Lecture (4): Some properties of the regression line equation

1. Sum of observed values = Sum of expected values $\sum Y_i = \sum \hat{Y}_i$

Proof :-

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$= \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i$$

$$\Rightarrow \hat{Y}_i = \bar{Y} + \hat{\beta}_1 (X_i - \bar{X})$$

$$\sum \hat{Y}_i = \sum \bar{Y} + \hat{\beta}_1 \sum (X_i - \bar{X})$$

$$\Rightarrow \sum \hat{Y}_i = n\,\bar{Y}$$

$$= n\frac{\sum Y_i}{n} \Rightarrow \sum \hat{Y}_i = \sum Y_i$$

2. Total random errors = zero $\sum e_i = 0$

Proof :-

$$e_i = Y_i - \hat{Y}_i$$

$$\sum e_i = \sum (Y_i - \hat{Y}_i)$$

$$= \sum Y_i - \sum \hat{Y}_i$$

$$= \sum Y_i - \sum \hat{Y}_i = 0$$

3. $\hat{\beta}_1 \, S_{XY} = \hat{\beta}_1 \, \frac{S_{XY} S_{XX}}{S_{XX}}$

$$= \hat{\beta}_1 \, \hat{\beta}_1 \, S_{XX} = \hat{\beta}_1^2 \, S_{XX}$$

4. The sum of the errors (or residuals) weighted by the corresponding values of x_i = zero, i.e. $\sum e_i x_i = 0$

Proof :-

$$\sum X_i (Y_i - \hat{Y}_i)$$

$$= \sum X_i \left[Y_i - \left(\bar{Y} + \hat{\beta}_1 \, (X_i - \bar{X}) \right) \right]$$

$$= \sum X_i \left[(Y_i - \bar{Y}) - \hat{\beta}_1 \, (X_i - \bar{X}) \right]$$

$$= \sum X_i(Y_i - \overline{Y}) - \hat{\beta}_1 \sum X_i(X_i - \overline{X})$$

$$= S_{xy} - \hat{\beta}_1 S_{xx}$$

$$= S_{xy} - \frac{S_{xy}}{S_{xy}} S_{xy} = 0 \rightarrow \sum e_i x_i = 0$$

5. **The sum of the residuals weighted by the Y_i values equals the sum of the squares of the errors. $\sum e_i Y_i = sse$**

Proof :-

$$\sum Y_i\, e_i = \sum Y_i\,(Y_i - \hat{Y}_i)$$

$$= \sum Y_i \left[Y_i - \left(\overline{Y} + \hat{\beta}_1\,(x_i - \overline{x})\right)\right] = \sum Y_i\,\left[(Y_i - \overline{Y}) - \hat{\beta}_1\,(x_i - \overline{x})\right]$$

$$= \sum Y_i\,\left(Y_i - \overline{Y}\right) - \hat{\beta}_1\,(x_i - \overline{x})Y_i = S_{YY} - \hat{\beta}_1\; S_{XY}$$

$$= sse \rightarrow \sum y_i e_i = sse = Residual\ sum\ of\ squares$$

$$= Error\ sum\ of\ squares$$

6. $\sum e_i \hat{Y}_i = 0$

Proof :-

$$\sum e_i \hat{Y}_i = \sum (Y_i - \hat{Y}_i)\hat{Y}_i$$

$$= \sum Y_i\,\hat{Y}_i - \sum \hat{Y}_i^2$$

$$= \sum Y_i \left(\overline{y} + \hat{\beta}_1\,(x_i - \overline{x})\right) - \sum \left[\overline{Y} + \hat{\beta}_1 (x_i - \overline{x})\right]^2$$

$$= \sum Y_i\,\overline{Y} + \hat{\beta}_1 \sum (x_i - \overline{x})\,Y_i - \sum \left[\overline{Y}^2 + 2\,\overline{Y}\,\hat{\beta}_1 (x_i - \overline{x}) + \hat{\beta}_1^2 (x_i - \overline{x})^2\right]$$

$$= \frac{(\sum Y_i)^2}{n} + \hat{\beta}_1\, S_{xy} - n\,\frac{(\sum Y_i)^2}{n^2} - 2\overline{Y}\,\hat{\beta}_1 \sum (x_i - \overline{x}) - \hat{\beta}_1^2 \sum (x_i - \overline{x})^2$$

$$= \frac{(\sum Y_i)^2}{n} + \hat{\beta}_1\, S_{xy} - \frac{(\sum Y_i)^2}{n} - \hat{\beta}_1^2\, S_{xx}$$

$$= 0 \Longrightarrow \sum e_i \hat{Y}_i = 0$$

7. $E(\hat{\beta}_1) = \beta_1$

Proof :-

$$E(\hat{\beta}_1) = E\left(\frac{S_{xy}}{S_{xx}}\right) = E\left[\frac{\sum(x_i-\overline{x})y_i}{\sum(x_i-\overline{x})^2}\right] = \frac{\sum(x_i-\overline{x})\sum y_i}{\sum(x_i-\overline{x})^2}$$

$$= \frac{\sum(x_i-\overline{x})\,(\beta_0+\beta_1 x_i)}{\sum(x_i-\overline{x})^2} = \frac{\beta_0\sum(x_i-\overline{x})\ +\beta_1\sum(x_i-\overline{x})\,x_i}{\sum(x_i-\overline{x})^2}$$

$$= \frac{\text{صفر}+\beta_1\sum(x_i-\overline{x})\,x_i}{\sum(x_i-\overline{x})^2} = \beta_1$$

