STA 410 L.#5 Jan 18/05

Computation of LS estimates

$$Y = X\beta + \varepsilon \qquad \underset{n\times 1}{\varepsilon} \sim (0, \sigma^2 I)$$

$$\hat\beta = (X'X)^{-1}X'y \iff (X'X)\hat\beta = X'y$$

: various methods of computation include

1. compute $(X'X)^{-1}$, multiply by $X'y$
2. solve $(X'X)\hat\beta = X'y$ $\quad (Az = b)$ Cholesky
   $X'X = U'U$
3. use $\perp$ methods Givens / Householder
4. singular value decomposition

1. The Cholesky decomp of a symmetric p.d. matrix is an upper triang. matrix U s.t. $U'U = A$

$$Az = b \Rightarrow U'Uz = b$$
$$U'w = b$$

U' is L.T. $\quad U$ is U.T. $\begin{bmatrix} - & & \cdot \\ & - & \vdots \\ 0 & & - \end{bmatrix}$

$U'w = b$

$$\begin{pmatrix} u_{11} & 0 & \cdots & 0 \\ u_{12} & u_{22} & \cdots & 0 \\ \vdots & & & \\ u_{1k} & u_{2k} & \cdots & u_{kk} \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}$$

$u_{11} w_1 = b_1$ ~~[illegible]~~ $w_1 = b_1 / u_{11}$

$u_{12} w_1 + u_{22} w_2 = b_2$ $\qquad w_2 = \dfrac{b_2 - u_{12} w_1}{u_{22}}$

$\vdots$

$w_i = b_i - \sum_{j=1}^{i-1} w_j u_{ij} / u_{ii}$ forward substitution.

$Uz = w$ $\quad \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ 0 & u_{22} & \cdots & u_{2k} \\ & & \ddots & \\ 0 & \cdots & 0 & u_{kk} \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix} = \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix}$

$z_k = w_k / u_{kk} \qquad u_{k-1,k-1} z_{k-1} + u_{kk} z_k = w_{k-1}$

$z_{k-1} = w_{k-1} - u_{kk} z_k / u_{k-1,k-1}$ ... etc. back substitution

- Finding U : see next pg

Finding the Cholesky decomposition

$$\begin{pmatrix} u_{11} & 0 & \cdots & 0 \\ u_{12} & u_{22} & \cdots & 0 \\ \vdots & & & \\ u_{1k} & u_{2k} & \cdots & u_{kk} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ 0 & u_{22} & \cdots & u_{2k} \\ & & \ddots & \\ 0 & & & u_{kk} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}$$

1st c.

$$\left[\begin{array}{ll} u_{11}^2 = a_{11} & u_{11} = \sqrt{a_{11}} \\ u_{11}u_{12} = a_{12} & u_{12} = a_{12}/u_{11} \\ \vdots & \\ u_{11}u_{1k} = a_{1k} & u_{1k} = a_{1k}/u_{11} \end{array}\right.$$

$$u_{12}^2 + u_{22} = a_{22} \qquad u_{22} = \sqrt{a_{22} - u_{12}^2}$$

$$u_{12}u_{13} + u_{22}u_{23} = a_{23} \qquad u_{23} = (a_{23} - u_{12}u_{13})/u_{22}$$

$$\vdots$$

for i from 1 to k do

$$u_{ii} \leftarrow \sqrt{a_{ii} - \sum_{k=1}^{i-1} u_{ki}^2}$$

for j from $i+1$ to k do

$$u_{ij} \leftarrow \left(a_{ij} - \sum_{k=1}^{i-1} u_{ki}u_{kj}\right)/u_{ii}$$

end

$$u_{11} = \sqrt{1.00001} = \sqrt{1 + 10^{-6}}$$

$$u_{12} = 1/u_{11}$$

$$u_{22} = \sqrt{1 + 10^{-6} - u_{12}^2}$$

$$u_{11} = (1 + 10^{-6})^{1/2} \sim 1 + \tfrac{1}{2} \times 10^{-6}$$

$$u_{12} \sim 1 - \tfrac{1}{2} \times 10^{-6}$$

$$u_{22} = \sqrt{1 + 10^{-6} - (1 - \tfrac{1}{2} \times 10^{-6})^2}$$

$$= (1 + 10^{-6} - (1 - 10^{-6}))^{\frac{1}{2}}$$

$$= (2 \times 10^{-6})^{1/2} = \sqrt{2} \times 10^{-3}$$

$$U = \begin{pmatrix} 1 + .5\times10^{-6} & 1 - .5\times10^{-6} \\ 0 & \sqrt{2} \times 10^{-3} \end{pmatrix}$$

