7/11

2910E936/B

Candidate's Seat No :_____

Intg. M.Sc. (DS) Sem.-5 (Supply.) Examination

CC-302 - Regression Theory

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

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

Q.1 **(a)** The following data are from a completely randomized design. **(07)**

	Treatment		
	A	**B**	**C**
	162	142	126
	142	156	122
	165	124	138
	145	142	140
	148	136	150
	174	152	128
Sample mean	156	142	134
Sample variance	164.4	131.2	110.4

1. Compute the sum of squares between treatments.
2. Compute the mean square between treatments.
3. Compute the sum squares due to error.
4. Compute mean squares due to error.
5. Set up the ANOVA table for this problem and give the appropriate conclusion using the F-test (where for α=0.05, the p-value for given degree of freedom is 3.68)

(b) A factorial experiment involving two levels of factor A and three levels of factor B resulted in the following data. **(07)**

Factor A	**Factor B**		
	1	**2**	**3**
Level 1	135 165	90 66	75 93
Level 2	125 95	127 105	120 136

Test for any significant main effects and any interaction using 5% level of significance. ($F_{2.05} = 0.2022, F_{4.06} = 0.0767$ and $F_{7.66} = 0.0223$ for the given degree of freedom)

OR

P.T.O

E936B-2

(a) To test whether the mean time needed to mix a batch of material is the same for machines produced by three manufacturers, the Narayan Chemical Company Ltd. Obtained the following data on the time (in minutes) needed to mix the material. **(07)**

Manufacturer		
1	**2**	**3**
20	28	20
26	26	19
24	31	23
11	17	22

1. Use these data to test whether the population mean time for mixing a batch of material differs for the three manufacturers.
2. At the $\alpha = 0.05$ level of significance, use Fisher's LSD procedure to test for the equality of the means for manufacturers 1 and 3. What conclusion can you draw after carrying out this test?
   $(F_{10.63} = 0.0043, t_{0.025} = 2.262)$

(b) The Scholastic Aptitude Test (SAT) contains three parts: Critical reading, Mathematics and Reading. Each part is scored on 800-point scale. A sample of SAT scores of six students follows. **(07)**

Students	Critical Reading	Mathematics	Reading
1	526	534	530
2	594	590	586
3	465	464	445
4	561	566	553
5	463	478	430
6	430	458	420

Using a 0.05 level of significance, do students perform differently on the three portions of SAT? $(p - value = 0.0231)$

Q.2 **(a)** The cost of a previously owned car depends upon factors such as make and model, model year, mileage, condition, and whether the car is purchased from a dealer or from a private seller. To investigate the relationship between the car's mileage and the sales price, data were collected on the mileage and the sale price for 10 private sales of model year 2000 Honda Accords. **(07)**

Miles (1000s)	90	59	66	87	90	106	94	57	138	87
Price ($1000s)	7.0	7.5	6.6	7.2	7.0	5.4	6.4	7.0	5.1	7.2

1. Use the least squares method to develop the estimated regression equation.
2. Provide the interpretation for the slope of the estimated regression equation.
3. Predict the sales price for a 2000 Honda Accord with 1,00,000 miles.

E-936B-3

(b) The following data gives the percentage women working in five companies in retail and trade industry. The percentage of management jobs held by women in each company is shown below. **(07)**

% Working	67	45	73	54	61
% Management	49	21	65	47	33

Develop the estimated regression equation by computing values of b_0 and b_1.

OR

(a) The following data are the monthly salaries y and the grade point averages x for the students who obtained a bachelor's degree in business administration with a major in information systems. The estimated regression equation for these data is $\hat{y} = 1790.5 + 581.1x$. **(07)**

GPA	Monthly Salary ($)
2.6	3300
3.4	3600
3.6	4000
3.2	3500
3.5	3900
2.9	3600

1. Compute SST, SSR and SSE.
2. Compute the coefficient of determination r^2. Comment on the goodness of fit.
3. What is the value of the sample correlation coefficient?

(b) Given are five observations for two variables, x and y. **(07)**

x_i	1	2	3	4	5
y_i	3	7	5	11	14

The estimated regression equation for these data is $\hat{y} = 0.20 + 2.60\,x$.

1. Compute the Mean square error.
2. Compute the standard error of the estimate.
3. Compute the estimated standard deviation of b_1.

Q.3 **(a)** A shoe store data used to develop the model came from a survey of 10 stores; for these data SST = 16,000 and SSR = 12,000. Moreover, the following estimated regression were being developed by equation relating sales to inventory investment and advertising expenditures. **(07)**

$$\hat{y} = 25 + 10x_1 + 8x_2$$

where

x_1 = inventory investment ($1000s)

x_2 = advertising expenditures ($1000s)

y = sales ($1000s)

P.T.O

E-9368-19

1. Estimate sales resulting from a \$15,000 investment in inventory and an advertising budget of \$10,000.
2. Interpret b_1 and b_2 in this estimated regression equation.
3. For the estimated regression equation given, compute R^2.
4. Compute R_a^2.
5. Does the model appear to explain a large amount of variability in the data? Explain.

