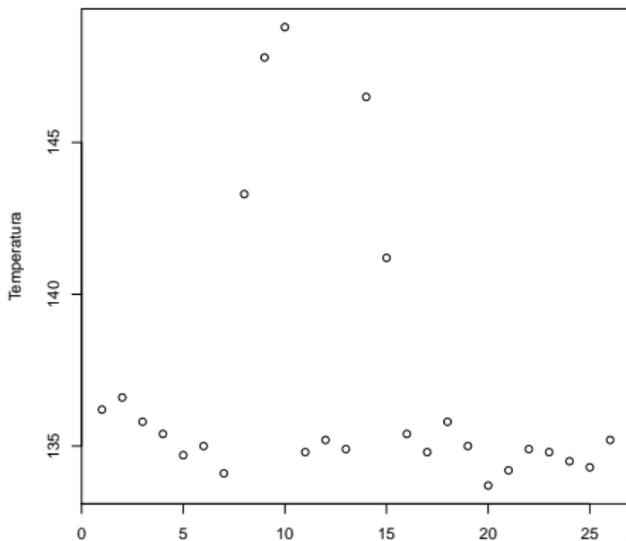




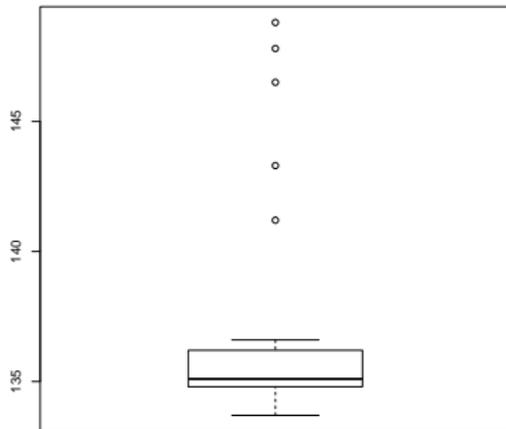
## Temperatura sublimación platino

26 mediciones de la temperatura de sublimación de platino en un experimento realizado por Hampson y Walker (1961).

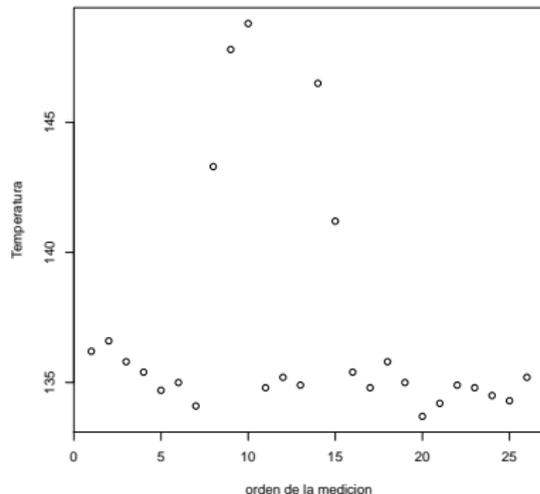
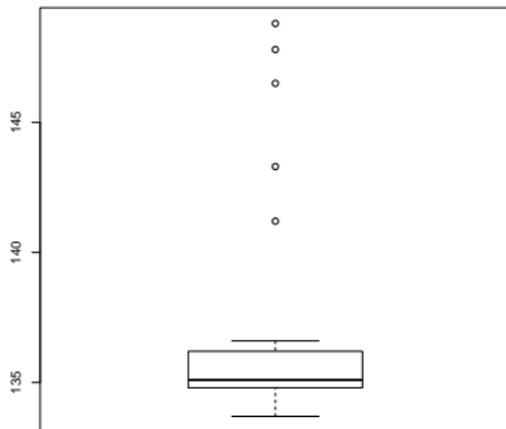
136.2	136.6	135.8	135.4	134.7	135.0	134.1	143.3	147.8
148.8	134.8	135.2	134.9	146.5	141.2	135.4	134.8	135.8
135.0	133.7	134.2	134.9	134.8	134.5	134.3	135.2	



# Boxplot Platino



## Boxplot Platino



Con todos los datos:  $\bar{x} = 137.05$

Sacando los datos anómalos:  $\bar{x} = 134.9$ .

Mediana  $M = 135.1$ ,  $d_I = 1.7$ .

## Diagrama de Hertzsprung–Russell

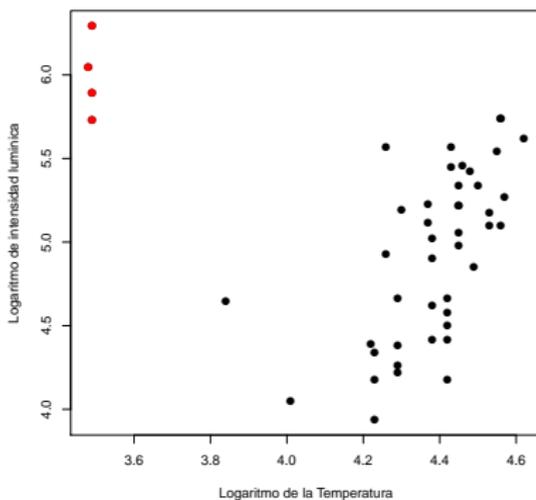
$X$  = logaritmo de la temperatura en la superficie de la estrella  
 $Y$  = logaritmo de intensidad lumínica.

$i$	$x_i$	$y_i$									
1	4.37	5.23	25	4.38	5.02	13	4.48	5.42	37	4.53	5.10
2	4.56	5.74	26	4.42	4.66	14	4.01	4.05	38	4.45	5.22
3	4.26	4.93	27	4.29	4.66	15	4.29	4.26	39	4.53	5.18
4	4.56	5.74	28	4.38	4.90	16	4.42	4.58	40	4.43	5.57
5	4.30	5.19	29	4.22	4.39	17	4.23	3.94	41	4.38	4.62
6	4.46	5.46	30	3.48	6.05	18	4.42	4.18	42	4.45	5.06
7	3.84	4.65	31	4.38	4.42	19	4.23	4.18	43	4.50	5.34
8	4.57	5.27	32	4.56	5.10	20	3.49	5.89	44	4.45	5.34
9	4.26	5.57	33	4.45	5.22	21	4.29	4.38	45	4.55	5.54
10	4.37	5.12	34	3.49	6.29	22	4.29	4.22	46	4.45	4.98
11	3.49	5.73	35	4.23	4.34	23	4.42	4.42	47	4.42	4.50
12	4.43	5.45	36	4.62	5.62	24	4.49	4.85			

# Diagrama de Hertzsprung–Russell

$$y_i = \theta_1 + \theta_2 x_i + \epsilon_i$$

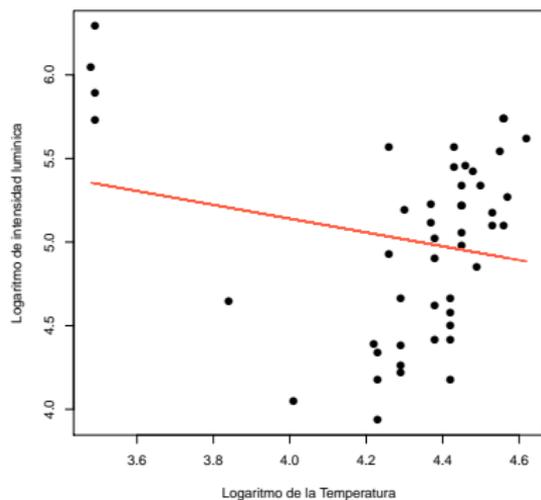
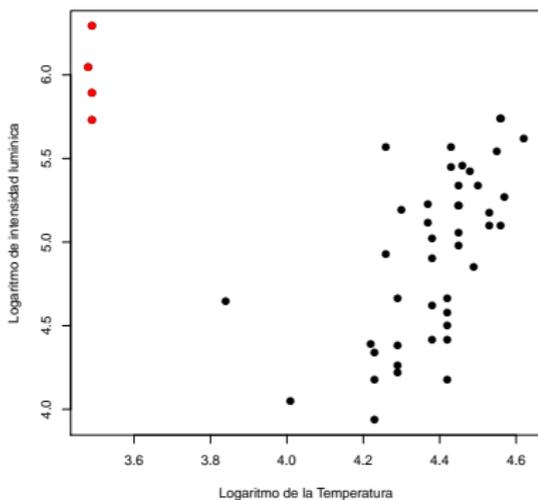
$$\text{LS: } (\hat{\theta}_1, \hat{\theta}_2) = 6.793, -0.413$$



