INDIAN STATISTICAL INSTITUTE

MS QMS

TEST ON MULTIVARIATE DATA ANALYSIS

Date: 04 May, 2015 Time: 3 hours Maximum Marks: 50

Answer as many questions as you can. Maximum you can score is 50 marks

1. An online retailer wants to enhance the mobile handset business. The company has conducted a conjoint analysis to obtain the customer preferences. The details of the conjoint analysis with aggregate ranking are given below. Analyze the data, compute part worth utilities and importance scores. What suggestions can you provide to the retailer to optimize the mobile handset business?

Combinations	Brand	Operating System	Screen Size	Camera Resolution	Score
1	Nokia	Windows	5.5 inches	Above 6 MP	1.2
2	Nokia	Android	5.0 inches	4 - 6 MP	1.5
3	Nokia	iOS	4.5 inches	2 - 4 MP	1.7
4	Sony	Windows	5.0 inches	2 - 4 MP	1.3
5	Sony	Android	4.5 inches	Above 6 MP	4.4
6	Sony	iOS	5.5 inches	4 - 6 MP	6.9
7	Apple	Windows	4.5 inches	4 - 6 MP	4.3
8	Apple	Android	5.5 inches	2 - 4 MP	6.8
9	Apple	iOS	5.0 inches	Above 6 MP	9.5

[12]

2.

a. Explain multicollinearity? How can it be detected? Briefly explain four different methods to handle multicollinearity?

b. Describe any two methods of cross validation commonly used in predictive modeling?

or

c. Describe at least one graphical technique for identifying the relationship between dependant variable (y) and independent variable (x) when

   i. both x and y are numeric

   ii. x is categorical and y is numeric

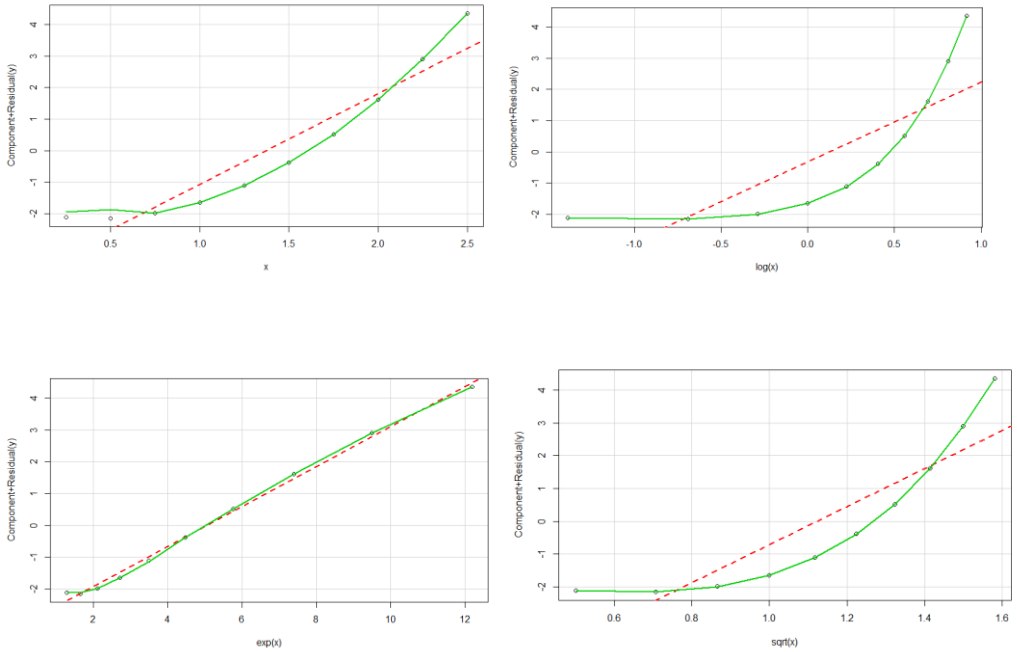
   iii. y is binary and x is numeric

   iv. both x and y are categorical

[12]

3. A model need to be developed for the weight loss (y) of a solid-fuel rocket propellant over time (x) in months of the form $y = \beta 0 + \beta f(x)$. The model coefficients and crPlots for different candidate models are given below. Choose the best model from the candidate models. Give reasons for your choice.

f(x)	β_0	β
x	-0.422	2.8778
log(x)	3.2184	2.5503
exp(x)	0.35476	0.62908
sqrt(x)	-2.979	5.798



Validate the model with the following data. Plot actual vs predicted, compute MSE, RMSE and give your interpretations?

x	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
y	1.42	1.39	1.55	1.89	2.43	3.15	4.05	5.15	6.43	7.89

[12]

