FAR WESTERN UNIVERSITY
FACULTY OF MANAGEMENT

MPhil/MPhil Leading to PhD Program in Management

FM: 100
Time: 4 Hrs.

Semester End Examination - 2024

Level: MPhil Semester: First Subject: Statistical Methods and Data Analysis

---

Candidates are required to give original, insightful and analytical answers in their own words as far as practicable.

Group A: Attempt any SEVEN questions (7 × 10 = 70)

1. A retail company is analyzing customer satisfaction survey data to understand the overall customer experience. The survey scores range from 1 (very dissatisfied) to 10 (very satisfied). The company has calculated the following statistics for the survey scores: 7, 8, 9, 6, 7, 8, 7, 6, 9, 8, 7, 5, 8, 7, 9, 7, 10, 8, 6, 8, 7, 8, 9, 7, 5, 8, 6, 7, 8, 9

   Calculate mean and standard deviation of satisfaction score also interpret the result. Find skewness and kurtosis of the data and interpret the result. Based on the above statistical measure what you recommend the company?
2. The lifespans of 1,000 batteries produced by a manufacturer are normally distributed with a mean lifespan of 500 hours and a standard deviation of 50 hours. Calculate: a. The proportion of batteries that last more than 550 hours. b. The number of batteries expected to last between 450 and 500 hours. c. The lifespan of the longest-lasting 5% of batteries.
3. In a company of 500 employees, a random sample of 100 employees was surveyed regarding their satisfaction with the new remote work policy. The survey results are tabulated below by department:

Department =>	HR	IT	Sales
Satisfied	15	20	25
Not Satisfied	10	5	25

   What is the probability that a randomly selected employee: a. Is satisfied with the policy? b. Works in the IT department and is satisfied? c. Is from the Sales department or is not satisfied? d. Given that the employee is from the HR department, what is the probability they are satisfied?
4. A supermarket chain conducted a survey to determine customer satisfaction with its loyalty program in two regions, Region A and Region B. Out of 200 customers surveyed in Region A, 150 reported satisfaction, while in Region B, 130 out of 160 customers were satisfied. Conduct a hypothesis test to determine if there is a significant difference in satisfaction levels between the two regions.
5. A manufacturing company conducted a study to evaluate the effectiveness of a new workflow optimization program designed to increase productivity. The study involved a sample of 50 production lines, and the increase in productivity was recorded for each line. The sample mean of increase in productivity was found to be 12.5%, with a standard deviation of 3.2%.

   i. Construct a 95% confidence interval for the mean increase in productivity.

   ii. What is the minimum sample size needed to ensure a margin of error of 2% with 95% confidence?
6. Explain the concept of cluster sampling. If you were tasked with conducting a study on consumer preferences for different brands of soft drinks across various cities, which sampling technique would you choose, and why?
7. A company wants to investigate the relationship between the number of hours employees work overtime per week and the number of customer complaints received. The data collected from 10 employees are as follows:

Overtime hours per week	2	4	6	8	10	12	14	16	18	20
Nos of Complaints	1	2	2	3	3	4	4	5	6	7

While using software the following partial results were obtained

ANOVA Table

	df	SS	MS	F
Regression	1	30.91		
Residual	8	1.19		

Coefficient table

	Coefficients	S.E.	t-Stat	P-value
Intercept	0.333	0.263	1.266	0.241
Nos of Complaints	0.306	0.021	14.43	0.000

a) Use this information to develop a regression model to predict the number of complaints by overtime hours per week. And hence predict the number of complaints if the overtime hours per week is 24.
b) Interpret the meaning of slope in the model obtained above.
c) Obtain standard error of estimate and interpret its meaning.
d) What percentage of the variation in number of complaints is explained by the overtime hours per week in the model above?
e) Is the overall model statistically significant?

8. You are analyzing the impact of different teaching methods (traditional, online, and hybrid) on student performance in Statistics. The scores of students taught using these methods are as follows:

   Traditional: 85, 88, 90, 87, 89, 92 Online: 78, 80, 83, 76, 79, 81 Hybrid: 82, 85, 88, 84, 87, 90

   After performing an ANOVA, the following results were obtained:

   Sum of Squares Between (SSB): 259 Sum of Squares Within (SSW): 101

   Determine whether there is a statistically significant difference in student performance across the different teaching methods.

