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

OLS – multiple predictors

Goal: developing a descriptive relationship

Data matrices: $\mathbf{y}$ is n by 1, $\mathbf{X}$ is n by p

Explain $\mathbf{y}$ via $\mathbf{X}\beta$, β is p by 1

Fit via

$$\min_\beta (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

normal equations (may overparametrize)

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

Solution

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$$

Write $\mathbf{y} - \mathbf{X}\beta = \mathbf{y} - \mathbf{X}\mathbf{b} + \mathbf{X}(\mathbf{b} - \beta)$

(Generalized inverse – more later)

Estimate $\mathbf{P}\beta$ by $\mathbf{P}\mathbf{b}$

Fitted values

$$\mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$$

hat matrix $\mathbf{H}$, n by n

$\mathbf{HX} = \mathbf{X}$, $\mathbf{H}^2 = \mathbf{H}$, $r(\mathbf{H}) = r(\mathbf{X})$

residuals

$\mathbf{r} = \mathbf{y} - \mathbf{Xb} = (\mathbf{I} - \mathbf{H})\mathbf{y}$

$\mathbf{X'r} = \mathbf{0}$

Evaluate univariate statistics, eg. stleaf

outliers?

SS identity

$\mathbf{y'y} = (\mathbf{Xb})'\mathbf{Xb} + \mathbf{r'r}$

degrees of freedom

$n = r(\mathbf{X}) + (n - r(\mathbf{X}))$, $r(\mathbf{X}) \le n, p$

Advantages of orthogonality

$\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$, with $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$

$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'\mathbf{y}$

$\mathbf{y'y} = (\mathbf{X}_1\mathbf{b}_1)'\mathbf{X}_1\mathbf{b}_1 + (\mathbf{X}_1\mathbf{b}_1)'\mathbf{X}_1\mathbf{b}_1 + \mathbf{r'r}$

ridge estimate

$(\mathbf{X'X} + \lambda\mathbf{I})^{-1}\mathbf{X'y}$

lsfit(), lm(), anova()

Residual plots

index

versus (possible) predictors

versus fitted values

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

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

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

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.

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}'$$

A m by n,
U orthogonal m by m,
Λ diagonal m by n,
V orthogonal n by n

Generalized inverse

$$\mathbf{A}^{-} = \mathbf{V}\Lambda^{-}\underline{\mathbf{U}'}$$

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

$AA^{-}A = A$

Solves consistent

$\mathbf{Ax} = \mathbf{b}$

$\mathbf{X} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z}$

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 median $\{|r_j|\}$

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

Use weights to locate outliers (w = 0)

Example – radioactive dating

Data - Bard et al. (1990). *Nature* 345, 405-410. n = 19

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

$\rho(r) \geq 0$, nondecreasing for $r \geq 0$
$\rho(0) = 0$
symmetric
cts for all but finite number of r

Robust regression

$\min_\beta \sum_i \rho((y_i - \mathbf{x}_i^T\beta)/s)$,   s scale value

E.g.

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

$= |r|^p$ (L_p)

$= .5r^2$ if $|r| \le H$ and $= H|r| - .5H^2$ otherwise (Huber)

$= [1 - (1 - r/B)^2]^3 B^2/6$ if $|r| \le B$ and $B^2/6$ otherwise (bisquare, biweight)

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

Fig 6. ρ and w

The IRLS approach

Seeking

$$\min_\beta \sum_i \rho((y_i - \mathbf{x}_i^T\beta)/s), \quad s \text{ scale value}$$

Differentiate wrt β and set to 0

$$\sum_i x_i^T \psi((y_i - x_i^T b)/s) = 0$$

A set of nonlinear equations

Solve for b iteratively

Need (good) set of starting values

Equations may be written

$$\sum_i w((y_i - x_i^T b)/s)\, x_i^T (y_i - x_i^T b) = 0$$

or

$$\sum_i w_i\, x_i^T (y_i - x_i^T b) = 0$$

with data dependent weights

$$w_i = w((y_i - x_i^T b_-)/s)$$

Use WLS until "convergence"

b_- comes from previous iteration

WLS

$$\min_\beta \sum_i w_i\, x_i^T\, (y_i - x_i^T \beta)^2$$

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 $x \equiv 1$