8. $\boldsymbol{E(\hat{\beta}_0) = \beta_0}$

Proof :-

$$E(\hat{\beta}_0) = E(\overline{Y}-\hat{\beta}_1\overline{X})$$

$$= E(\overline{Y}) - \overline{X}\,E(\hat{\beta}_1)$$

$$= E\left(\frac{1}{n}\sum Y_i\right) - \overline{X}\beta_1$$

$$= \frac{1}{n}\sum E(Y_i) - \overline{X}\beta_1$$

$$= \frac{1}{n}\sum(\beta_0+\beta_1 X_i) - \overline{X}\beta_1$$

$$= \frac{\sum\beta_0}{n} + \beta_1\frac{\sum X_i}{n} - \overline{X}\beta_1$$

$$= \frac{\cancel{n}\beta_0}{\cancel{n}} + \cancel{\beta_1\overline{X}} - \cancel{\overline{X}\beta_1} = \beta_0$$

9. $\boldsymbol{SSe = \sum e_i^2 = \sum(Y_i-\hat{Y}_i)^2}$

$$= \sum\left[y_i - \left(\overline{y}+\hat{\beta}_1(x_i-\overline{x})\right)\right]^2$$

$$= \sum\left[(y_i-\overline{y}) - \left(\hat{\beta}_1(x_i-\overline{x})\right)\right]^2$$

$$= \sum\left[(y_i-\overline{y})^2 - 2\hat{\beta}_1(x_i-\overline{x})(y_i-\overline{y}) + \hat{\beta}_1^2(x_i-\overline{x})^2\right]$$

$$= \sum(y_i-\overline{y})^2 - 2\hat{\beta}_1\sum(x_i-\overline{x})(y_2-\overline{y}) + \hat{\beta}_1^2(x_i-\overline{x})^2$$

$$= S_{yy} - 2\hat{\beta}_1 S_{xy} + \hat{\beta}_1^2 S_{xx}$$

$$= S_{yy} - 2\,\hat{\beta}_1 S_{xy} + \hat{\beta}_1 S_{xy} = S_{yy} - \hat{\beta}_1 S_{xy} = S_{yy} - \hat{\beta}_1^2 S_{xx}$$

Estimation of the variance of the regression coefficient $\hat{\beta}_1$

We have $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum(x_i-\overline{x})Y_i}{\sum(x_i-\overline{x})^2} = \sum\left(\frac{(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right)y_i$

Let $a = a_1\, y_1 + a_2\, y_2 + \cdots + a_n\, y_n$

$\Longrightarrow v(a) = a_1^2\, v\,(y_1) + a_2^2\, v\,(y_2) + \ \dots + \ a_n^2\, v\,(y_n)$

By taking advantage of Homoscedasticity property $v(y_1) = v(y_2) = \cdots v(y_n)$

Then $v(a) = (a_1^2\ + a_2^2\ +\ \dots +\ a_n^2)\, v\,(y)$

$$= \left(\sum a_i^2\right) v(y) = \left(\sum a_i^2\right)\hat{\sigma}^2$$

By comparing $\hat{\beta}_1$ with a, we have:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum\left(\frac{(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right)y_i$$

We have: $\sum\left(\frac{(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right)y_i\ = \underbrace{\frac{(x_1-\overline{x})}{S_{xx}}}_{a_1}\, y_1 + \underbrace{\frac{(x_2-\overline{x})}{S_{xx}}}_{a_2}\, y_2 + \cdots + \underbrace{\frac{(x_n-\overline{x})}{S_{xx}}}_{a_n}\, y_n$

$\Longrightarrow v(\hat{\beta}_1) = \sum\left(\frac{x_i-\overline{x}}{S_{xx}}\right)^2 .\hat{\sigma}^2 \ = \sum\frac{(x_i-\overline{x})^2}{(S_{xx})^2}.\hat{\sigma}^2 = \frac{S_{xx}}{(S_{xx})^2}.\hat{\sigma}^2 = \frac{\hat{\sigma}^2}{S_{xx}}$

$$\therefore v(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}}$$

The sample random error variance S_e^2 is an unbiased estimate of the population random error variance σ^2, i.e. $E\,(S_e^2) = \sigma^2$

Proof:

$$E\,(S_e^2) = E(\hat{\sigma}^2)\ = E\,\frac{SSE}{n-2} = \frac{E(SSE)}{n-2} = E(Mse)$$

