$$y = X\beta + \varepsilon \qquad \varepsilon \sim (0, \sigma^2 I) \qquad (*)$$

$Q_{n\times n}$ orthogonal $\quad Q'Q = I$

$$Qy = QX\beta + Q\varepsilon \qquad y^* = X^*\beta + \varepsilon^*$$

$$\varepsilon^* \sim (0, \sigma^2 I)$$

[if $\varepsilon \sim N(\ )$ then $y \sim N$ then $\hat\beta \sim N(\beta, \sigma^2 (X^T X)^{-1})$, used for p-values, confidence limits]

LS estimates $\hat\beta$ are unbiased ($E\hat\beta = \beta$) under (*)

best linear unbiased estimates of β

smallest variance

in y

$$\hat\beta = \underline{(X^T X)^{-1} X^T}\, \underline{\underline{y}} \qquad \text{(linear)}$$

Q chosen so $\quad X^* = \begin{pmatrix} X_1^* \\ 0 \end{pmatrix} \quad X_1^*$ upper triang.

(backsolve for $\hat\beta$)

$$X^* = QX \quad \Rightarrow \quad Q'X^* = X \qquad \text{bec. } Q' = Q^{-1}$$

called "QR" decomposition in literature

in R, lm uses ~~qr~~ Q·R decomp. of X to get $\hat\beta$

$$y^* = \begin{pmatrix} X_1^* \\ 0 \end{pmatrix} \beta + \varepsilon^*$$

$$= \underbrace{\begin{pmatrix} X_1^* \\ 0 \end{pmatrix}}_{n \times p} \begin{matrix}]_{p \times p} \\ ]_{n-p \times p} \end{matrix} \beta + \varepsilon^*$$

$$= \begin{pmatrix} X_1^* \\ 0 \end{pmatrix} V V' \beta + \varepsilon^* \qquad V_{p \times p} \perp \text{ matrix}$$

V also chosen so $X_1^* V =$ diagonal

$$= \begin{pmatrix} D \\ 0 \end{pmatrix} V' \beta + \varepsilon^* = \begin{pmatrix} D \\ 0 \end{pmatrix} \theta + \varepsilon^*$$

- completely diagonal system

$$D_{p \times p} = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{bmatrix}$$

Ass^n $n > p$; X has rank p ; $\Rightarrow X^T X$ invertible

(Still works if replace p by column rank of X)

$$\hat{\theta}_j = y_j^* / d_j \qquad \text{sol'n of LS eq'n is trivial}$$

$$\text{s.e}(\hat{\theta}_j) = \sigma / d_j$$

$$j = 1, \dots, p$$

$\cancel{X_2}\ X = X_1^* D V'$

$$QX = X^* \Rightarrow X = Q'X^*$$

$$X^* = DV' \quad \text{so} \quad X = \underline{Q'} D V'$$

$$\equiv \underline{\underline{U}} D V'$$

another change!

singular value decomposition of X

Any $n \times p$ matrix X can be (p.62 VR)

expressed as $\underset{n\times p}{X} = \underset{n\times p}{U}\ \underset{p\times p}{D}\ \underset{p\times p}{V'}$

where $U'U = I$, $V'V = I$

$D = \text{diag}(d_1, \ldots, d_p)$ $\quad$ $\cancel{\text{e.g.}}$ ① $d_i \geq 0$ ②

(the zeros at the bottom of X^* dropped)

d_i^2 is ith eigenvalue of $X^T X$

$\# \, d_i > 0 \quad = \text{rank of } X^T X$

or $\text{col}^{\underline{n}}$ rank of X

$n > p$

$\# \, \text{obs}^{\underline{n}}$ $\quad$ # predictors

$qr(X) = Q_1 X_1^*$

Diagnostics from regression §6.3

- plot residuals $y_i - \hat{y}_i$ vs. $\hat{y}_i$

(to check const. variance ass$^{\underline{n}}$)

- plot residuals on normal probability scale

```
>qqnorm(lm$residuals)
```

(to check if N ass$^{\underline{n}}$ ok)

- looking for <u>outliers</u> (model doesn't fit well) 2 in hills.

$\hat{\varepsilon}_i$ not independent (ε_i are)

$$\underset{n\times 1}{\hat{\varepsilon}} = y - \hat{y} = y - Hy = \cancel{y(\ )}\ \underbrace{(I-H)}_{tr\, =\, n-p} y$$

- not constant variance

$$\operatorname{var} \hat{\varepsilon}_i = \sigma^2 (1 - h_{ii})$$

1) ~~studentized~~ standardized residuals $\dfrac{\hat{\varepsilon}_i}{s\sqrt{1-h_{ii}}}$

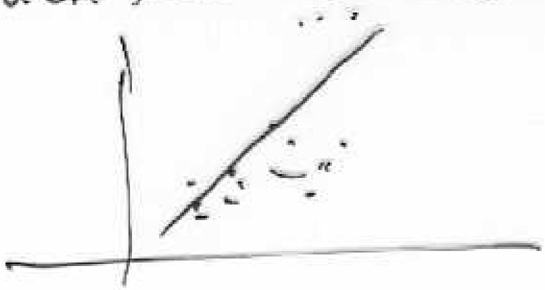
$$s^2 = \hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-p}$$ i const. var.

~~studres~~ > stdres

2) studentized residuals $\dfrac{\hat{\varepsilon}_i}{s_{(i)}\sqrt{1-h_{ii}}}$

$s_{(i)}^2$ is est. of σ^2 <u>without</u> y_i

- influential values (not always outliers)



one or 2 or 3 pts. that are really "different"

- if h_{ii} = ith diag. element of

$$H = X(X^TX)^{-1}X^T$$ is 'large'

then ith point 'influential'

$$\cancel{\hat{y}_i = H} \qquad \hat{y} = H y$$

$$E(\hat{\varepsilon}\hat{\varepsilon}^T) = E(y-\hat{y})(y-\hat{y})^T$$
$$= \cdots = \sigma^2(I-H)$$

Since $tr(H) = p$ (exercise)

average value of $h_{ii} \simeq \frac{p}{n}$ $\boxed{p. 152}$

'large' is usually $> \frac{3p}{n}$ $\left(\frac{2p}{n}\right)$ ...

Cook's distance combines outlier/influ.
into one measure

$$C_i = \frac{(\hat{\varepsilon}_i^{STD})^2 \cdot h_{ii}}{p(1-h_{ii})}$$

$$\hat{\varepsilon}_i^{STD} = \frac{\hat{\varepsilon}_i}{s\sqrt{1-h_{ii}}} \leftarrow$$

very rough guideline $C_i > \frac{8}{n-2p}$

"large"

$h > \frac{2p}{n}$ or $|\varepsilon| > 2$

>plot.lm automatically highlights 3 biggest ones

Use a different fitting method (robust)

§6.5 – doesn't get influenced by outliers

§5.5 – doesn't depend on normality