Basic statistics
Jan-2010

Dec. PR 201

Con. 6110-09.

BB-9212

(3 Hours) **[Total Marks : 60**

N.B. : (1) **All** questions are **compulsory**.
(2) Use of simple calculator is **allowed.**
(3) **Figures** to the **right** indicate **full** marks.
(4) Answer to the **two** sections should be written on **separate** answer books.

Section I

1. (a) The following data give the time (in minutes) that each of 20 students took to complete a statistics test. **10**

55	49	53	59	38	56	39	58	47	53
58	42	37	43	44	55	51	46	45	47

Construct a stem-and-leaf display for these data. Find Q_1, Q_2, Q_3 and interquartile range.

(b) Following table gives the classification of all employees of a company by gender and college degree.

	Graduate	**Not Graduate**
Male (M)	7	20
Female(F)	4	9

If one employee is selected at random find the following probabilities :–
(i) Graduate
(ii) Graduate given male
(iii) Male and not graduate
(iv) Not graduate or female.
(v) Are the events "Graduate" and "Male" independent ? Why ?

(c) Customers arrive at a checkout counter in a department store according to a Poisson process at an average of 7 per hour. What are the probabilities that–
(i) in first hour at least two customers arrive.
(ii) in first half an hour exactly three customers arrive.

2. (a) Suppose that Y has following p.d.f. **10**

$$f(y) = cy \quad , \quad 0 < y < 2,$$
$$= 0 \quad , \quad \text{otherwise}$$

(i) Find C
(ii) Find mean, variance of Y.
(iii) Find median of this distribution.

(b) 5% of all VCRs manufactured by a large electronics company are defective. Five VCRs are randomly selected from the production line. What is the probability that –
(i) No VCR is defective out of these five VCRs.
(ii) Exactly one VCR is defective out of these five VCRs.

(c) The number of typing errors made by a typist per page is denoted by X. Suppose the typist makes on an average 4 errors per page. Find MGF and PGF of X.

[TURN OVER

3. (a) Let (X, Y) have a following joint probability mass function. 10

Y \ X	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

(i) Find marginal and conditional p.m.f. of both X and Y.
(ii) Find covariance between X and Y.

(b) Let X and Y have the following pdf.

$f(x, y) = 1, \quad 0 < x < 1, \quad 0 < y < 1$
$= 0 \quad , \quad$ otherwise

(i) Are X and Y independent ? Why ?
(ii) Find $P(\cdot 1 < x < \cdot 3, \ 0 < y < \cdot 5)$
(iii) What is covariance between X and Y ?

(c) An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2·5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed ·5 inches.

Section II

4. (a) Let $x_1, \ldots, x_n$ denote a random sample from the p.d.f. 10

$f(x, \theta) = \theta x^{\theta - 1} \quad , \quad 0 < x < 1, \quad \theta > 0,$
$= 0 \quad , \quad$ otherwise

(i) Find MLE of θ and CRLB for the variance of an unbiased estimator of θ.
(ii) Estimate θ using method of moments. Is it unbiased ? Find its variance.

(b) Suppose that $x_1, \ldots, x_n$ denote a random sample from the B(n, p) distribution. (n known). Obtain moment estimator and MLE of P and CRLB for the variance of the unbiased estimator of P.

5. (a) A chemical process has produced, on the average 800 tons of chemical per day. The daily yields for the past week are 785, 805, 790, 793 and 802 tons. Do these data indicate that the average yield is less than 800 tons and hence that something is wrong with the process ? (use $\alpha = \cdot 05$). State the assumptions. Construct 95% confidence interval for the average yield μ. 10

(b) A randomly selected sample of 100 persons who suffer from allergies were asked during what seasons they suffer the most. The results are given in the following table.

Season	Fall	Winter	Spring	Summer
Persons allergic	18	13	31	38

Using $\alpha = \cdot 01$, test the null hypothesis that the proportions of all allergic persons are the same over the four seasons.

