STA 302 / 1001

Note Title 11/3/2011

Samuel Beatty In-Course Scholarships
- see announcements on Blackboard

Reminders
- Mon-Tues is Fall Break.
- my office hours on Mon Nov. 7 are cancelled
- no lecture on Tues Nov. 8
- Andriy in RW 107 & 109 to help with A2:
  Tues Nov 8 10-12
  Wed Nov 9 10-12

$$\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 \begin{pmatrix} \sum X_i^2 / nS_{xx} & -\frac{\bar{X}}{S_{xx}} \\ -\frac{\bar{X}}{S_{xx}} & \frac{1}{S_{xx}} \end{pmatrix}_{2\times 2} = \sigma^2 (X'X)^{-1}$$

$\hat{\beta}$: vector of length 2

$$\mathrm{Var}(\hat{\beta}_0 \mid X) = \sigma^2 \frac{\sum X_i^2}{nS_{xx}}$$

$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1 \mid X) = -\sigma^2 \frac{\bar{X}}{S_{xx}}$$

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \sigma^2 / S_{xx}$$

Fitted Values and Residuals

$$\hat{\underset{\sim}{e}} = \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \vdots \\ \hat{e}_n \end{pmatrix} = Y - \hat{Y} = Y - Xb, \quad b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$b = (X'X)^{-1}X'Y$$

$$\hat{Y} = Xb = X(X'X)^{-1}X'Y$$

$$= HY$$

where $H = X(X'X)^{-1}X'$ is the "hat" matrix

h_{ii}, h_{ij} are the elements of H

H is <u>idempotent</u> (See website for facts about idempotent matrices.)

a matrix A is idempotent if $A^2 = A$

i.e. $H^2 = HH = X(X'X)^{-1}\underbrace{X'X}\underbrace{(X'X)^{-1}}X' = X(X'X)^{-1}X' = H$

H is <u>symmetric</u>

$$H' = \Big(X\underbrace{(X'X)^{-1}}_{\text{symmetric}}X'\Big)' = X(X'X)^{-1}X' = H$$

$$\hat{e} = Y - \hat{Y} = Y - HY = (I-H)Y$$

$I-H$ idempotent? YES

$$(I-H)(I-H) = I - H - H + H^2 = I - H - H + H = I - H$$

$I-H$ symmetric YES

$$E(\hat{e}|X) = E\left[(I-H)Y \,\middle|\, X\right]$$

Model: $Y = X\beta + e$

$$= (I-H)\,E(Y|X)$$

$$= (I-H)\,X\beta$$

$$= X\beta - \left(X(X'X)^{-1}X'\right)X\beta$$

$$= X\beta - X\beta$$
$$= \underset{\sim}{0}$$

$$\text{Var}(\hat{e}|X) = E\left[(I-H) Y Y' (I-H) \middle| X \right]$$

since $(I-H)$ is symmetric

$$= E\left[(\hat{e} - E(\hat{e}))(\hat{e} - E(\hat{e}))' \right]$$

definition of Var matrix

$$= (I-H)\, E(YY')\, (I-H)$$

$$E(\underset{(n\times 1)}{Y}\underset{(1\times n)}{Y'}) = E\left((X\beta+e)(X\beta+e)' \mid X\right)$$

$$= E\left((X\beta+e)(\beta'X'+e') \mid X\right)$$

$$= E\left(\underset{(n\times 2)(2\times 1)(1\times 2)(2\times n)}{X\beta\beta'X'} + \underbrace{X\beta e'}_{n\times 2\ \ 2\times 1\ \ 1\times n} + \underset{n\times 1\ \ 1\times 2\ \ 2\times n}{e\beta'X'} + \underset{n\times n}{ee'} \mid X\right)$$

$$= E(X\beta\beta'X' \mid X) + \underbrace{E(ee' \mid X)}_{Var(e)\ \text{since}\ E(e)=0} \qquad E(e)=\underset{\sim}{0}$$

$$E(YY') = X\beta\beta' X' + \sigma^2 I$$

$$Var(\hat{e} \mid X) = (I-H)\left(X\beta\beta' X' + \sigma^2 I\right)(I-H)$$

$$= \left(X\beta\beta' X' + \sigma^2 I - X\underbrace{(X'X)^{-1}}\underbrace{X' X}\beta\beta' X' - \sigma^2 H\right)(I-H)$$

$$= \sigma^2 (I-H)(I-H)$$

$$\boxed{Var(\hat{e} \mid X) = \sigma^2 (I-H)}$$

For $A = A_1 + A_2$ and A, A_1, A_2 are all idempotent then $\text{rank}(A) = \text{rank}(A_1) + \text{rank}(A_2)$