# Diagrama de Hertzsprung–Russell

$$y_i = \theta_1 + \theta_2 x_i + \epsilon_i$$

$$\text{LS: } (\hat{\theta}_1, \hat{\theta}_2) = 6.793, -0.413$$

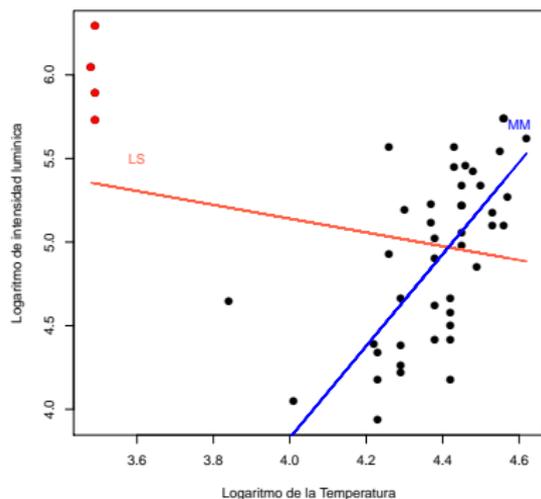
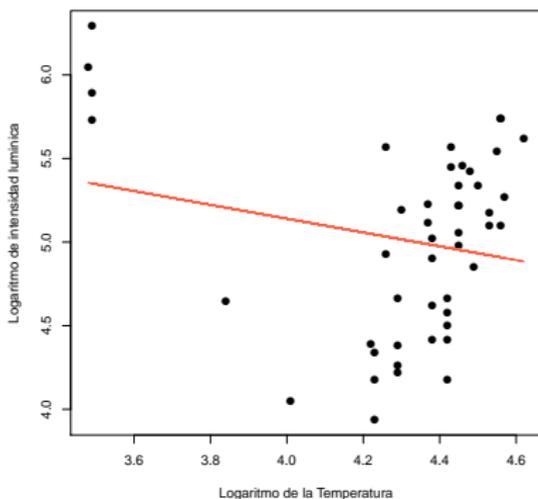


# Diagrama de Hertzsprung–Russell

$$y_i = \theta_1 + \theta_2 x_i + \epsilon_i$$

LS:  $(\hat{\theta}_1, \hat{\theta}_2) = (6.793, -0.413)$

MM:  $(\hat{\theta}_1, \hat{\theta}_2) = (-7.153, 2.745)$





## El Problema de Robustez

- Si  $x_1, \dots, x_n$  i.i.d.  $u_j \sim F \in V_\epsilon \implies$ 
  - a) Una proporción  $(1 - \epsilon)$  de las observaciones estarán dadas por  $x_i = \mu + \sigma u_i$  con  $u_i \sim \Phi$ ,
  - b) Una proporción  $\epsilon$  estarán dadas por  $x_i = \mu + \sigma u_i$  con  $u_i \sim H$ . Estas últimas observaciones serán denominadas **puntos atípicos** o *outliers*, .
- Si  $\mathbb{E}_H(u) = 0$  entonces

$$\text{var}_F(\bar{x}) = \frac{\sigma^2}{n} \quad \text{var}_F(u) = \frac{\sigma^2}{n} ((1 - \epsilon) + \epsilon \text{var}_H(u))$$





# El Problema de Robustez

Procedimientos estables frente a las siguientes perturbaciones:

- Presencia de un pequeño porcentaje de observaciones con errores groseros

## El Problema de Robustez

Procedimientos estables frente a las siguientes perturbaciones:

- Presencia de un pequeño porcentaje de observaciones con errores groseros
- Presencia de errores de redondeo

## El Problema de Robustez

Procedimientos estables frente a las siguientes perturbaciones:

- Presencia de un pequeño porcentaje de observaciones con errores groseros
- Presencia de errores de redondeo
- Apartamientos de la independencia

## El Problema de Robustez

Procedimientos estables frente a las siguientes perturbaciones:

- Presencia de un pequeño porcentaje de observaciones con errores groseros
- Presencia de errores de redondeo
- Apartamientos de la independencia
- Modelo es una aproximación de la realidad

## El Problema de Robustez

Procedimientos estables frente a las siguientes perturbaciones:

- Presencia de un pequeño porcentaje de observaciones con errores groseros
- Presencia de errores de redondeo
- Apartamientos de la independencia
- Modelo es una aproximación de la realidad

Consideramos **estimadores**  $T_n$  que se escriben como  $T_n = T(F_n)$  donde  $F_n$  indica la empírica y  $T$  es un funcional sobre el espacio de las medidas de probabilidad.

## El Problema de Robustez

Procedimientos estables frente a las siguientes perturbaciones:

- Presencia de un pequeño porcentaje de observaciones con errores groseros
- Presencia de errores de redondeo
- Apartamientos de la independencia
- Modelo es una aproximación de la realidad

Consideramos **estimadores**  $T_n$  que se escriben como  $T_n = T(F_n)$  donde  $F_n$  indica la empírica y  $T$  es un funcional sobre el espacio de las medidas de probabilidad. La noción de robustez está relacionada con la continuidad de  $T$  respecto de la distancia de Prohorov.







## Métrica de Lévy y de Prohorov

$\Omega$  espacio métrico polaco,  $A^\delta = \{x \in \Omega : \inf_{y \in A} d(x, y) \leq \delta\}$

**Distancia de Prohorov:**

$$\Pi(P, Q) = \inf\{\epsilon > 0 : \forall A \in \mathcal{B} \quad P(A) \leq Q(A^\epsilon) + \epsilon\}$$

Si  $\Omega = \mathbb{R}$ :  $d_L \leq \Pi$

## Métrica de Lévy y de Prohorov

$\Omega$  espacio métrico polaco,  $A^\delta = \{x \in \Omega : \inf_{y \in A} d(x, y) \leq \delta\}$

**Distancia de Prohorov:**

$$\Pi(P, Q) = \inf\{\epsilon > 0 : \forall A \in \mathcal{B} \quad P(A) \leq Q(A^\epsilon) + \epsilon\}$$

Si  $\Omega = \mathbb{R}$ :  $d_L \leq \Pi$  **La distancia de Prohorov metriza la topología débil.**  $\mathcal{M}$  resulta un espacio métrico separable y completo.

## Métrica de Lévy y de Prohorov

**Teorema de Strassen** Los siguientes dos enunciados son equivalentes

- $P(A) \leq Q(A^\delta) + \epsilon, \forall A \in \mathcal{B}$
- Existen variables aleatorias  $X$  e  $Y$  sobre  $\Omega$  tales que  $X \sim P$ ,  $Y \sim Q$ ,  $(X, Y) \sim R$

$$R(d(X, Y) \leq \delta) > 1 - \epsilon$$

## $M$ —estimadores de posición

- $\mathbf{x}_i = \boldsymbol{\theta} + \sigma \mathbf{u}_i$ ,  $u_i \sim F$  i.i.d.  $\sigma = \sigma_0$  conocido.  $\Rightarrow x_i \sim F_{\theta\sigma_0}$
- $f$  densidad de  $F$ ,  $f_{\theta\sigma_0}(x) = \frac{1}{\sigma_0} f\left(\frac{x - \theta}{\sigma_0}\right)$ ,

## $M$ —estimadores de posición

- $x_i = \theta + \sigma u_i$ ,  $u_i \sim F$  i.i.d.  $\sigma = \sigma_0$  conocido.  $\Rightarrow x_i \sim F_{\theta\sigma_0}$
- $f$  densidad de  $F$ ,  $f_{\theta\sigma_0}(x) = \frac{1}{\sigma_0} f\left(\frac{x - \theta}{\sigma_0}\right)$ ,
- Función de verosimilitud correspondiente a  $x_1, \dots, x_n$

$$L(\theta) = \frac{1}{\sigma_0^n} \prod_{i=1}^n f\left(\frac{x_i - \theta}{\sigma_0}\right)$$

## $M$ —estimadores de posición

- $\mathbf{x}_i = \boldsymbol{\theta} + \boldsymbol{\sigma} \mathbf{u}_i$ ,  $u_i \sim F$  i.i.d.  $\sigma = \sigma_0$  conocido.  $\Rightarrow x_i \sim F_{\theta\sigma_0}$
- $f$  densidad de  $F$ ,  $f_{\theta\sigma_0}(x) = \frac{1}{\sigma_0} f\left(\frac{x - \theta}{\sigma_0}\right)$ ,
- Función de verosimilitud correspondiente a  $x_1, \dots, x_n$

$$L(\theta) = \frac{1}{\sigma_0^n} \prod_{i=1}^n f\left(\frac{x_i - \theta}{\sigma_0}\right)$$

- El EMV  $\hat{\theta}_f$  minimiza

$$\mathbf{S}(\theta) = \frac{1}{n} \sum_{i=1}^n \rho_f\left(\frac{x_i - \theta}{\sigma_0}\right), \quad (1)$$

$$\rho_f(\mathbf{u}) = -\log f(\mathbf{u}) + \log f(\mathbf{0}).$$

## M—estimadores de posición

- Huber (1964) definió los M-estimadores para el modelo de posición como el valor  $\hat{\theta}$  que minimiza

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i - \theta}{\sigma_0} \right), \quad (2)$$

## M—estimadores de posición

- Huber (1964) definió los **M-estimadores para el modelo de posición** como el valor  $\hat{\theta}$  que minimiza

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i - \theta}{\sigma_0} \right), \quad (2)$$

- $\rho$  se elige independientemente de  $f$  y de tal manera que tenga las propiedades deseadas:
  - El estimador es **altamente eficiente** cuando  $f$  corresponde a la distribución  $N(0,1)$ .
  - El estimador es **poco sensible a contaminación por outliers**, en particular es altamente eficiente para toda  $f$  correspondiente a una distribución de  $\mathcal{V}_\epsilon$ .