9. The satisfaction levels of employees at a company before and after the introduction of a new management style are recorded below:

   Before: 70, 65, 60, 75, 80, 85, 90, 92 After: 75, 70, 65, 78, 85, 90, 95, 98

   Perform a paired t-test to determine if the new management style has led to a significant improvement in employee satisfaction.

10. A marketing survey was conducted among 1,200 individuals to evaluate the effectiveness of an advertisement campaign across different age groups. The results are displayed in the cross-tabulation below:

	Perception of effectiveness		
Age Group	Not effective	Somewhat effective	Very effective
Under 25	100	150	250
25-40	120	180	200
Above 40	90	130	180

Test the hypothesis that there is no association between age group and perception of the advertising campaign's effectiveness at a 5% significance level.

Group B: Practical using SPSS, Attempt any TWO questions (2 × 15 = 30)

11. A dataset named `employee_performance.sav` has been provided to you, containing data on various performance metrics for 100 employees. Use SPSS to address the following:
    a. Determine the mean, median, and standard deviation of employee performance ratings, categorized by department (e.g., Sales, Marketing, IT).
    b. Calculate the mean, median, and standard deviation of employee salaries based on different education levels (e.g., High School, Bachelor's, Master's).
    c. Conduct a hypothesis test to check if there is a significant difference in average job satisfaction scores between employees working in teams versus those working individually.
    d. Perform a cross-tabulation analysis between employee turnover (resigned/stayed) and gender. Test if there is a significant association between these variables.
    e. Use ANOVA to determine if the distribution of salary differs significantly across different job performance categories (e.g., High Performer, Average Performer, Low Performer).
12. A large retail company is conducting an extensive study to determine the factors affecting monthly sales revenue (in Rs. lakhs). Data was collected from 100 stores over the past year. A data set named `revenue.sav` The factors being analyzed include:

    Number of customer service representatives (x1): The number of staff assigned to customer service.

    Average product rating (x2): Average rating of products sold, based on customer reviews (out of 5).

    Monthly marketing expenditure (x3): The amount spent on marketing in Rs. thousands.

    Store foot traffic (x4): The number of customers visiting each store in a month

    The company wants to explore the relationship between these factors and **monthly sales revenue (Y)** in Rs. lakhs. The following sample dataset contains observations for 100 stores

    a. Obtain the unstandardized regression coefficients of sales revenue on the number of customer service representatives, average product rating, marketing expenditure, and store foot traffic, along with the standard error, t-value, and significance. Interpret the values of the regression coefficients so obtained and discuss the significance of these values.
    b. Discuss the overall model fit on the basis of the ANOVA table. What can be inferred about the model's ability to explain sales revenue?
    c. Obtain the fitting regression equation for this data and find the expected sales revenue in a store with 19 customer service representatives, an average product rating of 4.5, a monthly marketing expenditure of Rs. 145,000, and foot traffic of 1,100 customers.
    d. What percentage of variation in sales revenue is explained by this equation?
    e. Is there any multicollinearity among the independent variables? Discuss this by checking the Variance Inflation Factor (VIF) in the regression output.
13. You have been given a dataset containing responses from 50 individuals regarding their purchasing behavior and related factors. The dataset includes responses to the following five variables:
    1. Monthly_Expenditure: Amount spent on purchases per month (in USD).
    2. Shopping_Frequency: Number of shopping trips made in the past month.
    3. Brand_Preference: Level of brand preference (measured on a scale of 1-10).
    4. Product_Interest: Interest in new products (measured on a scale of 1-10).
    5. Loyalty_Programs: Participation in loyalty programs (measured on a scale of 1-10).

    Additionally, the dataset contains a categorical variable:

    Buyer_Status: Indicates whether the individual is a Buyer or Non-Buyer.

The complete data `purchasing_behavior.sav` is provided to you. Use SPSS and answer the following questions:

1. Perform a Discriminant Analysis to classify individuals as Buyers or Non-Buyers using the five continuous variables. Provide the discriminant function coefficients.
2. Interpreting the group means so obtained, explain how does these means of the discriminant function differ between Buyers and Non-Buyers?
3. On the basis of the value of Wilks' Lambda and the associated p-value, Interpret the significance of the results in terms of the discriminant function's ability to differentiate between Buyers and Non-Buyers.
4. Calculate the classification accuracy of the discriminant function by creating a classification table and report the percentage of correct classifications. How accurately does the model predict Buyer status?
5. Analyze the structure matrix to determine the most influential variables.