(b) Consider the following data for a dependent variable y and two independent variables, x_1 and x_2. **(07)**

x_1	x_2	y
10	5	0
20	7	12
30	10	20
40	14	24
50	15	40
60	21	36
70	20	60

1. Develop an estimated regression equation relating y to x_1. Estimate y if $x_1 = 90$
2. Develop an estimated regression equation relating y to x_2. Estimate y if $x_2 = 25$
3. Find SSE, SST and SSR for both x_1 and x_2.
4. Compute R^2 and R_a^2 for both x_1 and x_2.

OR

(a) A 10-year study conducted by the American Heart Association provided data on how blood pressure, and smoking relates. Assume that the following data are from a portion of that study. For the smoking variable, a dummy variable with 1 indicating a smoker and 0 indicating a nonsmoker is already defined **(07)**

Sr. No.	Blood Pressure	Smoker
1	152	0
2	163	0
3	155	0
4	177	1
5	196	0
6	189	1
7	155	1
8	120	0
9	135	1
10	98	0
11	152	0
12	173	1

Find equations to predict smokers and non-smokers.

(b) Data for two variables, x and y, follow: **(07)**

x_i	22	24	26	28	40
y_i	12	21	31	35	70

1. Develop the estimated regression equation for these data.
2. Compute the leverage values for these data. Do there appear to be any influential observations in this data? Explain.
3. Compute Cook's distance measure for these data. Are any observations influential?

Q.4 **(a)** Consider the following time series data. **(07)**

Week	1	2	3	4	5	6
Value	18	13	16	11	17	14

1. Construct a time series plot. What type of patten exists in the data?
2. Develop the two-week moving average forecasts for this time series. Compute MSE and a forecast for week 7.
3. Use $\alpha = 0.2$ to compute the exponential smoothing forecasts for the time series. Compute MSE and a forecast for week 7.

(b) Consider the following time series data. **(07)**

Period	**Value**
1	6
2	11
3	9
4	14
5	15

Construct a time series plot. What is the forecast for 6^{th} period and also develop the linear trend equation.
Also use Holt's linear exponential smoothing method with $\propto = 0.2$ and $\beta = 0.2$ to develop a forecast for 6^{th} Month.

OR

(a) Collected the following information on the number of tips he has collected from parking cars the last seven nights. **(07)**

Day	**Tips**
1	18
2	22
3	17
4	18
5	28
6	20
7	12

P.T.O

1. Compute the 3-day moving averages for the time series.
2. Compute the mean square error for the forecasts.
3. Forecast John's tips for day 7.

(b) The weekly demand for a particular brand of automatic dishwasher detergent for a chain of grocery stores located in Columbus follows. **(07)**

Week	Demand
1	22
2	18
3	23
4	21
5	17
6	24
7	20
8	19
9	18
10	21

1. Use three weeks moving average to develop a forecast for week 11.
2. Use exponential smoothing with a smoothing constant of $\propto = 0.2$ to develop a forecast for week 11.
3. Which of the two methods do you prefer? Why?

Q.5 Attempt any **SEVEN** out of **TWELVE** (Each carries **TWO** marks): **(14)**

(1) The interval estimates of an individual value of y for a given value of x is
a. prediction interval estimate
b. confidence interval estimate
c. average regression
d. x versus y correlation interval

(2) Write down formula for Cook's distance measure for multiple linear regression.

(3) In a regression and correlation analysis if $r^2 = 1$, then
a. SSE must also be equal to one
b. SSE must be equal to zero
c. SSE can be any positive value
d. SSE must be negative

(4) Which of the following is the cyclic behavior of time series?
a. Level
b. Trend
c. Seasonality
d. Noise

(5) The last period's forecast was 70 and demand was 60. What is the simple exponential smoothing forecast with alpha of 0.4 for the next period?

(6) A regression analysis is inappropriate when _____
a. you have two variables that are measured on an interval or ratio scale.
b. you want to make predictions for one variable based on information about another variable.
c. the pattern of data points forms a reasonably straight line.
d. there is heteroscedasticity in the scatter plot.

(7) When we can say that time series is stationary time series?

(8) For the ANOVA table

Source of variations	Sum of squares	Degree of freedom
Between treatment	45	3
Error	32	16
Total	99	19

The F-statistics is _____.

(9) Write an equation to predict dependent variable for logistic regression.

(10) If $t_{\frac{\alpha}{2}} = 2.305$, $s_{pred} = 14.69$ and $\hat{y}^* = 110$ then find the value for prediction interval.

(11) If SSE=1,530 and SSR=14,200 then what is the value of F-statistics if degree of freedom is 8 and 1 respectively.

(12) If the demand is 100 during October 2016, 200 in November 2016, 300 in December 2016, 400 in January 2017. What is the 3-month simple moving average for February 2017?

**** ****