Estimating Model Variance

Thm: Under SLR assumptions,

$$e_i = y_i - \hat{y}_i \sim \overset{①}{N}\left(\overset{②}{0}, \overset{③}{\sigma^2}\left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSX}\right]\right),$$

where $SSX = \sum_{j=1}^{n} (x_j - \bar{x})^2$.

① Since e_i is a linear comb. of 2 Normal RVs, it also has a Normal dist.

② $E(e_i) = E(y_i - \hat{y}_i)$

$$= E(y_i) - E(\hat{y}_i)$$

$$= \beta_0 + \beta_1 x_i - (\beta_0 + \beta_1 x_i) = 0.$$

③ $\text{Var}(e_i) = \text{Var}(y_i - \hat{y}_i)$

$$= \text{Var}(y_i) + \text{Var}(\hat{y}_i) + 2\text{cov}(y_i, -\hat{y}_i)$$

$$= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{SSx}\right] - 2\,\mathrm{Cov}(y_i, \hat{y}_i)$$

$$\mathrm{Cov}(y_i, \hat{y}_i) = \mathrm{Cov}\left(y_i, \sum_{j=1}^{n} c_j y_i\right)$$

where $c_j = \frac{1}{n} + (x_i - \bar{x}) k_j$

and $k_j = (x_j - \bar{x}) / SSx$.

$$= \mathrm{Cov}(y_i, c_1 y_1 + c_2 y_2 + \cdots + c_i y_i + \cdots c_n y_n)$$

$$= \mathrm{Cov}(y_i, c_i y_i) \quad b/c \ y_i \perp\!\!\!\perp y_j, \ i \neq j.$$

$$= c_i\,\mathrm{Cov}(y_i, y_i) = c_i\,\mathrm{Var}(y_i)$$

$$= c_i \sigma^2$$

$$= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{SSx}\right] - 2 c_i \sigma^2$$

$$= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_i-\bar{x})^2}{SSX}\right] - 2\sigma^2\left[\frac{1}{n}+\frac{(x_i-\bar{x})^2}{SSX}\right]$$

$$= \sigma^2\left[1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{SSX} - 2\left(\frac{1}{n}+\frac{(x_i-\bar{x})^2}{SSX}\right)\right]$$

$$= \sigma^2\left[1-\frac{1}{n}-\frac{(x_i-\bar{x})^2}{SSX}\right]$$

We can use the sample sum of squares of the residuals to estimate σ^2.

Observe that

$$\frac{\sum_{i=1}^{n} e_i^2}{\sigma^2} \sim \chi^2_{n-2}$$

<u>Pf</u>: OH

Hint: $E(\chi^2_{n-2}) = n-2$

$$E\left(\frac{\sum_{i=1}^{n} e_i^2}{\sigma^2}\right) = \underline{\underline{n-2}}$$

$$\frac{E\left(\sum_{i=1}^{n} e_i^2\right)}{\sigma^2} = n-2$$

$$\Rightarrow \frac{E\left(\sum_{i=1}^{n} e_i^2\right)}{n-2} = \sigma^2$$

$$\Rightarrow E\left(\frac{\sum_{i=1}^{n} e_i^2}{n-2}\right) = \sigma^2$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

"sum of sq. errors"

$$SSE = \sum_{i=1}^{n} e_i^2$$

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-p} \quad (p=2)$$

p = # of param. = # of pred + 1

Hyp. Testing & CIs for SLR

Testing the Slope

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2 / \sum_{i=1}^{n}(x_i-\bar{x})^2\right)$$

Problem: σ^2 is unknown!!

Solution: just plug in $\hat{\sigma}^2$ for it!

$$\mathrm{Var}(\hat{\beta}_1) = \sigma^2/SSX \Rightarrow$$

$$\widehat{\mathrm{Var}}(\hat{\beta}_1) = \hat{\sigma}^2/SSX$$

$$\widehat{SE}(\hat{\beta}_1) = \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)}$$

standard error

The quantity we will base our tests / CIs on is:

$$t = \frac{\hat{\beta}_1 - \beta_1^{(\cdot)}}{\widehat{SE}(\hat{\beta}_1)} \overset{H_0}{\sim} t_{n-p}$$

if we're testing

$H_0: \beta_1 = \beta_{10}$ vs

$H_1: \beta_1 \neq \beta_{10}$

Most commonly:

$H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

So the test statistic becomes:

$$t = \frac{\hat{\beta}_1 - 0}{\widehat{se}(\hat{\beta}_1)}$$

Decision Rule:

if $|t| > t^{*}_{n-2}(1-\frac{\alpha}{2})$

$\Rightarrow$ reject H_0 $\Rightarrow$

"X is a significant pred. of Y"

CI for β_1:

$$\hat{\beta}_1 \pm \underbrace{t^*_{n-2}\left(1-\frac{\alpha}{2}\right)} \widehat{SE}(\hat{\beta}_1)$$