TA Notes - Week 6: Selected Discussion Questions

Conor Foley*

Econ 103 (Randall Rojas) Winter 2018

Contents

1 Question 5.7

What are the standard errors of the least squares estimates $\boldsymbol{b_2}$ and $\boldsymbol{b_3}$ in the regression model $\boldsymbol{y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + e}$ when $\boldsymbol{N = 202}$, $\boldsymbol{SSE = 11.12389}$, $\boldsymbol{r_{23} = -0.114255}$, $\boldsymbol{\sum_{i=1}^{N} (x_{i2} - \bar{x}_2)^2 = 1210.178}$, and $\boldsymbol{\sum_{i=1}^{N} (x_{i3} - \bar{x}_3)^2 = 30307.57}$.

See section 5.3.1 of textbook. We make use of the following formulas for the case where we have two right-hand side variables x_2 and x_3 together with a constant (x_1):

$$var\left(b_2\right) = \frac{\sigma^2}{\left(1 - r_{23}^2\right) \sum_{i=1}^{N} \left(x_{i2} - \bar{x}_2\right)^2}$$

$$r_{23} = \frac{\sum \left(x_{i2} - \bar{x}_2\right)\left(x_{i3} - \bar{x}_3\right)}{\sqrt{\left[\sum \left(x_{i2} - \bar{x}_2\right)^2\right]\left[\sum \left(x_{i3} - \bar{x}_3\right)^2\right]}}$$

Since (a) the order of the variables is arbitrary, and (b) $r_{23} = r_{32}$, we can use the identical formula above for $var\left(b_3\right)$. Also recall that:

$$\hat{\sigma}^2 = \frac{1}{N-K} \sum \hat{e}_i^2 = \frac{1}{N-K} SSE$$

*For errors or corrections, please email me at conor.teaches.econ@gmail.com.

Taking these values together, we have:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{202-3}11.12389 = 0.055899 \\
se(b_2) &= \left(\frac{\hat{\sigma}^2}{(1-r_{23}^2)\sum_{i=1}^{N}(x_{i2}-\bar{x}_2)^2} \right)^{1/2} \\
&= \left(\frac{0.055899}{(1-(-0.114255)^2)(1210.178)} \right)^{1/2} = 0.0068412 \\
se(b_3) &= \left(\frac{\hat{\sigma}^2}{(1-r_{23}^2)\sum_{i=1}^{N}(x_{i3}-\bar{x}_2)^2} \right)^{1/2} \\
&= \left(\frac{0.055899}{(1-(-0.114255)^2)(30307.57)} \right)^{1/2} = 0.0013670
\end{aligned}$$

2 Question 6.3

Consider the model

$$\boldsymbol{y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + e}$$

and suppose that the application of least squares to 20 observations on these variables yields the following results ($\widehat{cov(b)}$ denotes the estimated covariance matrix):

$$\begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{b_2} \\ \boldsymbol{b_3} \end{bmatrix} = \begin{bmatrix} \mathbf{0.96587} \\ \mathbf{0.69914} \\ \mathbf{1.77690} \end{bmatrix}, \quad \widehat{\boldsymbol{cov(b)}} = \begin{bmatrix} \mathbf{0.218120} & \mathbf{0.019195} & \mathbf{-0.050301} \\ \mathbf{0.019195} & \mathbf{0.048526} & \mathbf{-0.031223} \\ \mathbf{-0.050301} & \mathbf{-0.031223} & \mathbf{0.037120} \end{bmatrix}$$

$$\boldsymbol{\hat{\sigma}^2 = 2.5193} \qquad\qquad \boldsymbol{R^2 = 0.9466}$$

(a) **Find the total variation, unexplained variation, and explained variation for this model.**

Using the definition of SSE (unexplained variation) and some algebra, we have:

$$\begin{aligned}
SSE &= \sum_{i=1}^{N} \hat{e}_i^2 \\
&= (N-K)\left(\frac{1}{N-K}\sum_{i=1}^{N} \hat{e}_i^2 \right) \\
&= (N-K)\hat{\sigma}^2 \\
&= (20-3)(2.5193) = 42.8281
\end{aligned}$$

Next, using the definition of R^2, we can calculate SST (total sum of squares/total variation) as:

$$\begin{aligned} 1 - \frac{SSE}{SST} &= R^2 \\ SST - SSE &= (SST)R^2 \\ (1 - R^2)SST &= SSE \\ SST &= \frac{SSE}{1 - R^2} \\ SST &= \frac{42.8281}{1 - 0.9466} = 802.0243 \end{aligned}$$

Finally, we can use the following equation (the accounting identity for SSR, SSE, and SST) to find SSR (regression sum of squares/explained variation) as:

$$\begin{aligned} SSR + SSE &= SST \\ SSR &= SST - SSE \\ SSR &= 802.0243 - 42.8281 = 759.1962446 \end{aligned}$$

Note - numbers may not add due to rounding.

(b) **Find the 95% interval estimates for $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$.**

First, let's find the 95% critical value for a 2-sided test with (20-3) = 17 degrees of freedom. Since this is a 2-sided values, we look for the $\left(1 - \frac{\text{confidence level}}{2}\right)$ value in the t subscript, or $t_{0.975,df}$. Going to the table from the textbook, and using the row for 17 degrees of freedom, we have $t^c = 2.110$.