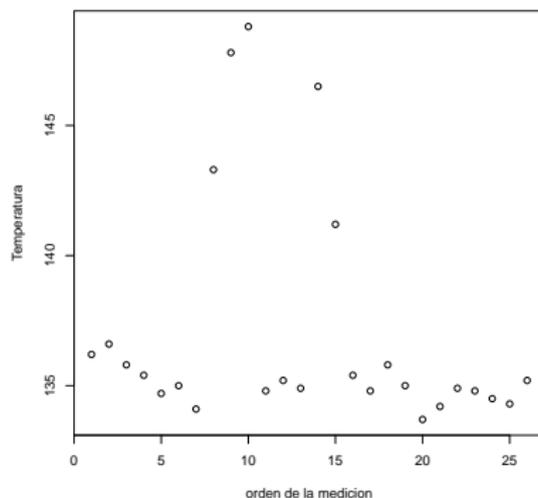


## M—estimadores de posición

- Propiedades de la función  $\rho$  que define al M-estimador
  - A1.  $\rho$  es derivable con derivada  $\psi = \rho'$ .
  - A2.  $\rho$  es par.
  - A3.  $\rho(u)$  es monótona no decreciente en  $|u|$ .
  - A4.  $\rho(0) = 0$ .
- Huber propuso una familia de funciones  $\rho_k^H$

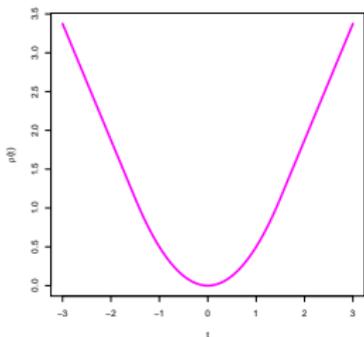
$$\rho_k^H(u) = \begin{cases} -ku - \frac{k^2}{2} & \text{si } u < -k \\ \frac{u^2}{2} & \text{si } |u| \leq k \\ ku - \frac{k^2}{2} & \text{si } u > k \end{cases} .$$

# Temperatura de sublimación del platino



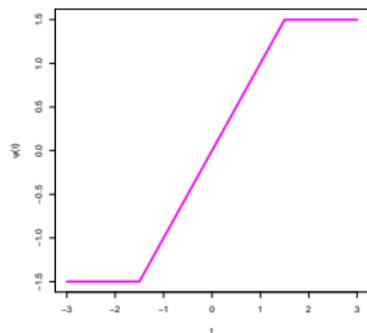
$M$ -estimador con $k = 1.5$	135.38
Mediana	135.1
$\bar{x}$	137.05
$\bar{x}$ sin atípicos	134.9

$$\rho(t) = t^2/2\mathbb{I}_{|t|\leq c} + (ct - c^2/2)\mathbb{I}_{|t|>c}$$

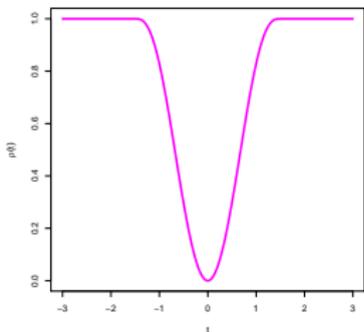


Huber

$$\psi(t) = t\mathbb{I}_{|t|\leq c} + c\operatorname{signo}(t)\mathbb{I}_{|t|>c}$$

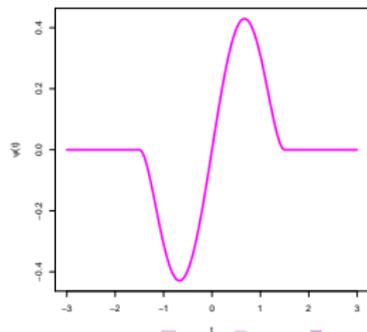


$$\rho(t) = \min(3(t/c)^2 - 3(t/c)^4 + (t/c)^6, 1)$$



Tukey

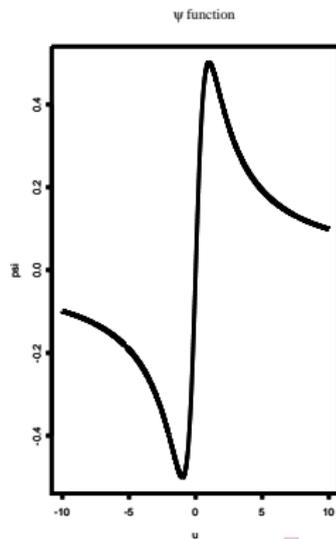
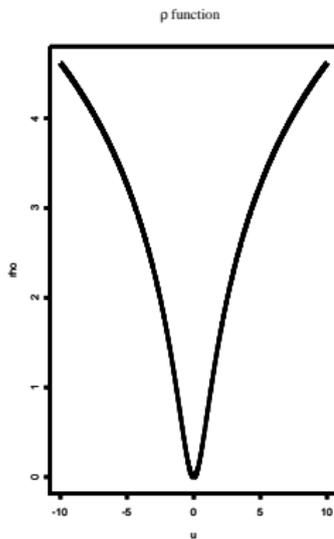
$$\psi(t) = (6/c^2)(1 - (t/c)^2)^2\mathbb{I}_{|t|\leq c}$$



## M—estimadores de posición: $\psi$ redescendiente

Otra familia de funciones usadas es la asociada los EMV respecto de la  $\mathcal{T}_\nu$

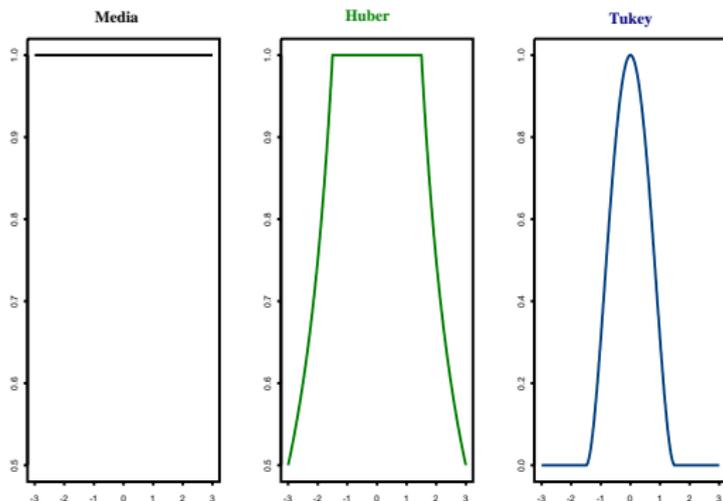
$$f_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{x^2}{\nu}\right]^{\frac{\nu+1}{2}} \quad \psi_\nu(u) = \frac{x}{x^2 + \nu}$$



## $M$ —estimadores de posición

Un  $M$ —estimador de posición puede verse como un promedio pesado. Sea

$$W(x) = \begin{cases} \frac{\psi(x)}{x} & \text{si } x \neq 0 \\ \psi'(0) & \text{si } x = 0 \end{cases} .$$



## M—estimadores de posición

$$\begin{aligned}\lambda_n(\hat{\theta}) &= \mathbf{0} \Leftrightarrow \\ \frac{1}{n} \sum_{i=1}^n \psi\left(\frac{\mathbf{x}_i - \hat{\theta}}{\sigma_0}\right) &= \mathbf{0} \Leftrightarrow \\ \frac{1}{n} \sum_{i=1}^n \left(\frac{\mathbf{x}_i - \hat{\theta}}{\sigma_0}\right) \mathbf{W}\left(\frac{\mathbf{x}_i - \hat{\theta}}{\sigma_0}\right) &= \mathbf{0} \Leftrightarrow\end{aligned}$$

$$\hat{\theta} = \frac{\sum_{i=1}^n \mathbf{W}_i \mathbf{x}_i}{\sum_{i=1}^n \mathbf{W}_i} \quad \text{con} \quad \mathbf{W}_i = \mathbf{W}\left(\frac{\mathbf{x}_i - \hat{\theta}}{\sigma_0}\right)$$