6. (a) Alpha Airlines claims that only 15% of its flights arrive more than 10 minutes late. Let p be the proportion of all of Alpha's flights that arrive more than 10 minutes late. Consider the hypothesis test. 10

$H_0 : p = \cdot 15$ against $H_1 : p > \cdot 15$

Suppose we take a random sample of 50 flights by Alpha Airlines and agree to reject H_0 if 9 or more of them arrive late. Find the level of significance for this test. Find the probability of type II error when $p = \cdot 17$.

(b) Following table gives the blood pressures of seven adults before and after the completion of a special dietary plan.

Before	210	180	195	220	231	199	224
After	193	186	186	223	220	183	233

Let μ_d be the mean of the population difference (before – after). Using $\alpha = \cdot 05$ test $H_0 : \mu_d = 0$ against $H_1 : \mu_d > 0$. Also construct 95% confidence interval for μ_d. State the assumptions you made to answer this question.

V-Ex-II-09-Scan-D-51

2 marketing Research

Con. 6113-09. January 2010 **BB-9227**

(3 Hours) **[Total Marks : 60**

N.B. : (1) **All** questions are **compulsory**.
(2) Use of simple calculators is allowed.
(3) **Figures** to the **right** indicate **full** marks.
(4) Answers to the **two** sections should be written on **separate** answer-books.

Section I

1. Explain :— **10**
   (a) Importance of Marketing Research.
   (b) For the variety of marketing decisions listed below, please assign type of research which can be undertaken for the following :—
   (i) Target markets (ii) Product (iii) Distribution (iv) Promotion (v) Pricing.

2. Explain how SWOT analysis helped "Wal-Mart" in planning and marketers to focus on key issues. **10**

3. Explain in detail "Overview of syndicated services". **10**

Section II

4. (a) Explain Qualitative Marketing Research. **3**
   (b) What is a measure of central tendency ? State two such measures of central tendency. **2**
   (c) State three reasons why it is necessary to carry out univariate analysis before complex analysis. Explain dependant and independent variables in multi-variate analysis. State two appropriate techniques when the data contains only independent variables. **5**

5. (a) What is the use of Bartlett's test of sphericity and KMO measure in factor analysis ? **4**
   (b) State 2 guidelines to decide the number of dimensions in MDS analysis. **2**
   (c) A discriminant analysis having the 5 independent variables, stated below was conducted :— **4**
   (1) Respondent' sex
   (2) Age of the respondent
   (3) Highest year of school completed by the respondent
   (4) Total family income of the respondent and
   (5) Whether the respondent thinks of himself as liberal or conservative Based on the output below, interpret :—
   (i) The canonical correlation.
   (ii) How many groups are there of the dependent variable ?
   (iii) Which of the above variables is the most discriminating variable ?

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.164	100	100	0.376

Standardized Canonical Discriminant Function Coefficients	
	Function
	1
Respondent's Sex	0.011
Age of Respondent	0.657
Highest Year of School Completed	0.712
Total family Income	0.423
Think of Self as Liberal or Conservative	0.018

6. (a) Which are the 3 types of clustering procedures'? Explain Hierarchical clustering. Which are the 3 types of linkage methods in Hierarchical clustering ? **4**
   (b) State the steps in conducting Conjoint Analysis. **2**
   (c) State any two reasons of why Report/Presentation is important for any marketing research project. **4**
   What is Executive Summary in the marketing research report ?

———S———

197 : 2nd Exm.09-DD-(A)

III Regression and Linear model

Con. 6115-09. Jan. 2010. **BB-9233**

(3 Hours) **[Total Marks : 60**

N.B. (1) **All** questions are **compulsory**.
(2) Use of **simple** calculators is **allowed**.
(3) **Figures** to the **right** indicate **marks**.
(4) Answer to the **two** sections should be written on **separate** answer books.

