Question Paper

Exam Date & Time: 22-Nov-2018 (10:00 AM - 01:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES

SECOND SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA ANALYTICS)

Multiple Linear Regression and Logistic Regression [BDA 606]

Marks: 100 **Duration: 180 mins.**

END SEMESTER DEGREE EXAMINATION NOVEMBER 2018

Answer all the questions.

1) 1A) Obtain the oridinary least square estimators of all the parameters involved in the simple linear regression model. (10)

1B) Briefly explain why the estimator of unknown variance in the simple linear regression model is known to be 'model dependent'?. (5+5=10 marks)

2) 2A) Prepare a detail note on the topic 'Coefficient of Determination'. (10)

2B) Describe how you will test the overall significance of a multiple linear regression model.

(5+5=10 marks)

3) 3A) What do you mean by 'outliers'? Why outlier detection is very important in regression analysis? Identify the major classifications of outliers. (10)

3B) Briefly explain the concept of multicollinearity? Explain any two methods of detecting multicollinearity.

(5+5=10 marks)

4) 4A) Briefly explain the method of 'backward elimination'. (10)

4B) Explain the use of Mallows's Cp statistic as a criterion for evaluating subset regression models.

(5+5=10 marks)

5) 5A) With suitable examples explain the use of dummy variables in multiple linear regression. (10)

5B) Briefly explain how the plots of residuals against the fitted values is useful for detecting several common types of model inadequacies in linear regression analysis.

(5+5=10 marks)

6) 6 Explain binary logistic regression under the heading (10)

6A) Regression model
6B) Estimation of parameters
6C) Confidence interval for paramters
6D) Role of confidence interval in test of significance
(1+5+3+1=10 marks)

7) 7) In context of binary logistic regression explain briefly a (10)

7A) Test of significance of the individual regressors

7B) Test of goodness of fit
(5+5=10 marks)

8) 8) Answer the following (10)

8A) Compare between simple linear regression and logistic regression based on

a) Link function
b) Linear relationship
c) Response variable

8B) Show that in logistic regression, the exponential of regression coefficients are odds ratios.

8C) Interpretation of odds ratio (3+5+2 = 10 marks)

9) 9) Explain the procedure of finding the optimum cut off for classifying subjects estimated probability as having event and not having event in logistic regression analysis. (10)
(10 marks)

10) 10 A) What are the assumptions to use logistic regression? (10)

10 B) Write the logistic regression model for prediction and find the probability of liver cirrhosis given the subject is smoker and alcoholic, Where
Constant $\beta 0 = -1.43$
$\beta 1 = -0.210$ for alcohol
$\beta 2 = 0.154$ for smoking

Smoking status= 1 for smoker,
= 0 for non-smoker

Alcohol status = 1 for alcoholic
= 0 for non-
alcoholic. (3+7 = 10 marks)

-----End-----