## M—estimadores de posición

$$\lambda_n(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \psi \left( \frac{x_i - \hat{\theta}}{\sigma_0} \right) = 0. \quad (3)$$

**Teorema.** *Supongamos que  $\psi$  es continua impar, no decreciente y para algún  $a$  se tiene  $\psi(a) > 0$ . Entonces*

- (i) *La ecuación (3) tiene al menos una raíz.*
- (ii) *Toda raíz de (3) corresponde a un mínimo de  $S_n(\theta)$ .*
- (iii) *Las raíces de (3) forman un intervalo.*
- (iv) *Si  $\psi$  es estrictamente creciente hay una única raíz de (3).*

## $M$ —estimadores: Funcional asociado

Sea  $\mathbf{T}(\mathbf{F})$  la solución de

$$\lambda_{\mathbf{F}}(\mathbf{t}) = \mathbb{E}_{\mathbf{F}} \psi \left( \frac{\mathbf{x} - \mathbf{t}}{\sigma_0} \right) = \mathbf{0}.$$

- Fisher—consistencia:  $T(F_{\theta\sigma_0}) = \theta$

$$\mathbb{E}_{F_{\theta\sigma_0}} \left( \psi \left( \frac{\mathbf{x} - \theta}{\sigma_0} \right) \right) = \mathbf{0},$$

- Es equivalente a  $\mathbb{E}_F(\psi(u)) = 0$ .
- El  $M$ —funcional es Fisher—consistente si  $F$  es simétrica respecto de 0 y  $\psi$  es impar.

## Propiedades asintóticas de los $M$ -estimadores

**Teorema** Sean  $x_1, \dots, x_n$  i.i.d.  $\mathbf{x}_i = \boldsymbol{\theta} + \sigma_0 \mathbf{u}_i$ ,  $u_i \sim F$ . Sea  $\hat{\boldsymbol{\theta}}_n$  solución de (3), donde  $\psi$  y  $F$  satisfacen  $\mathbb{E}_F(\psi(u)) = 0$ . Luego  $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}$  en cualquiera de los siguientes casos

1. La función  $\psi$  es estrictamente creciente.
2. La función  $\psi$  es no decreciente,  $\psi(u) > \psi(0)$  y  $F(u) > F(0)$  para todo  $u > 0$ .

Un resultado más general que incluye el caso de las funciones  $\psi$  redescendientes puede verse en Boos y Serfling (1980).

## Propiedades asintóticas de los $M$ -estimadores

**Teorema.**  $x_1, \dots, x_n$  i.i.d.  $\mathbf{x}_i = \theta + \sigma_0 \mathbf{u}_i$ ,  $u_i \sim F$ . Sea  $\hat{\theta}_n$  solución de (3), donde  $\psi$  y  $F$  satisfacen  $\mathbb{E}_F(\psi(u)) = 0$ . Supongamos que  $\hat{\theta}_n \xrightarrow{P} \theta$ , y que además  $\psi$  tiene dos derivadas continuas y  $\psi''$  es acotada. Entonces,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, \sigma_0^2 V(\psi, F)),$$

donde

$$V(\psi, F) = \frac{\mathbb{E}_F \psi^2(\mathbf{u})}{(\mathbb{E}_F \psi'(\mathbf{u}))^2}.$$

$V(\psi, F)$  se minimiza en

$$\psi_0(t) = -\frac{f'(t)}{f(t)} \quad \text{y} \quad V(\psi_0, F) = \frac{1}{I_f}$$

## $M$ —estimador minimax

$$\mathcal{F}_{1,\epsilon} = \{F : F = (1 - \epsilon)\Phi + \epsilon H \text{ con } H \text{ simétrica}\}.$$

- La mayor varianza posible en este entorno del  $M$ —estimador basado en  $\psi$  está dada por  $V(\psi) = \sup_{F \in \mathcal{F}_{1,\epsilon}} V(\psi, F)$ .
- Huber eligió el  $M$ —estimador con función  $\tilde{\psi}$ ,  
 $V(\tilde{\psi}) = \min_{\psi} V(\psi)$ .
- Estos estimadores se denominarán **minimax** (minimizan la máxima varianza asintótica en el entorno de contaminación  $\mathcal{F}_{1,\epsilon}$ ).
- Huber (1964) mostró que  $\tilde{\psi}$  está en la familia  $\psi_k^H$ , donde  $k$  depende de la cantidad de contaminación  $\epsilon$ .

## $M$ -estimador minimax

**Teorema.** Sea  $V(\psi) = \sup_{F \in \mathcal{F}_{1,\epsilon}} V(\psi, F)$  donde

$\mathcal{F}_{1,\epsilon} = \{F : F = (1 - \epsilon)\Phi + \epsilon H \text{ con } H \text{ simétrica}\}.$

El  $M$ -estimador asociado a  $\psi_k^H$  satisface

$$\psi_k^H = \underset{\psi}{\operatorname{argmin}} V(\psi)$$

o sea,

$$\sup_{F \in \mathcal{F}_{1,\epsilon}} V(\psi_k^H, F) \leq \sup_{F \in \mathcal{F}_{1,\epsilon}} V(\psi, F)$$

donde  $k$  y  $\epsilon$  cumplen

$$\frac{1}{1 - \epsilon} = \Phi(k) - \Phi(-k) + \frac{2}{k}\varphi(k) \quad \text{con} \quad \varphi = \Phi'$$

## Algunos estimadores de escala equivariantes

Sea  $x_i = \theta + \sigma u_i$ ,  $u_i \sim F_0$ , o sea,  $F_{\theta,\sigma}(x) = F_0((x - \theta)/\sigma)$ .

### Estimador de dispersión

$$\hat{\sigma}(\mathbf{x} + c) = \hat{\sigma}(\mathbf{x}) \quad \text{y} \quad \hat{\sigma}(c \mathbf{x}) = |c| \hat{\sigma}(\mathbf{x})$$

Dos estimadores usados

$$\text{mad}(\mathbf{x}) = \text{mediana}_{1 \leq i \leq n} \left| x_i - \text{mediana}_{1 \leq j \leq n}(\mathbf{x}_j) \right|$$

$$\text{IQR}(\mathbf{x}) = x^{(n-m+1)} - x^{(m)} \quad m = \left\lceil \frac{n}{4} \right\rceil$$

Si  $F_{\theta,\sigma} = N(\theta, \sigma^2)$ ,

$$\text{mad}(F_{\theta,\sigma}) = \Phi^{-1} \left( \frac{3}{4} \right) \sigma$$

$$\text{IQR}(F_{\theta,\sigma}) = 2\Phi^{-1} \left( \frac{3}{4} \right) \sigma$$

## Algunos estimadores de escala equivariantes

Sea  $x_i = \theta + \sigma u_i$ ,  $u_i \sim F_0$ , o sea,  $F_{\theta,\sigma}(x) = F_0((x - \theta)/\sigma)$ .

Luego, para obtener estimadores Fisher-consistentes bajo la normal

$$\text{MADN}(\mathbf{x}) = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \text{MAD}(\mathbf{x})$$
$$\text{IQRN}(\mathbf{x}) = \frac{1}{2 \Phi^{-1}\left(\frac{3}{4}\right)} \text{IQR}(\mathbf{x})$$

## Algunos estimadores de escala equivariantes

Sea  $x_i = \theta + \sigma u_i$ ,  $u_i \sim F_0$ , o sea,  $F_{\theta,\sigma}(x) = F_0((x - \theta)/\sigma)$ .

Un estimador de dispersión que no centra los datos y más robusto que IQRN fue propuesto por Rousseeuw (1992) y Croux y Rousseeuw (1993)

- $d_{ij} = x^{(i)} - x^{(j)}$ ,  $i > j$
- $d^{(1)} \leq d^{(2)} \leq \dots \leq d^{(m)}$  con  $m = \binom{n}{2}$

$$Q_n = d^{(k)} \quad \text{donde} \quad k = \binom{\left\lfloor \frac{n}{2} \right\rfloor + 1}{2}$$

$$2.222 \quad Q_n \xrightarrow{a.s.} \sigma$$

Tiene eficiencia 0.82

## Algunos estimadores de escala equivariantes

Sea  $x_i = \theta + \sigma u_i$ ,  $u_i \sim F_0$ , o sea,  $F_{\theta, \sigma}(x) = F_0((x - \theta)/\sigma)$ .

Martin y Zamar (1993) consideraron un estimador basado en el shorth.

