5 Oct 2001

Resistant.

Insensitive to localized misbehavior in the data

eg. median

~~mean~~

Robust.

Insensitive to departures from assumptions surrounding an underlying probability model

mean

trimmed mean

5 Oct 2001

M - estimates

The Biweight

$$y^* = \frac{\sum w_i y_i}{\sum w_i}$$

$$w_i = \left(1 - \left(\frac{y_i - y^*}{cS}\right)^2\right)^2 \qquad \left(\frac{y_i - y^*}{cS}\right)^2 < 1$$

$$= 0 \qquad \text{otherwise}$$

$$S = \text{median}\{|y_i - y^*|\}$$

or $\frac{1}{2}$ IQR

S estimates "$\frac{2}{3}\sigma$"

eg $c = 6$, residuals count up to 4σ

Iterate to convergence

12 Oct. 01

The heuristics of robust regression. Consider a weight function like

$$w(r) = (1-r^2)^2 \qquad |r| \le 1$$
$$ = 0 \qquad \text{otherwise}$$

Consider a preliminary estimate $\hat{\beta}^-$ and residuals $\hat{\epsilon}^- = Y - X\hat{\beta}^-$

Regress y_i on x_i with weight $w\left(\frac{\hat{\epsilon}_i^-}{\sigma}\right) = w_i$

i.e. $\hat{\beta}$ to satisfy

$$\sum_i (y_i - x_i^T \hat{\beta}) w_i \overset{k\times 1}{x_i} = \overset{k\times 1}{0} \qquad \text{cp.}$$

In the limit

$$\sum_i (y_i - x_i^T \hat{\beta}) w\left(\frac{y_i - x_i^T\hat{\beta}}{\sigma}\right) x_i = 0 \qquad \text{cp. } (Y - X\hat{\beta})^T W X = 0$$

i.e.

$$\psi(r) = r\,w(r) \qquad \text{or} \qquad \boxed{w(r) = \psi(r)/r}$$
$$= r(1-r^2)^2 \qquad |r| \le 1$$
$$= 0 \qquad \text{otherwise}$$

$$\rho(r) = \int_0^r \tfrac{1}{2}(1-s^2)^2\, ds^2 = -\int_0^{r^2} \tfrac{1}{2}(1-t)^2\, d(1-t) \qquad t = s^2$$

$$= \int_{1-r^2}^{1} \tfrac{1}{2}u^2\, du = \tfrac{1}{2}\,\frac{u^3}{3}\Bigg|_{1-r^2}^{1} \qquad u = 1-t$$

$$\tfrac{1}{2}(1-r^2)^2\, 2r$$

$$\boxed{\rho(r) = \tfrac{1}{6}\left[1 - (1-r^2)^3\right] \quad |r| \le 1 \qquad = \tfrac{1}{6} \quad |r| \ge 1}$$

7 Oct. 01

This approach suggests the approximate variance matrix

$\mathrm{var}(\hat{\beta} - \beta) \sim \sigma^2 (X^T \hat{W} X)^{-1}$

e.g. $\hat{\sigma} = \frac{3}{4} \mathrm{IQR}$, $\mathrm{median}\{|\hat{\epsilon}_i|\}/.6745$

Asymptotic normality from M-estimate theory

Start eg. with L_1 estimate

12 Oct. 01

Robust / Resistant Estimates

Would like an automatic way to handle outliers and long-tailed distributions.

Give up full efficiency, eg. in normal case, but obtain protection against outliers and nonnormality.

$$y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \epsilon_i$$

M-estimate

$$\min_{\underset{\sim}{\beta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \overset{1\times k}{\underset{\sim}{x}_i^T} \overset{k\times 1}{\underset{\sim}{\beta}}}{\sigma}\right)$$

or to satisfy

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \overset{1\times k}{\underset{\sim}{x}_i^T} \overset{k\times 1}{\hat{\underset{\sim}{\beta}}}}{\sigma}\right) \overset{k\times 1}{\underset{\sim}{x}_i} = \overset{k\times 1}{\underset{\sim}{0}}$$

$$\psi = \rho'$$

Chapter 8 in W & R

How to compute ?