Also recall that by definition, the diagonal elements of the covariance matrix is the variance of the respective b_k estimate.

Using these values and the typical formula for confidence intervals, we have:

$$\begin{aligned} \beta_2 \text{ Confidence Interval: } &= b_2 \pm (t^c)se(b_2) \\ &= 0.69914 \pm (2.11)(0.048526)^{1/2} = [0.234336, 1.163944] \\ \beta_3 \text{ Confidence Interval: } &= 1.77690 \pm (2.11)(0.03712)^{1/2} = [1.370376, 2.183424] \end{aligned}$$

(c) **Use a $\boldsymbol{t}$-test to test the hypothesis $\boldsymbol{H_0}$: $\boldsymbol{\beta_2 \geq 1}$ against the alternative $\boldsymbol{H_1}$:$\boldsymbol{\beta_2 < 1}$.**

The t-statistic is:

$$tstat = \frac{b_2 - 1}{se(b_2)} = \frac{0.69914 - 1}{(0.048526)^{1/2}} = -1.36577$$

Given that we are doing a LHS test, this is the correct sign for a rejection. Comparing this t-stat to the values in the row for 17 from the table in the textbook, we see that

1.36577 is between the values in the $t_{0.90}$ (1.333) and $t_{0.95}$ (1.740). This means that our p-value for this 1-sided test is between 0.1 and 0.05. We would reject the null at $\alpha = 10\%$ but fail to reject the null at $\alpha = 5\%$.

(d) **Use your answers in part (a) to test the joint hypothesis $\boldsymbol{H_0}$: $\boldsymbol{\beta_2 = 0, \beta_3 = 0}$.**

Since this is a joint hypothesis test, we need to calculate the F-statistic. Since β_2 and β_3 are the only betas in the regression (besides the constant), this exercise is "testing the significance of the model" as discussed in section 6.1.2 of the textbook. That means we can use equation (6.8), which is a specialization of the general equation (6.4) to the case of testing the null that all the betas (except for the constant) are equal to zero.

Plugging into equation (6.8), we have:

$$\begin{aligned} F\text{-stat} &= \frac{(SST - SSE)/(K-1)}{SSE/(N-K)} \\ &= \frac{(802.0243 - 42.8281)/(3-1)}{42.8281/(20-3)} = 150.676 \end{aligned}$$

Strictly speaking, we didn't even need to do the calculations in part (a) to calculate the F-stat. If we rearrange equation (6.8), we get the following expression for the F-stat in when we "test the regression":

$$\begin{aligned} F\text{-stat} &= \frac{(SST - SSE)/(K-1)}{SSE/(N-K)} \\ &= \left(\frac{SST - SSE}{SSE}\right)\left(\frac{N-K}{K-1}\right) \\ &= \left(\frac{SST}{SSE} - 1\right)\left(\frac{N-K}{K-1}\right) \\ &= \left[\left(1 - R^2\right)^{-1} - 1\right]\left(\frac{N-K}{K-1}\right) \end{aligned}$$

Thus, to calculate the F-stat when we are testing the full regression, all we need is R^2, N, and K.

Now we turn to finding the critical value and the p-value for this test. We have 2 numerator degrees of freedom (= # of restrictions) and 17 denominator degrees of freedom (= $N - K$). Looking at tables 4 and 5 of Appendix D, the number of restrictions specifies the column while $N - K$ specifies the row. If $N - K$ is very large, then we can also use the Chi-square table (Table 3) with df equal to the number of restrictions.

Table 4 tells us that the 95% confidence level critical value is between 3.49 and 3.68, while Table 5 tells us that the 99% confidence level critical value is between 5.85 and 6.36. Given that our F-stat of 150.676 is much higher than any of these values, meaning we can reject the null that $\beta_2 = 0$ and $\beta_3 = 0$ with extremely high confidence.

(e) **Test the hypothesis $\boldsymbol{H_0}$: $\boldsymbol{2\beta_2 = \beta_3}$**

This null hypothesis is equivalent to testing:

$$\lambda = 2\beta_2 - \beta_3 = 0$$

This is a case of testing a linear combination of parameters. First, calculating the point estimate for λ, we have:

$$\hat{\lambda} = 2b_2 - b_3 = 2(0.69914) - 1.7769 = -0.37862$$

Next, the estimated variance of $\hat{\lambda}$ is:

$$\begin{aligned}\widehat{var(\hat{\lambda})} &= \left(2^2\right) var(b_2) + var(b_3) + (2)(2)(-1)cov(b_2, b_3)\\ &= 4(0.048526) + 0.037120 - (4)(-0.031223)\\ &= 0.356116\end{aligned}$$

With these two values, we can calculate the t-stat for our test:

$$\begin{aligned}t\text{-stat} &= \frac{\hat{\lambda} - 0}{se(\hat{\lambda})}\\ &= \frac{-0.37862}{(0.356116)^{1/2}}\\ &= -0.63447\end{aligned}$$

Finally, we can conduct our test by comparing this t-stat to the values in the t-distribution table. Recall that we have $N - K = 17$ degrees of freedom. Going across the $df = 17$ row in Table 2, we see that 0.63447 is less than even the smallest value in the row for 17 (1.333). This means that we fail to reject the null that $2\beta_2 = \beta_3$.