OLS – multiple predictors

Goal: developing a descriptive relationship

Data matrices: \(y\) is n by 1, \(X\) is n by p

Explain \(y\) via \(X\beta\), \(\beta\) is p by 1

Fit via

\[
\min_{\beta} \ (y - X\beta)' (y - X\beta)
\]

normal equations (may overparametrize)

\[
X'Xb = X'y
\]

Solution

\[
b = (X'X)^{-1}X'y
\]

Write \(y - X\beta = y - Xb + X(b - \beta)\)

(Generalized inverse – more later)

Estimate \(P\beta\) by \(Pb\)

Fitted values

\[
Xb = X(X'X)^{-1}X'y = Hy
\]

hat matrix \(H\), n by n
\[ HX = X, \ H^2 = H, \ r(H) = r(X) \]

residuals

\[ r = y - Xb = (I - H)y \]

\[ X' r = 0 \]

Evaluate univariate statistics, eg. stleaf

outliers?

**SS identity**

\[ y'y = (Xb)'Xb + r'r \]

degrees of freedom

\[ n = r(X) + (n - r(X)), \ r(X) \leq n, \ p \]

Advantages of orthogonality

\[ X = [X_1 \ X_2], \text{ with } X_1'X_2 = 0 \]

\[ b_1 = (X_1'X_1)^{-1}X_1'y \]

\[ y'y = (X_1b_1)'X_1b_1 + (X_1b_1)'X_1b_1 + r'r \]

ridge estimate

\[ (X'X + \lambda I)^{-1}X'y \]

lsfit(), lm(), anova()
Residual plots
index
versus (possible) predictors
versus fitted values

$H_{ii}$ leverage or influence of $y_i$ on fitted $y_i$

fitted $y_i = H_{ii} y_i + \sum_{j \neq i} H_{ij} y_j$

$0 \leq H_{ii} \leq 1$

leverage point $H_{ii} > 2r/n$

e.g. $r = 2$

$H_{ii} = 1/n + (x_i - \bar{x})^2 / \sum_j (x_j - \bar{x})^2$

Relates directly to how near $x_i$ is to $\bar{x}$

`lm.influence()$hat` [library(MASS)]

lurking variable – has an important effect, yet not included in predictors

Some $x$’s might be dummy variables or factors
SVD.

\[ A = U \Lambda V' \]

- **A** m by n,
- **U** orthogonal m by m,
- **\Lambda** diagonal m by n,
- **V** orthogonal n by n

**Generalized inverse**

\[ A^{-} = V \Lambda^{-} U' \]

\[ \Lambda^{-} = \text{diag}\{1/\lambda_j \mid \lambda_j \neq 0\} \]

\[ AA^{-}A = A \]

Solves consistent

\[ Ax = b \]

\[ X = A^{-}b + (I - A^{-}A)z \]

solve(), svd()

---

Statistics 215a – 9/17/03 – D. R. Brillinger

**Robust/resistant fitting of a straight line**

**resistant** – not strongly affected by outliers
Robust – remains effective with departures from assumptions

Often the two go together

Methods – $L_1$, $L_p$, three groups, bisquare, ...

Functions
   `l1fit(MASS)`, `rbiwt(MASS)`, `rreg(MASS)`,
   `rlm(MASS)`
   `line(eda)`, `hubers(MASS)`, `rlm(MASS)`

Researchers – Bolt (1960), Huber (1973),
   Beaton & Tukey (1974), Tukey (197?), ...

Three groups line. Tukey

Groups – low, middle, high

Summary points – medians
Slope and intercept
(Iterate)

Bisquare.

Iterative (IWLS)

Residuals, $r_i$

$u_i = r_i / 6 \text{ median } \{ |r_j| \}$

$w(u) = (1 - u^2)^2$, $|u| \leq 1$. Graph
Computation discussed below

Use weights to locate outliers (w = 0)

**Example** - radioactive dating


Barbados corals dating back 125k yrs (bp)

Two dating methods: uranium-thorium (U-Th) and radiocarbon, $^{14}$C

U-Th is more accurate

Fig. 1. $^{14}$C age vs. U-Th age

$y = x$ line

Calibrate $^{14}$C ages

Fig. 2. $^{14}$C age vs. U-Th age

OLS line

influence plot

Fig. 3. $^{14}$C age vs. U-Th age

OLS line

residual plot

Fig. 4. $^{14}$C age vs. U-Th age

bisquare line

residual plot
Fig. 5. Index plot of weights

Fit

RC = 1.106 + .763Th

Calibration equation

age = (RC - 1.106)/.763

(Estimates of dating errors available)

M-estimates.

loss function ρ

ρ(r) ≥ 0, nondecreasing for r ≥ 0
ρ(0) = 0

symmetric

cts for all but finite number of r

Robust regression

\[ \min_\beta \sum_i \rho((y_i - x_i^T\beta)/s), \quad s \text{ scale value} \]

E.g.

\[ \rho(r) = r^2 \text{ (OLS)} \]

\[ = |r|^p \ (L_p) \]

\[ = .5r^2 \text{ if } |r| \leq H \text{ and } = H|r| - .5H^2 \text{ otherwise (Huber)} \]

\[ = [1 - (1 - r/B)^2]^3B^2/6 \text{ if } |r| \leq B \text{ and } B^2/6 \text{ otherwise (bisquare, biweight)} \]

Evaluate by IRLS with \( \psi = \rho' \), \( w(r) = \psi(r)/r \)

Fig 6. \( \rho \) and \( w \)

**The IRLS approach**

Seeking

\[ \min_\beta \sum_i \rho((y_i - x_i^T\beta)/s), \quad s \text{ scale value} \]

Differentiate wrt \( \beta \) and set to 0

\[ \sum_i x_i^T \psi((y_i - x_i^Tb)/s) = 0 \]

A set of nonlinear equations

Solve for \( b \) iteratively

Need (good) set of starting values
Equations may be written

\[ \sum_i w((y_i - x_i^T b)/s) \cdot x_i^T(y_i - x_i^T b) = 0 \]

or

\[ \sum_i w_i \cdot x_i^T(y_i - x_i^T b) = 0 \]

with data dependent weights

\[ w_i = w((y_i - x_i^T b_\text{old})/s) \]

Use WLS until “convergence”

b_\text{old} comes from previous iteration

WLS

\[ \min_\beta \sum_i w_i \cdot x_i^T (y_i - x_i^T \beta)^2 \]

What can sequences do?
converge
diverge
have limit points
cycle
be chaotic

Mallows (1979). “A simple and useful strategy is to perform one’s analysis both robustly and by standard methods and to compare the results. If the differences are minor, either set can be presented. If the differences are not, one must perforce consider why not, and the robust analysis is already at hand to guide the next steps.”

Note: location problem corresponds to $x \equiv 1$