CHAPTER 3
THE TWO-VARIABLE MODEL: HYPOTHESIS TESTING

QUESTIONS

3.1. (***a***) In the regression context, the method of least squares estimates the regression parameters in such a way that the sum of the squared difference between the actual *Y* values (i.e., the values of the dependent variable) and the estimated *Y* values is as small as possible.

(***b***) The estimators of the regression parameters obtained by the method of least squares.

(***c***) An estimator being a random variable, its variance, like the variance of any random variable, measures the spread of the estimated values around the mean value of the estimator.

(***d***) The (positive) square root value of the variance of an estimator.

(***e***) Equal variance.

(***f***) Unequal variance.

(***g***) Correlation between successive values of a random variable.

(***h***) In the regression context, TSS is the sum of squared difference between the individual and the mean value of the dependent variable *Y*, namely, $\sum (Y_i - \bar{Y})^2$.

(***i***) ESS is the part of the TSS that is explained by the explanatory variable(s).

(***j***) RSS is the part of the TSS that is not explained by the explanatory variable(s), the *X* variable(s).

(***k***) It measures the proportion of the total variation in *Y* explained by the explanatory variables. In short, it is the ratio of ESS to TSS.

(***l***) It is the standard deviation of the *Y* values about the estimated regression line.

STUDENTS-HUB.com Uploaded By: anonymous

(***m***) BLUE means best linear unbiased estimator, that is, a linear estimator that is unbiased and has the least variance in the class of all such linear unbiased estimators.

(***n***) A statistical procedure of testing statistical hypotheses.

(***o***) A test of significance based on the *t* distribution.

(***p***) In a one-tailed test, the alternative hypothesis is one-sided.

For example: $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$ or $\mu < \mu_0$, where μ is the mean value.

(***q***) In a two-tailed test, the alternative hypothesis is two-sided.

(***r***) It is a short-hand for the statement: reject the null hypothesis.

3.2. (***a***) *False*. It minimizes the sum of residuals squared, that is, it minimizes $\sum e_i^2$.

(***b***) *True*.

(***c***) *True*.

(***d***) *False*. The OLS does not require any probabilistic assumption about the error term in estimating the parameters.

(***e***) *True*. The OLS estimators are linear functions of u_i and will follow the normal distribution if it is assumed that u_i are normally distributed. Recall that any linear function of a normally distributed variable is itself normally distributed.

(***f***) *False*. It is ESS / TSS.

(***g***) *False*. We should reject the null hypothesis.

(***h***) *True*. The numerator of both coefficients involves the covariance between *Y* and *X*, which can be positive or negative.

(***i***) *Uncertain*. The *p value* is the exact level of significance of a computed test statistic, which may be different from an arbitrarily chosen level of significance, α.

3.3. (***a***) *t* (***b***) se(b_2) (***c***) 0 and 1 (***d***) -1 and +1

(***e***) ESS (***f***) ESS (***g***) the standard error of the estimate

(***h***) $\sum (Y_i - \bar{Y})^2$ (***i***) $b_2^2 \sum x_i^2 + \sum e_i^2$

STUDENTS-HUB.com Uploaded By: anonymous

3.4. The answers to the missing numbers are in boxes:

$$\hat{Y}_i = -66.1058 + 0.0650\, X_i \qquad r^2 = 0.9460$$

$$\text{se} = (10.7509) \quad (0.0035) \qquad n = 20$$

$$t = (-6.1489) \quad (18.73)$$

The critical t value at the 5% level for 18 d.f. is 2.101 (two-tailed) and 1.734 (one-tailed). Since the estimated t value of 18.73 far exceeds either of these critical values, we reject the null hypothesis. A two-tailed test is appropriate because no *a priori* theoretical considerations are known regarding the sign of the coefficient.

