Akshay

lec 9

Date:
Page No.

Regression

$$\hat{\beta} = (Z'Z)^{-1}Z'Y \quad , \hat{y} = Z\hat{\beta} = \underbrace{Z(Z'Z)^{-1}Z'}_{H}y = Hy$$

$$\hat{\epsilon} = y - \hat{y} = y - Z\hat{\beta} = (I-H)y$$

→ symmetric
→ Idempotent

$$Z'\hat{\epsilon} = \underset{\sim}{0} \quad , \hat{y}'\hat{\epsilon} = 0$$

→ **Sum of squares decomposition:**

$$\hat{y}'\hat{\epsilon} = 0$$

$$y_1^2 + y_2^2 + y_3^2 + \cdots + y_n^2$$

$$\text{TSS: } y'y = \sum_{j=1}^{n} y_j^2 = (\hat{y} + y - \hat{y})'(\hat{y} + y - \hat{y})$$

$$= (\hat{y} + \hat{\epsilon})'(\hat{y} + \hat{\epsilon}) = \hat{y}'y + \hat{\epsilon}'\hat{\epsilon}$$

$$1'\hat{\epsilon} = 0$$ because mean of ϵ is zero

$$\sum_{j=1}^{n} \hat{\epsilon}_j = \sum_{j=1}^{n} (y_j - \hat{y}_j) = \sum_{1}^{n} y_j - \sum_{1}^{n} \hat{y}_j = 0.$$

$$= \boxed{\bar{y} = \bar{\hat{y}}}$$ →averages are same.

Scanned with CamScanner

$Cov(\underset{\sim}{X}) = E[(\underset{\sim}{x} - \underset{\sim}{\mu})$

Date
Page No.

$$y'y - n\bar{y}^2 = \hat{y}'\hat{y} - n(\bar{\hat{y}})^2 + \hat{\epsilon}'\hat{\epsilon}$$

or

$$\underbrace{\sum_{j=1}^{n} (y_j - \bar{y})^2}_{\text{TSS about mean}} = \underbrace{\sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2}_{\text{RSS}} + \underbrace{\sum_{j=1}^{n} \hat{\epsilon}_j^2}_{\text{Residual sum of Sqr.}}$$

RSS: Regression sum of sqr

Coefficient of determination

What proportion of the total variation is explained by the ~~var~~ predictor variables. $\Rightarrow$

$$R^2 = 1 - \frac{\text{Residual SS}}{\text{TSS}}$$

$$= 1 - \frac{\sum_{j=1}^{n} \epsilon_j^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2}$$

Used to check whether model is good fit or not.

$$0 \le R^2 \le 1$$

$$= \frac{\sum (\hat{y}_j - \bar{y}_j)^2}{\sum (y_j - \bar{y}_j)^2}$$

$R^2 \approx 1$ good fit

$R^2 \sim 0$ Not good fit

Date:
Page No.

Sampling Properties. $\hat{\beta}, \hat{\epsilon}$

$$E(\hat{\beta}) = E\left[(Z'Z)^{-1}Z'Y\right]$$

$$E\left[(Z'Z)^{-1}Z'(Z\beta + \epsilon)\right]$$

$$= \underbrace{(Z'Z)^{-1}Z'Z}_{I}\beta + E(\epsilon) \rightarrow 0$$

$$= \beta$$

OLS $\hat{\beta}$ is unbiased.

$$Cov(\hat{\beta}) = \sigma^2 (Z'Z)^{-1}$$

$$Cov(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\right]$$

$$\hat{\beta} = (Z'Z)^{-1}Z'(Z\beta + \epsilon) \qquad E\left[\cancel{(Z'Z)^{-1}Z'\epsilon}\right]$$

$$= I\beta + (Z'Z)^{-1}Z'\epsilon$$

$$\hat{\beta} - \beta = (Z'Z)^{-1}Z'\epsilon$$

$$Cov(\hat{\beta}) = E\left[(Z'Z)^{-1}Z'\epsilon\epsilon'Z(Z'Z)^{-1}\right]$$

$$= (Z'Z)^{-1}Z' E(\epsilon\epsilon') Z(Z'Z)^{-1}$$

$$Cov(\epsilon) = \sigma^2 I$$

$$E\left[(\epsilon - 0)(\epsilon - 0)^T\right]$$

$$= \sigma^2 I.$$

Date:
Page No.

$$= \sigma^2 I (Z'Z)^{-1} Z'Z (Z'Z)^{-1} = \sigma^2 (Z'Z)^{-1}$$

$$\boxed{E(\hat{\beta}) = \beta, \quad Cov(\hat{\beta}) = \sigma^2 (Z'Z)^{-1}}$$

$$\boxed{E(\hat{\epsilon}), \quad Cov(\hat{\epsilon}) = \sigma^2 (I - H)}$$

$$\hat{\epsilon} = (I-H)Y = (I-H)(\cancel{Z}\underset{\sim}{\beta} + \underset{\sim}{\epsilon})$$

$$(Z\underset{\sim}{\beta} + \underset{\sim}{\epsilon})$$

$$= (\cancel{Z} - Z(Z'Z)^{-1} Z'Z)\underset{\sim}{\beta} + (I-H)\underset{\sim}{\epsilon}$$

$$= \underset{\sim}{0} + (I-H)\underset{\sim}{\epsilon}$$

$$E(\hat{\epsilon}) = \underset{\sim}{0}$$

$$Cov(\hat{\epsilon}) = Cov(I-H)Y$$

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

* $\hat{\beta}$ and $\hat{\epsilon}$ are uncorrelated

Gauss least square theorem

OLS is BLUE — Best linear unbiased estimator

$$\hat{\beta} = CY$$

$$\hat{\beta}_v = c_1 y_1 + c_2 y_2 + \dots + c_n y_n$$

Variance of $\hat{\beta}$ is minium. among all other linear estimators.