Section I

1. (a) Relationship between Y and X is given by — **2**

$Y = 50 - X^2$

Considering values of X as –4, –3, –2,–1, 0, 1, 2, 3, 4 generate values of Y and find correlation coefficient r between Y and X. Give explanation for value of r and relationship between Y and X.

(b) An executive (HR) of a life insurance company desires to analyse the salaries of 30 acturial staff employees in the company. The salary should be based on the number of acturial papers cleared and years of professional experience in life insurance company. Suggest a suitable statistical model defining the symbols used. **2**

(c) Multiple linear regression is studied using 25 observations. Give procedure for testing the following hypotheses in the model — **6**

$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_5 X_5 + e$

(i) $\beta_1 = \beta_2 = \ldots\ldots = \beta_5 = 0$

(ii) $\beta_1 = \beta_4 = 0$

(iii) $\beta_1 - \beta_3 = 1$ given $\beta_2 = \beta_4 = \beta_5 = 0$

2. (a) In simple regression if roles of X and Y are interchanged show that value of R^2 remains same. **2**

(b) What is heteroscedasticity ? How do you remove it if $Var(e) = K^2 X^2$? **2**

(c) Describe how VIF values can be used to detect multicollinearity and comment when — **6**

(i) $Rj^2 = 0$ and (ii) $Rj^2 = 0{\cdot}9$

where j stands for, predictor variable number. Also state criteria in terms of eigen values.

3. (a) In multiple linear regression model how do you express fitted values in terms of observed responses using hat matrix. State the equations and elements of hat matrix in case of simple regression. **5**

(b) To study relationship between continuous random variable age in years (X) and presence or absence of coronary heart disease (CHD) 100 subjects were selected. Dependent variable Y is defined as — **5**

$$Y = \begin{cases} 1 \text{ if CHD is present} \\ 0 \text{ if CHD is absent} \end{cases}$$

Computer output of logistic regression is given below :—

Variable	Coeff	S.e.	z	p value
Age	0·111	0·0241	4·61	< 0·001
Constant	–5·309	1·1337	–4·68	< 0·001

Estimate $\pi(X)$ when X = 26, 55 and 75. How would you classify following observations :—

Age X	CHD Y
26	0
55	0
75	1

[TURN OVER

Section II

4. The time of reaction to different stimuli (S_1, S_2, S_3,S_4) was being studied in a CRD. The reaction time measured in seconds are as follows :—

	S_1	S_2	S_3	S_4	
	1·7	1·6	2·2	0·9	
	1·5	1·7	1·9	0·8	
	1·5	1·8	2·3	0·8	
	1·4	1·6	2·1	0·9	
Total	6·1	6·7	8·5	3·4	24·7

Sum of squares of all observations : 40·525

(a) Test for the significant differences between the means of stimuli at 5% level of significance. (F table value at 5% is 3·49). **3**

(b) Give one procedure for testing the homogenity of variance. **2**

(c) Explain one post hoc test to determine the significant pair of stimuli. **2**

(d) Write SAS program to carry out the ANOVA. **3**

5. The CEO of large chain of super market in the city is concerned about the waiting time for the customers checking out. The dependent variable is waiting time for each customer before a cashier begins service. A total of 90 customers being studied were evenly divided among three stores (S_1, S_2, S_3) and three timing (t_1, t_2, t_3).

(a) Write down two-way classification model to analyse the data. What are the hypotheses and what do you conclude ? What difference if any should concern the CEO ? **3**

(b) Outline computational procedure to carryout the analysis given in (a), namely ANOVA table and post hoc procedure to detect significant difference. **4**

(c) Write NINITAB program to carryout the ANOVA. **3**

6. Suppose we want to study the effects of two factors A and B set at a and b levels on a response variable Y. There are r observation per cell :—

(a) Write appropriate model when both factors are random and state the assumptions. **2**

(b) Give the procedure to various sum of squares and write down ANOVA table. **3**

(c) Obtain expected mean square and estimate of variance components. **3**

(d) State the hypothesis which can be tested and give the procedure to carryout the same. **2**

IV Decision making and Forecasting
Jan. 2016

Con. 6117-09.

