Management
UNIVERSITY OF TORONTO
SCARBOROUGH

MGEB12: Quantitative Methods in Economics-II

Problem Set-6

Chapter 15: 2,4,12,14,15,19,22,23,28,29,32, 34,38

Supplemental

1. The following is the partial results of a multiple regression

ANOVA

Source	DF	SS	MS	F
Regression	2	?	?	?
Error	?	2259	?	
Total	47	?		

Parameter Estimates

Variable	Coefficient	SE	t	P-value
Intercept	-12.65	2.35	?	
X1	?	1.29	2.41	
X2	10.1	3.755	?	

The p-value for the estimated slope parameter on X_2 is:

A. 5%
B. 4%
C. 3%
D. 2%
E. 1%

2. For the following regression equation, $\hat{y} = 15 + 6x_1 + 5x_2 + 4x_1x_2$ and $x_2 = -1.5$ a unit increase in x_1 is associated with average changes in the value of $\hat{y}$ equal to:

A. 0
B. 1
C. -6
D. -1
E. -3

Questions 3– 4: The following regression is based on a sample of 59 homes. The dependent variable is the sale price of the home (in thousands of dollars). The independent variables measure characteristics of each home. The variables are:

Price: price of the house (in thousands of dollars)
Age: age of the house (in number of years)
Ac: (1 if central air, 0 otherwise)
Pool: (1 if with pool, 0 otherwise)
Sqft: area of the house (in square feet)
View: (1 if coastal view, 0 otherwise)

Variable	Coefficient	Standard Error
Intercept	72.91	33.14
Age	-2.57	1.06
AC	30.21	25.41
Pool	35.17	21.15
Sqft	0.11	0.22
View	41.32	18.42
View*Pool	3.17	1.41
Adj-R^2=46.21%		

3. The F-statistic for the significance of regression is approximately?

A. 6.72
B. 7.45
C. 8.17
D. 9.30
E. 10.70

4. What would be the expected difference in price between a 2,000-square foot home with central air conditioning and a 1,800-square foot home with no central air conditioning (given that the values for the other *X* variables are the same for the two homes)?

A. 10,238
B. 30,210
C. 40,196
D. 52,210
E. 57,471

5. The following relationship between executive salaries at a firm, the years they were employed, and their gender is estimated:

$$\hat{Y} = 20{,}000 + 999X_1 + 1000X_2$$

Y: Dollar salary
X_1: Years employed
X_2: Gender (=1 if male and =0 if female)

What is the correct interpretation of these results?

A. Females and males with similar seniority are predicted to receive $21,999
B. Being a female, versus a male with similar seniority, on average means an additional $1000
C. Being a male, versus a female with similar seniority, on average means an additional $1000
D. Females and males with no seniority can be expected to receive $20,000
E. On average executives with 10 years of experience earn $29,990

Questions 6 – 8: Consider a regression model to explain the number of students enrolled in colleges. Explanatory variables include the percent female enrollment (PercFem), the percent of students from within province (PercProv), the rank of the college (Tier_Two = 1 if in second tier, 0 otherwise; Tier_Three = 1 if in third tier, 0 otherwise), whether the college offers a business degree (Business = 1 if business degree is offered, 0 otherwise), and whether the college is located in the East (East = 1 if in the East, 0 otherwise). The sample size is 93.

	Coefficient	SE
Intercept	2013.0	227.5
PercFem	6.13	2.60
PercProv	-3.82	1.12
Tier_Two	-156.5	63.5
Tier_Three	-307.2	105.7
Business	163.8	68.0
East	350.9	445.9

6. The coefficient on which variable is NOT statistically significant?

A. PercFem
B. PercProv
C. Tier_Two
D. East
E. None of the above

7. What is the predicted enrollment for a first tier college in the East with 60 percent females and 40 percent from within province that does not offer a business degree?