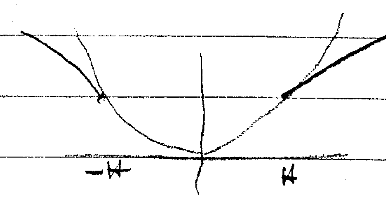
17 Oct '01

Examples of ρ:

1. OLS $\quad \rho(r) = r^2$

2. L_1 $\quad \rho(r) = |r|$

3. Huber $\quad \rho(r) = r^2 \qquad |r| \le H$



$$= H(2|r| - H)$$

H: large : normal
H small : L_1

H = 1.345 gives 95% efficiency at normal
ρ: convex so math convenient

4. Bisquare $\quad \rho(r) = \frac{B^2}{6}\left[1 - \left(1 - \left(\frac{r}{B}\right)^2\right)^3\right] \qquad |r| \le B$

$$= \frac{B^2}{6} \text{ otherwise}$$

8 Oct. 2001

~~Example~~. Detecting circular arcs in image. Roads.

Fitting a circular arc when outliers severely perturb fits

C: $(x-a)^2 + (y-b)^2 = R^2$

$$x^2+y^2 = (-2a)x + (-2b)y + (a^2+b^2-R^2)$$

$$z = \theta_1 x + \theta_2 y + \theta_3$$

CAE: circular arc estimator

DRO: Duda road operator
- nonlinear line detector

Theil-Sen: median of $\binom{N}{2}$ slopes

RM: repeated median

LSM: least median of squared residuals

MM: variant of M-estimator
- estimates scale too

SW: sliding window

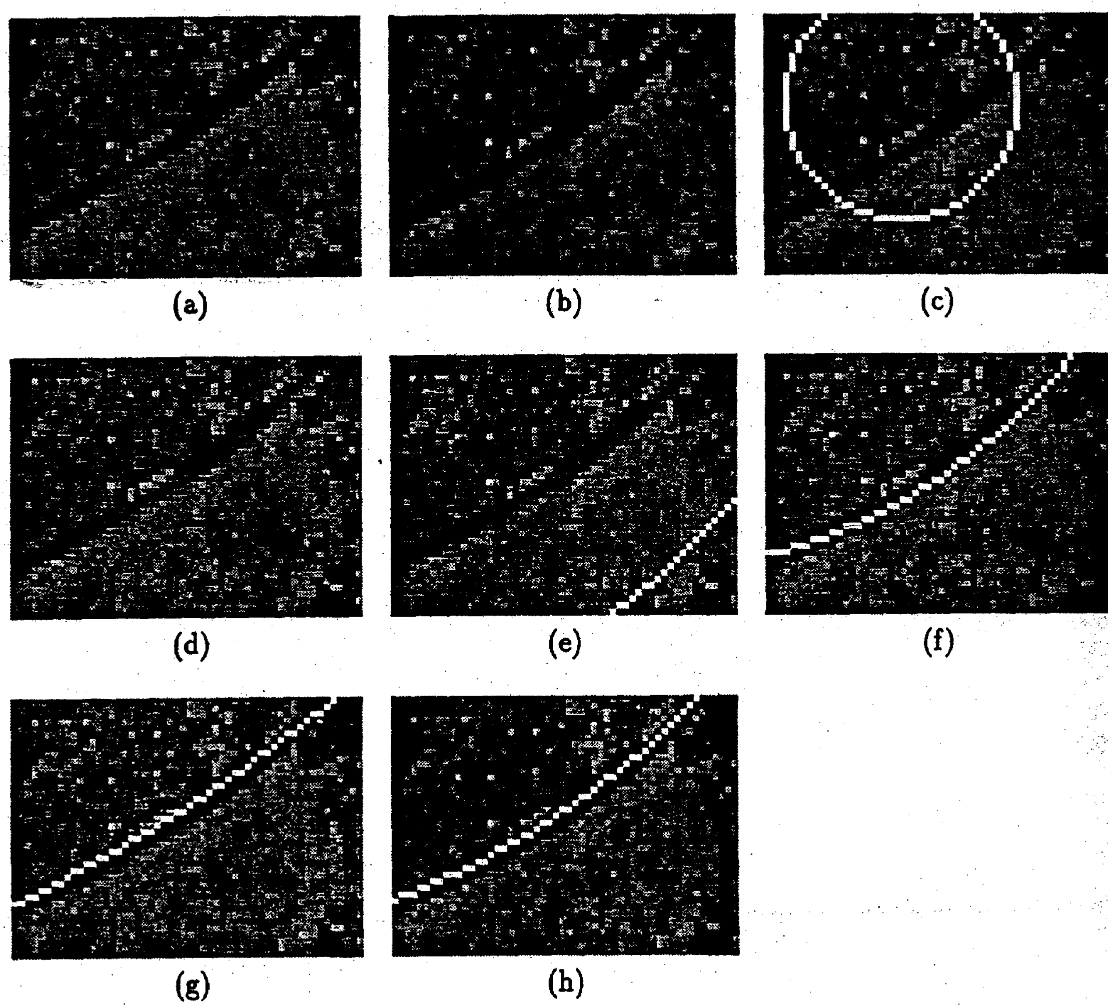


Figure 15: Data set 2: (a) Input image; (b) DRO output; (c) OLS fit; (d) Theil-Sen CAE fit (not contained in the displayed window); (e) RM CAE fit; (f) LMS fit; (g) nonlinear MM fit; (h) SW fit using an unbiased OLS estimator.

18 Oct. 01

Robust estimates.

M-estimates. Return to the regression problem

$$\underset{\sim}{y} \sim \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

Consider the problem

$$\min_{\beta} \sum_{j=1}^{n} \rho\left(\frac{y_j - \underset{\sim}{x}_j^T \underset{\sim}{\beta}}{s}\right) \qquad \underset{\sim}{x}_j \quad j\text{-th row}$$

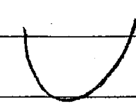
s: a known or estimated scale parameter, eg.