- $[a_n, b_n]$  intervalo más corto que contiene a la mitad de los datos
- **Shorth de posición:**  $\hat{\theta} = \frac{a_n + b_n}{2}$

$$\hat{\theta} = \underset{t \in \mathbb{R}}{\operatorname{argmin}} \operatorname{mediana}_{1 \leq i \leq n} |x_i - t|$$

## Estimador de Dispersión

$$\hat{\sigma} = b_n - a_n$$

## Algunos estimadores de escala equivariantes

Si  $F_{\theta, \sigma} = N(\theta, \sigma^2)$ ,

- $a(F_{\theta, \sigma}) = \mu - c \sigma$
- $b(F_{\theta, \sigma}) = \mu + c \sigma$ , con

$$\Phi(c) - \Phi(-c) = \frac{1}{2} \quad \text{o sea} \quad c = \Phi^{-1} \left( \frac{3}{4} \right)$$

Luego,

$$\hat{\sigma} \xrightarrow{\text{a.s.}} 2\Phi^{-1} \left( \frac{3}{4} \right) \sigma$$

y podemos normalizarlo, para ser Fisher-consistente bajo la normal

**Estimador de Dispersión Normalizado**

$$\hat{\sigma} = \frac{1}{2\Phi^{-1} \left( \frac{3}{4} \right)} (b_n - a_n)$$





## M—estimadores de escala

Sea  $\mathbf{x}_i = \sigma \mathbf{u}_i$ ,  $u_i \sim F_1$ ,  $F_1$  con densidad  $f_1$ ,  $F_\sigma(x) = F_1(x/\sigma)$ .

**Una función  $\rho$**  es una función que cumple

- $\rho(x)$  es no decreciente en  $|x|$
- $\rho(0) = 0$
- $\rho(x)$  es creciente si  $x > 0$  y  $\rho(x) < \|\rho\|_\infty$
- Si  $\rho$  es acotada, se supondrá  $\|\rho\|_\infty = 1$

**Un M— estimador de escala  $\hat{\sigma}$**  es una solución de

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i}{\hat{\sigma}} \right) = \delta, \quad 0 < \delta < \|\rho\|_\infty$$

- Si  $\#\{x_i : x_i = 0\} > n(1 - \delta)$ , no existe solución y se define  $\hat{\sigma} = 0$

## Fisher-consistencia

Sea  $F_1 = N(0, 1)$  y  $c_{\rho, F_1}$  tal que

$$\mathbb{E}_{N(0,1)} \rho \left( \frac{x}{c_{\rho, F_1}} \right) = \delta$$

Sea  $\rho_c = \rho(\cdot/c)$ .

Luego, la solución  $\hat{\sigma} \xrightarrow{a.s.} \sigma(F)$

$$\mathbb{E}_F \rho_{c_{\rho, F_1}} \left( \frac{x}{\sigma(F)} \right) = \delta$$

y  $\sigma(F)$  es Fisher-consistente en  $N(0, 1)$

## M—estimadores de escala: Ejemplos

- Un estimador muy usado es el asociado a la función de Tukey

$$\rho_1^T(x) = \min\{1 - (1 - x^2)^3, 1\}$$

con  $\delta = 1/2$ .

La constante  $c_\rho = 1.56 \Rightarrow$  Fisher-consistencia en  $N(0, 1)$

- Si  $\rho_c(t) = I(|t| > c)$  y  $\delta = 0.5 \Rightarrow \hat{\sigma} = \text{mediana } |x_i|/c$  que no es el EMV para ninguna distribución

## Estimador preliminar de escala

$x_i = \theta + \sigma u_i$ ,  $u_i \sim F$  i.i.d.  $\sigma$  des conocido.  $\Rightarrow x_i \sim F_{\theta\sigma}$

Si  $\sigma = \sigma_0$  conocido definimos  $\hat{\theta}$  como la solución de

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i - \theta}{\sigma_0} \right).$$

Luego, el **M**—estimador de posición con estimador preliminar de escala  $\hat{\sigma}_n$  se define como el valor  $\hat{\theta}$  que minimiza

$$\tilde{S}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i - \theta}{\hat{\sigma}_n} \right).$$

- $\hat{\sigma}_n$  un estimador robusto y consistente de  $\sigma$
- Por ejemplo, la MAD,  $\hat{\sigma}_n^2 = \text{MADN}_{1 \leq i \leq n}(x_i)$

## M—estimador de posición con escala desconocida

$$\sum_{i=1}^n \psi \left( \frac{x_i - \theta}{\hat{\sigma}_n} \right) = 0. \quad (4)$$

**Teorema.** Sean  $x_1, \dots, x_n$  i.i.d.  $\mathbf{x}_i = \theta + \sigma \mathbf{u}_i$ ,  $\mathbf{u}_i \sim \mathbf{F}$ . Sea  $\hat{\theta}_n$  la solución de (4), con  $\psi$  impar y **F simétrica respecto de 0**.

Supongamos que  $\hat{\theta}_n \xrightarrow{P} \theta$ ,  $\hat{\sigma}_n \xrightarrow{P} \sigma$ , y que  $\psi \in \mathcal{C}^2$ ,  $\psi''$  es acotada. Luego se tiene que

$$\mathbf{n}^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathbf{N}(0, \sigma^2 \mathbf{V}(\psi, \mathbf{F})),$$

donde  $\mathbf{V}(\psi, \mathbf{F})$  está dada por

$$\mathbf{V}(\psi, \mathbf{F}) = \frac{\mathbf{E}_{\mathbf{F}} \psi^2(\mathbf{u})}{(\mathbf{E}_{\mathbf{F}} \psi'(\mathbf{u}))^2}.$$

## $M$ —estimador de posición con escala desconocida

$$\left| \hat{\theta}(x_0, \dots, x_0, x_{m+1}, \dots, x_n) - \hat{\theta}(x_1, \dots, x_n) \right| \quad n = 20 \quad x_0 = 1000$$

$m$	$\bar{x}$	mediana	$H_{MAD_p}$	$H_{MAD_s}$	$H_{SD}$	$\bar{x}_\alpha$	$T_{MAD_p}$	MAD	IQR
1	50	0.00	0.03	0.04	16.06	0.04	-0.02	0.12	0.08
2	100	0.01	0.10	0.11	46.78	55.59	0.04	0.22	0.14
2	200	0.21	0.36	0.37	140.5	166.7	0.10	0.46	0.41
5	250	0.34	0.62	0.95	202.9	222.3	0.15	0.56	370.3
7	350	0.48	1.43	42.66	350.0	333.4	0.21	1.29	740.3
9	450	0.76	3.23	450.0	450.0	444.5	0.40	2.16	740.2
10	500	500.5	500.0	500.0	500.0	500.0	500.0	793.3	740.2

$\alpha = 0.085$   $k = 1.37$  para el  $M$ —estimador con función de Huber  
 Tabla reproducida de Maronna *et al.* (2019)



## Intervalos de confianza

Sean  $x_i \sim N(\theta, \sigma^2)$  el IC se basa en  $T = (\bar{x} - \theta) / (s/\sqrt{n})$

$$\mathcal{I} = \left[ \bar{x} - \frac{s}{\sqrt{n}} t_{n-1, \frac{\alpha}{2}}, \bar{x} + \frac{s}{\sqrt{n}} t_{n-1, \frac{\alpha}{2}} \right]$$

Si  $x_i \sim F((x - \theta)/\sigma)$  con  $F$  simétrica con  $\mathbb{E}_F|X| < \infty$ ,

- $\mathcal{I}$  estará centrado
- **la longitud de  $\mathcal{I}$  puede verse muy incrementada si  $F \in \mathcal{F}_\epsilon = \{(1 - \epsilon)\Phi + \epsilon H : H \text{ simétrica } E_H|X| < \infty\}$ .**

## Intervalos de confianza

Si  $F \in \mathcal{F}_\epsilon = \{(1 - \epsilon)\Phi + \epsilon H : H \text{ no simétrica } E_H|X| < \infty\}$ ,

Entonces,  $\mathbb{E}_{F_{\theta, \sigma}} \bar{X} \neq \theta$  por lo tanto

- $\mathcal{I}$  no estará centrado
- la longitud de  $\mathcal{I}$  puede incrementarse.

Una manera es basarse en la distribución asintótica del  $M$ -estimador  $\hat{\theta}_n$ .