3.5. $r^2 = (\sum y_i^2 - \sum e_i^2) / \sum y_i^2$

$= \sum \hat{y}_i^2 / \sum y_i^2$

$= b_2^2 \sum x_i^2 / \sum y_i^2$, following Equations (3.34) and (3.35)

In proving the last equality, note that $\sum y_i \hat{y}_i = b_2 \sum y_i x_i$. Then the result follows by substitution.

3.6. $\sum e_i = n\,\bar{Y} - n\,(\bar{Y} - b_2\bar{X}) - n\,b_2\bar{X} = 0$. See also Problem 2.22.

PROBLEMS

3.7. **(*a*)** The d.f. here are 14. Therefore, the 5% critical t value is 2.145. So, the 95% confidence interval is:

$$3.24 \pm 2.145(1.634) = (-0.2649, 6.7449)$$

(*b*) The preceding interval does include B_2. Therefore, do not reject the null hypothesis.

(*c*) $t = 3.24 / 1.634 = 1.9829$. Since car sales are expected to be positively related to real disposable income, the null and alternative hypotheses should be: $H_0 : B_2 \leq 0$ and $H_1 : B_2 > 0$. Therefore, an one-tailed t test is appropriate in this case. The 5% one-tailed t value for 14 d.f. is 1.761. Since the computed t value of 1.9829 exceeds the critical value, reject the null hypothesis (one- and two tail tests sometimes give different results).

STUDENTS-HUB.com Uploaded By: anonymous

3.8. (***a***) The slope coefficient of 1.0598 means that during the 1956 –1976 period a percentage point increase in the market rate of return lead to about 1.06 percent points increase in the mean return on the IBM stock. In the same period, if the market rate of return were zero, the average rate of return on the stock would have been about 0.73 percent, which may not make economic sense.

(***b***) About 47 percent of the variation in the mean return on the IBM stock was explained by the (variation) in the market return.

(***c***) $H_0 : B_2 \leq 1$, $H_1 : B_2 > 1$. Hence:

$$t = \frac{(1.0598 - 1)}{0.0728} = 0.8214.$$

For 238 d.f, this t value is not statistically significant at the 5% level on the basis of the one-tailed t test. Thus, during the study period, the *beta coefficient* of IBM was not statistically different from unity, suggesting that the IBM stock was not volatile or aggressive.

3.9. (***a***) $b_1 = 21.22;\ b_2 = 0.5344$

(***b***) $\text{se}(b_1) = 8.5894;\ \text{se}(b_2) = 0.0484$

(***c***) $r^2 = 0.9385$

(***d***) 95% CI for B_1: 1.4128 to 41.0272

95% CI for B_2: 0.4228 to 0.6460

(***e***) Reject H_0, since the preceding CI does not include $B_2 = 0$.

3.10. (***a***) The answers to the missing numbers are in boxes:

$$\text{GNP}_t = -995.5183 + 8.7503\, M_{1t} \qquad r^2 = 0.9488$$

$$\text{se} = (\boxed{260.2128}) \quad (0.3214)$$

$$t = (-3.8258)\ (\boxed{27.2230})$$

(***b***) $H_0 : B_2 \leq 0$, $H_1 : B_2 > 0$. The null hypothesis can be rejected.

(***c***) No particular economic meaning can be attached to it.

(***d***) $\hat{\text{GNP}}_{2007} = -995.5183 + 8.7503(750) \approx 5{,}567$ billion.

STUDENTS-HUB.com Uploaded By: anonymous

3.11. (***a***) Negative.

(***b***) Yes. Here, n = 14 (14 presidential elections starting in 1928 and ending in 1980) and therefore d.f. = 12. The computed t value of -2.67 is statistically significant at the five percent level (one-tailed test).

(***c***) Probably. But in the 1984 elections the personal popularity of Ronald Reagan was an important factor.

(***d***) Since $t = b_i / \text{se}(b_i)$ under the null hypothesis that the true B_i is zero, $se(b_i) = \dfrac{b_i}{t}$. In the present example these standard errors are 1.5572 and 0.6367, respectively.