$$1.48\left(\operatorname*{med}_j \left| (y_j - \underset{\sim}{x}_j \hat{\beta}_{OLS}) - \operatorname{med}(y_j - \underset{\sim}{x}_j \hat{\beta}_{OLS}) \right| \right)$$

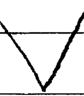
1.48 works for normal

Here $\rho(r)$ is a (robust) loss function

eg. $\rho(r) = r^2$ OLS

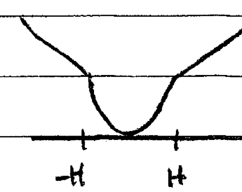


$\rho(r) = |r|^p$ L_p regression



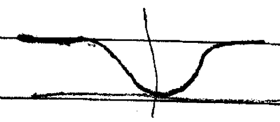
$p=1$

$$\rho(r) = \begin{cases} r^2 & |r| \le H \\ H|r| - \frac{1}{2}H^2 & |r| > H \end{cases}$$ Huber



8 Oct 01

$$\rho(r) = \begin{cases} \frac{B^2}{6}\left[1-\left(1-\left(\frac{r}{B}\right)^2\right)^3\right] & |r| \le B \\ \frac{B^2}{6} & |r| > B \end{cases}$$



biweight
bisquare

Might pick tuning constants $H, B, \ldots$ so 95% efficient in the normal error case.

Note. 1. The location problem is a particular case with $\underset{\sim}{X} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

L_2 regression gives sample mean
L_1 " " " median

2. $\exists$ other rules for estimating scale

Properties of $\rho(.)$
i) $\rho(r) \geq 0$, $\rho(r)$ non decreasing for $r \geq 0$
ii) $\rho(0) = 0$
iii) $\rho(-r) = \rho(r)$
iv) $\rho(.)$ continuous for all but a finite number of r

8. Oct. 01

Differentiate wrt $\underset{\sim}{\beta}$ and write $\boxed{\psi = \rho'}$

$$\sum_{j=1}^{n} \underset{\sim}{x}_j^T \psi\left(\frac{y_j - \underset{\sim}{x}_j \hat{\underset{\sim}{\beta}}}{s}\right) = \underset{\sim}{0} \qquad (*)$$

A set of nonlinear equations
- solve iteratively
- good ~~starting~~ values needed

L_1 (linear programming)

Write $\boxed{w(r) = \psi(r)/r}$

$(*)$ becomes

$$\sum_{j=1}^{n} w\left(\frac{y_j - \underset{\sim}{x}_j \hat{\underset{\sim}{\beta}}}{s}\right) \underset{\sim}{x}_j^T (y_j - \underset{\sim}{x}_j^T \hat{\underset{\sim}{\beta}}) = \underset{\sim}{0}$$

