Wednesday, September 12, 2012
12:38 PM

LAST WEEK (WEEK 4), WHAT WE HAVE DONE ARE

- GET $\hat{\beta}_1$ AND $\hat{\beta}_2$ BOTH IN THE LEVEL FORM AND IN THE DEVIATION FORM
- CHECK THAT $\hat{\beta}_1$ AND $\hat{\beta}_2$ ARE UNBIASED, i.e.,

  $E(\hat{\beta}_1) = \beta_1$ AND $E(\hat{\beta}_2) = \beta_2$
- OBTAIN $\text{var}\ \hat{\beta}_2 = \dfrac{\sigma_u^2}{\sum_{i=1}^{n} x_i^2}$

KEEP IN MIND THAT σ_u^2 IS TRUE VARIANCE OF THE DISTURBANCE TERM BUT UNKNOWN

SO ? $\Rightarrow$ WE HAVE TO FIND AN ESTIMATOR FOR σ_u^2.

TODAY WE START w/ "$\text{var}(\hat{\beta}_1)$".

$$\text{var}(\hat{\beta}_1) = E\left[\hat{\beta}_1 - E(\hat{\beta}_1)\right]^2 \Rightarrow$$

POSSIBLE DEVIATION OF $\hat{\beta}_1$ FROM ITS AVERAGE VALUE

$$= E\left[\hat{\beta}_1 - \beta_1\right]^2$$

$\hat{t} = \dfrac{\bigcirc}{\bigcirc}$

s.e.

$\text{s.e.} = \sqrt{\text{var}}$

FROM $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \Leftrightarrow \hat{\beta}_1 = \sum_{i=1}^{n} \left[\frac{1}{n} - \bar{X} k_i\right] Y_i$

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X} k_i\right) u_i$$

$$\hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X} k_i\right) u_i .$$

SO, $\text{var}\ \hat{\beta}_1 = E(\hat{\beta}_1 - \beta_1)^2$

$$= E\left[\sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X} k_i\right) u_i\right]^2 .$$

$$\boxed{\text{var}\ \hat{\beta}_1 = \frac{\sigma_u^2 \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2}}$$

TRUE BUT UNKNOWN, SO WE HAVE TO ESTIMATE IT.

PURPOSE ?

GET $\text{var}\ \hat{\beta}_1 \rightarrow$ s.e. of $\hat{\beta}_1$

$\downarrow$

COMPUTE $\hat{t} = \dfrac{\hat{\beta}_1}{\text{s.e. of } \hat{\beta}_1}$

$\downarrow$

DO HYPOTHESIS ...

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = E\left[\hat{\beta}_1 - E(\hat{\beta}_1)\right]\left[\hat{\beta}_2 - E(\hat{\beta}_2)\right]$$

$\hookrightarrow$ MEASURES INTERRELATIONSHIP BETWEEN $\hat{\beta}_1$ AND $\hat{\beta}_2$

↳ MEASURES INTERRELATIONSHIP BETWEEN $\hat{\beta}_1$ AND $\hat{\beta}_2$

$$= E[\hat{\beta}_1 - \beta_1][\hat{\beta}_2 - \beta_2]$$

$$= -\bar{X} \operatorname{var} \hat{\beta}_2 \,.$$

THEREFORE,

$$\boxed{\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X} \frac{\sigma_u^2}{\sum_{i=1}^{n} x_i^2}.}$$

s.e. of β_1

DO HYPOTHESES TESTING

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

(FOR EXAMPLE)

SUMMARY : MODEL $\quad Y_i = \beta_1 + \beta_2 X + u_i$

VIA A SET OF ASSUMPTIONS, WE USE OLS TO ESTIMATE β_1 AND β_2

OLS ESTIMATOR FOR $\beta_1 \Rightarrow \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

$$\operatorname{var} \hat{\beta}_1 = \frac{\sigma_u^2 \sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} x_i^2}$$

OLS ESTIMATOR FOR $\beta_2 \Rightarrow \hat{\beta}_2 = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$

$$\operatorname{var} \hat{\beta}_2 = \frac{\sigma_u^2}{\sum_{i=1}^{n} x_i^2}$$

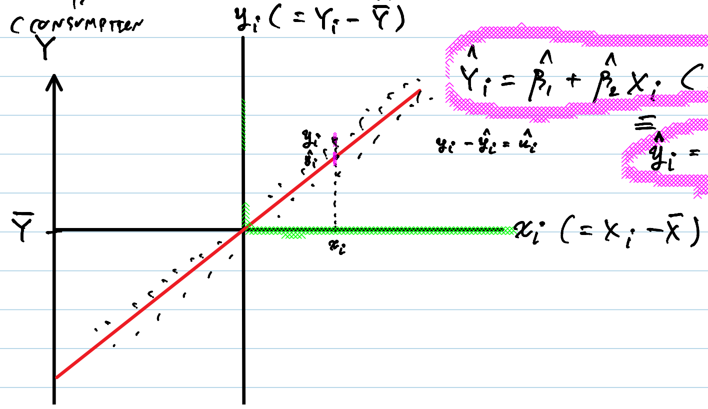
$$\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\bar{X}\sigma_u^2}{\sum_{i=1}^{n} x_i^2}$$