3.12. (***a***) It could be negative or positive. As more output is produced as a result of increased capacity, price increases (i.e., inflation) will slow down. However, if capacity utilization is at its optimal value, and if demand pressures continue, inflation may actually rise.

(***b***) The output in *EViews* format is as follows:

Dependent Variable: INFLATION
Sample: 1960 2007

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.953476	7.762046	0.766998	0.4470
CAPACITY	-0.021545	0.095918	-0.224615	0.8233
R-squared	0.001096			

(***c***) The estimated slope coefficient is negative but also statistically insignificant, for the estimated p value is quite high.

(***d***) Yes it is, for under the null hypothesis that the true slope coefficient is 1, the estimated t value is

$$t = \frac{-0.0215 - (1)}{0.0959} = -10.652$$

The probability of obtaining such a t value is practically zero.

(***e***) To get this, solve 5.9535 – 0.0215C = 0, which gives $C \approx 276.91$, which may be called the "natural" rate of capacity utilization.

STUDENTS-HUB.com Uploaded By: anonymous

Note: The results of the above regression are virtually insignificant. Plus, the "natural" rate of capacity utilization that was found to be 276.91 may be problematic because the measure of capacity utilization does not exceed 100. The reason for the regression breakdown is the fact that the data include the decade of the 1970s with its high rates of inflation and the mid 1970s stagflation. Running the regression over a period that excludes the 1970s, say 1982-2001, will produce more reasonable and statistically significant results. In fact, if the regression covers the 1982-2001 period, the reader can easily verify that the "natural" rate of capacity utilization is approximately 93.90.

3.13. ***(a)*** The *EViews* regression results are as follows:

Dependent Variable: CAPACITY
Sample: 1960 2007

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	81.0226	1.147241	70.6238	0.0000
INFLATION	-0.05085	0.226396	-0.22462	0.8223
R-squared	0.00011			

Note: This regression is also insignificant for the reason discussed above.

(*b*) Multiplying the two slope coefficients, we obtain the value of 0.0011 which is equal to the R^2 value obtained from either equation. This result is not surprising in view of Problem 2.21.

(*c*) By way of another example, let Y = salary and X = qualifications for a group of men and women. As Maddala notes, the direct regression will answer the question whether men and women with the same X value get the same Y value. The reverse regression will answer the questions whether men and women with the same Y value will have the same X value. Reverse regression is advocated for wage discrimination cases.

(*d*) No.

3.14. **(*a*)** Positive.

(*b*) and **(*c*)** The scattergram will show that the relationship between the two is generally positive, although there are a few outliers.

STUDENTS-HUB.com Uploaded By: anonymous

(***d***) The regression results are as follows:

$$\hat{Y}_t = \quad 373.3014 \quad + \quad 0.4199\, X_t$$

$$\text{se} = \ (9530.3786) \quad (0.1154)$$

$$t = \quad (0.0392) \qquad (3.6406) \qquad r^2 = 0.5464$$

(***e***) 99% CI: $0.0615 \leq B_2 \leq 0.7783$.

Since the preceding interval does not include zero, we can reject the null hypothesis.

3.15. (***a***) The regression results are:

$$\widehat{\text{MATHM}}_t = 198.7370 + 0.6705\ \text{MATHFM}_t$$

$$\text{se} = (12.8754) \quad (0.0265)$$

$$t \ = (15.4354) \quad (25.3325) \qquad r^2 = 0.9497$$

(***b***) Reject the null hypothesis, since the computed t value of 25.3325 far exceeds the critical value even at the 0.001 level of significance.

(***c***) $\widehat{\text{MATHM}}_{2008} = 527.282 \approx 527$

(***d***) CI: (526.7373, 527.8090)

3.16. (***a***) The regression results are as follows:

$$\widehat{\text{MaleCR}}_t = 132.7778 + 0.75\ \text{FemaleCR}_t$$

$$\text{se} = (33.7245) \quad (0.0670)$$

$$t \ = (3.9371) \ (11.1873) \qquad r^2 = 0.7864$$

(***b***) Reject the null hypothesis, since the computed t value is very high.