A. 566
B. 943
C. 1,936
D. 2,579
E. 3,101

8. What is the predicted difference in enrollment between a comparable second tier and third tier college?

A. The second tier college is predicted to enroll 151 more students
B. The second tier college is predicted to enroll 157 less students
C. The second tier college is predicted to enroll 307 more students
D. The second tier college is predicted to enroll between 157 and 307 more students
E. None of the above

Questions 9-10: A researcher is studying factors related to income. The researcher has 76 observations of individual income measured in 1000's of dollars per year INCOME) and some other variables. Two of the variables in the data available to the researcher are CANCIT and NCANCIT. CANCIT = 1 if person is a Canadian citizen and 0 otherwise. NCANCIT = 1 if person does not have Canadian citizenship and 0 otherwise. The researcher considers the following simple linear regression model:

$INCOME_i = \alpha + \beta * CANCIT_i + \varepsilon_i$

The following simple linear regression results are obtained. INCOME_HAT is the predicted (fitted) value of income and the standard error of the slope coefficient is in parentheses.

$$\text{INCOME\_HAT} = 35.07 + \underset{(2.39)}{4.51} * \text{CANCIT}$$

9. What would the estimates be if instead of CANCIT the researcher includes NCANCIT?

A. INCOME_HAT = 39.58 - 4.51*NCANCIT
B. INCOME_HAT = 39.58 - 0.22*NCANCIT
C. INCOME_HAT = 35.07 - 4.51*NCANCIT
D. INCOME_HAT = 35.07 + 3.51*NCANCIT
E. INCOME_HAT = 35.07 + 4.51*NCANCIT

10. Consider whether there is a statistically significant difference in income between Canadian and non-Canadian citizens. Which is the most plausible conclusion about the difference?

A. The difference is statistically significant at the 1% level
B. The difference is statistically significant at the 5% level
C. The difference is statistically significant at the 10% level
D. The difference is not statistically significant
E. There is no difference

Questions 11 – 12: Consider a multiple regression that investigates the relationship between the price of a one-night-stay in a hotel and the amenities. The data is a randomly selected cross-section of 278 hotels. The dependent variable is the price of a one-night-stay in dollars. The table below reports the coefficient estimates and the standard errors (se) of the coefficient estimates in parentheses along with descriptions of all of the included regressors.

Variable	Estimate (se)	Variable Description
constant	-10.58 (1.22)	Constant term (intercept)
square_met	1.25 (0.99)	Size of room measured in square meters
ac	19.16 (5.28)	= 1 if room has air conditioning, = 0 otherwise
air_km	-0.36 (0.09)	Kilometers drive from nearest major airport
air_km_2	-0.08 (0.01)	= air_km squared
stars	10.89 (2.58)	Number of stars (0 – 5) hotel awarded by a travel guide
stars_2	2.44 (0.78)	= stars squared
pool	-2.30 (3.25)	= 1 if hotel has a swimming pool, = 0 otherwise
down	14.23 (4.26)	= 1 if hotel located downtown, = 0 otherwise
stars*down	5.69 (1.67)	= stars * down

11. How much of a price differential is typically observed between a 4-Star downtown hotel and a 4-Star non-downtown hotel (other things equal, which means that other independent variables in the model are held constant)?

A. $14.23
B. $19.92
C. $22.76
D. $26.41
E. $36.99

12. How much of a price differential is typically observed between a hotel with air conditioning but no pool and one without air conditioning but with two pools (other things equal, which means that other independent variables in the model are held constant)?

A. $14.56
B. $16.86
C. $19.16
D. $21.46
E. $23.76

Long Questions

Question-1: The objective of a study is to produce a multiple regression model to predict sales of cotton fabric. The explanatory variables are

X_1	Whole sale price index
X_2	Quantity of Imported Fabric
X_3	Quantity of Exported Fabric
X_4	Time

Part of a computer output from the estimated regression based on 28 observations is shown below:

Predictor	Coeff	StdDev
Constant	8876	2295
X_1	-24	25
X_2	-6	2.5
X_3	0.5	0.2
X_4	-63	70

Analysis of Variance

Source	SS
Regression	21080
Error	1426
Total	22506

a) Write down the estimated regression equation and interpret the estimated regression coefficients for X_2 and X_3.
b) Test if the overall regression model is significant or not using a 0.05 level of significance.
c) Test the significance of each regression coefficient. Use significance level 0.05.
d) Suppose we drop the variable X_4 from the model. The reduced regression model has X_1, X_2 and X_3 as explanatory variables. The reduced model has R^2=0.9. Test if the reduced regression model is significant or not at significance level 0.05.
e) Suppose you believe that the quantity of cotton fabric sold may also be impacted by how cold the weather is. Furthermore, suppose you have classified the weather in three categories, cold, normal, warm. Design a regression model that allows for the new variables and show how you test your hypothesis.

