Statistics 215a – 9/24/03 – D. R. Brillinger

Residual analysis.

data = fit + residual

discrepancy between data and fit

fit might be OLS or robust/resistant or ?

patterns suggest fit can be improved, e.g. by transformations, and there are often surprises

Procedure: improvement of fit by stages

Types of residuals

"ordinary", $r_i = y_i - x_i^T b$

standardized, $r_i / s\sqrt{(1-h_{ii})}$

cross-validation, $y_i - x_i^T b_{-i}$

Uses of residual plots

improvement of fit

identification of outliers

behavior of techniques on the data to portray adequacy of fit

Types of residual plots

residuals, r_i, vs. fitted values, $x_i^T b$

residuals vs. explanatories, x_{ij}

residuals vs. functions of the x's, e.g. products

residuals vs. new variables, e.g. time

$|r_i|$ vs. $x_i^T b$

smoothed residuals vs. …

Some patterns

sloping band – include linear term

curved band – add quadratic or somesuch

wedging – variability increasing
(use weights in fitting)
- but, perhaps result of point density increasing (add lowess lines)

- wedging may be one-sided if |r| used

Plot of response vs. fit can be missleading

Partial residual plots

Plot partial residual of j-th explanatory vs. its values

i.e. $r_i^j = y_i - \sum_{k \neq j} b_k x_{ik}$ versus x_{ij}

have “removed” the linear effects of the other x’s

the slope is b_j

the residuals are the r_i

transform the j-th explanatory?

Fitting by stages/one variable at a time

Two explanatories case

(x_1, x_2, y)

1) fit y by x_1

$y_{.1} = y - c_1 x_1$

2) fit x_2 by x_1

$x_{2.1} = x_2 - d_1x_1$

3) fit $y_{.1}$ by $x_{2.1}$, including intercept

Because of the orthogonalities this gives the one step result

$b_0 + b_1x_1 + b_2x_2$

now written as

$b_0 + c_1x_1 + c_2(x_2 - d_1x_1)$

One can graph (x_1,y), (x_1,x_2), $(x_{2.1},y_{.1})$ along the way as well as residuals

Robust/resistant fitting could be used

R. D. Cook and S. Weisberg (1982). *Residuals and Influence in Regression*. Chapman and Hall.

Statistics 215a – 9/24/03 – D. R. Brillinger

The x-values.

1. location parameter, $p = 1$, $x \equiv 1$

2. factor

object taking on values from a set of levels

often created via x-variables taking on the values 0, 1

e.g. for q levels set

x^1 = 1 for level 1 and 0 otherwise

x^2 = 1 for level 2 and 0 otherwise

.

x^{q-1} = 1 for level q-1 and 0 otherwise

x^q = 0

provides a coding

The X-matrix is made up of 0’s and 1’s

factor(), contrasts()

Example – the RBC data

```
A: day

B: run

lm(y ~ -1 +A +B)
```