(***c***) $\widehat{\text{MaleCR}}_{2008} \approx 511.528$

(***d***) CI: (510.5676, 512.4879)

3.17. (***a***) There is a positive relationship between real return on the stock price index this year and the dividend price ratio last year: Per unit increase in the latter, the mean real return goes up by 5.26 percentage points. The intercept has no viable economic meaning.

(***b***) If the preceding results are accepted, it has serious implications for the efficient market hypothesis of modern finance.

STUDENTS-HUB.com Uploaded By: anonymous

3.18 **(*a*)** The *EViews* regression output is as follows:

Dependent Variable: AVGHWAGE Sample: 1 13				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.014453	0.874624	-0.016525	0.9871
YEARSSCH	0.724097	0.069581	10.40648	0.0000
R-squared	0.907791			

(*b*) On the basis of the *t* test this hypothesis can be easily rejected, for the computed *t* value is highly significant; its *p* value is practically zero.

(*c*) Here $t = \dfrac{0.7240 - 1}{0.0695} = -3.9712$. This *t* value is also highly significant, leading to the conclusion that the education coefficient is statistically different from 1. The *p* value of obtaining the computed *t* value is 0.0011 (two-tail test).

3.19 *Note*: This Problem is an extension of Problem 2.17.

(*a*) Based on the regression, we need to calculate new variables based on the real GDP (RGDP) and the unemployment rate (UNRATE). These calculations, based on the data in Table 2-13, are as follows:

CHUNRATE = Change in UNRATE = UNRATE – UNRATE(-1)

PCTCRGDP = % Change in RGDP = [RGDP / RGDP(-1)]*100-100

Using *EViews*, the regression results are:

Dependent Variable: PCTCRGDP Sample (adjusted): 1960 2006				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.31911	0.166734	19.90654	0.0000
CHUNRATE	-1.86297	0.185785	-10.02758	0.0000
R-squared	0.6956			

Note: The sample is adjusted to start in 1961 instead of the initial observation of 1960 because we are calculating percentage changes (RGDP) and changes (UNRATE): This causes the loss of the first observation.

STUDENTS-HUB.com Uploaded By: anonymous

(***b***) Yes, for the estimated slope coefficient has a *t* value of -10.028 whose *p* value is practically zero.

(***c***) The intercept term is also statistically significant. The interpretation here is that if the change in the unemployment rate were zero, the growth in the real GDP will be about 3.3%, which may be called the long-term, or steady-state, rate of growth of GDP.

3.20 The regression results, using *EViews*, are:

Dependent Variable: SP500
Sample: 1980 1999

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	404.4067	128.6475	3.1435	0.0041
1/MTB3	996.8656	404.2324	2.4661	0.0206
R-squared	0.273968			

As these results show, the slope coefficient is statistically significant at about the 2% level; the intercept is also significant at a 0.4% level. Any minor differences with the regression shown in the text are solely due to rounding.

3.21. The *EViews* regression results for (2.27) are as follows:

Dependent Variable: PRICE
Sample: 1 32

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-191.6662	264.4393	-0.724802	0.4742
AGE	10.48562	1.793729	5.845711	0.0000
R-squared	0.532509			

The estimated slope coefficient is highly statistically significant, for the *p* value of obtaining a *t* statistic of 5.8457 or greater under the null hypothesis of a zero true population slope coefficient is practically zero. In contrast, the estimated intercept coefficient is statistically insignificant since its *p* value is relatively high.

Likewise, the *EViews* results of regression (2.28) are:

STUDENTS-HUB.com Uploaded By: anonymous

Dependent Variable: PRICE Sample: 1 32				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	807.9501	231.0921	3.496226	0.0015
NOBIDDERS	54.57245	23.26605	2.345582	0.0258
R-squared	0.154971			

Here both the coefficients are individually statistically significant, as their *p* values are quite low.