OR

$$\sum_{j=1}^{n} w_j \, \underset{\sim}{x}_j^T (y_j - \underset{\sim}{x}_j^T \hat{\underset{\sim}{\beta}}) = \underset{\sim}{0}$$

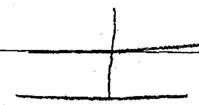
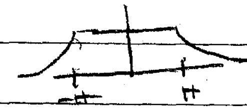
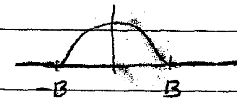
with data dependent weights

$$w_j = w\left(\frac{y_j - \underset{\sim}{x}_j \hat{\underset{\sim}{\beta}}_-}{s}\right)$$

IRLS regress $\sqrt{w_j}\, y_j$ on $\sqrt{w_j}\, x_{ij}$

until "convergence"

18. Oct. 01

$w(r)$			
2			OLS
$1/\lvert r\rvert$			L_1
1	$\lvert r\rvert \le H$		Huber
$H/\lvert r\rvert$	$\lvert r\rvert > H$		
$(1-(r/B)^2)^2$	$\lvert r\rvert \le B$		Bisquare
0	$\lvert r\rvert > B$		
			Trisquare

Notes.

1. ρ: nondecreasing $\rightarrow$ estimate exists and unique
2. $\rho(r)$ = constant for $|r| \geq$ constant
   $\Longrightarrow$ possible deletion of observations

Classification

$\psi(r) = 0$ for $|r|$ sufficiently large, eg. Bisquare
hard redescender

$\psi(r) \rightarrow 0$ as $|r| \rightarrow \infty$, eg. Cauchy
soft redescender $\quad w(r) = 1/(1+(r/c)^2)$

Huber

nondecreasing

8 Oct 01

Biweight estimate of location

$$\hat{\mu} = \sum_i w_i y_i \Big/ \sum_i w_i$$

$$w_i = \left(1 - \left(\frac{y_i - \hat{\mu}_-}{cs}\right)^2\right)_+^2$$

$$s = \text{med}\{|y_i - \hat{\mu}_-|\}$$

$$c = 6, 9$$

Seems to converge quickly

Freedman and Diaconis (1982) Ann Statist 10, 454-461
inconsistency results

8 Oct 01

Heuristics in location case (Jeffries)

Y_i from $f\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma}$ σ: "known"

log likelihood $\sum_i \log f\left(\frac{y_i-\mu}{\sigma}\right)$

$\frac{\partial}{\partial\mu}$: $-\frac{1}{\sigma}\sum_i \frac{1}{f\left(\frac{y_i-\hat{\mu}}{\sigma}\right)} f'\left(\frac{y_i-\hat{\mu}}{\sigma}\right) = 0$

or $\sum_i w_i\,(y_i-\hat{\mu}) = 0$

$$w_i = -\frac{1}{f\left(\frac{y_i-\hat{\mu}}{\sigma}\right)}\, f'\left(\frac{y_i-\hat{\mu}}{\sigma}\right)\frac{1}{(y_i-\hat{\mu})}$$

Cauchy: $f = \frac{1}{1+y^2}$, $f' = -\frac{2y}{(1+y^2)^2}$

$$w = -\frac{1}{f}f'\frac{1}{y} = (1+y^2)\frac{2y}{(1+y^2)^2}\frac{1}{y} = \frac{2}{(1+y^2)}$$

Jeffries took mixture of normal and uniform

8 Oct. 01

Approximate distribution

Suppose $\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ distribution $\frac{\varepsilon_j}{\sigma}$: P

$$\hat{\underset{\sim}{\beta}} \sim N_p\left(\underset{\sim}{\beta},\ \frac{E_P(\psi^2)}{[E_P(\psi')]^2}(\underset{\sim}{X}^T\underset{\sim}{X})^{-}\sigma^2\right)$$

Estimate of the covariance matrix

$$\frac{(ns)^2}{n-p}\ \frac{\sum_{j=1}^{n}\psi\left(\frac{\hat{\varepsilon}_j}{s}\right)^2}{\left[\sum_{j=1}^{n}\psi'\left(\frac{\hat{\varepsilon}_j}{s}\right)\right]^2}(\underset{\sim}{X}^T\underset{\sim}{X})^{-} \qquad \psi(r) = r\,w(r)$$