$I - H$, I, H are all idempotent

$\text{rank}(I) = n$

$\text{rank}(H) = 2$

$\text{rank}(I-H) = n-2$

$= df$ for error

$\text{rank}(AB) \le \min\{\text{rank}(A), \text{rank}(B)\}$

$H = X(X'X)^{-1}X'$

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$\text{rank}(X) = 2$

Analysis of Variance in Matrix Terms

$$SST = SSReg + RSS$$

$$SST = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$$

$$= Y'Y - \frac{1}{n} Y'JY$$

where J is $n \times n$ matrix of 1's

Exercise: show $\frac{1}{n} Y'JY = n\bar{y}^2$

$$SST = Y'\left(I - \frac{1}{n}J\right)Y$$

$I - \frac{1}{n} J$ is symmetric

For a vector Y, A symmetric, $Y'AY$ is a quadratic form

I idempotent, $\frac{1}{n} J$ is idempotent,

$I - \frac{1}{n} J$ is idempotent

$$\text{rank}\left(I - \tfrac{1}{n} J\right) = \text{rank}(I) - \text{rank}\left(\tfrac{1}{n} J\right)$$

$$= n - 1$$

$$= \text{df for SST}$$

Take SST and decompose it as SSReg + RSS

$$SST = Y'Y - \frac{1}{n} Y'JY$$

$$Y'Y = (\underline{Y - Xb} + Xb)'(\underline{Y - Xb} + Xb)$$

$$= (Y - Xb)'(Y - Xb) + (Y - Xb)'Xb + (Xb)'(Y - Xb) + (Xb)'(Xb)$$

$$= \hat{e}'\hat{e} + \underbrace{\hat{e}'Xb + (Xb)'\hat{e}}_{\text{equal, because scalar and one is transpose of other}} + b'X'Xb$$

and $= 0$

since $X'\hat{e} = X'(I-H)Y$

$$= X'Y - \underbrace{X'X}\,\underbrace{(X'X)}^{-1}X'Y$$

$$= \underset{\sim}{0}$$

$$Y'Y = \hat{e}'\hat{e} + b'X'Xb$$

$$SST = Y'Y - Y'\tfrac{1}{n}JY = \underbrace{\hat{e}'\hat{e}}_{RSS} + \underbrace{b'X'Xb - Y'\tfrac{1}{n}JY}_{SSReg}$$

$$RSS = \hat{e}'\hat{e} \qquad , \quad \hat{e} = (I-H)Y$$

$$= Y'(I-H)'(I-H)Y$$

$$= Y'(I-H)Y$$

$I-H$ symmetric, idempotent

I, H idempotent

Another quadratic form in Y

$$\text{rank}(I-H) = \text{rank}(I) - \text{rank}(H)$$

$$= n-2$$

$= df$ for error.

$$SSReg = b'X'Xb - Y'\tfrac{1}{n}JY \qquad , \quad b = (X'X)^{-1}X'Y$$

$$= Y'X(X'X)^{-1}\underbrace{X'X}\underbrace{(X'X)^{-1}}X'Y - Y'\frac{1}{n}JY$$

$$= Y'\underbrace{X(X'X)^{-1}X'}_{H}Y - Y'\frac{1}{n}JY$$

$$= Y'\left(H - \frac{1}{n}J\right)Y$$

Another quadratic form in Y since $H - \frac{1}{n}J$ is symmetric

H, $\frac{1}{n}J$, $H - \frac{1}{n}J$ are all idempotent

$$\text{rank}\left(H - \frac{1}{n}J\right) = \text{rank}(H) - \text{rank}\left(\frac{1}{n}J\right) = 2 - 1 = 1$$