WHERE $\quad x_i = X_i - \bar{X} \quad$ (DEVIATION FORM)

NOTE: ALL WE DISCUSS TODAY IS STILL IN CH. 3.

AS WE DO NOT KNOW σ_u^2, WE HAVE TO FIND A WAY TO ESTIMATE IT.

$\Rightarrow$ FIND UNBIASED ESTIMATOR OF σ_u^2.



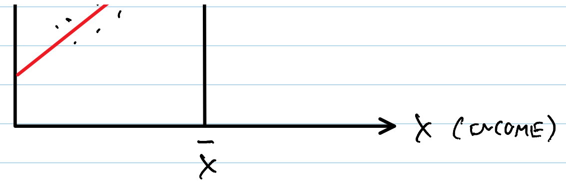
$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ (SRF)

$\equiv$

$\hat{y}_i = \hat{\beta}_2 x_i$ (SRF)

$y_i - \hat{y}_i = \hat{u}_i$

$u_i \Rightarrow$ DISTURBANCE TERM OR



$u_i \Rightarrow$ DISTURBANCE TERM OR ERROR TERM (POPULATION SENSE)

$\hat{u}_i \Rightarrow$ RESIDUAL TERM (SAMPLE SENSE)

$\hat{u}_i = y_i - \hat{y}_i$ AND $\hat{y}_i = \hat{\beta}_2 x_i$ (OUR SRF IN DEVIATION FORM)

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad -(1)$$

$$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \bar{u} \qquad -(2)$$

(1) − (2) GIVES :

$$Y_i - \bar{Y} = \beta_2 (X_i - \bar{X}) + (u_i - \bar{u})$$

$$y_i = \beta_2 x_i + (u_i - \bar{u})$$

NOW

$$\hat{u}_i = y_i - \hat{y}_i$$

$$= y_i - \hat{\beta}_2 x_i$$

$$\hat{u}_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i$$

$$\hat{u}_i = (u_i - \bar{u}) - (\hat{\beta}_2 - \beta_2) x_i$$

$$\hat{u}_i^2 = (u_i - \bar{u})^2 + (\hat{\beta}_2 - \beta_2)^2 x_i^2 - 2(u_i - \bar{u})(\hat{\beta}_2 - \beta_2) x_i$$

$(a-b)^2 = a^2 - 2ab + b^2$

$$\sum_{i=1}^{n} \hat{u}_i^2 = \underbrace{\sum_{i=1}^{n} (u_i - \bar{u})^2}_{\mathrm{I}} + \underbrace{(\hat{\beta}_2 - \beta_2)^2 \sum_{i=1}^{n} x_i^2}_{\mathrm{II}} \underbrace{- 2(\hat{\beta}_2 - \beta_2) \sum_{i=1}^{n} (u_i - \bar{u}) x_i}_{\mathrm{III}}$$

SEE IF $E\left[\sum_{i=1}^{n} \hat{u}_i^2\right] = \sigma_u^2$?

(I) $E\left[\sum_{i=1}^{n} (u_i - \bar{u})^2\right] = (n-1)\sigma_u^2$

(II) $E\left[(\hat{\beta}_2 - \beta_2)^2 \sum_{i=1}^{n} x_i^2\right] = \sum_{i=1}^{n} x_i^2 \, E[(\hat{\beta}_2 - \beta_2)]^2 = \sum_{i=1}^{n} x_i^2 \, \mathrm{var}(\hat{\beta}_2) = \cancel{\sum_{i=1}^{n} x_i^2} \cdot \dfrac{\sigma_u^2}{\cancel{\sum_{i=1}^{n} x_i^2}} = \sigma_u^2$.