BB-9239

(3 Hours) **[Total Marks : 60**

N.B. (1) **All** questions are **compulsory.**
(2) Use of **simple** calculators is **allowed.**
(3) **Figures** to the **right** indicate **marks.**
(4) Answers to the **three** sections should be written on **separate** answer books.

Section I

1. A ten-year old luxury hotel wants to plan a business strategy for the next year. How will you proceed to forecast its quarter-wise sales next year ? **10**

2. Which of the following five statements are correct ? Mention your opinion as 'CORRECT' or 'FALSE' against each statement number. Also, give reasons to support your opinion. **10**
Statement Number 2.1 : Leading indicator model of regression can be effectively tried to forecast the quantum of a bank's deposits as of 31st March 2012 from the number of its branch offices on 31st March 2009.
Statement Number 2.2 : The minimum mean square error test in a multiplicative time series model: Y = T x S x C x E should check whether the error term E is close to zero.
Statement Number 2.3 : Strong correlation between two variables does not necessarily mean that one of them is the cause for the other.
Statement Number 2.4 : Moving averages of period 5 (five) should effectively smoothen the quarterly sales figures of steel.
Statement Number 2.5 : Forecasting future values of a variable is possible even if past quantitative data on it are not available.

3. (a) Describe the Six Thinking Hats technique for making business decision in a complex real life situation. **10**
(b) Apply this technique to a software development company which wants to make long term decisions about undertaking unrelated diversification projects.

Section II

4. Describe three techniques of inventory control with selective control. Explain ABC analysis and how it is useful in inventory management. **10**
Give one example of ABC analysis.

5. For a fixed order quantity system, find out the Economic Order Quantity (E.O.Q.), optimum buffer stock, normal lead time consumption and re-order level with the following data; **10**

Annual consumption in units	= 6400
Cost in Rs. one unit	= 1
C_3 (Set up cost) in Rs. per production run	= 12
C_1 (holiding cost) in Rs. per unit	= 0·24

Past lead times : 10 days, 20 days, 8 days, 9 days, 25 days, 12 days.

[TURN OVER

Section III

6. A Ghana based soft drink company is planning to establish a subsidiary company in India to produce mineral water. **10**
Based on the estimated annual sales of 40,000 bottles of the mineral water, cost studies produced the following estimates for the Indian subsidiary.

Particulars	Total Annual Cost	Percentage of Total Annual Cost which is variable
Material	2,10,000	100%
Labour	1,50,000	80%
Factory overheads	92,000	60%
Administration Expenses	40,000	35%

The Indian production will be sold by the manufacturer's representatives who will receive a commission of 8% of the sales price. No portion of the Ghana office expenses is to be allocated to the Indian subsidiary

Required :

(a) Compute the sales price per bottle to enable the management to realise an estimated 10% profit on sale proceeds in India.

(b) Calculate the Break Even Point in Rupee sales as also in number of bottles for the Indian subsidiary on the assumption that the sales price is Rs. 14/- per bottle.

V Statistical Process Control

January 2010

Con. 6167–09. BB–9132

(3 Hours) [Total Marks : 60

N.B. (1) **All** questions are **compulsory**.
(2) Use of **simple** calculators is **allowed**.
(3) **Figures** to the **right** indicate marks.
(4) Answer to the **two** sections should be written on **separate** answer-books.

Section I

1. A medical company produces a part that has a hole measuring 0·5" +/–0·050". The tooling used to make the hole is worn and needs replacing, but management doesn't feel it necessary since it still makes "good parts". All parts pass QC, but several parts have been rejected by assembly. A failure cost per part is $0·45. Using the loss function, explain why it may be to the benefit of the company and customer to replace or sharpen the tool more frequently. Find k ? **10**

2. A hospital may have a greater cost for nonconformance than a shopping mall. A ratio of 0·005 is used for the purposes of this paper. This would produce a cost of $5,000 for non-conformance for every $1,000,000 of project cost. Specification is established as 1·00 + / –0·02. Calculate : Cost to the owner if the project went 4 weeks beyond the planned. **10**

Section II

N.B. (1) Use of **Mini Tab**, **Excel**, **Excel based templates** and **scientific** calculator is **allowed**.
(2) Write down Null and Alternative Hypothesis for the test, name of the test used, basis of test and conclusion apart from analysis output from session window.