3. Sol'n using orthogonal transformation

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

Suppose $Q_{n\times n}$ is orthog i.e. $Q'Q = I$, $Q' = Q^{-1}$, $QQ' = I$ $\left(|Qx| = |x| \ \forall x \in \mathbb{R}^n \right)$

- $Qy = QX\beta + Q\varepsilon$

  $y^* = X^*\beta + \varepsilon^*$ say $\qquad E\varepsilon^* = 0 \quad \operatorname{cov} \varepsilon^* = \sigma^2 I$ as before

- choose Q so that

$$X^* = \begin{bmatrix} x_{11}^* & \cdots & & x_{1p}^* \\ 0 & x_{22}^* & \cdots & x_{2p}^* \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots\ 0 & x_{pp}^* \\ 0 & & & 0 \\ \vdots & & & \vdots \\ 0 & - & - & 0 \end{bmatrix}_{n\times p} \qquad \begin{bmatrix} X_1^* {}_{p\times p} \\ 0_{n-p\times p} \end{bmatrix}$$

- LS est min $(y - X\beta)^T(y - X\beta) = \varepsilon^T\varepsilon = \underline{\varepsilon}^{*T}\underline{\varepsilon}^* = \sum_{i=1}^n \varepsilon_i^{*2}$

  choose $\varepsilon_{p+1}^* \cdots \varepsilon_{p+n}^* = 0 \qquad \hat\beta = (X_1^*)^{-1} y_1^* \qquad y = \begin{pmatrix} y_1^* \\ y_2^* \end{pmatrix}$

- but more directly $\hat\beta_p = y_p^* / x_{pp}^*$

$$\hat\beta_{p-1} = (y_{p-1}^* - x_{p-1,p}^*\hat\beta_p) / x_{p-1,p-1}^*$$

$$\vdots$$

back substitution again

- We have $\underset{n\times n\ \ n\times p}{X = Q'X^*} = \underset{n\times p\ \ p\times p}{Q_1'X_1^*} = QR$ in more usual not^n — qr decomposition

18/5

- How to find the matrix Q? Built up from specially chosen $\perp$ matrices $Q = Q_1 Q_2 \cdots Q_p$

- the component pieces are called <u>Householder</u> transformations

- ~~the~~ product $QR = Q_1' X_1^*$ is returned with the lm object eg ~~hills~~ hills.lm$qr

- note that $(X_1^{*'} X_1^*) = \cancel{(QX)'QX = X'X}$ $(X^{*'}X^*) = (QX)'QX$
$= X'X$

  and X_1^* is upper triangular
  $\therefore$ is also $= U$ of the Cholesky decomposition

- so Householder transf. give a different way to construct Cholesky decomp.
  seems to be more stable & accurate; works directly on X
  instead of first computing X^TX

- in matrix not^n $\hat\beta = X_1^{*-1} y_1^* = X_1^{*-1} Q_1' y$

- note that the hat matrix H $(Hy = \hat y \quad X\hat\beta y = \hat y$
  $H = X\hat\beta = X(X^TX)^{-1}X^T$

$X = Q_1'R \qquad H = Q_1'R(R^TQ_1Q_1^TR)^{-1}R^TQ_1$

$= Q_1'R(R^TR)^{-1}R^TQ_1 \qquad = Q_1'Q_1$

and can also show that sequential SS can be obtained
from this QR decomposition

- and easily extended to forward stepwise regression i.e. easy to update the model

```
> hills.lm1 <- lm(time ~ dist $, data = hills)
> hills.lm <- update(hills.lm1, . ~ . + climb)
```

Finally ~~QR~~ singular value decomposition

recall $y = X\beta + \varepsilon$

$$Qy = QX\beta + \underline{Q\varepsilon}$$
$$y^* = X^*\beta + \varepsilon^*$$
$$= \begin{pmatrix} X_1^*\beta \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_1^* \\ \varepsilon_2^* \end{pmatrix}$$

X_1^* upper triang

Q orthog.

