

Total No. of Pages 3

Roll No.....

II SEMESTER

MBA(Business Analytics)

END SEMESTER EXAMINATION May/June-2019

PAPER CODE: MB207& TITLE OF PAPER: Predictive Modelling

Time: 3:00 Hours

Max. Marks: 60

Note: Answer any five questions out of the given six questions.
All questions carry equal marks.
Assume suitable missing data, if any.

Q.1[a] Differentiate between classification and regression algorithms in machine learning. (6)

[b] The average room prices (in US\$) paid at hotel "Homeaway" by various nationalities while travelling abroad (away from their home country) in 2018 were recorded as follows:

174 161 167 171 163 141 147 154

- i) Compute Z score. Are there any outliers? Explain. (3)
ii) Describe the shape of the above data set. (3)

Q.2[a] What is the difference in calculating average value through the various measures of central tendencies (mean, median and mode)? What are the advantages and disadvantages of each? (6)

[b] Analysis of the monthly wages of two firms gives the following data:

	Firm A	Firm B
Number of workers	500	600
Variances	81	100
Average Wages (in INR)	186	175

- i) Which firm pays out a larger wage bill? (2)
ii) In which firm does greater variability occur? (2)
iii) What is the average wage if firm A and firm B are combined? (2)

Q.3 An agent for a residential real estate company in a suburb located outside of Washington, DC; has the business objectives of developing more accurate estimates of the monthly rental cost for apartments. Towards that goal, the agent would like to use the size

139

of an apartment, as defined by square footage to predict the monthly rental cost. The agent collects the following data of 7 one-bedroom apartments:

size (in '00sq.ft.)	3	5	4	7	2	1	9
monthly rental cost (in '000\$)	6	9	8	11	4	3	14

- i) Construct a scatter plot. (3)
- ii) Use least square method to determine the regression model for the above problem. (3)
- iii) Interpret the meaning of regression coefficients b_0 and b_1 in this problem. (3)
- iv) Predict the monthly rent for an apartment that has size as 800 sq. ft. (3)

Q.4[a] The residuals for 10 consecutive time periods are as follows:

Time Period	1	2	3	4	5	6	7	8	9	10
Residual	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5

Compute Durbin-Watson statistic to check the autocorrelation among the residuals. (6)

[b] The following ANOVA summary table is for multiple regression model with two independent variables. Complete the table and find Adjusted R^2 : (6)

Source	Degree of freedom	Sum of squares	Mean squares	F statistic
Regression	2	60		
Error	18	120		
Total	20	180		

Q.5[a] A prospective MBA student wanted to predict starting salary upon graduation, based on program per-year tuition. If $RSS=21.8$ and $TSS=64$ for this problem; determine the coefficient of determination R^2 and interpret its meaning. (6)

[b] Consider the following logistic regression equation:

$$\ln(\text{Estimated odds ratio}) = 0.1 + 0.5X_1 + 0.2X_2$$

- i) Interpret the meaning of the above logistic regression coefficients. (2)
- ii) If $X_1=2$ and $X_2=1.5$, compute the estimated odds ratio. (2)
- iii) On the basis of results of (ii), compute the estimated probability of an event of interest. (2)

Q.6 Consider the following correlation matrix, related to three variables with unities in the diagonal spaces:

Variables	Variables		
	A	B	C
A	1.000	0.709	0.204
B	0.709	1.000	0.051
C	0.204	0.051	1.000

Use the Principal Component Method of factor analysis to determine the factor loadings the first Principal Component for the above matrix. (12)

40