## Intervalos de confianza

Sea  $x_i = \theta + \sigma u_i$  con  $F \in \mathcal{F}_\epsilon = \{(1 - \epsilon)\Phi + \epsilon H : H \text{ simétrica}\}$

$$\hat{\theta}_n \approx N\left(\theta, \frac{V^2}{n}\right) = N\left(\theta, \frac{\sigma^2}{n} \frac{\mathbb{E}_F \psi^2(u)}{(\mathbb{E}_F \psi'(u))^2}\right),$$

Luego podemos estimar la varianza asintótica por

$$\widehat{V}^2 = \widehat{\text{ASVAR}} = \hat{\sigma}^2 \frac{\frac{1}{n} \sum_{i=1}^n \psi^2\left(\frac{x_i - \hat{\theta}_n}{\hat{\sigma}_n}\right)}{\left[\frac{1}{n} \sum_{i=1}^n \psi'\left(\frac{x_i - \hat{\theta}_n}{\hat{\sigma}_n}\right)\right]^2}$$

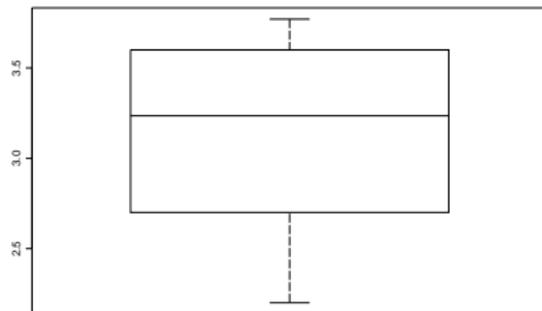
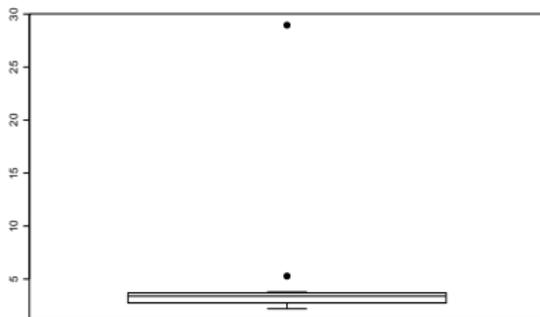
Un intervalo aproximado será

$$\mathcal{I} = \left[\hat{\theta}_n - \frac{\widehat{V}}{\sqrt{n}} z_{\frac{\alpha}{2}}, \hat{\theta}_n + \frac{\widehat{V}}{\sqrt{n}} z_{\frac{\alpha}{2}}\right]$$

# Intervalos de confianza: Ejemplo, Maronna *et al.*(2019)

24 determinaciones del contenido de cobre en harina

2.20	2.20	2.40	2.40	2.50	2.70	2.80	2.90
3.03	3.03	3.10	3.37	3.40	3.40	3.40	3.50
3.60	3.70	3.70	3.70	3.70	3.77	5.28	28.95



# Intervalos de confianza: Ejemplo, Maronna *et al.*(2019)

Estimador	$\hat{\theta}_n$	$\hat{V}$	$\mathcal{I}$
$\bar{x}$	4.280	5.297	[2.161,6.400]
$M$ — estimador con bicuadrada	3.144	0.637	[2.885, 3.404]
$\bar{x}_\alpha$	3.269	0.416	[3.103,3.435]

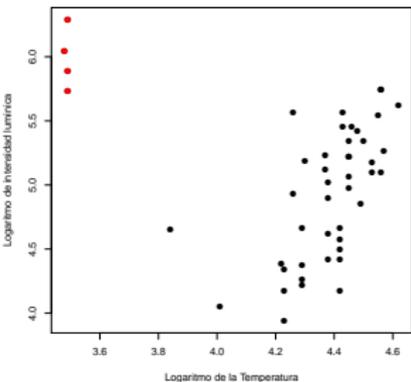
Intervalos de confianza: Ejemplo, Maronna *et al.*(2019)

Estimador	$\hat{\theta}_n$	$\hat{V}$	$\mathcal{I}$
$\bar{x}$	4.280	5.297	[2.161, 6.400]
$M$ — estimador con bicuadrada	3.144	0.637	[2.885, 3.404]
$\bar{x}_\alpha$	3.269	0.416	[3.103, 3.435]
$\bar{x}$ sin outliers	3.114	0.530	[2.892, 3.335]

# Inferencia robusta

- Logaritmo de la intensidad de luz de la estrella
- Logaritmo de la temperatura efectiva en la superficie

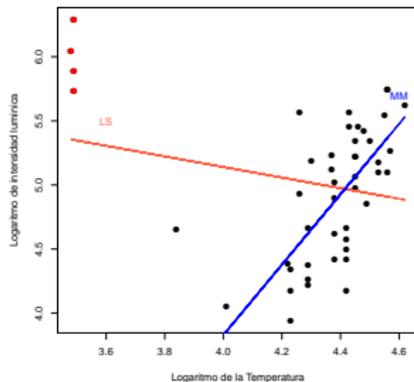
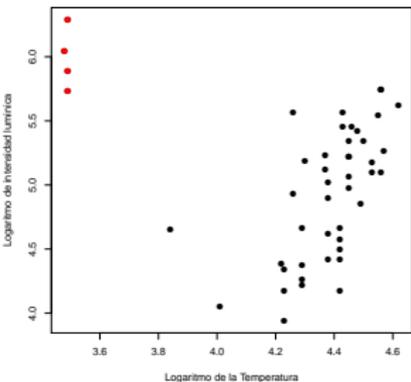
correspondientes al diagrama de Hertzsprung-Russell del Cúmulo de estrellas CYG OB1



# Inferencia robusta

- Logaritmo de la intensidad de luz de la estrella
- Logaritmo de la temperatura efectiva en la superficie

correspondientes al diagrama de Hertzsprung-Russell del Cúmulo de estrellas CYG OB1



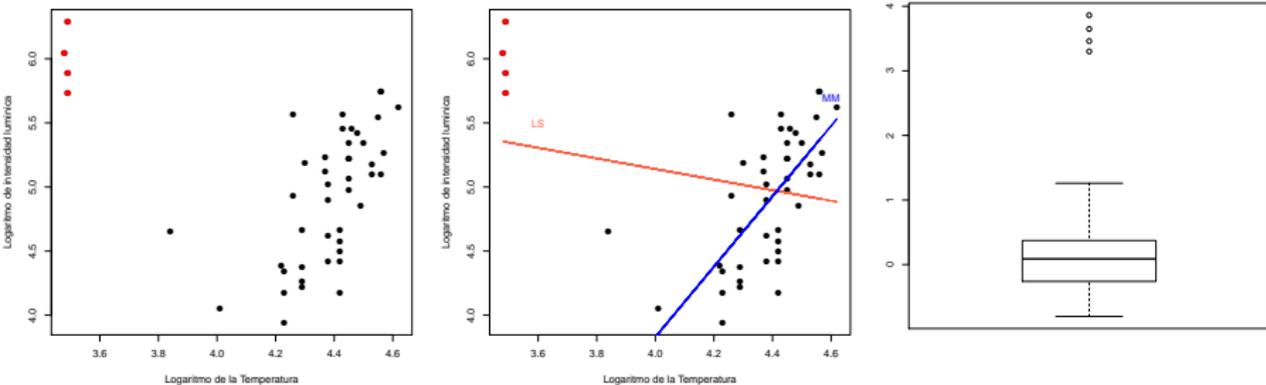
MM—estimador de regresión en azul<sup>a</sup>.

Estimador de mínimos cuadrados en rojo

# Inferencia robusta

- Logaritmo de la intensidad de luz de la estrella
- Logaritmo de la temperatura efectiva en la superficie

correspondientes al diagrama de Hertzsprung-Russell del Cúmulo de estrellas CYG OB1



**MM—estimador de regresión en azul<sup>a</sup>.**

**Estimador de mínimos cuadrados en rojo**

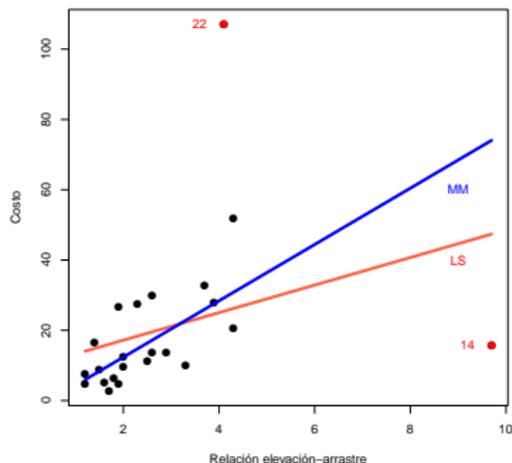
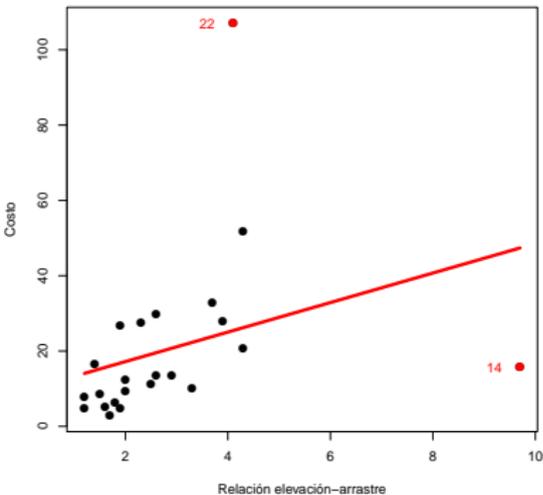
<sup>a</sup>Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, **15**, 642-656.



# Estimador de mínimos cuadrados

23 aviones monomotores construidos durante los años 1947-1979.