Sometimes $s^2(\underset{\sim}{X}^T\hat{W}\underset{\sim}{X})^{-}$

e.g. $s^2 = \frac{3}{4}\,\mathrm{IQR}$

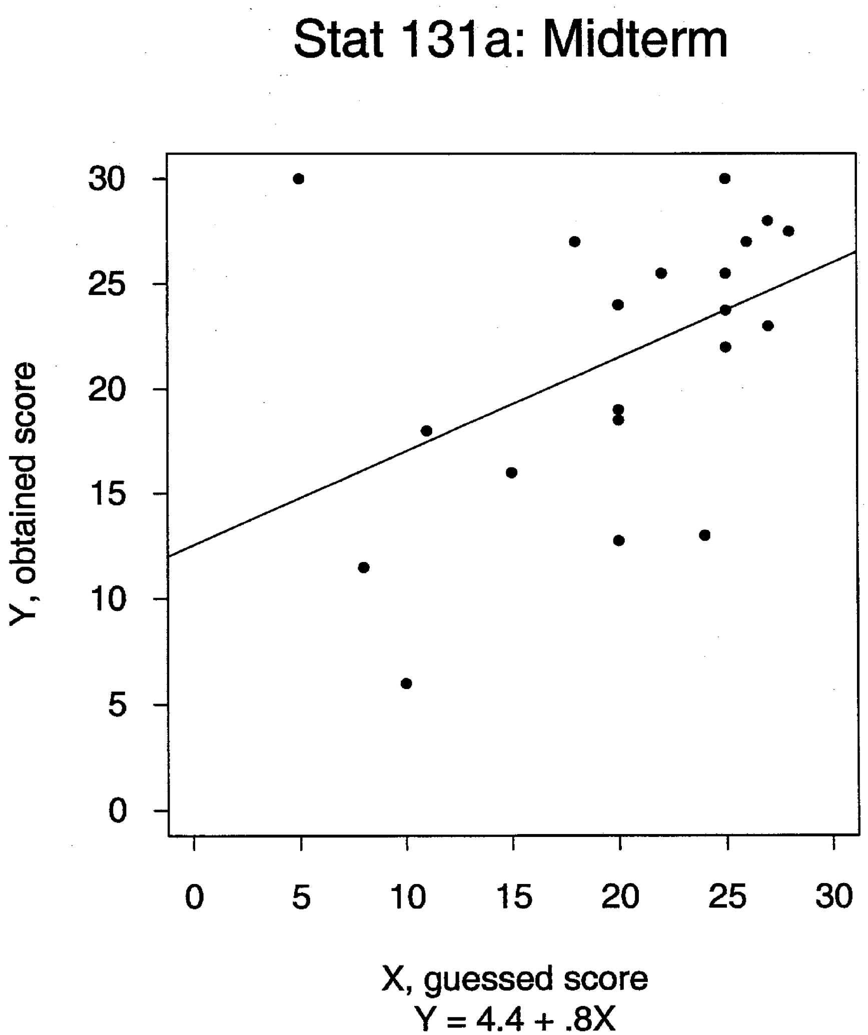
8 Oct 01

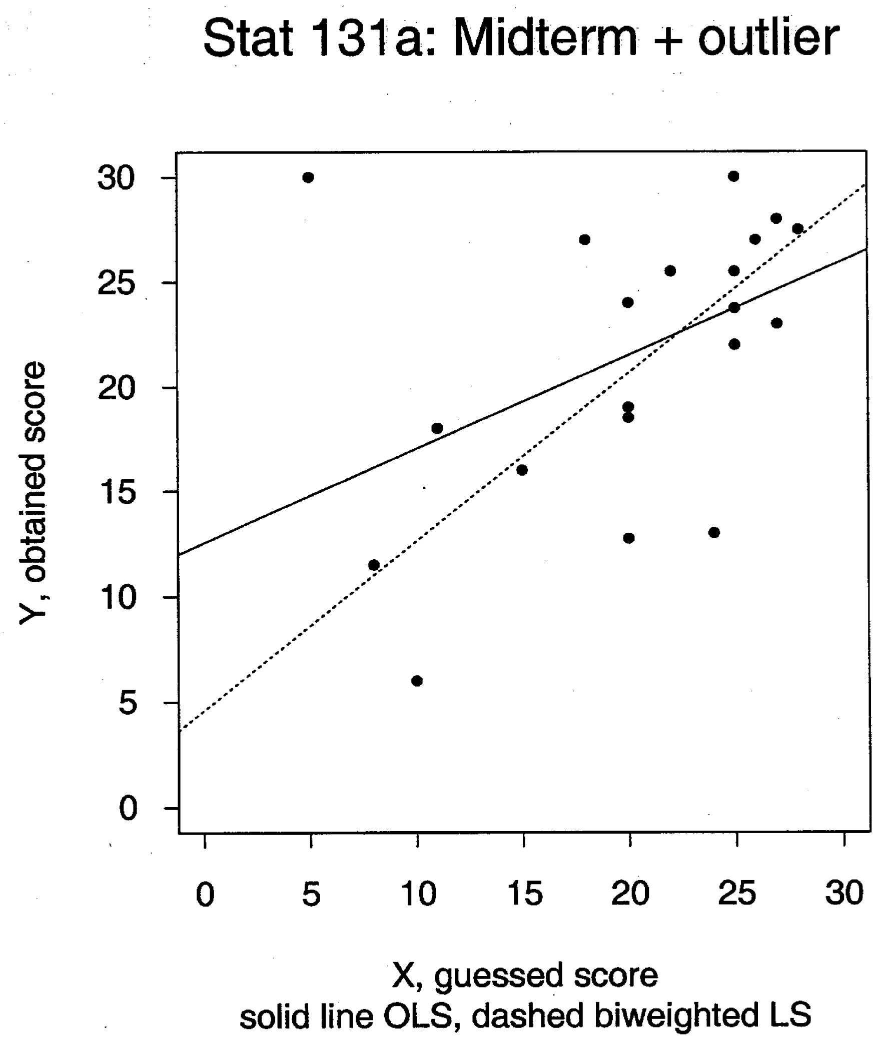