3.22. *Note*: The regression results presented here are identical to those of Problem 2.16.

(*a*) The results, using *EViews*, are as follows:

Dependent Variable: ASP Sample: 1 64				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-88220.49	76638.60	-1.1511	0.2541
GPA	55227.44	22697.53	2.4332	0.0179
R-squared	0.0872			

As these results suggest, GPA has a positive impact on ASP, and it is statistically very significant, as the *p* value of the estimated coefficient is very small.

(*b*) The results for GMAT are as follows:

Dependent Variable: ASP Sample: 1 65				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-241386.6	29464.86	-8.19236	0.0000
GMAT	511.7207	44.35705	11.5364	0.0000
R-squared	0.6822			

These results show that GMAT has a positive and statistically significant impact on ASP.

(*c*) The results for annual tuition are as follows:

STUDENTS-HUB.com Uploaded By: anonymous

Dependent Variable: ASP Sample: 1 65				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	42878.33	5502.0635	7.79314	0.0000
TUITION	1.634784	0.156924	10.4177	0.0000
R-squared	0.6364			

Tuition (perhaps reflecting the quality of education) has a positive and statistically significant impact on ASP.

Incidentally, it can also be shown that the impact of recruiter rating has a positive and highly significant impact on ASP, as it can be seen from the following *EViews* output:

Dependent Variable: ASP Sample: 1 65				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-29943.60	10973.495	-2.72872	0.0089
RECRUITER	37300.30	3020.5187	12.34897	0.0000
R-squared	0.7644			

3.23. The regression results of expenditure on imported goods (Y) and personal disposable income (X), using *EViews*, are as follows:

Dependent Variable: Y Sample: 1959 2006				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-136.1649	23.56509	-5.77825	0.0000
X	0.208248	0.005467	38.0911	0.0000
R-squared	0.9693			

These results suggest that personal disposable income has a very significant positive impact on expenditure on imported goods, an unsurprising finding. The p value for the slope is virtually zero, and the null hypothesis is therefore rejected.

3.24. If we let $w_i = \dfrac{x_i}{\sum x_i^2}$, we can write $b_2 = \sum w_i Y_i$, that is, b_2 is a linear estimator, i.e., a linear function of the Y values. Note that we are treating

STUDENTS-HUB.com Uploaded By: anonymous

X as non-stochastic. Follow similar steps to show that b_1 is also a linear function of the Y values.

Now:

$$b_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} = \frac{\sum x_i (B_1 + B_2 X_i + u_i)}{\sum x_i^2}$$

$$= B_1 \frac{\sum x_i}{\sum x_i^2} + B_2 \frac{\sum x_i X_i}{\sum x_i^2} + \frac{\sum x_i u_i}{\sum x_i^2}$$

$$= B_2 + \frac{\sum x_i u_i}{\sum x_i^2}$$

This is in view of the fact that $\sum x_i = \sum (X_i - \overline{X}) = 0$ and $\frac{\sum x_i X_i}{\sum x_i^2} = 1$.

Therefore, $E(b_2) = E\left[B_2 + \frac{\sum x_i u_i}{\sum x_i^2} \right] = B_2$

Note: $E\left(\frac{\sum x_i u_i}{\sum x_i^2} \right) = \frac{1}{\sum x_i^2} E(\sum x_i u_i)$, since $\sum x_i^2$ is a constant and since X and u are uncorrelated by OLS assumption. Follow similar steps to prove that b_1 is also unbiased.

3.25. Squaring Equation (7.33) and summing, we obtain:

$$\sum y_i^2 = b_2^2 \sum x_i^2 + \sum e_i^2 + 2 b_2 \sum x_i e_i$$

$$= b_2^2 \sum x_i^2 + \sum e_i^2$$

since $\sum x_i e_i = 0$.

STUDENTS-HUB.com Uploaded By: anonymous