$y =$  costo (en unidades de \$ 100,000)      $x =$  relación elevación-arrastre



- Estimador de mínimos cuadrados en rojo
- El estimador de mínimos cuadrados minimiza

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

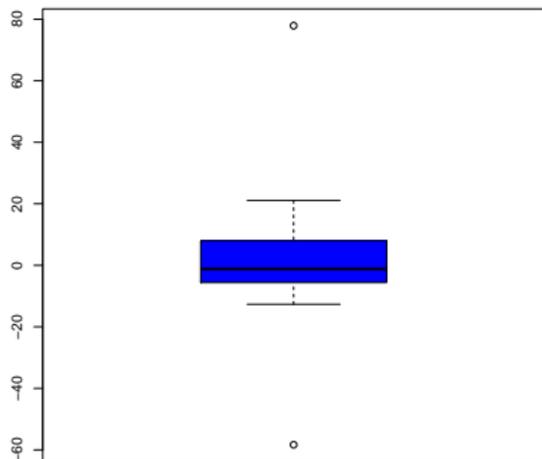
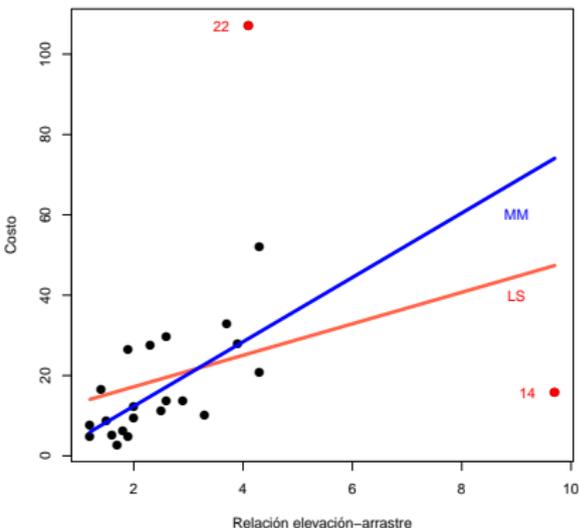
- MM—estimador de regresión en azul
- El MM—estimador\* minimiza

$$\sum_{i=1}^n \rho \left( \frac{y_i - \alpha - \beta x_i}{\hat{\sigma}} \right)$$

# Estimador de mínimos cuadrados

23 aviones monomotores construidos durante los años 1947-1979.

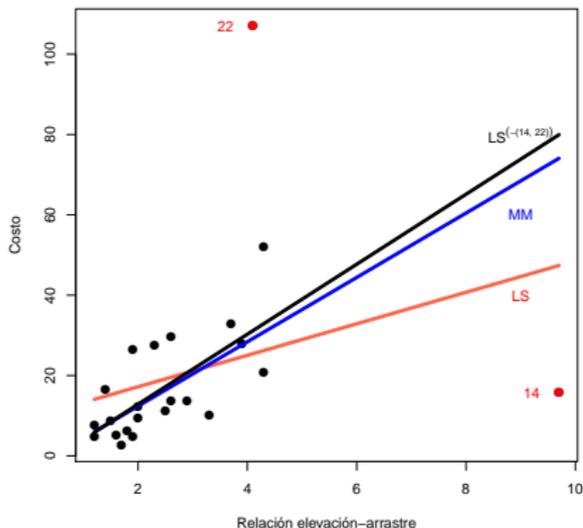
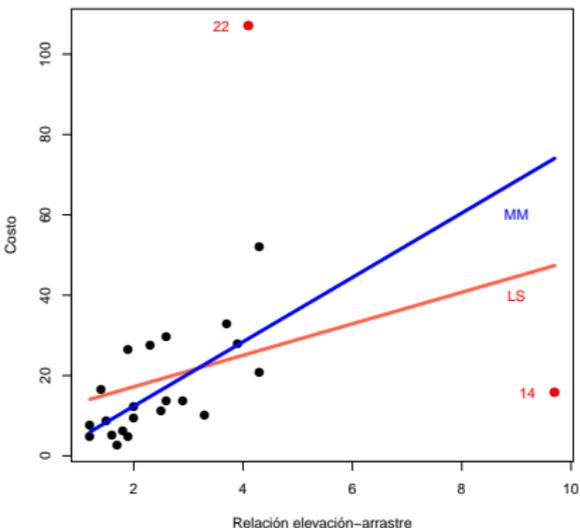
$y =$  costo (en unidades de \$ 100,000)     $x =$  relación elevación-arrastre



# Estimador de mínimos cuadrados

23 aviones monomotores construidos durante los años 1947-1979.

$y =$  costo (en unidades de \$ 100,000)     $x =$  relación elevación-arrastre



## M-estimadores

Para combinar robustez y eficiencia y teniendo en cuenta que

$$\hat{\beta}_{\text{LS}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n r_i^2(\beta)$$

podemos definir los  $M$ -estimadores como

$$\hat{\beta}_M = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho \left( \frac{r_i(\beta)}{\hat{\sigma}_n} \right)$$

donde

- $\hat{\sigma}_n$  es un estimador de la escala del error
- $\rho$  es una función  $\rho$

## M-estimadores

### M-estimadores

$$\hat{\beta}_M = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho \left( \frac{r_i(\beta)}{\hat{\sigma}_n} \right)$$

donde

- $\hat{\sigma}_n$  es un estimador de la escala del error
- $\rho$  es una función  $\rho$

Un caso particular, es el estimador  $L^1$  que corresponde a  $\rho(t) = |t|$

$$\hat{\beta}_{L^1} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n |r_i(\beta)|$$

## M—estimadores

Son

- **Regresión equivariante**

$$\widehat{\beta}(\mathbf{X}, \mathbf{y} + \mathbf{X}\gamma) = \widehat{\beta}(\mathbf{X}, \mathbf{y}) + \gamma$$

- **Escala equivariante**

$$\widehat{\beta}(\mathbf{X}, \lambda \mathbf{y}) = \lambda \widehat{\beta}(\mathbf{X}, \mathbf{y})$$

- **Afín equivariante**

$$\widehat{\beta}(\mathbf{X}\mathbf{A}, \lambda \mathbf{y}) = \mathbf{A}^{-1} \widehat{\beta}(\mathbf{X}, \mathbf{y})$$

$\mathbf{A} \in \mathbb{R}^{p \times p}$  no singular

## M-estimadores

Como

$$\hat{\beta}_M = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \beta}{\hat{\sigma}_n} \right),$$

entonces,  $\hat{\beta}_M$  satisface

$$\frac{1}{n} \sum_{i=1}^n \psi \left( \frac{y_i - \mathbf{x}_i^T \hat{\beta}_M}{\hat{\sigma}_n} \right) \mathbf{x}_i = 0,$$

## M—estimadores: Propiedades

Si

- $\hat{\sigma}_n \xrightarrow{P} \sigma$
- $\mathbb{E}_F \psi(u/\sigma) = 0$
- $\max_{1 \leq j \leq n} \mathbf{x}_j^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_j \rightarrow 0$

entonces

$$\hat{\beta}_n \xrightarrow{a.s.} \beta$$

## M—estimadores: Propiedades

Sea  $\mathbf{S} \in \mathbb{R}^{p \times p}$  tal que  $\mathbf{S}\mathbf{S}^T = \mathbf{X}^T\mathbf{X}$ . Si Si

- $\hat{\sigma}_n \xrightarrow{p} \sigma$
- $\mathbb{E}_F \psi(u/\sigma) = 0$
- $\max_{1 \leq j \leq n} \mathbf{x}_j^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{x}_j \rightarrow 0$

entonces

$$\mathbf{S}^T (\hat{\beta}_n - \beta) \xrightarrow{D} N(0, \sigma^2 V(\psi, F) \mathbf{I})$$

con

$$V(\psi, F) = \frac{\mathbb{E}_F \psi^2(u)}{(\mathbb{E}_F \psi'(u))^2}.$$

Luego, la eficiencia no depende de la distribución de  $\mathbf{x}$ ,

$$\text{eff}(\hat{\beta}_n) = 1/V(\psi, F).$$

## Cómo elegimos $\hat{\sigma}_n$ ?

Una opción es

1. Calcular el estimador  $L^1$ ,  $\hat{\beta}_{L^1}$

2. Definir  $\hat{r}_i = r_i(\hat{\beta}_{L^1})$

3.

$$\hat{\sigma}_n = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \operatorname{mediana}_{\substack{1 \leq i \leq n \\ \hat{r}_i \neq 0}} |\hat{r}_i|$$

## Cómo elegimos $\hat{\sigma}_n$ ?

Es fácil ver que  $\hat{\sigma}_n$  es

- **Regresión invariante**  $\hat{\sigma}_n(\mathbf{X}, \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}) = \hat{\sigma}_n(\mathbf{X}, \mathbf{y})$
- **Afín invariante**  $\hat{\sigma}_n(\mathbf{XA}, \lambda\mathbf{y}) = \hat{\sigma}_n(\mathbf{X}, \mathbf{y})$ ,  $\mathbf{A} \in \mathbb{R}^{P \times P}$  no singular
- **Escala equivariante**  $\hat{\sigma}_n(\mathbf{X}, \lambda\mathbf{y}) = |\lambda| \hat{\sigma}_n(\mathbf{X}, \mathbf{y})$

con lo cual el  $M$ —estimador resultante es regresión, afín y escala equivariante.