We have $E(SSE) = E\left[\sum_{i=1}^{n}(Y_i-\overline{Y})^2 - \hat{\beta}_1^2\sum\left(X_i-\overline{X}\right)^2\right]$

$$= \ E\left[\sum_{i=1}^{n}(Y_i^2 - 2\overline{Y}\,Y_i + \overline{Y}^2) + \ \hat{\beta}_1^2\sum\left(X_i-\overline{X}\right)^2\right]$$

$$= E\left[\sum {Y_i}^2 - 2\frac{\sum Y_i}{n}\sum Y_i + n\overline{Y}^2 - \hat{\beta}_1^2 \sum (X_i - \overline{X})^2\right]$$

$$= E\left[\sum {Y_i}^2 - 2\frac{(\sum Y_i)^2}{n} + n\overline{Y}^2 - \hat{\beta}_1^2 \sum (X_i - \overline{X})^2\right]$$

$$= E\left[\sum {Y_i}^2 - 2n\left(\frac{\sum Y_i}{n}\right)^2 + n\overline{Y}^2 - \hat{\beta}_1^2 \sum (X_i - \overline{X})^2\right]$$

$$= E\left[\sum {Y_i}^2 - n\overline{Y}^2 - \hat{\beta}_1^2 \sum (X_i - \overline{X})^2\right]$$

$$= \sum E\,({Y_i}^2) - nE\left(\overline{Y}^2\right) - \sum (X_i - \overline{X})^2 E(\hat{\beta}_1^2)$$

$\because \overline{Y} = \beta_0 + \beta_1\overline{X} + \overline{e} = \overline{Y} - \beta_1\overline{X} + \beta_1\overline{X} + \overline{e} = \overline{Y}$

$V(X) = E(X^2) - [E(X)]^2 \Rightarrow E(X^2) = V(X) - [E(X)]^2$

$\therefore E(Y_i^2) = V(Y_i) + \left(E(Y_i)\right)^2 = \sigma^2 + \left(\beta_0 + \beta_1\overline{X}\right)^2$

$$E\left(\overline{Y}^2\right) = V(\overline{Y}) + \left(E(\overline{Y})\right)^2$$

$$= \frac{\sigma^2}{n} + \left(\beta_0 + \beta_1\overline{X}\right)^2$$

$\therefore E(\beta_1^2) = V(\beta_1) + \left(E(\hat{\beta}_1)\right)^2 = \dfrac{\sigma^2}{\sum (X_i - \overline{X})^2} + \beta_1^2$

$$E(SSE) = \sum_{i=1}^{n}[\sigma^2 + (\beta_0 + \beta_1X_i)^2] - n\left[\frac{\sigma^2}{n} + (\beta_0 + \beta_1X_i)^2\right] - \sum (X_i - \overline{X})^2\left[\frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} + \beta_1^2\right]$$

$$= n\sigma^2 + \textstyle\sum_{i=1}^{n}(\beta_0 + \beta_1X_i)^2 - \sigma^2 - n(\beta_0 + \beta_1X_i)^2 - \sigma^2 - \beta_1^2\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= (n-2)\sigma^2 + \textstyle\sum_{i=1}^{n}(\beta_0 + \beta_1X_i)^2 - n(\beta_0 + \beta_1\overline{X})^2 - \beta_1^2\sum_{i=1}^{n}(X_i - \overline{X})^2$$

We have $\sum_{i=1}^{n}(\beta_0 + \beta_1X_i)^2 = \sum(\beta_0^2 + 2\beta_0\beta_1X_i + \beta_1^2{X^2}_i) = n\beta_0^2 + 2\beta_0\beta_1\sum_{i=1}^{n}X_i + \beta_1^2\sum_{i=1}^{n}{X^2}_i$

And $n(\beta_0 + \beta_1\overline{X})^2 = n[\beta_0^2 + 2\beta_0\beta_1\overline{X} + \beta_1^2\overline{X}] = n\beta_0^2 + 2\beta_0\beta_1\sum X_i + \beta_1^2\frac{(\sum X_i)^2}{n}$

$$\Rightarrow \sum(\beta_0 + \beta_1 X_i)^2 - n(\beta_0 + \beta_1\overline{X})^2$$

$$= n\beta_0^2 + 2\beta_0\beta_1\sum X_i + \beta_1^2\sum_{i=1}^{n} X_i^2 - n\beta_0^2 - 2\beta_0\beta_1\sum X_i - \beta_1^2\frac{(\sum X_i)^2}{n}$$

$$= \beta_1^2\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right) = \beta_1^2\sum\left(X_i - \overline{X}\right)^2$$

$$\Rightarrow E(SSE) = (n-2)\sigma^2 + \beta_1^2\sum\left(X_i - \overline{X}\right)^2 - \beta_1^2\sum\left(X_i - \overline{X}\right)^2 = (n-2)\sigma^2$$

$$\Rightarrow E(S^2 e) = \frac{E(SSE)}{n-2} = \frac{(n-2)\sigma^2}{(n-2)} = \sigma^2$$

$S^2 e = \sigma^2 \therefore$ is an unbiased estimate of $S^2 e = \sigma^2 \therefore$

Note:

$$\hat{\sigma}^2 = S^2 = S_{Y/X}^2 = MSE$$