata de obtener estimadores con el mayor punto de colapso posible (ver definición 7.3.1).

En este sentido, hay algunos estimadores de regresión con un punto de colapso cercano al 50 por ciento. Uno de esos estimadores de "influencia acotada" es el de la regresión mínimo-cuadrada recortada o LTS (*least-trimmed squares*).

Supongamos que los residuos cuadráticos de la regresión son

$$e_i^2 = (y_i - \mathbf{x}_i'\mathbf{b})^2 \qquad i = 1, \ldots, n$$

y que los ordenamos de forma creciente $e_{(1)}^2, e_{(2)}^2, \ldots, e_{(n)}^2$. El estimador LTS debe seleccionar los coeficientes de regresión **b** que minimicen la suma de los m menores residuos cuadráticos

$$\mathrm{LTS}(\mathbf{b}) = \sum_{i=1}^{m} e_{(i)}^2$$

donde m es un entero entre $n/2$ y n. Una posibilidad es elegir $m = \lfloor n/2 \rfloor + \lfloor (k+2)/2 \rfloor$, es decir, un poco más que $n/2$, donde los paréntesis inferiores representan el ajuste al entero inferior más cercano.

Ahora bien, aunque la descripción de este método es relativamente sencilla, la verdad es que los mecanismos para obtener los estimadores LTS son muy complicados. Además, este tipo de estimadores puede producir resultados inesperados bajo ciertas circunstancias y también resultan un tanto ineficientes.

Actualmente hay dos formas de obtener una regresión con un alto punto de colapso y al mismo tiempo una alta eficiencia. La primera es considerar una regresión con M-estimadores que considere como pesos los elementos h_{ii} de la diagonal de la matriz **P**. La otra es una regresión MC ponderada con los pesos equivalentes al tamaño de los residuos a partir de una regresión LTS. La investigación en este campo es muy activa y se esperan nuevas soluciones en un futuro próximo.

10.3. Ejemplos con S-PLUS

En primer lugar vamos a utilizar los métodos de regresión robusta del programa S-PLUS. Para la regresión con M-estimadores dispone de la función `rreg` entre cuyos argumentos destacan los diferentes métodos o funciones de ponderación a elegir (ver tabla 10.2).
Consideremos el ejemplo de los niños de la tabla 7.1 cuyos datos son:

```
> edad<-c(109,113,115,116,119,120,121,124,126,
+         129,130,133,134,135,137,139,141,142)
> altura<-c(137.6,147.8,136.8,140.7,132.7,145.4,
+           135.0,133.0,148.5,148.3,147.5,148.8,
+           133.2,148.7,152.0,150.6,165.3,149.9)
> ninos.df<-data.frame(edad,altura)
> ninos.df
   edad altura
 1  109  137.6
 2  113  147.8
 ...
18  142  149.9
```

Procedemos a calcular los resultados del modelo lineal mínimo-cuadrático ordinario.

```
> mod.mc<-lm(altura ~ edad, data=ninos.df)
> summary(mod.mc)

Call: lm(formula = altura ~ edad, data = ninos.df)
Residuals:
    Min      1Q Median   3Q   Max
 -15.01 -2.223   1.22 2.55 13.51

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) 79.6962 21.2511     3.7502  0.0017
       edad  0.5113  0.1670     3.0608  0.0075

Residual standard error: 7.029 on 16 degrees of freedom
Multiple R-Squared: 0.3693
```

```
F-statistic: 9.369 on 1 and 16 degrees of freedom, the p-value is 0.00747

Correlation of Coefficients:
     (Intercept)
edad -0.997
```

En este caso se observan algunos residuos especialmente grandes:

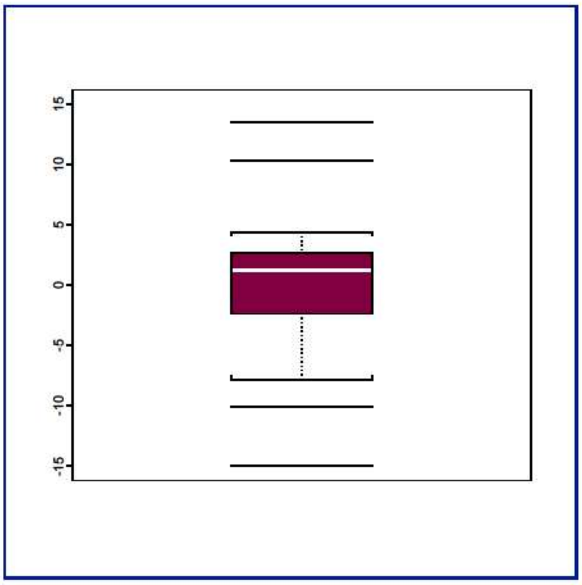
```
> mod.mc$resid
        1        2         3        4         5        6
 2.173504 10.32836 -1.694217 1.694496 -7.839364 4.349349

         7        8        9       10        11       12
 -6.561938 -10.0958 4.381628 2.647767  1.336481 1.10262

        13           14       15          16        17         18
 -15.00867 -0.01995356 2.257473 -0.1651009 13.51233 -2.398961

> summary(mod.mc$resid)
      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
 -15.00867  -2.22278   1.21955   0.00000   2.55019  13.51233

> boxplot(mod.mc$resid,ylim=c(-15,15))
> stem(mod.mc$resid)
> stem(mod.mc$resid,fence=1.5)
```




```
N = 18   Median = 1.21955
Quartiles = -2.39896, 2.647767
Decimal point is 1 place
to the right of the colon

Low:  -15.00867

  -1 : 0
  -0 : 87
  -0 : 2200
   0 : 11222344
   0 :
   1 : 0

High: 13.51233
```

Figura 10.2: Boxplot y diagrama de tallo y hojas de los residuos mínimo-cuadráticos ordinarios.