- if we diagonalize X_1^* then sol^n will be v. easy

$$y^* = \begin{pmatrix} X_1^* \\ \underset{\sim}{0} \end{pmatrix}\beta + \varepsilon^*$$
$$= \begin{pmatrix} X_1^* \\ \underset{\sim}{0} \end{pmatrix} VV'\beta + \varepsilon^*$$
$$= \begin{pmatrix} D \\ \underset{\sim}{0} \end{pmatrix}\theta + \varepsilon^*$$

$D = \text{diag}(d_1, \ldots, d_p)$

V is $p \times p$ $\perp$

- now switch not^n! $\underset{n \times n}{Q} \to \underset{n \times n}{U'}$

$$U'y = \underbrace{(U'XV)}_{= \begin{pmatrix} D \\ 0 \end{pmatrix}}\theta + \varepsilon^*$$

$\hat{\theta}_i = y_i^* / d_i \quad SE(\hat{\theta}_i) = \sigma / d_i$

18/7

Diagnostics

- plot residuals $y_i - \hat{y}_i$ vs. x's, vs $\hat{y}_i$ & qqnorm( )

  $\uparrow$ large values ⇒ outliers

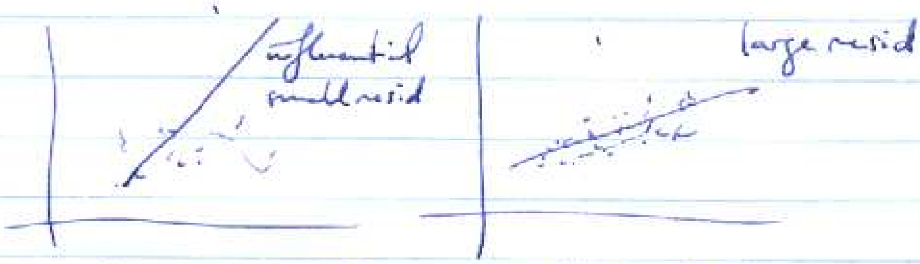
- note that $E(\hat{\varepsilon}\hat{\varepsilon}^T) = E(y - X\hat{\beta})(y - X\hat{\beta})^T$

$$= E(y - X(X^TX)^{-1}X^Ty)(y - Hy)^T$$

$$= E[(I-H)y][(I-H)y]^T$$

$$= \sigma^2(I-H) \quad \text{; note not iid}$$

- if h_{ii} is 'large' then y_i is said to be 'influential'; pulls fit towards itself



- if $h_{ii} \uparrow$, $\text{var}(\hat{\varepsilon}_i) \downarrow$ $\quad \hat{\varepsilon}_i / s\sqrt{1-h_{ii}}$ are called std'd resids (all have var 1) stdres

$$\varepsilon_i^* = \frac{y_i - \hat{y}_{(i)}}{\sqrt{\text{var}(y_i - \hat{y}_{(i)})}}$$ call studentized resid, $s \leftarrow s_{(i)}$, studres

- since $tr(H) = p$, $h_{ii} > 2$ or $3 \times \frac{p}{n}$ is 'large'

for Hills data $p = 3$ $n = 35$ ~~$p/n \approx 1/12$~~ see p 152

Bens of Jura has high leverage + high residual See p 153 top
Lairig Ghru has high leverage but " "
Knock Hill has high residual but low leverage

- Cook's distance is an attempt to summarize both in one plot

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

(C_i large if stand. resid large or h_{ii} large or both)

rough guide $C_i > 8/(n-2p)$ $\quad (h > \frac{2p}{n}, |\varepsilon^{STD}| > 2)$

$p = \text{rank}(X^T X) = 3$

$8/35-6 = 8/29 \simeq 0.28$

- plot.lm automatically labels the 3 largest values