Enrolment No. _____

AHMEDABAD UNIVERSITY

AMRUT MODY SCHOOL OF MANAGEMENT

~~UNDERGRADUATE~~ / GRADUATE PROGRAMMES

2024 MONSOON SEMESTER

END SEMESTER EXAMINATION

TOD601 ANOVA AND REGRESSION

Date: 03-12-2024 Total Marks: 75

Time: 2:00 pm – 4:00 pm No. of printed pages: 03

INSTRUCTIONS

1. *All questions are compulsory.*
2. *The data to be used is in the file called TOD601-M24-FX-DATA.XLSX (provided separately).*
3. *Create a new Word.docx file and provide all your answers in this file with appropriate labelling for your answers as 1a, 2c, etc. Along with your answers, this file must contain all relevant items (charts, tables, etc), copy-pasted from the various software you have used.*
4. *In addition to this main answer file in Word, please also submit, per answer, the software output file (eg, jamovi, Excel, etc). If you use R-software, please submit the .R and .RData per answer.*
5. *You must name all your files as Name Surname_Roll No_ TOD601M24FX_ABC and upload them on LMS. ABC = Main, A1, A2, A3, A4, and A5. There will be a total of SIX files (plus associated data files for those using R software).*
6. *Present your results up to TWO decimal places only and p-values up to THREE decimals.*
7. *Your laptop must be on AIRPLANE MODE at all times.*
8. *Use 5% as the level of significance.*

Question 1 [15 marks]

The data in the attached file (worksheet: SOLIDWASTE) shows solid waste generated in grams/capita/day by household income categories and seasons. Using an appropriate statistical technique, analyse this and answer the following questions (referring to the table(s)/chart(s)) from your software outputs, as applicable).

a) Does INCOME have a statistically significant effect on SOLIDWASTE GENERATED? [3 marks]
b) Does SEASON have a statistically significant effect on SOLIDWASTE GENERATED? [3 marks]
c) Does the interaction of INCOME and SEASON have a statistically significant effect on SOLIDWASTE GENERATED? [3 marks]
d) To support your answer in (c), plot a graph of the interaction effect and comment on it (and copy-paste it into your answer book) [3 marks]
e) Develop a regression model for predicting SOLIDWASTE GENERATED using INCOME and SEASON and write the prediction model. Take Low INOCME and Monsoon SEASON as your reference categories. Given that, what does your model predict for middle-income households in winters. [2+1 marks]

Question 2 [15 marks]

A study assessed the effectiveness of different executive education training programmes on employees' performance. Employees' performance was measured using a scale of 0–100, which included SALES volume, CUSTOMER satisfaction, and SUPERVISOR rating. The employees were randomly assigned to the training programmes. The training programmes were categorised as follows, and the data is available in the attached file (worksheet: EXECUTIVE EDUCATION).

1. Traditional: in-person classroom-based teaching from industry experts
2. Experimental: heavy components of hands-on workshops, team-building exercises, and some classroom teaching from industry experts
3. Hybrid: a carefully balanced mix of lectures from industry experts, hands-on workshops, and unmoderated inspirational videos for participants to see.

Please answer the following questions:

a) Using an appropriate model(s) and justifying its choice, determine if there is a difference in the educational training programmes in terms of their impact on the employees' performance. Is any training programme more impactful than others? [10 marks].
b) Construct a table of the means in various categories accompanied by a plot of the means, such that training programmes are on the x-axis, and comment on it. [3 marks]
c) Draw nine box plots in one chart, each showing the rating scores per training method, and comment on it. [2 marks]

Question 3 [25 marks]

The data in the attached file (worksheet: BOLLYWOOD) shows data for Bollywood movies, and the description of the variables is shown in Table 1.

Table 1: Bollywood movie data description

S#	*Variable*	*Type*	*Remarks*
1	BOX OFFICE COLLECTION	Numeric	₹ crores
2	TOTAL PRODUCTION COST	Numeric	₹ crores
3	YOUTUBE VIEWS	Numeric	YouTube views of the trailer before the movie was released (in millions)
4	LEAD ACTOR	Categorical 1. Not star 2. Star	"Not star" is an actor who is a debutant or is relatively a newcomer.
5	DIRECTOR	Categorical 1. Novice 2. Established 3. Highly reputed	-
6	ITEM SONG	Categorical 1. No 2. Yes	Whether an item song was included in the movie

Note: Take the first item listed as the reference category.

Using this data, build an appropriate regression model. You may use any software. First, create a dependent variable called "SUCCESSFUL" (that signifies whether the movie was successful), which will be defined as "Yes" if the "BOX OFFICE COLLECTION" is more than 120% "TOTAL PRODUCTION COST", "No" otherwise (and take "No" as the reference category). After you have built the regression model using the five independent variables (DO NOT include "BOX OFFICE COLLECTION" in the model), answer the following questions, and in doing so, paste the relevant table(s)/chart(s) from the software in your answer book:

a) Report and comment on the model coefficients and other important outputs. [6 marks]
b) Assuming the model represents a good fit, what are the top three factors more likely to make a film successful? Explain your answers. [6 marks]
c) Comment on the likelihood ratio statistic. [2 marks]
d) Report the accuracy, true negative rate, true positive rate, and error rate.[4 marks]
e) Report and comment on the AUC for ROC. [2 marks]
f) Explain how the ROC is constructed. [4 marks]
g) Comment on multicollinearity in the model and state the basis for your answer. [1 mark]

Question 4 [15 marks]

Using the data in the attached file (worksheet: SALARY), develop a regression model using your preferred software. Take Marketing, Female, and Small as the reference categories.

a) Comment on the overall model fit and explain the meaning of r-squared [2 marks], and overall significance [2 marks], referring to the values from your outputs.
b) Report your regression model equation [1 mark] and comment on the slope parameters referring to the values from your outputs. [4 marks]
c) What salary would your model predict for a "Male" aged 33 years in the "FinTech" sector working in a "Medium" (size) city? [2 marks]
d) Comment on multicollinearity in the model and state the basis for your answer. [2 marks]
e) List the independent variables in order of their impact on the dependent variable. [2 marks]

Question 5 [5 marks]

The data in the attached file (worksheet: DATAVIZ) shows a survey of 200 people. They were asked whether you engage in sustainable practices in your daily routine. Their responses "Yes" or "No" were recorded along with their GENDER and INCOME levels. Create as many data visualisations as you wish to explain the patterns in the data. Well-labelled and aesthetically pleasing charts will carry more credentials.

End of paper