(III) $E\left[-2(\hat{\beta}_2 - \beta_2) \sum_{i=1}^{n} (u_i - \bar{u}) x_i\right] = -2\sigma_u^2$

THEREFORE $E\left[\sum_{i=1}^{n} \hat{u}_i^2\right] = (n-1)\sigma_u^2 + \sigma_u^2 - 2\sigma_u^2$

$$= [(n - \cancel{1}) \cancel{+1} - 2]\, \sigma_u^2$$

$$= (n-2)\sigma_u^2 \rightarrow \text{WHICH IS NOT UNBIASED.}$$

SO IF WE USE $\sum_{i=1}^{n} \hat{u}_i^2$ AS AN ESTIMATOR FOR $\mathrm{var}(u_i) = \sigma_u^2$,

THEN $E\left[\sum \hat{u}_i^2\right] = \sigma_u^2 (n-2) \neq \sigma^2$ ☹

SO IF WE USE $\sum_{i=1} \hat{u}_i$ AS AN ESTIMATOR FOR $var(u_i) = \sigma_u^2$,

THEN $E\left[\frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}\right] = \frac{\sigma_u^2 (n-2)}{n-2} = \sigma_u^2$:)

$\therefore$ WE DEFINE $\boxed{\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}}$ IS AN UNBIASED ESTIMATOR FOR σ_u^2.

(14.9.2012)

AFTER WE OBTAIN ALL STUFFS, i.e.,

$$\left.\begin{array}{l} \hat{\beta}_1 = \dots \\ \hat{\beta}_2 = \dots \\ var(\hat{\beta}_1) = \dots \rightarrow s.e.(\hat{\beta}_1) = \dots \\ var(\hat{\beta}_2) = \dots \rightarrow s.e.(\hat{\beta}_2) = \dots \\ cov(\hat{\beta}_1, \hat{\beta}_2) = \dots \\ \hat{\sigma}_u^2 = \dots \end{array}\right\} \text{FROM OLS ESTIMATOR} \downarrow \text{(TOOL)}$$

ESTIMATOR FOR σ_u^2

WE NEXT DISCUSS ABOUT "STATISTICAL PROPERTIES OF" OLS ESTIMATOR.

① $\hat{\beta}_1$ AND $\hat{\beta}_2$ ARE EXPRESSED IN TERM OF OBSERVABLE VARIABLES (i.e., X AND Y FROM YOUR SAMPLE)

$\hat{\beta}_1 \xrightarrow{\text{TO MAKE AN INFERENCE ON}}$ β_1 (population intercept coefficient)

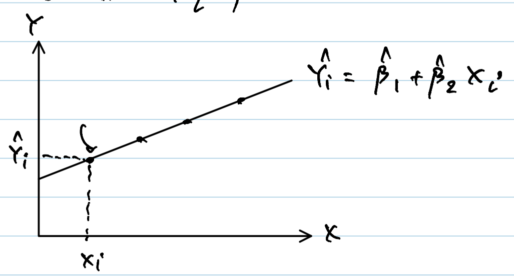
$\hat{\beta}_2 \longrightarrow$ β_2 (population slope coefficient)

SO, THEY CAN BE COMPUTED EASILY.

② THEY ARE POINT ESTIMATORS.

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

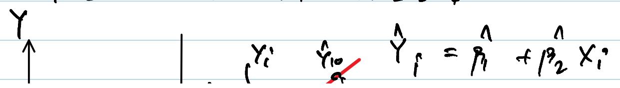
i.e, GIVEN X_i, WE CAN GET ONE POINT OF $\hat{Y}_i$



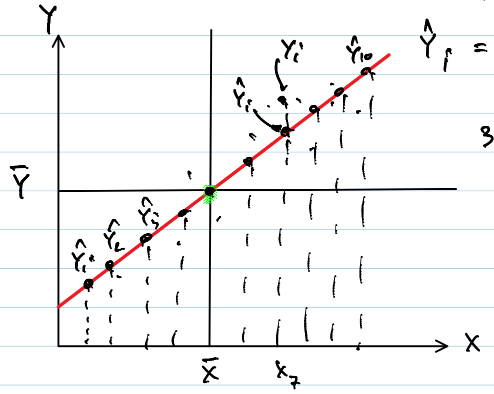
OF COURSE, YOU CAN ALSO FIND "INTERVAL ESTIMATORS" IF YOU'D LIKE

GIVE YOU A RANGE OF POSSIBLE VALUES FOR THE UNKNOWN PARAMETERS.