Question-2: You are interested in determining what factors influence the price of a stock. After some examination, a statistician hypothesized that a stock price (Y in $) would be affected by its quarterly dividends (X1 in $), its price/earnings ratio (X2), and the interest rate of treasury bills (X3 in %). The values of the relevant variables were observed for a period of 40 quarters. When the data were run on Excel, the following partial printout was created.

Multiple Regression Analysis

Parameter	Estimate	SE	t-Statistic
Constant	17.3925	5.5254	?
X1	?	7.5016	5.5054
X2	-0.4158	0.5228	?
X3	0.57091	?	1.3982

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio
Model	?	?	?	?
Residual	180.095	?	?	
Total	697.040	?		

R-squared = ?
Adjusted R-Squared = ?

a) What is the value of b1? Carefully explain what this number means.
b) What is the value of SSR? Carefully explain what this number means.
c) Test at the 5% level of significance the hypothesis that there is a linear and negative relation between price/earning (X2) and the stock price. State your conclusions in the context of the problem.
d) Test the following hypothesis, use a=.99. State your conclusions in the context of the problem.

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

Question-3: You are interested in the determinants of labour income. You obtained a random sample of workers and collected data on:

(1) income earned by the individual, measured in thousands of dollars (INCOME);
(2) the number of years the individual was in school (EDUC);
(3) the number of years of the individual's job experience (JOBEXP)
(4) the individual's age, in years (AGE).

You used Excel to analyze the relationship among these variables, the partial output of which appears below.
Anova

Source	DF	SS	MS	F
Regression	3	1720	?	?

Error	?	?	?
Total	19		

R-square = 0.9448
Adj R-sq = ?

Variable	Estimates	SE	t
Intercept	-2.98	2.36	?
EDUC	?	0.13	15.82
JOBEX	1.20	0.39	?
AGE	-0.31	?	-5.38

a)What is the estimated regression equation? Provide an interpretation of the estimated slope coefficient on EDUC.
b) Does job experience have a significant effect on the Income at the 1% level?
b1) Does Job experience have a significant and positive effect on the Income at 1% level?
b2) Your friend claims that 1 year addition to the experience increases the income by more than $1000. Use the regression to test this hypothesis at 1% level.
c) Is the multiple regression significant at the 1% level?
d) When the statistician estimated a previous regression with INCOME as the dependent variable and just EDUC and JOBEXP as the independent variables, he obtained an adjusted R2 of 0.8342. Was the regression improved by the addition of AGE as the third independent variable?

Question-4: Data on car theft rates (annual percentage), unemployment rates (annual percentage) is collected for 21 years. Below is the Excel output for estimating car theft rates with unemployment (X_1) and time trend (X_2) as the independent variables. [Note: The time tend is 1 for the first year, 2 for the second year,…, and 21 for the last]

Regression Results

Multiple R	0.5807			
R Square				
Adjusted R Square				
Standard Error	2.9327			
Observations	21			
		ANOVA		
	df	SS	MS	F
Regression				
Residual		154.81		
Total	20			
		Estimates		
	Coefficients	Standard Error	t Stat	P-value
Intercept	6.7131	3.179	2.1117	
Unemployment rate	1.6188	0.5366		
Time trend	-0.4115	0.2373		

a) Write out the regression equation. Carefully interpret the meaning of the coefficients in the regression.
b) Compute the adjusted coefficient of determination $\overline{R}^2$, and explain what it means.

c) What is the sum of the squares for the regression, SSR? Is the Model significant at 1%? What about 5%?
d) Find the t-statistic for and the p-value of (X_2)? Is X_2 significant? Why? How about X_1?
e) After adding another independent variable (X_3) to the model you have re-estimated the regression. R^2 of this new regression is 0.37. Did the addition of this new variable improve the model?

