Tuesday, February 21, 2012
9:07 AM

WEEK 7 (23 FEB 2012)

35%

READ: ① GUJARATI CH 1 − 4 + CH.5 (5.1 − 5.8)

② TOPICS IN LECTURE NOTE

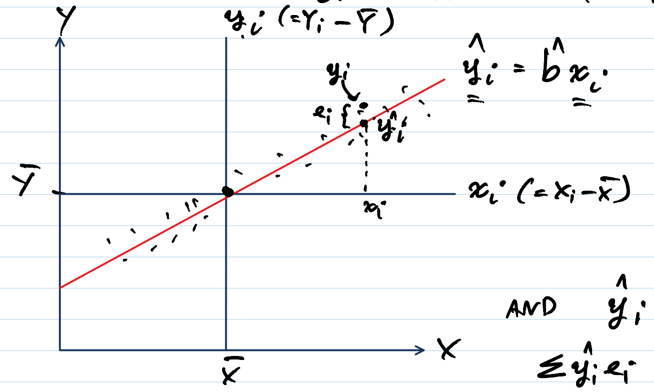
3 MAIN QUESTIONS [. PROOF / INTERPRETATION

DIY: ◦ EXERCISES ON EACH CHAPTER

• (WHY?)

• THE COEFFICIENT OF DETERMINATION (R^2)

= DEGREE OF EXPLANATORY POWER.



$\hat{y}_i = \hat{b} x_i$

FROM $y_i = \hat{y}_i + e_i$

$y_i^2 = \hat{y}_i^2 + e_i^2 + 2\hat{y}_i e_i$ $\qquad e_i = \hat{u}_i$ IN GUJARATI

$$\boxed{\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 + 2\sum \hat{y}_i e_i}$$

AND $\hat{y}_i = \hat{b} x_i$

$$\sum \hat{y}_i e_i = \sum \hat{b} x_i e_i = \hat{b} \underbrace{\sum x_i e_i}_{0} = 0$$

$$\therefore \boxed{\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2}.$$

$\sum (Y_i - \bar{Y})^2$: TOTAL SUM OF SQUARES (TSS)

$\sum \hat{y}_i^2$: EXPLAINED SUM OF SQUARES (ESS)

$\sum e_i^2$: RESIDUAL SUM OF SQUARE (RSS)

$$\therefore \boxed{TSS = ESS + RSS}$$

DEFINE $R^2 = \dfrac{ESS}{TSS} = \dfrac{\text{EXPLAINED SUM OF SQUARE}}{\text{TOTAL SUM OF SQUARE}}$

(TO SEE HOW MUCH VARIATION IN Y COULD BE EXPLAINED BY X)

FROM $TSS = ESS + RSS$,

DIVIDE THROUGH OUT W/ TSS:

$$\frac{TSS}{TSS} = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

$$1 = R^2 + \frac{RSS}{TSS} = R^2 + \frac{\sum e_i^2}{\sum y_i^2}$$

2 ∴ .

$$1 = R^2 + \frac{RSS}{TSS} = R^2 + \frac{\sum e_i^2}{\sum y_i^2}$$

SO $$\boxed{R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2}}$$

- $R^2 = 1$ (MAX. VALUE) WHEN $\sum e_i^2 = 0$ (each error $= 0$)

  SO WE CALL "PERFECT FIT"

- $R^2 = 0$ (MIN VALUE) WHEN ERROR TERM EQUALS TO VARIATION AROUND y_i, i.e., $\sum e_i^2 = \sum y_i^2$

  IMPLIES THAT THE REGRESSION LINE CANNOT EXPLAINED ANY VARIATION IN Y.

R^2 : MEASURES THE GOODNESS OF FIT.

FOR EXAMPLE : $R^2 = 0.8 \Rightarrow$ 80% VARIATION IN Y CAN BE EXPLAINED BY THE REGRESSION LINE.

THE GREATER R^2 IS, THE HIGHER DEGREE OF EXPLANATORY POWER OF THE REGRESSION LINE.