3. Describe following :— **10**
   (a) Use of FMEA technique and how to estimate rpn no.
   (b) Team and meeting effectiveness.

4. FMCG company wants to know if there is a difference in the price of a bath soap at 3 different types of stores. The price of the soap was checked in a sample of 5 at each type of store and the results are given below. Help the company to conclude the same. **10**

Store A	12	13	14	12	15
Store B	15	17	14	18	17
Store C	19	17	16	20	19

5. An accounts firm knows that homogeneity in work experience affects effectiveness of account executives. It collects the following data on years of experience for its executives in office 1 and 2. Which office is more effective ? **10**

Office 1	4	3	6	5	5	2	5	4	6	4
Office 2	7	2	3	4	7	4	2	4		

6. The yield of a chemical process is related to the concentration of the recent and the operating temperature. Given below are the results of an experiment. Find out the significant factors and levels for next level DOE. **10**

Yield	Concentration	Temperature
81	1	150
89	1	180
83	2	150
91	2	180
79	1	150
87	1	180
84	2	150
90	2	180

6. medical statistics
Jan. 2010

Con. 6163–09. BB–9125

(3 Hours) [Total Marks : 60

N.B. (1) **All** questions are **compulsory**.
(2) Use of **simple** calculators is **allowed**.
(3) **Figures** to the **right** indicate marks.

1. Briefly explain the application of following tests/methods in analysis of clinical research data : (Attempt any **four** of five and **2.5** marks **each**) **10**
   (a) Fisher's exact test
   (b) Wilcoxon signed rank test
   (c) Logistic regression
   (d) Log rank test
   (e) Two way ANOVA

2. Describe in detail, role of a statistician at various stages of execution of a clinical trial project. **10**

3. Answer following questions (**5** marks **each**) :— **10**
   (a) Explain P-value, confidence interval, type II error and power of a test.
   (b) What is the difference between crossover design and parallel design ? Describe benefits and drawbacks of each.

4. What are ICH guidelines ? Briefly discuss the purpose of ICH E9, E6 and ICH E3 guideline. **10**

5. Describe the various data gathered on a Case Report form [CRF] during the course of a clinical trial. Provide an overview of safety data collected in a clinical trial. **10**

6. Select the **correct choice** for the following questions. **Each** question has **one** correct choice :— **10**
   (i) Which one of the following is not a valid SAS variable name ?
      (a) _2var
      (b) 2_var
      (c) Var_2
      (a) i
      (b) ii
      (c) iii
      (d) all are valid SAS variable name
      (e) none of the above.

   (ii) A value for the numeric variable SiteNum is 32·3. Which statement correctly converts the values of SiteNum to character values when creating the variable SiteChar ?
      (a) Sitechar=input sitenum;
      (b) Sitechar=input(sitenum,4.1);
      (c) Sitechar=put sitenum;
      (d) Sitechar=put(sitenum,4.1);

[TURN OVER

(iii) Identify the missing.

```
Data exam;
    Number=roll/10;
Run;

Proc sort data=exam;
_____number;
Run;
```

(a) Set
(b) Merge
(c) Freq
(d) By

(iv) A trial was conducted at three different geographies to determine the effect of two doses of a new therapeutic agent MCGX on *frequent nighttime heartburn.* There were three treatment arms as follows MCGX 15mg, MCGX 30mg and Placebo. The response was measured using the PSQI **(Pittsburgh Sleep Quality Index) score.** Which of the following test is most appropriate to examine the question – Does the primary response differ significantly in the three treatment arms and amongst the three geographies ?