Además,  $\hat{\sigma}_n \xrightarrow{P} \sigma$ .

## Comentario

- Como en posición, estimadores con  $\psi$  redescendientes dan peso 0 a residuos grandes lo que permite tener alta eficiencia bajo normalidad y bajo colas pesadas.
- Los  $M$ —estimadores con  $\psi$  monótona se recomiendan para diseños 0 – 1 y quizás para diseños uniformes, pero no para diseños donde pueden aparecer puntos de alta palanca.
- Para combinar eficiencia y robustez, se recomienda usar la bicuadrada para calcular el  $M$ —estimador con el estimador preliminar de escala descripto.

## Cómputo de $M$ —estimadores de regresión

Como en posición un  $M$ —estimador puede verse como un estimador de mínimos cuadrados pesado.

Sea

$$W(t) = \frac{\psi(t)}{t}$$
$$\hat{w}_i = W\left(\frac{r_i(\hat{\beta}_n)}{\hat{\sigma}_n}\right)$$

Luego,

$$\frac{1}{n} \sum_{i=1}^n \psi\left(\frac{y_i - \mathbf{x}_i^T \hat{\beta}_n}{\hat{\sigma}_n}\right) \mathbf{x}_i = 0 \Leftrightarrow$$
$$\frac{1}{n} \sum_{i=1}^n \hat{w}_i (y_i - \mathbf{x}_i^T \hat{\beta}_n) \mathbf{x}_i = 0$$

## Cómputo de $M$ -estimadores de regresión

Podemos aplicar el siguiente proceso iterativo **IRWLS**:

- a) Calculamos el estimador  $L^1$ ,  $\hat{\beta}^{(1)} = \hat{\beta}_{L^1}$ ,

$$\hat{\sigma}_n = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \operatorname{mediana}_{\substack{1 \leq i \leq n \\ r_i(\hat{\beta}_{L^1}) \neq 0}} |r_i(\hat{\beta}_{L^1})|$$

- b) Para  $k = 1, 2, \dots$ , dado  $\hat{\beta}^{(k)}$

i) Sea  $\hat{r}_{i,k} = r_i(\hat{\beta}^{(k)}) = y_i - \mathbf{x}_i^T \hat{\beta}^{(k)}$  y  $\hat{w}_{i,k} = W(\hat{r}_{i,k}/\hat{\sigma}_n)$

- ii) Calcule  $\hat{\beta}^{(k+1)}$  resolviendo

$$\frac{1}{n} \sum_{i=1}^n \hat{w}_{i,k} (y_i - \mathbf{x}_i^T \beta) \mathbf{x}_i = 0$$

- c) Pare cuando  $\max_i |\hat{r}_{i,k} - \hat{r}_{i,k+1}| / \hat{\sigma}_n < \epsilon$

## MM—estimadores

Una propuesta para obtener estimadores robustos, cuando  $\mathbf{x}$  e  $y$  pueden contener outliers, consiste en

- tomar una función  $\rho$  acotada

combinada con

- un estimador preliminar de escala asociado a un estimador inicial de alto punto de ruptura.

## MM–estimadores

Consideraremos un  $M$ –estimador de regresión,  $\hat{\beta}$ , solución de

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \beta}{\hat{\sigma}_n} \right),$$

donde

- $\hat{\sigma}_n$  es un estimador de la escala del error asociado a un estimador inicial de alto punto de ruptura y
- $\rho$  es una función  $\rho$  **acotada**.

## MM—estimadores

Si  $\rho$  es derivable

$$\frac{1}{n} \sum_{i=1}^n \psi \left( \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}_n} \right) \mathbf{x}_i = \mathbf{0}$$

Esta ecuación tendrá varias soluciones y generalmente sólo una corresponde al mínimo global.

Si  $\hat{\sigma}_n$  es **regresión y afín invariante** y **escala equivariante**,  $\hat{\boldsymbol{\beta}}$  será regresión, afín y escala equivariante.

## MM—estimadores

Una manera de obtener un estimador robusto y con alta eficiencia es encontrar un *buen* mínimo local de

$$L_\rho(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\hat{\sigma}_n} \right)$$

## MM—estimadores

Una manera de obtener un estimador robusto y con alta eficiencia es encontrar un *buen* mínimo local de

$$L_{\rho}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\hat{\sigma}_n} \right)$$

Sean  $\rho_1$  y  $\rho_0$  dos funciones  $\rho$  acotadas, tales que  $\rho_1 \leq \rho_0$

## MM—estimadores

Una manera de obtener un estimador robusto y con alta eficiencia es encontrar un *buen* mínimo local de

$$L_{\rho}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\hat{\sigma}_n} \right)$$

Sean  $\rho_1$  y  $\rho_0$  dos funciones  $\rho$  acotadas, tales que  $\rho_1 \leq \rho_0$

Por ejemplo,

$$\begin{aligned} \rho_0(t) &= \rho_{\text{TUK}}(t/c_0) & \rho_1(t) &= \rho_{\text{TUK}}(t/c_1) \\ \rho_{\text{TUK}}(t) &= \min\{1 - (1 - t^2)^3, 1\} \end{aligned}$$

con  $c_1 > c_0$ .

## MM—estimadores

- a) Calcule un estimador inicial de alto punto de ruptura,  $\hat{\beta}_{\text{INI}}$ , pero eventualmente poco eficiente bajo normalidad

- Para cada  $\beta$  sea  $\hat{\sigma}(\beta)$  la solución de

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{r_i(\beta)}{\hat{\sigma}(\beta)} \right) = \frac{1}{2}$$

- Defina  $\hat{\beta}_{\text{INI}} = \operatorname{argmin}_{\beta} \hat{\sigma}(\beta)$

- b) Calcule la escala robusta,  $\hat{\sigma}_n$  de los residuos  $r_i(\hat{\beta}_{\text{INI}})$ , o sea,

$$\hat{\sigma}_n = \hat{\sigma}(\hat{\beta}_{\text{INI}})$$

con lo cual

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{r_i(\hat{\beta}_{\text{INI}})}{\hat{\sigma}_n} \right) = \frac{1}{2}$$

Si elegimos  $\rho_0(t) = \rho_{\text{TUK}}(t/c_0)$  y  $c_0 = 1.56$ ,  $\hat{\sigma}_n \xrightarrow{P} \sigma$  si  $u_i \sim N(0, \sigma^2)$ .

## MM-estimadores

c) Sea  $\psi_1 = \rho'_1$ . Busque una solución  $\hat{\beta}_{MM}$  de

$$\frac{1}{n} \sum_{i=1}^n \psi_1 \left( \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\hat{\sigma}_n} \right) \mathbf{x}_i = 0 \quad (5)$$

tal que  $L_{\rho_1}(\hat{\beta}_{MM}) \leq L_{\rho_1}(\hat{\beta}_{INI})$ .  $\hat{\beta}_{MM}$  se llama un MM-estimador.

Para el cálculo de  $\hat{\beta}_{MM}$  se utiliza IRWLS partiendo de  $\hat{\beta}_{INI}$ , con  $W(t) = \psi_1(t)/t$ .

Eficiencia	0.8	0.85	0.9	0.95
$c_1$	3.14	3.44	3.88	4.68

## Métodos numéricos: Algoritmo de Remuestreo

Para encontrar una solución aproximada de  $\hat{\sigma}(\hat{\beta}) = \underset{\beta}{\operatorname{argmin}} \hat{\sigma}(\beta)$  que sirva de *buen punto inicial* se utiliza remuestreo

- Tomemos  $N$  submuestras  $\mathcal{I}_k$ ,  $1 \leq k \leq N$  al azar de tamaño  $p$ . Si una submuestra es colinear reemplacela por otra.
- Para cada submuestra  $\{(\mathbf{x}_i, y_i), i \in \mathcal{I}\}$  sea  $\hat{\beta}_{\mathcal{I}}$  tal que

$$\mathbf{x}_i^T \hat{\beta}_{\mathcal{I}} = y_i \quad i \in \mathcal{I}$$

- Calcule

$$k^* = \underset{1 \leq k \leq N}{\operatorname{argmin}} \hat{\sigma}(\hat{\beta}_{\mathcal{I}_k})$$

- $\hat{\beta}_{\mathcal{I}_{k^*}}$  es el candidato a punto inicial