Brief Article

The Author

October 8, 2015

1 Problem 1

1.1

For OLS, we can derive the following properties:

$Cov[X, Y] = \hat{\beta}_1/\hat{\beta}_1 Cov[X, Y]$

$= Cov[\hat{\beta}_1 X, Y]/be\hat{t}a_1$

$= Cov[\hat{\beta}_1 X + \hat{\beta}_0, Y]/\hat{\beta}_1$

$= Cov[\hat{Y}, Y]/\hat{\beta}_1$

$\sigma[X] = \hat{\beta}_1/\hat{\beta}_1 \sigma[X]$

$= \sigma[\hat{\beta}_1 X]/\hat{\beta}_1$

$= \sigma[\hat{\beta}_1 X + \hat{\beta}_0]/\hat{\beta}_1$

$= \sigma[\hat{Y}]/\hat{\beta}_1$

$\Rightarrow \rho_{xy} = Cov[X, Y]/\sigma[X]\sigma[Y] = Cov[\hat{Y}, Y]/\sigma[\hat{Y}]\sigma[Y]$

$= \sum(\hat{y}_i - \overline{y})(y_i - \overline{y})/\sqrt{\sum(\hat{y}_i - \overline{y})^2}\sqrt{\sum(y_i - \overline{y})^2}$

$= \sum(\hat{y}_i - \overline{y})(y_i - \hat{y}_i + \hat{y}_i - \overline{y})/\sqrt{\sum(\hat{y}_i - \overline{y})^2}\sqrt{\sum(y_i - \overline{y})^2}$

$= \sum[(\hat{y}_i - \overline{y})(y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})(\hat{y}_i - \overline{y})]/\sqrt{\sum(\hat{y}_i - \overline{y})^2}\sqrt{\sum(y_i - \overline{y})^2}$

$= \sum[(\hat{y}_i - \overline{y})\hat{u}_i + (\hat{y}_i - \overline{y})^2]/\sqrt{\sum(\hat{y}_i - \overline{y})^2}\sqrt{\sum(y_i - \overline{y})^2}$

$= \sum[(\hat{y}_i - \overline{y})^2]/\sqrt{\sum(\hat{y}_i - \overline{y})^2}\sqrt{\sum(y_i - \overline{y})^2}$ $\quad \because \sum(\hat{y}_i - \overline{y})\hat{u}_i = 0$

$= \sqrt{\sum[(\hat{y}_i - \overline{y})^2]}/\sqrt{\sum(y_i - \overline{y})^2}$

$= \sqrt{SSE/SST}$

$= \sqrt{R^2}$

$\Rightarrow \rho_{xy}^2 = R^2$ for single variate OLS

1.2

In general, $R^2 \neq \rho_{xy}^2$ in the multivariate case. Observe that in the proof for the single variable case, a transformation is made from X to $\hat{Y}$. However, in the multivariable case, the covariance terms would not reduce to 0 after making the transformation, and therefore the two values would not converge.

2 Problem 2

2.1

For a general linear model of the form $y_i = \sum_{j=0}^{k} \beta_j x_{ij} + u_i$
If MLR.1-MLR.6 are satisfied

$\Rightarrow \hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2/[SST_j(1-R_j^2)])$

For simplicity let $SST_j' = SST_j(1-R_j^2)$

$se[\hat{\beta}_j] = \sqrt{s^2/SST_j'}$

$s^2 = \frac{1}{n-k-1}\sum \hat{u}_i^2$

By MLR.6 $\hat{u} \sim \mathcal{N}(0,\sigma^2) \Rightarrow \hat{u}/\sigma \sim \mathcal{N}(0,1) \Rightarrow \sum \hat{u_i}^2 \sim \sigma^2 \mathcal{X}^2$

$\Rightarrow s^2(n-k-1)/\sigma^2 \sim \mathcal{X}_{n-k-1}^2$

$$\Rightarrow (\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j) = (\hat{\beta}_j - \beta_j)/\sqrt{s^2/SST_j'}$$
$$= (\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2/SST_j'}/\sqrt{s^2/\sigma^2} \qquad \because \text{ multiplying by 1}$$
$$= (\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2/SST_j'}/\sqrt{s^2(n-k-1)/[\sigma^2(n-k-1)]} \qquad \because\text{multiplying by 1}$$
$$\sim \mathcal{N}(0,1)/\sqrt{\mathcal{X}_{n-k-1}^2/(n-k-1)}$$
$$\sim t_{n-k-1}$$

2.2

Based on results from previous problem, $s^2(df)/\sigma^2 \sim \mathcal{X}_{df}^2$
Also in general, $SSR = (df)s^2 \sim \sigma^2 \mathcal{X}_{df}^2$
Since unrestricted model has $n-k-1$ degrees of freedom$\Rightarrow$ restricted model has $n-k-1-q$ degrees of freedom
This implied: $\quad (SSR_{ur} - SSR_r)/q \sim \sigma^2(\mathcal{X}_{n-k-1}^2 - \mathcal{X}_{n-k-1-q}^2)/q$

$\sim \sigma^2(\mathcal{X}_q^2/q)$

And

$SSR_{ur}/(n-k-1) \sim \sigma^2 \mathcal{X}_{n-k-q}^2$

The σ^2 in top and bottom will cancel

$\Rightarrow [(SSR_{ur} - SSR_r)/q]/[SSR_{ur}/(n-k-1)] \sim (\mathcal{X}_q^2/q)/(\mathcal{X}_{n-k-1}^2/(n-k-1))$

$\sim F_{q,n-k-1}$

3 Problem 3

4 Problem 4

$\log wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$

4.1

i. Another year of general work experience has as much effect on $\log wage$ as another year of tenure with current employer

$H_0 : \hat{\beta}_2 = \hat{\beta}_3$

However, this can also be written as:

$H_0 : \theta = \hat{\beta}_2 - \hat{\beta}_3 = 0$

And the model equation will be transformed to:

$\log(wage) = \beta_0 + \beta_1 educ + \theta exper + \beta_2(exper + tenure)$

ii. Calculating the t-statistic of the regression:

5 Problem 5

$\log(psoda) = \beta_0 + \beta_1 prpblck + \beta_2 \log(income) + \beta_3 prppov + u$

6 Problem 6

The second model is preferable since it has a higher $\overline{R}^2$ value than that of the other two models. Without