③ THE REGRESSION LINE ($\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$) HAS THE FOLLOWING FEATURES:



FOLLOWING FEATURES :



$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

3.1 THE REGRESSION LINE WILL PASS THROUGH THE SAMPLE MEANS OF Y AND X (i.e., $\bar{Y}$ AND $\bar{X}$)

$$\boxed{\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}}$$

$\hat{Y}_i$

3.2 THE MEAN VALUE OF THE ESTIMATED Y IS EQUAL TO THE MEAN VALUE OF THE ACTUAL Y !

$$\bar{\hat{Y}} = \bar{Y}$$

PROOF :

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i$$

$$= \bar{Y} - \hat{\beta}_2 \bar{X} + \hat{\beta}_2 X_i$$

$$= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})$$

SUMMING UP ALL $\hat{Y}_i$ AND DIVIDING BY SAMPLE SIZE (n) OR # OF OBSERVATIONS,

WE WILL GET

$$\frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i = \frac{1}{n}\sum_{i=1}^{n}\left[\bar{Y} + \hat{\beta}_2 (X_i - \bar{X})\right]$$

$$\bar{\hat{Y}} = \bar{Y} + \frac{\hat{\beta}_2 \overbrace{\sum_{i=1}^{n}(X_i - \bar{X})}^{=0}}{n}$$

$$\boxed{\bar{\hat{Y}} = \bar{Y}}$$

(READ OUT LOUD)

MEAN VALUE OF ESTIMATED $\hat{Y}_i$

MEAN VALUE OF ACTUAL Y_i

EX: $\boxed{1, 2, 3, 4}$

$\bar{X} = \frac{1+2+3+4}{4}$

$= \frac{10}{4} = 2.5$

$X_1 - \bar{X} = 1 - 2.5 = -1.5$

$X_2 - \bar{X} = 2 - 2.5 = -.5$

$X_3 - \bar{X} = 3 - 2.5 = .5$

$X_4 - \bar{X} = 4 - 2.5 = 1.5$

$\sum_{i=1}^{4} (X_i - \bar{X}) = -1.5 - .5 + .5 - 1.5$

$= 0$

3.3 MEAN VALUE OF RESIDUALS, $\hat{u}_i$ IS ZERO :

$$\boxed{\bar{\hat{u}} = \frac{\sum_{i=1}^{n} \hat{u}_i}{n} = 0}$$

3.4. THE RESIDUALS $\hat{u}_i$ ARE UNCORRELATED WITH THE PREDICTED Y_i, $\hat{Y}_i$

MATHEMATICALLY, $\boxed{\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0}$ (IN DEVIATION FORM)

NOTE : SEE THE LITTLE PROOF IN GUJARATI. **

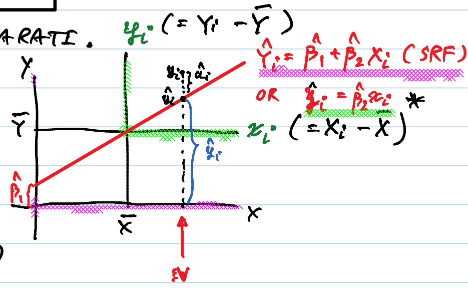
3.5 THE RESIDUALS $\hat{u}_i$ ARE UNCORRELATED WITH X_i

MATHEMATICALLY, $\boxed{\sum_{i=1}^{n} X_i \hat{u}_i = 0.}$

NOTE: SEE THE PROOF IN GUJARATI.

3.6 $\boxed{y_i = \hat{y}_i + \hat{u}_i}$.

WHERE DOES IT COME FROM MATHEMATICALLY ?



SEE ⟹ $Y_i = \hat{Y}_i + \hat{u}_i$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad \text{— (1)}$$

BE CAREFUL HERE...

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X} + \bar{\hat{u}} \quad \text{— (2)}$$

$= 0$ FROM THE PROPERTY WE HAVE ABOVE.

(1) − (2) GIVES $Y_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) + \hat{u}_i$

$$y_i = \underbrace{\hat{\beta}_2 x_i}_{= \hat{y}_i} + \hat{u}_i .$$

THEREFORE $y_i = \hat{y}_i + \hat{u}_i$ WHERE $\hat{y}_i = \hat{\beta}_2 x_i$

SEE THE GRAPH ABOVE

ACTUAL y_i — PREDICTED y_i — RESIDUALS

(IN DEVIATION FORM)

THAT'S ALL FOR "STATISTICAL PROPERTIES" OF OLS ESTIMATORS.

NEXT, BEFORE WE START TO USE IT EMPIRICALLY, WE SHOULD UNDERSTAND "THE ASSUMPTIONS" UNDERLYING THE OLS METHOD.

WHY WE HAVE TO TALK ABOUT "ASSUMPTIONS" ?

ANSWERS

① OUR OBJECTIVE IS NOT ONLY TO OBTAIN $\hat{\beta}_1$ AND $\hat{\beta}_2$ BUT ALSO

- WE WANT TO BE ABLE TO KNOW HOW CLOSE $\hat{Y}_i$ IS TO OUR TRUE $E(Y \mid X_i)$.
- WE ALSO WANT TO KNOW HOW $\hat{\beta}_1 \xrightarrow{\text{CLOSE TO}} \beta_1$

  HOW $\hat{\beta}_2 \xrightarrow{\text{CLOSE TO}} \beta_2$