$= df$ for SS_{Reg}

Want to show $s^2 = \frac{RSS}{n-2}$ is an unbiased estimator for σ^2

Will show: $E(RSS) = E\left(\sum \hat{e}_i^2\right) = (n-2)\sigma^2$

I'll assume X is not random

$E(RSS) = E(\hat{e}'\hat{e})$ $\qquad \hat{e} = (I-H)Y$

$= E(Y'(I-H)Y)$ $\qquad$ since $I-H$ is symmetric & idempotent

$$= E\left(\operatorname{trace}\left(Y'(I-H)Y\right)\right)$$ since $Y'(I-H)Y$ is a scalar

$$= E\left(\operatorname{trace}\left((I-H)YY'\right)\right)$$ $\operatorname{trace}(AB) = \operatorname{trace}(BA)$

$$= \operatorname{trace}\left[(I-H)\,E(YY')\right]$$

Now $E(YY') = X\beta\beta'X' + \sigma^2 I$ (done earlier today).

so $$E(RSS) = \operatorname{trace}\left[(I-H)\left(\sigma^2 I + X\beta\beta'X'\right)\right]$$

$$= \text{trace}\left[(I-H)\sigma^2 + X\beta\beta' X' - X(\underbrace{X'X})^{-1}\underbrace{X'X}\beta\beta' X' \right]$$

$$= \text{trace}(I-H)\,\sigma^2$$

$$\text{trace}(I-H) = \text{trace}(I) - \text{trace}(H)$$

since $\text{trace}(A+B) = \text{trace}(A) + \text{trace}(B)$

$$\text{trace}(I) = n$$

$$\text{trace}(H) = \sum h_{ii} = 2$$

$$= \mathrm{trace}\left(X(X'X)^{-1}X' \right)$$

$$= \mathrm{trace}\left(X'X\,(X'X)^{-1} \right)$$

$(\mathrm{trace}(AB) = \mathrm{trace}(BA))$

$$= \mathrm{trace}\left(I_{2\times 2} \right) = 2$$

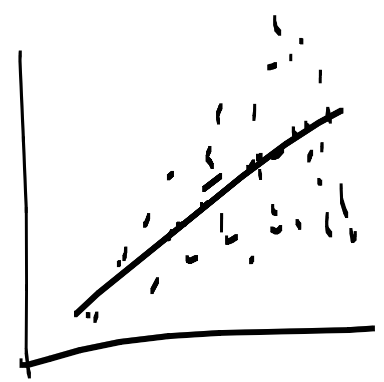
$$E(RSS) = (n-2)\,\sigma^2$$

$\rightarrow$ end of SLR in Matrix Form $\leftarrow$

Chapter 4 : Weighted Least Squares
(A very brief introduction)

Why do weighted least squares? heteroskedasticity
i.e. non-constant variance

When to use it?



- linear relationship
- distribution of residuals is not skewed
- But non-constant variance

The Idea: down weight points which have more variability

Find least squares estimates by minimizing:

Weighted residual sum of squares $= WRSS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2$

where $w_i \propto \frac{1}{\sigma_i^2}$ or $Var(e_i) = \sigma_i^2 \propto \frac{1}{w_i}$

"weights"

where σ_i^2 = error variance for i^{th} observation

How to <u>find weights</u>:

1. If have multiple observations for each value of x, can estimate σ_i^2

2. Common choices in SLR, $\underbrace{w_i = \frac{1}{x_i} \quad \text{or} \quad w_i = \frac{1}{x_i^2}}_{\text{variance increasing proportional to the mean or mean}^2}$

3. Often weights are chosen as $\hat{y}_i$ in an iterative fitting process.

4. Often get unequal variance if y_i is average (or median)

of n_i observations. Then $\text{Var}(Y_i) \propto \frac{1}{n_i}$ so use $w_i = n_i$

Minimizing WRSS to weighted least squares estimates

$$b_{1w} = \frac{\sum w_i(x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum w_i(x_i - \bar{x}_w)^2}$$

and $b_{0w} = \bar{y}_w - b_{1w}\bar{x}_w$

where $\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}$ and $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$

(weighted averages)

In matrix terms, if $Var(e) = \sigma^2 V$

where V is diagonal matrix with i^{th} entry $= 1/w_i$

$$\begin{pmatrix} b_{0w} \\ b_{1w} \end{pmatrix} = b_w = (X'V^{-1}X)^{-1} X'V^{-1}y$$

and $Var(\hat{\beta}_w) = \sigma^2 (X'V^{-1}X)^{-1}$