In practice run OLS and rob/res in parallel.

If results dissimilar, indication on an invalid assumption (of classical regression)

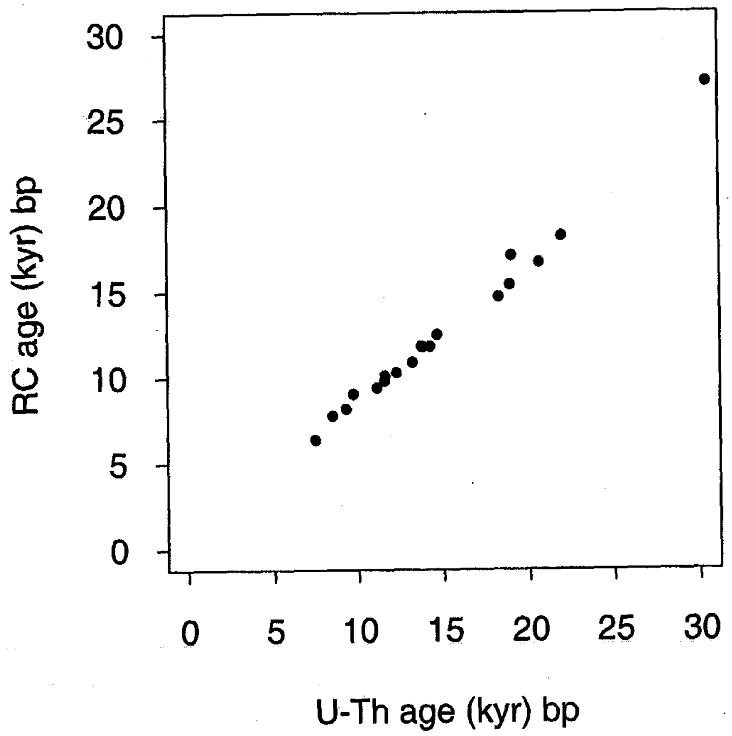
Look at residuals for possible outliers
eg. index plot