En el boxplot de la figura 10.2 aparecen 4 residuos atípicos (*outliers*), mientras que en el gráfico de tallo y hojas sólo se señalan los dos más alejados. Esto se debe a que el factor del rango intercuartílico por defecto de la función `stem` es 2, cuando habitualmente se considera como razonable un factor `fence=1.5`.

Podemos recalcular el modelo eliminando los dos datos de mayor residuo:

```
> mod.mc.2<-update(mod.mc,subset=-c(13,17))
> summary(mod.mc.2)

Call: lm(formula = altura ~ edad, data = ninos.df, subset =  - c(13, 17))
Residuals:
```

```
    Min      1Q Median     3Q   Max
 -10.27 -1.781  1.126 2.759 9.548

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) 86.6853 16.9427     5.1164  0.0002
       edad  0.4563  0.1346     3.3902  0.0044

Residual standard error: 5.196 on 14 degrees of freedom
Multiple R-Squared: 0.4508
F-statistic: 11.49 on 1 and 14 degrees of freedom, the p-value is 0.0044

Correlation of Coefficients:
     (Intercept)
edad -0.9971
```

Se observa una gran modificación de los coeficientes estimados por lo que se procede a utilizar un método robusto como el siguiente:

```
> rreg(edad,altura,method=wt.huber)
$coefficients:
 (Intercept)          x
    82.80675 0.4906192

$residuals:
 [1]   1.3157618   9.5532851   -2.4279533   0.9814275  -8.4904301
 [6]   3.7189507  -7.1716684  -10.6435260   3.8752356   2.2033781
[11]   0.9127589   0.7409013  -15.3497179  -0.3403371   1.9784246
[16]  -0.4028138  13.3159478   -2.5746714

$fitted.values:
 [1] 136.2842 138.2467 139.2280 139.7186 141.1904 141.6810
 [7] 142.1717 143.6435 144.6248 146.0966 146.5872 148.0591
[13] 148.5497 149.0403 150.0216 151.0028 151.9841 152.4747

$w:
 [1] 1.0000000 0.5192744 1.0000000 1.0000000 0.5867797 1.0000000
 [7] 0.6947497 0.4676898 1.0000000 1.0000000 1.0000000 1.0000000
[13] 0.3239894 1.0000000 1.0000000 1.0000000 0.3735146 1.0000000

$int:
[1] T

$conv:
[1] 0.059493679 0.018831865 0.001970252

$status:
[1] "converged"
```

También se puede utilizar la función de ponderación bicuadrada:

```
> rreg(edad,altura,method=wt.bisquare)
$coefficients:
 (Intercept)          x
    82.97594 0.4902345

$residuals:
 [1]    1.1885086    9.4275707   -2.5528982   0.8568674  -8.6138360
 [6]    3.5959295   -7.2943049  -10.7650083   3.7545228   2.0838194
```

```
[11]   0.7935849   0.6228815 -15.4673529  -0.4575874   1.8619437
[16]  -0.5185252  13.2010059  -2.6892286

$fitted.values:
 [1] 136.4115 138.3724 139.3529 139.8431 141.3138 141.8041
 [7] 142.2943 143.7650 144.7455 146.2162 146.7064 148.1771
[13] 148.6674 149.1576 150.1381 151.1185 152.0990 152.5892

$w:
 [1] 0.99163379 0.53486719 0.96071253 0.99561954 0.60084361 0.92332564
 [7] 0.70359985 0.42138479 0.91633460 0.97374445 0.99613100 0.99758601
[13] 0.07641784 0.99878906 0.97884351 0.99845775 0.22189085 0.95720077

$int:
[1] T

$conv:
[1] 0.063100738 0.029412101 0.005141062 0.001374108

$status:
[1] "converged"
```

Por último podemos utilizar el método de regresión robusta mínimo-cuadrada recortada a través de la función `ltsreg`.

```
> ltsreg(altura ~ edad)
Method:
Least Trimmed Squares Robust Regression.

Call:
ltsreg(formula = altura ~ edad)

Coefficients:
 Intercept    edad
 85.3356    0.4670

Scale estimate of residuals:  7.044

Total number of observations:  18
Number of observations that determine the LTS estimate:  16
```

Las ventajas de uno u otro método ya se han comentado en los apartados anteriores. Con este ejemplo no es apropiado hacer una discusión de los resultados ya que el número de datos es muy reducido.

Alternativamente, se pueden calcular los M-estimadores de Huber con la función `rlm` de la librería `MASS` de R. La misma función permite opcionalmente los cálculos con la función de ponderación bicuadrada, si se añade la librería `lqs`. Además, esta última contiene una función `ltsreg` para la regresión robusta recortada. Ver los ejemplos en el apéndice de regresión robusta del libro de J. Fox [30].

10.4. Ejercicios

Ejercicio 10.1

Calcular la regresión con M-estimadores y diferentes métodos o funciones de ponderación a elegir de la tabla 10.2 sobre los datos del ejercicio 6.10. Discutir su necesidad frente a la recta mínimocuadrática.

Realizar también una regresión robusta mínimocuadrática recortada.

Ejercicio 10.2

Calcular la regresión con M-estimadores y diferentes métodos o funciones de ponderación a elegir de la tabla 10.2 sobre los datos (c) de la tabla 6.4. Comparar el resultado con el de la tabla 6.5 y con la recta de regresión mínimocuadrática sin la observación 16.

Realizar también una regresión robusta mínimocuadrática recortada.

Ejercicio 10.3

Con los datos del ejercicio 8.4 realizar algunas regresiones robustas y comparar el resultado con la regresión MC ordinaria.