5/31

0105N110

Candidate's Seat No :_____

MSc Sem.-2 Examination

407

AMS

Time : 2-30 Hours] **May-2025** **[Max. Marks : 70**

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** A pharmaceutical company tests a new cholesterol-lowering drug. They measure the cholesterol levels of 10 patients before and after taking the drug for 8 weeks. Has the drug significantly changed cholesterol levels? Apply appropriate nonparametric test for below datasets. Table Value=1.96 **(07)**

Patient	1	2	3	4	5	6	7	8	9	10
Before Drug	220	210	230	200	215	225	240	195	205	235
After Drug	200	190	225	180	215	205	230	180	190	225

(b) A researcher is testing the effectiveness of three different pain-relief techniques (Technique A, Technique B, and Technique C) on the same group of 10 patients. Each patient tries all three techniques on different days, and the pain levels are recorded after each treatment. Pain is measured on a numeric scale (0–10) where lower is better. Apply appropriate nonparametric test for below datasets. Table Value=5.991 **(07)**

Patient	1	2	3	4	5	6	7	8	9	10
Technique A	6	7	5	6	8	7	6	5	7	6
Technique B	4	5	3	5	6	4	3	4	6	5
Technique C	2	4	2	3	4	3	2	3	5	4

OR

(a) A pharmaceutical researcher wants to compare the effectiveness of two pain relief medications: Drug A and Drug B. The effectiveness is measured as pain reduction score (higher is better) after 30 minutes. Apply appropriate nonparametric test for below datasets. Table value: 1.96 **(07)**

Subject	1	2	3	4	5	6	7
Drug A	8	6	7	5	9	4	6
Drug B	7	9	5	6	8	4	3

(b) A researcher is studying the effect of three different diets (Diet A, Diet B, and Diet C) on weight loss. After 6 weeks, the weight loss (in kg) for participants in each group is recorded. Apply appropriate nonparametric test for below datasets. Table value:5.991 **(07)**

Diet A	3.1	2.8	3	2.7	3.3
Diet B	4.5	4.2	3.9	4	4.1
Diet C	5	4.9	5.1	4.8	5.2

(P.T.O)

N110-2

Q.2 **(a)** i. Write down the normal equations of simple linear regression equation obtained using least square method. **(07)**

ii. Write down the relationship between SST, SSE, and SSR.

iii. Write down the t-test statistic for the hypothesis testing of significant relation between regressor variable and response variable, and criterion for rejecting H_0.

iv. Write down the $100(1-\alpha)\%$ confidence interval on the slope of the simple linear regression equation.

(b) Given are the five observations of two variables x and y. **(07)**

x_i	3	12	6	20	14
y_i	55	40	55	10	15

i. Use the least square method to find the simple linear regression equation of the form $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x$.

ii. Use the F-test to test hypothesis for a significant relationship between x and y. Present result in an ANOVA table. $(Take\ \alpha = 0.05, F_\alpha(1,3) = 10.13)$

OR

(a) Given are the five observations of two variables x and y. **(07)**

x_i	2	6	9	13	20
y_i	7	18	9	26	23

The estimated regression equation is $\hat{y} = 7.6 + 0.9x$.

i. Calculate the value of sample correlation coefficient.

ii. What conclusions can be drawn from this value?

iii. How the above value relates to the coefficient of determination R^2 ?

(b) i. Obtain the log likelihood function, which has to be maximized for the logistic regression model. **(07)**

ii. Does F-test and t-test always give the same conclusion for hypothesis testing of relationship between regressor variable(s) and response variable? Discuss it for simple linear and multiple linear regression models. Establish any relation if there is between two test statistics.

Q.3 **(a)** i. For the multiple regression model, $Y = X\beta + \epsilon$, Show that the least square estimator of β is $\hat{\beta} = (X^TX)^{-1}X^TY$. **(07)**

ii. Obtain the least-square normal equations for the polynomial regression model of the form $y_i = b_0 + b_1x_i + b_2{x_i}^2 + \epsilon_i\ for\ each\ i \in \{0,1,2,3,\cdots,n\}$.

(b) i. Write down the assumptions and the problems of the multiple regression model **(07)**

ii. Define the terms: Correlation matrix, Multicollinearity, Autocorrelation.