Look at index plot of weights

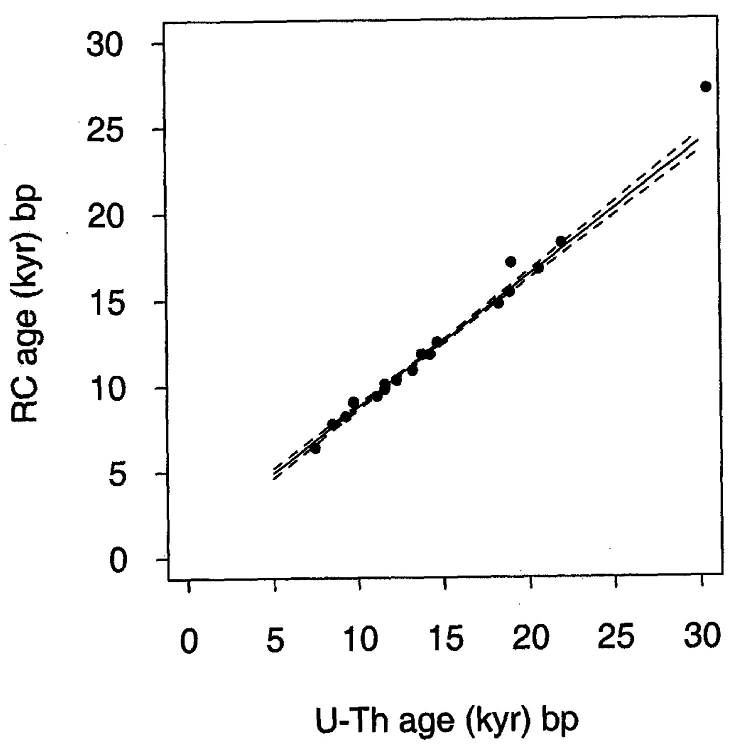

Stat 131a: Midterm
Y, obtained score
0
5
10
15
20
25
30
X, guessed score
Y = 4.4 + .8X


Stat 131a: Midterm + outlier
Y, obtained score
0
5
10
15
20
25
30
X, guessed score
solid line OLS, dashed biweighted LS