MC(Solutions)

1. E
2. A
3. D
4. D
5. C
6. D
7. D
8. A
9. A
10. C
11. E
12. D

Long Questions (Solutions)

Question-1:

a) Let Y be the sales of cotton fabric. The estimated regression equation is

$$Y = 8876 - 24X_1 - 6X_2 + 0.5X_3 - 63X_4$$

The regression coefficient for X_2 is -6. It means that the value of Y deceases (increases) by 6 units when the value of X increases (decreases) by 1 unit; when all other explanatory variables are kept at fixed levels.

The regression coefficient for X_3 is 0.5. It means that the value of Y deceases (increases) by 0.5 units when the value of X decreases (increases) by 1 unit; when all other explanatory variables are kept at fixed levels.

b)

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$; $H_1:$ Not all $\beta_i = 0,\ i = 1,2,3,4$

At the 0.05 significance level, reject H_0 if $F \geq 2.8$, do not reject H_0 if $F < 2.8$.

The ANOVA table is

Source	SS	df	MS	F
Regression	21080	4	5270	85
Error	1426	23	62	
Total	22506	27		

Since F = 85 falls into the rejection region. Reject H_0 and conclude that the model is significant.

c)

$H_0: \beta_i = 0,\ i = 1,2,3,4$; $H_1: \beta_i \neq 0,\ i = 1,2,3,4$.

At the 0.05 significance level, reject H_0 if $t \geq 2.069$ or $t \leq -2.069$. Do not reject H_0 if $-2.069 < t < 2.069$.

Predictor	Coeff	StdDev	t
Constant	8876	2295	3.8675
X_1	-24	25	-0.96
X_2	-6	2.5	-2.4
X_3	0.5	0.2	2.5
X_4	-63	70	0.9

From the above table, regression coefficients for X_2 and X_3 are significant. The regression coefficients for X_1 and X_4 are not significant.

d) The reduced model has 3 explanatory variables. At 0.05 significance level, reject H_0 if $F \geq 3.01$, do not reject H_0 if $F < 3.01$.

The test statistic is $F = \frac{MSR}{MSE} = \frac{n-1-k}{k}\frac{R^2}{1-R^2} = \frac{28-1-3}{3}\frac{0.9}{1-0.9} = 72$

This falls into the rejection region. The conclusion is that the reduced model is significant.

e) You need to have two indicator variables. You can choose any two. Let's choose cold and warm as the two indicator variables:

I_1= 1, if warm, 0 if not
I_2= 1, if cold, 0 if not

Estimate a regression that includes I_1, and I_2 as well the other variables. Do a t test in line with what you did in part (c) Test whether the coefficients of these two variables are zero or not. I.e.

$H_0 : \beta_i = 0,\ \ i = 5,6$; $H_1 : \beta_i \neq 0,\ \ i = 5,6$

If the weather has no impact neither of the null hypotheses should be rejected.

Question-2:

a)

t = (b1-B1) / SEb1: 5.5054 = (b1-0) / 7.5016
b1 = 41.30

Meaning: For each $1.00 increase in quarterly dividends, the stock price is expected to be $41.30 higher.

b)

SSR + SSE = SST; SSR + 180.095 = 697.040;
SSR = 516.945

Meaning: This means that of the total variation in stock price (697.040), 516.945 of this amount is explained by the regression model. Or put another way, the percentage of the total variation in stock price that is explained by the model is 516.945 / 697.040 = .7416 = 74.16% = R2.

$$\text{Adjusted - } R^2 = 1-(1-R^2)\frac{n-1}{n-k-1} = 1-(1-0.7416)\frac{40-1}{40-3-1} \approx 0.7202$$

c)

test statistic: t = (b_2 – B_2) / SE(b_2)= (-.4158 -0) / .5228 = -.7953

Critical Value: t(a, dof=36) = -1.6883
Since the test statistic (absolute value) is not greater than the critical value, we cannot reject the null hypothesis. "Holding quarterly dividends and interest rates constant, at a 5% level of significance, there is insufficient statistical evidence to conclude that the price earnings ratio has a negative and linear relation to stock price,

d)