OR

(a) Consider the multiple regression model, $y = b_0x_1 + b_1x_2 + \epsilon$ with following data: **(07)**

x_1	1	2	3	4
x_2	-2	0	3	5
y	2	6	9	13

i. Calculate SST, SSR, SSE.
ii. Calculate coefficient of determination R^2.
iii. Calculate adjusted coefficient of determination R_a^2.

(b) For a sample data of multiple regression model, the following values are obtained: **(07)**

$$S_{yy} = 175.089,\ S_{1y} = 4215.372,\ S_{2y} = 2.33,$$
$$S_{11} = 184710.16,\ S_{22} = 0.047755,\ S_{12} = 51.2873$$

i. Find the correlation matrix for this problem.
ii. Calculate the standardized regression coefficients.

Q.4 **(a)** The enrolment data (in thousands) for a state university over the past six years are shown. **(07)**

Year	1	2	3	4	5	6
Enrolment	20.5	20.2	19.5	19.0	19.1	18.8

Develop the equation for the linear trend component of this time series. Comment on what is happening to enrolment at this institution.

(b) Find the exponential smoothing forecasts for periods 2-10 using the following data, where α=0.10. Assume $s_1=d_1$. **(07)**

Year	Demand
1	820
2	775
3	680
4	655
5	750
6	802
7	798
8	689
9	775
10	?

OR

(a) The values of Vastrapur building contracts (in millions of dollars) for 12-month periods follow. **(07)**

240 350 230 260 280 320 220 310 240 310 240 230

a) Compare a 3-month moving averages forecast with an exponential smoothing forecast and use alpha=0.2. Which provides the better forecasts?

b) What is the forecast for the next month?

(P.T.O)

(b) A lady wants to compute the estimated price of the onion for tomorrow based on an average of the last ten days. He believes there is a 10% upwards trend because of rising fuel prices. Also, he believes that the prices of onions fluctuate based on moving averages. The last ten days' prices of the onion per kg are 15, 17, 22, 25, 21, 23, 25, 22, 20, and 22. Compute the projected onion price on day 11. **(07)**

Q.5 Attempt any **SEVEN** out of **TWELVE:** **(14)**

(1) Which test helps identify *which means are different* after ANOVA?
- A) Bartlett's Test
- B) F-test
- C) Tukey's HSD Test
- D) t-test for mean

(2) The primary assumption of ANOVA is:
- A) Independence of treatments
- B) Equal means
- C) Homogeneity of variances
- D) Small sample sizes

(3) Correlation matrix is used to detect the presence of multicollinearity in regression model. Which of the following does **NOT** indicate the greater degree of multicollinearity?
- A) Determinant of matrix is close to 1.
- B) Determinant of matrix is close to 0.
- C) One of the eigenvalues of matrix is close to 0.
- D) None of these

(4) The Kolmogorov-Smirnov (K-S) test is used for:
- A) Testing equality of means
- B) Testing normality of data
- C) Testing independence
- D) Testing homogeneity of variances

(5) Which of the following is assumption to the error term of the linear regression model?
- A) Error term is normally distributed with zero mean.
- B) Error term for particular value of x can be related to any other value of x.
- C) Error term is normally distributed with non-zero mean.
- D) Error term has non-constant variance for all values of x.

(6) Which of the following is correct for simple linear least-square regression models?
- A) The F-test and the t-test yield different conclusions when testing whether the slope differs from zero.

B) The sample correlation coefficient r_{xy} is numerically equal to the coefficient of determination R^2.

C) The square of sample correlation coefficient r_{xy}^2 is numerically equal to coefficient of determination R^2.

D) The sum of squares due to regression (SSR) is different from the mean square due to regression (MSR).

(7) A Durbin–Watson statistic of approximately 2 suggests which of the following?

A) Positive autocorrelation in the residuals.

B) Negative autocorrelation in the residuals.

C) No first-order autocorrelation in the residuals.

D) None of these

(8) Which statement about the Durbin–Watson statistic is correct?

A) It tests for first-order autocorrelation in the residuals of a regression model.

B) A Durbin–Watson value close to 0 indicates no autocorrelation.

C) A Durbin–Watson value close to 4 indicates no autocorrelation.

D) It requires normally distributed errors to be valid.

(9) For a multiple regression model with n observations and p regression variables, the degree of freedom for the error/residual is:

A) p

B) $n - p$

C) $n - p - 1$

D) n

(10) What is autocorrelation?

(11) What are the components of time series?

(12) What is the difference between Stationary and Non-Stationary Time Series?

_____ ******** _____ ******** _____