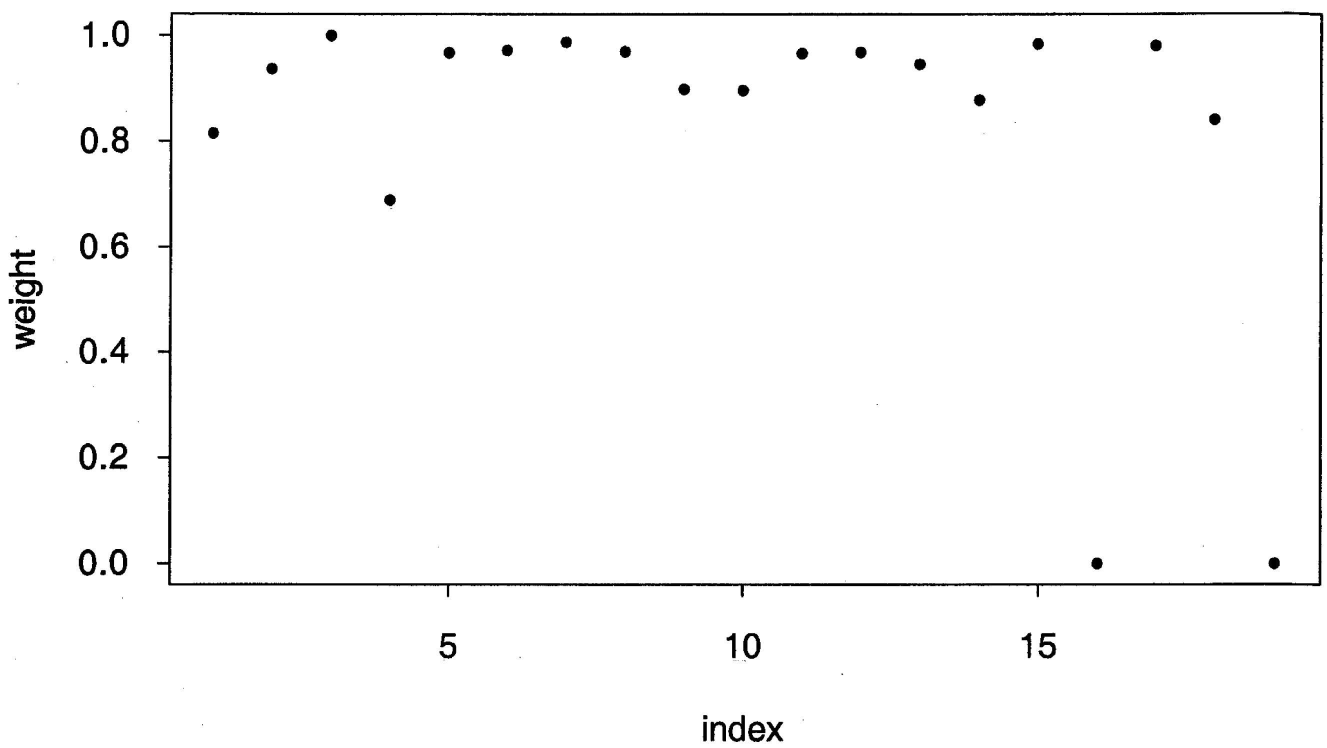
Radiocarbon age vs. uranium-thorium age



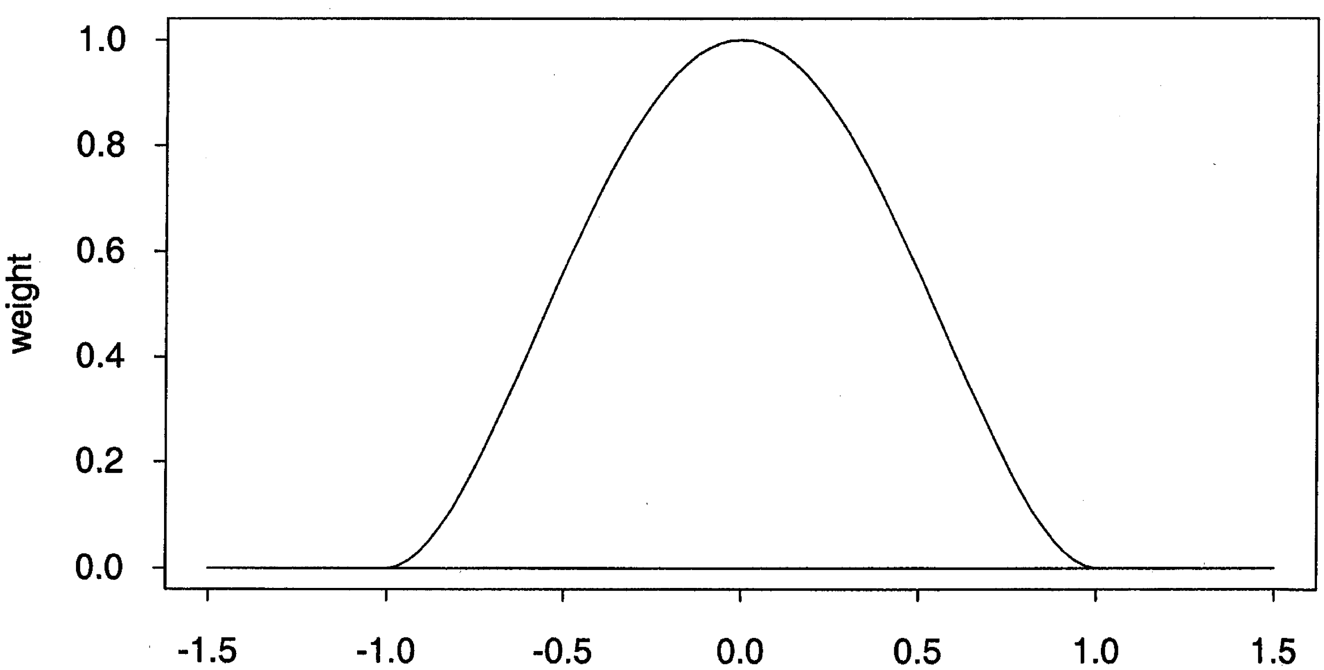
Radiocarbon age vs. uranium-thorium age



Weights plot



Bisquare weights



8 Oct 01

Summary of Robust / Resistant Methods

1. Can lead to identification of outliers
2. Can suggest where present model does not apply.
3. Leads to parameter estimates which are not sensitive to arbitrary changes in any small part of the data
   - breakdown point

Discuss anomalies with subject matter experts

8 Oct. 01

Functions in Splus.

location.m( ) bisquare, huber

l1fit( ) L_1

rreg( ) converged Huber followed by bisquare

rbiwt( ) bisquare, single x

ltsreg( ) trims half $|\hat{\varepsilon}_i|$

robust( ) glm(, family = robust)

gam( )