RAJARATA UNIVERSITY OF SRI LANKA, MIHINTALE

B.Sc. (General Degree) in Applied Sciences

Third Year Semester I– End Semester Examination – November 2014

MAT 3203-Regression Analysis

Answer All Questions in Part A &BTime : 2 hours

Part A

01. Regression modeling is a statistical framework for developing a mathematical equation that describes how
- I. one explanatory and one or more response variables are related
- II. several explanatory and several response variables response are related
- III. one response and one or more explanatory variables are related
- IV. All of these are correct.

02. If there is a very weak correlation between two variables, then the coefficient of determination must be
- I. much larger than 1, if the correlation is positive
- II. much smaller than 1, if the correlation is negative
- III. much larger than one
- IV. none of the above.

03. In least squares regression, which of the following is not a assumption about the error term ε?
- I. The expected value of the error term is one
- II. The variance of the error term is the same for all values of x
- III. The values of the error term are independent
- IV. The error term is normally distributed

04. If the coefficient of correlation is a negative value, then the coefficient of determination
- I. must also be negative
- II. must be zero
- III. can be either negative or positive
- IV. must be positive

05. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained.

$$\hat{y} = -0.0127 + 0.0180\,x$$

The above equation implies that
- I. each beer consumed increases blood alcohol by 1.27%
- II. on average it takes 1.8 beers to increase blood alcohol content by 1%
- III. each beer consumed increases blood alcohol by an average of amount of 1.8%
- IV. each beer consumed increases blood alcohol by exactly 0.018

06. Regression analysis was applied to determine the relationship between thedemand for a product (Y) and the price of the product (X).The following estimated regression equation was obtained.

$\hat{Y} = 120 - 10\,X$

Based on the above estimated regression equation, if price is increased by 2 units, then demand is expected to

88

[06.Cont…]

I. increase by 120 units
II. increase by 20 units
III. decrease by 20 units
IV. decrease by 120 units

07. SSE can never be
I. larger than SST
II. smaller than SST
III. equal to 1
IV. equal to zero

08. In a regression analysis if SSE = 200 and SSR = 300, then the coefficient of determination is
I. 0.6667
II. 0.6000
III. 0.4000
IV. 1.5000

09. Interpret the *XL* output given below related to simple linear regression model : $\hat{Y} = A + B\,X$

Regression Statistics	
Multiple R	0.8814144
R Square	0.7768914
Adjusted R Square	0.7717028
Standard Error	14.441584
Observations	45

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	31227.86	31227.86	149.7313	1.34E-15
Residual	43	8968.053	208.5594		
Total	44	40195.91			

10. Given below is the result output by the *XL* for multiple linear regression model. Comment on the output.

Regression Statistics	
Multiple R	0.752715
R Square	0.56658
Adjusted R Square	0.507477
Standard Error	1.317266
Observations	51

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	99.80499	16.63417	9.586367	9.73E-07
Residual	44	76.34834	1.73519		
Total	50	176.1533			

	Coefficients	*t Stat*	*P-value*
Intercept	-31.2495801	-3.01236126	0.0042862
X Variable 1	-0.05808865	-1.71944313	0.09256209
X Variable 2	0.000796204	2.410550275	0.02017381
X Variable 3	-0.06430005	-3.42540598	0.00134045
X Variable 4	1.081326469	5.492437327	1.8688E-06
X Variable 5	0.080875806	1.581742562	0.12087093
X Variable 6	0.011003724	1.629317165	0.11038553

Part B

1. For an experiment the pair (x_i, y_i), and the following summary statistics are given;

$n=13 \quad \bar{X} = 198 \quad \bar{Y} = 15 \quad \sum Y_i^2 = 3226 \quad \sum X_i^2 = 625625 \sum X_i Y_i = 43950$

$$SSE = 16.67$$

a). Find the variance of the least square estimators.

b). By identifying the matrix $X'X$ and vector $X'Y$,write down the normal equation in matrixform.

c). Prepare the ANOVA table and interpret the results.

2. The values of a response variable Y for five observations corresponding to levels of an independent variable are given below.

Observation		1	2	3	4	5
Independent variable	X	-2	-1	0	1	2
Response variable	Y	6	4	3	1	-1

a). Write down the above data in the form $Y = X\beta + \varepsilon$

b). Using matrix operations, obtain the least square estimators $\underline{b}$for the parameters $\underline{\beta}$

c). Find the sum of squares SST ,SSE and SSR

3. A regression model is to be developed to predict the average sales price of a one-family house in the given housing development: The prices at which a random sample of 8 one-family houses were sold recently have been taken with two regresses ;
X_1 = the number of bed rooms
X_2 = the number of baths

We wish to fit the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 .+ \varepsilon$.

Some necessary information are given below.

$$(X'X)^{-1} = \frac{1}{84}\begin{pmatrix} 107 & -20 & -17 \\ -20 & 32 & -40 \\ -17 & -40 & 71 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 637{,}000 \\ 2{,}031{,}100 \\ 1{,}297{,}000 \end{pmatrix} \sum Y_i = 637{,}000$$

a). Determine least square estimates of the multiple regression coefficients.

b). What is the predictive value of a house when there are 5 bed rooms and 2 baths.

c). If $Y'Y = 50{,}907{,}080{,}000$ is given, determine the estimate of$\hat{\sigma}$.

d). Test the null hypothesis $\beta_1 = \$\,3500$ against the alternative hypothesis $\beta_1 > \$\,3500$ at the 0.05 level of significance.