

1654

8

(i) Precision

(ii) Recall

(iii) Accuracy

(iv) Specificity (8)

7. Write short notes on any three of the following:

(a) K-means Clustering

(b) Decision Tree approach for Classification and Regression

(c) Testing of Assumptions in Linear Regression Analysis

(d) Web and Social Media Analytics (5*3=15)

(500)

[This question paper contains 8 printed pages.]

Your Roll No.....

आपका अनुक्रमांक.....

Sr. No. of Question Paper : 1654

I

Unique Paper Code : 2922062301

Name of the Paper : Introduction to Business Analytics

Name of the Course : Bachelor of Management Studies (BMS)

Semester : III

Duration : 3 Hours

Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt any six questions.
3. All questions carry equal marks.
4. Use of scientific calculator is allowed.
5. Show your working clearly in your answer sheet.

P.T.O.

1. What are the differences between descriptive, predictive, and prescriptive analytics? Explain how these types of analytics can be combined to provide comprehensive business insights, using an example from a business context. (15)

2. (a) What is Big Data Analytics and why is it important? What are the main challenges and benefits of working with Big Data in business? (7)

(b) Consider a dataset named `inventory_data` which contains information on stock levels, sales and lead times for products in a retail store. One of the variables is `days_out_of_stock`, which tracks the number of days a product was unavailable for sale. Write a suitable Python or R code to calculate the following for the `days_out_of_stock` variable:

(i) Mean

(ii) Range

(iii) Standard Deviation (8)

6. (a) The following is a dataset template of students who wish to take admission to a prestigious university:

GMAT Score	CGPA	Work Experience	Admitted

In this dataset, the outcome 'Admitted' is coded as 1 for admission and 0 for rejection. Write down the steps and provide Python or R code for building a logistic regression model using this dataset. (7)

(b) Consider the following confusion matrix obtained from a classification model: (7)

Pre dic tio	Positive	10	4
	Negative	3	13
		Positive	Negative

Actual

Regarding the above table, calculate the following:

(b) The box office success of Bollywood movies was analyzed using a logistic regression model. In this analysis, the target variable 'Box Office Success' (Y) is coded as 1 for success and 0 for failure. The independent variable (X) is the budget, measured in crores of rupees. The relevant coefficients and intercept values obtained from the model are presented in the following table.

	Coef.	Std.Err.	Wald	P> z
constant	1.620	0.503	10.39	0.005
Budget	-0.016	0.008	3.825	0.045

(i) A production house is making a movie with a 100-crore budget. What is the probability of success for this movie? Reference values for e^x :

X	0.18	0.19	0.2	0.21	0.22
e^x	1.197	1.209	1.221	1.233	1.246

(ii) Calculate the budget for which box office success and failure are equally likely. (8)

3. (a) What are the main differences between simple regression models and multiple regression models? How do you interpret the coefficients (β) in Simple Linear Regression compared to those in Multiple Linear Regression? (7)

(b) The following is the dataset template of companies producing automobiles.

S.No.	Marketing Expense (in lacs)	Sales (in Crores)

Suppose the name of the dataset is 'auto'. Build a suitable regression model to predict the sales of a company. Write down the steps with python/R codes for building the model using the above data template. (8)

4. We aim to build a model to estimate sales based on the advertising budget allocated to YouTube, Facebook, and Newspapers. After running the regression analysis, we obtained the following output:

Call:

`lm(formula = Sales ~ YouTube + Facebook +`

P.T.O.

1654

4

Newspaper, data = marketing)

Residuals:

Min	1Q	Median	3Q	Max
-10.59	-1.07	0.29	1.43	3.40

Coefficients:

coefficients Std. Error t value Pr(>|t|)

(Intercept) 3.52667 0.37429 9.42 <2e-16 ***

YouTube 0.04576 0.00139 32.81 <2e-16 ***

Facebook 0.18853 0.00861 21.89 <2e-16 ***

Newspaper -0.00104 0.00587 -0.18 0.86

Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1" 1

Residual standard error: 2.02 on 196 degrees of freedom Multiple R-squared: 0.897, Adjusted R-squared: 0.896 F-statistic: 570 on 3 and 196 DF, p-

1654

5

value: <2e-16

(a) Write the regression equation, clearly specifying the regressors, target variable and the sample size.

(b) Provide an interpretation of the intercept in the regression model. Additionally, check its statistical significance and discuss whether the intercept has any meaningful interpretation in this context.

(c) Interpret the coefficients for 'YouTube', 'Facebook', 'Newspaper'. Analyze how each of these variables relates to the response variable 'Sales', and assess their statistical significance.

(d) Does the model have a good fit? Make useful interpretations in this respect.

(e) Using the regression model, predict the value of Sales given the advertising budgets for YouTube, Facebook, and newspapers are *1,40,000, 25,000, and 10,000 respectively. (15)

5. (a) What is logistic regression and how does it differ from linear regression? Discuss the assumptions of logistic regression and explain how it is used to predict binary outcomes. (7)

P.T.O.