4. A model is developed to classify the customer seeking overdraft from their bank account. The response variable is whether the customer will repay the overdraft (1: success) or not (0:failure). The predictor variables are Individual expected level of activity score (x1), Transaction speed score (x2) and Peer comparison score in terms of transaction volume (x3). The table of coefficient for best fit model is given below

	Estimate	Std. Error	z	p value
(Intercept)	-35.5062	4.4061	-8.058	0.00000
x1	2.7957	0.3553	7.868	0.00000
x2	2.7532	0.3427	8.033	0.00000
x3	3.5153	0.4343	8.095	0.00000

Identify the model for predicting the probability of success p(y). Write down the criteria for classifying whether a customer will repay the overdraft or not based on p(y) value. Validate the model with the following test data

x1	6.2	2.6	9.5	2.6	10	0.2	2.5	0.4	0.2	8.9	9.3	9.4
x2	9.3	2.2	1.5	5	7.7	2.3	2.8	1.6	9.8	6.1	5.2	2.1
x3	7.4	8.7	8.2	0.4	7.2	3.4	4.6	5.3	0.8	9.2	2.1	4.1
y	1	1	1	0	1	0	0	0	0	1	1	1

[12]

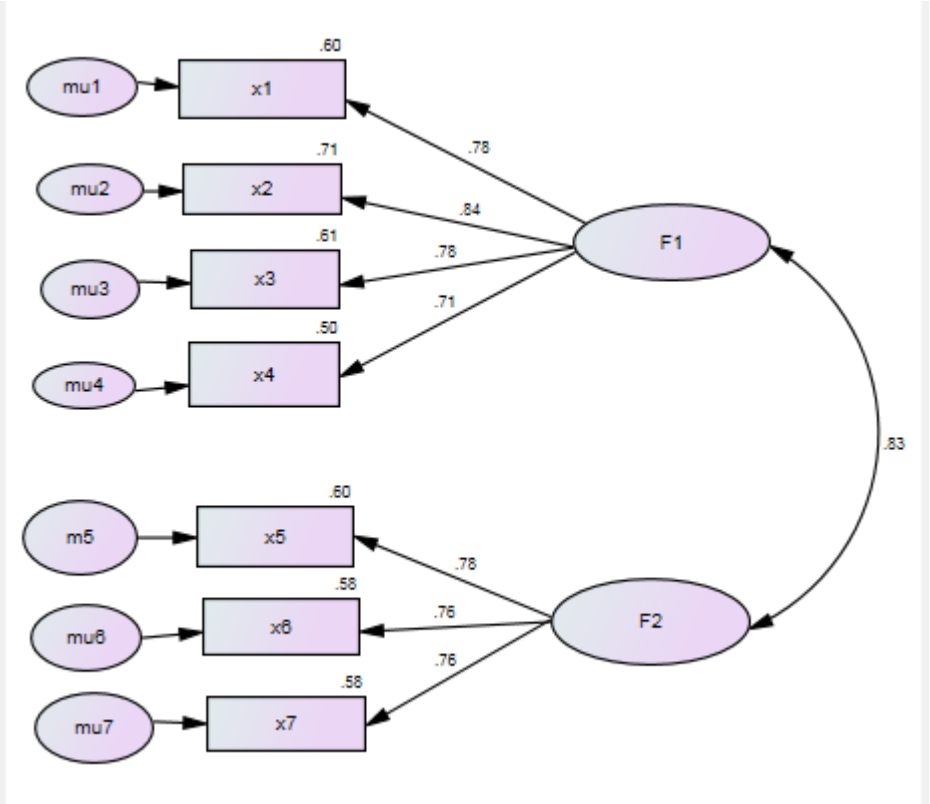
5.

a. Explain confirmatory factor Analysis? How it is different from exploratory factor analysis?

b. An ITeS company feels that the customer satisfaction depends on two unobservable or latent factors namely Operational Efficiency (F1) and SLA Management (F2). The customers are requested to rate in a 5 point scale (1: Very Dissatisfied to 5: Very Satisfied) four questions related to operational efficiency and three questions related to SLA management. The questions are given below.

Operational Efficiency (F_1)	Ability to deliver seamlessly in the event of changes (volume fluctuations, resource movement etc) (x_1)
	Operational performance(x_2)
	Application of process knowledge(x_3)
	Effectiveness of the reviews around operations delivery(x_4)
Service Level Agreement Management (F_2)	Ability to meet service level agreements(x_5)
	Communication with you(x_6)
	Flexibility and responsiveness to special service requests(x_7)

The output of the structural equation modeling for the above integrated model is given below: Interpret the results. Give suggestions to improve the model



Goodness of fit test p value	GFI	AGFI	RMSE
0.077	0.946	0.883	0.027

[12]