(a) ANCOVA
(b) Repeated Measures ANOVA
(c) Cross-Over ANOVA
(d) Two Way ANOVA

(v) A parallel design study of a new lipid lowering medication NO-LIP, was planned to compare NO-LIP and conventional therapy. 25 patients in each treatment arm were studied for 180 days of treatment. Change in Cholesterol (mg/100 ml) was observed at end of therapy in both the treatment arms. Is there any difference in response among the two groups ? How to analyze the data to come to any conclusion ?

(a) One sample t-test
(b) Log rank test
(c) Wilcoxon Signed rank test
(d) Two sample t-test.

VII Multivariate Techniques
January 2010

Con. 6165–09. BB–9128

(OLD & REVISED COURSE)

(3 Hours) [Total Marks : 60

N.B. : (1) **All** questions are **compulsory**.
(2) Use of **simple** calculators is **allowed**.
(3) **Figures** to the **right** indicate **marks**.
(4) Answer to the **two** sections should be written on **separate** answer books.

Section I

1. (a) A set of observations on two variables (X_1, X_2) give sample variances as $S_{11} = 4$ and $S_{22} = 1$. X_1 and X_2 are independent. 4
   (i) Define statistical distance of point (X_1, X_2) from origin.
   (ii) Give coordinates of five points which are at unit distance from origin.

   (b) Five observations on three variables X_1, X_2 and X_3 are given below. Find (i) Mean vector (ii) variance - covariance matrix (divisor n – 1) and correlation matrix— 6

X_1	X_2	X_3
9	12	3
2	8	4
6	6	0
5	4	2
8	10	1

2. (a) Write p.d.f. $f(X_1, X_2)$ of bivariate normal distribution having mean vector μ and var-cov. matrix Σ. 2

   (b) Write conditional distribution of X_1 given $X_2 = 5$. 3

   (c) X is distributed as $N_3(\mu, \Sigma)$. Find the distribution of $\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix}$. 5

3. (a) Describe method of tasting normality of given data using (i) Q – Q plot (ii) χ^2 – plot. 5

   (b) Why is transformation of data necessary ? State names of the transformations known to you. 2

   (c) Given sample $X_1, X_2, \ldots X_n$ from $N_p(\mu, \Sigma)$ obtain likelihood ratio test for mean vector of the distribution. 3

4. (a) Data on two variables X_1, X_2 is given below. Test the hypothesis $\mu_0 : \mu = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ 5

   against the alternative $H_1 : \mu \neq \begin{bmatrix} 9 \\ 5 \end{bmatrix}$. Use $\alpha = \cdot 05$.

$$X = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}.$$

$F_{2,\,1}(\cdot 05) = 18 \cdot 51$.

   (b) Using the given data, estimate regression coefficients when two response variables are modelled using single predictor variable. 5

X	0	1	2	3	4
Y_1	1	4	3	8	9
Y_2	– 1	– 1	2	3	2

[TURN OVER

Section II

5. (a) (i) Write down the orthogonal factor model with two factors. 3

(ii) Explain the terms communality, specific variance and factor loadings. 3

(b) What are principal components ? 4
Explain how they are useful in Factor Analysis.

6. (a) If π_1 and π_2 are two p variate normal populations $N(\mu_1, \Sigma)$ and $N(\mu_2, \Sigma)$ respectively.

(i) Define Δ^2 the distance between two populations. 1

(ii) State Fisher's classification rule. 2

(iii) If $\mu_1' = [2 \;\; 3] \;\; \mu_2' = [1 \;\; 4]$ 3

$$\Sigma = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$$

Compute Δ^2 and Fisher's classification rule.

(b) (i) Explain the problem of cannonical correlation with the help of an example. 2

(ii) Explain how MDS (Multidimensional scaling) is useful in marketing. 2

OR

(ii) Define Optimum Error Rate (OER) in the classification problem. 2

———v———