REMEMBER THAT WE EARLIER DISCUSSED ABOUT r : CORRELATION COEFFICIENT WHICH MEASURES THE ASSOCIATION BETWEEN ANY TWO VARIABLES.

ADJUSTED-R^2

FROM $$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2} \times \frac{\sqrt{\frac{\sum y_i^2}{n}}}{\sqrt{\frac{\sum y_i^2}{n}}}$$

$= \sqrt{\sum x^2} \cdot \sqrt{\sum x^2}$

$$= \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}} \cdot \frac{\sqrt{\frac{\sum y_i^2}{n}}}{\sqrt{\frac{\sum x_i^2}{n}}}$$

$$= \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \cdot \frac{S_y}{S_x} = r \cdot \frac{S_y}{S_x}$$

$$\therefore r = \hat{b}\,\frac{S_x}{S_y} \Rightarrow r^2 = \hat{b}^2 \cdot \frac{S_x^2}{S_y^2}$$

$$= \hat{b}^2 \cdot \frac{\sum x_i^2}{\sum y_i^2}$$

$$= \frac{\sum \hat{b}^2 x_i^2}{\sum y_i^2} = \frac{\sum (\hat{b} x_i)^2}{\sum y_i^2} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = R^2$$

SO $\boxed{r^2 = R^2}$ TRUE ONLY IN SIMPLE

$\sum y_i^2 \quad \sum y_i^2 \quad \sum y_i^2$

SO, $\boxed{r^2 = R^2}$ TRUE ONLY IN SIMPLE REGRESSION. (TWO-VARIABLE REGRESSION)

INTERVAL ESTIMATION AND HYPOTHESES TESTING [CHAPTER 5]

BIG PICTURE :

CLASSICAL STATISTICS
- ESTIMATION
  - POINT ESTIMATION
  - INTERVAL ESTIMATION → RELATED WITH
- HYPOTHESES TESTING

POINT ESTIMATION:

$Y = a + bX_i + u_i$ → PRF

$\hat{Y} = \hat{a} + \hat{b}X_i + \hat{u}_i$ → SRF

POINT ESTIMATOR.

BY USING OLS

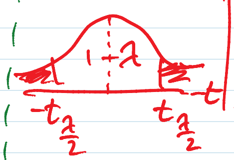
Ex: $C = 70 + 0.82\,Y$

$\hat{MPC} = 0.82$

estimated coefficient

5000 FAMILIES

n = 30 FAMILIES



Ex:

$H_0: b = 0$

$H_1: b \neq 0$

two-tail test

OR

$H_0: b \leq 0$

$H_1: b > 0$

one-tailed test

WE USE HYPOTHESIS TESTING TO TELL THAT $\hat{a}$ AND $\hat{b}$ ARE STATISTICALLY DIFFERENT FROM ZERO.

ex: IF b = MPC ⇒

$H_0: b \leq 0$

$H_1: b > 0$

(ONE-TAILED TEST)

FROM MODEL : $Y_i = a + bX_i + u_i$

YOU GET $\hat{a} = \bar{Y} - \hat{b}\bar{X}$

AND

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \sum_{i=1}^{n} k\, Y_i$$

IF WE ASSUME THAT $u_i \sim N(0, \sigma_u^2)$

" THIS IS NORMALITY ASSUMPTION

OF u_i (READ CH. 4)

$u_i \sim N(0, \sigma_u^2) \Rightarrow$ MEAN: $E(u_i) = 0$

VARIANCE: $E(u_i - E(u_i))^2 = E(u_i)^2$

$= \sigma_u^2$

COVARIANCE: $E(u_i - E(u_i))(u_j - E(u_j))$

$= E(u_i u_j) = 0$

QUESTION: WHY DO YOU NEED NORMALITY ASSUMPTION OF u_i?

(EXAMPLE QUESTION YOU MIGHT SEE IN THE MIDTERM)

So, u_i IS iid $N(0, 1)$ WHERE

iid = independently and identically distributed