Test statistic: F = MSR/MSE = 172.315 / 5.003 = 34.44
MSR = 516.945 / 3 = 172.315, MSE = 180.095 / 36 = 5.003

Alternatively:

$$F_{k,n-k-1} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}} = \frac{\frac{0.7416}{3}}{\frac{0.2584}{36}} = 34.44$$

Critical Value: F(=.01; numerator dof = 3; denominator dof = 36) = 4.4
Because the test statistic is greater than the critical value, we reject the null hypothesis, and conclude the alternative hypothesis. At a 1% level of significance, there is sufficient statistical evidence to conclude that the price earnings ratio and/or the dividend and/or interest rates have a relation to stock prices. [i.e. the regression is valid {significant)]

Question-3

a)

$$\widehat{Income} = -2.98 + 2.06EDUC + 1.2JOBEX - 0.31AGE$$

EDUC: Everything equal one more year of school is associated with $2060 increase in income.
JOBEX: Everything equal one more year of experience is associated with $1200 increase in income.
AGE: Everything equal one more increase in age is associated with $310 decline in income.
Constant: Has no meaning here because EDUC, JOBEX, and AGE all together cannot be zero.

b)

t-stat from the table is:
$t_{0.005,\,16} = 2.921$
$t_{0.01,\,16} = 2.583$

$$t = \frac{1.2}{0.39} = 3.08$$

3.08> 2.291. Reject the null. Experience have a significant effect.

b1)

$$t = \frac{1.2}{0.39} = 3.08$$

3.08> 2.583. Reject the null. Experience have a significant effect.

b2)

$$H_o: \beta_2 = 1$$

$$H_1: \beta_2 > 1$$

$$t = \frac{1.2 - 1}{0.39} = 0.513$$

0.513> 2.583. Cannot reject the null.

c)

$$F_{3,16} = \frac{MSR}{MSE} = \frac{R^2}{1 - R^2}\left(\frac{n - k - 1}{k}\right) = \frac{0.9448}{1 - 0.9448}\left(\frac{16}{3}\right) = 91.29$$

The F from the table 6.3. Therefore, reject the null, regression is significant.

d)

The Adjusted R^2 for the initial model is:

$$AdjR^2 = 1 - \frac{\frac{SSE}{(n - k - 1)}}{\frac{SST}{n - 1}} = 1 - (1 - R^2)\left(\frac{n - 1}{n - k - 1}\right) = 0.93445$$

The Adjusted R^2 for the new model is 0.8342. Conclusion: The addition of AGE improved the model.

Question-4:

a)

$$\widehat{Theft} = 6.7131 + 1.6188 Unemployment\ Rate - 0.4115 Time\ Trend$$

Unemployment Rate: Everything equal one percent increase in Unemployment Rate is associated with 1.6188 percent increase in Theft rate.
Time Trend: Everything equal there is an annual decline in Theft rate of 0.,4115%.
Constant: Has no meaning here because Unemployment Rate, and Time Trend both together cannot be zero.

b)

$$AdjR^2 = 1 - (1 - 0.5807^2)\left(\frac{20}{18}\right) = 0.2636$$

Adjusted R^2 is R adjusted for the degrees of freedom. The adjustment is done since R^2 always increases with the addition of more independent variables even when they are irrelevant.

c)

$$R^2 = r^2 = 0.3372$$

$$R^2 = 1 - \frac{SSE}{SST} = 0.3372 = 1 - \frac{154.81}{SST} => SST = 233.57, SSR = 78.76$$

$$F_{2,18} = \frac{MSR}{MSE} = \frac{78.76/2}{154.81/18} = 4.579$$

$F_{1\%,2,18} = 6.012$
$F_{5\%,2,18} = 3.555$

Conclusion: The model is significant at 5% but not at 1%.

d)

$$t_{18} = \frac{-0.4115}{0.2373} = -1.734$$

The p-Value is 10%

$$t_{18} = \frac{1.6188}{0.5366} = 3.017$$

The p-value is less than 1%

e)

$$AdjR^2 = 1 - (1 - 0.37^2)(\frac{20}{17}) = 0.2588$$

The Adjusted R^2 is lower so the inclusion of X_3 has not improved the model.