

B.E. FULL-TIME DEGREE END SEMESTER EXAMINATIONS, APRIL 2011

INDUSTRIAL ENGINEERING

SIXTH SEMESTER- (REGULATION-2008)

IE9351 MULTI-VARIATE STATISTICAL ANALYSIS

(Use of Statistical Tables are permitted)

Time: 3 hr

Max. Mark: 100

Answer ALL Questions

Part-A (10 X 2 = 20 Mark)

1. What are the two assumptions used in interpreting the standard error of estimate in regression analysis?
2. How will you measure dispersion around the multiple regression plane?
3. List few examples for the usage of hypothesis in multivariate analysis?
4. What are the properties of multi-variate normal distribution?
5. What are the underlying assumptions in the structure of the data for one way MANOVA?
6. What is average linkage clustering method?
7. Define Factor Scores
8. How the weights are chosen in the PCA?
9. List few examples for Discriminant Analysis.
10. Define Discriminant Ratio.

Part-B (5 X 16 = 80 Marks)

11. How will you assess the assumptions of normality? How will you construct the Q-Q plot for the ten samples of ordered observations: -1.00, 0.10, 0.16, 0.41, 0.62, 0.80, 1.26, 1.54, 1.71 and 2.30. Explain the steps involved in evaluating bivariate normality and constructing a chi-square plot? (16)

12 (a) Explain the procedure involved in the computation of Multi-Variate Two Way Fixed Effects Model With Interaction? (16)

OR

12 (b) (i) Write a brief note on ROC Algorithm? (4)

(ii) In a survey, the number of years of experience of employee in two different skills is summarized in table 1. Perform cluster analysis using the agglomerative hierarchical clustering method. (12)

Table 1: Survey data about Number of Years of Experience on Skills

Employee	Skill-X	Skill-Y
1	2	8
2	8	15
3	3	6
4	6	9
5	8	7
6	10	10

13 (a) In finance, it is of interest to look at the relationship between Y, a stock's average return, and X, the overall market return. The slope coefficient computed by linear regression is called the stock's beta by investment analysts. A beta greater than 1 indicates that the stock is relatively sensitive to changes in the market; a beta less than 1 indicates that the stock is relatively insensitive. For the following data, compute the beta and test to see whether it is significantly less than 1. Use alpha equal to 0.05. (8)

Y (%)	10	12	8	15	9	11	8	10	13	11
X (%)	11	15	3	18	10	12	6	7	18	13

(ii) in a regression problem with a sample size of 17, the slope was found to be 3.73 and the standard error of estimate 28.654. The quantity $(\sum X^2 - n\bar{X}^2) = 871.56$

- Find the standard error of the regression slope coefficient (5)
- Construct a 98 percent confidence interval for the population slope (3)

OR

13 (b) Below are some data (table 2) on consumption expenditures, CONSUMP, disposable income, INCOME; and sex of the head of household, SEX of 12 randomly chosen families. The variable GENDER has been coded:

$$\text{GENDER} = \begin{cases} 1 & \text{if SEX = 'M' (male)} \\ 0 & \text{if SEX = 'F' (female)} \end{cases} \quad (16)$$

- Determine the best fitting regression
- If disposable income is held constant, is there a significant difference in consumption between households headed by a male versus those where the head of household is female? State explicit hypothesis test them at the 0.10 level, and state an explicit conclusion.
- Give an approximate 95 percent confidence interval for consumption for a household with disposable income of \$40,000 headed by a male.

14 (a) The potential customers of a computer company rate the product of the company into good or bad based on the time to respond to breakdown calls (X1) & the percentage discount on product price (X2). The ratings by customers are presented in the following table 3 (16)

- (a) Design the Discriminant function
- (b) Compute discriminant ratio, K and identify the variable which is more important in relation to the other variable
- (c) Validate the discriminant function using the given data by forming groups based on the critical discriminant score
- (d) Test whether the group means are equal in importance at a significance level of 0.05

Table 3: Customer data

Customer	Rating	X1 (in hours)	X2 (%)
1	Good	24	5
2	Good	12	8
3	Bad	36	4
4	Bad	12	0
5	Bad	36	3
6	Good	36	10
7	Good	24	3
8	Bad	48	4
9	Bad	96	5
10	Good	36	12

OR

14 (b) (i) Campus stores has been selling the Believe It or Not: Wonders of Statistics Study Guide for 12 semesters and would like to estimate the relationship between sales and number of sections of elementary statistics taught in each semester. The following data have been collected are shown in Table 4:

- (a) Develop the estimating equation that best fits the data (4)
- (b) Calculate the sample coefficient of determination and the sample coefficient of correlation (4)
- (ii) Write down the procedure for P mean Clustering Algorithm (8)

15 (a) In a survey on a product, data on three variable have been collected, as shown in the following table 5. Perform factor analysis using centroid method to identify two factors which can represent the variables of the study. (16)

OR

15 (b) Explain the Step-by-Step Method of PCA in detail? (16)

Table 2

Consump	Income (\$)	Sex	Gender
37,070	45,100	M	1
22,700	28,070	M	1
24,260	35,000	F	0
30,420	35,000	M	1
17,360	18,860	F	0
33,520	41,270	M	1
26,960	32,940	M	1
19,360	21,440	F	0
35,680	44,700	M	1
22,360	24,400	F	0
28,640	33,620	F	0
39,720	46,000	M	1

Table 4: sales data

Sales (units)	Number of sections
33	3
38	7
24	6
61	6
52	10
45	12
65	12
82	13
29	12
63	13
50	14
79	15

Table 5: Survey data

Respondent	Variable		
	X1	X2	X3
1	5	7	8
2	3	3	5
3	3	1	5
4	1	1	5
5	3	2	4
6	3	3	5
7	2	2	1
8	6	6	5
9	4	2	1
10	4	3	2