

- 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) The following data give the time (in minutes) that each of 20 students took to complete a statistics test— 4

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

- (i) Construct a stem-and leaf display for these data.
(ii) Determine the mean, median and mode.
- (b) Of the items produced daily by a factory, 40% come from line I and 60% from line II. Line I has a defect rate of 80%, where as line II has a defect rate of 10%. If an item is chosen at random from the day's production, find the probability that it will not be defective. 3
- (c) In a bolt factory machines A, B and C manufacture respectively 25%, 35% and 40%. Of the total of their output 5, 4, 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machine A, B or C ? 3
2. (a) Suppose the r.v Y has a following p.d.f. 3

$$f(y) = \begin{cases} \frac{3}{32}(y-2)(6-y) & 2 \leq y \leq 6 \\ 0 & \text{elsewhere} \end{cases}$$

Find mean and variance of Y.

- (b) Ten percent of the engines manufactured on an assembly line are defective. If engines are randomly selected one at a time and tested. What is the probability that the third non defective engine will be found— 3
- (i) on the fifth trial ?
(ii) on or before the fifth trial.
- (c) Let Y denote a Poisson random variable with mean λ . Find the probability-generating function for Y and use it to find $E[Y]$ and $V[Y]$. 4

[TURN OVER

3. (a) Two discrete random variable x, y have the following joint probability function—

y	x		
	1	2	3
1	0.2	0	0.2
2	0	0.2	0
3	0.2	0	0.2

Determine— (i) $E[x]$

(ii) Probability distribution of $y/x=1$

(iii) Are x and y correlated?

(iv) Are x and y independent?

- (b) The service times for customers coming through a checkout counter in a retail store are independent random variable with mean 1.5 minutes and variance 1.0. Approximate the probability that 100 customers can be served in less than 2 hours of total service time. 3

- (c) Let x and y have joint density function given by— 3

$$f(x, y) = \frac{3}{5} x(x+y) \quad 0 < x < 1, \quad 0 < y < 2$$

Determine the conditional expectation $E\left[\frac{y}{X} \mid X=x\right]$.

Section II

4. (a) Let x_1, \dots, x_n denote a random sample from the probability density function— 5

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Obtain MLE of θ .

(ii) Obtain moment estimator of θ .

Is it unbiased? Find its variance.

- (b) Let x_1, \dots, x_n be a random sample from $B(n, p)$ distribution. Obtain MLE of p and CRLB of variance of any unbiased estimator of p . Obtain moment estimator of P . Is it unbiased? Find its variance. 5

5. (a) A private agency claims that the crash course it offers significantly increases the writing speed of secretaries. The following table gives the scores of eight secretaries before and after they attended this course. 6

Before	81	75	89	91	65	70	90	64
After	97	72	93	110	78	69	115	72

Using $\alpha = .05$ can you conclude that attending this course increases the writing speed of secretaries? Also construct 95% confidence interval for μ_d ($d = \text{before} - \text{after}$). State the assumption you make.

- (b) Two hundred working people who had recently had heart attacks were surveyed. The day on which their heart attacks occurred appear in the following table. —

Sun	Mon	Tue	Wed	Thur	Fri	Sat
24	36	27	26	32	26	29

Test using $\alpha = .05$ whether the percentages of heart attacks that occur on different days of the week are same.

6. (a) The manager of super market wanted to know the percentage of shoppers who prefer to by name brand products. A random sample of 20 shoppers who shopped at this store were asked this question. Seven persons said 'yes'. Construct a 99% confidence interval for the percentage of all shoppers who prefer to by name brand products. State the assumption you make. 4
- (b) A sample of 12 observations taken from a normally distributed population produced the following data— 6
- 13, 15, 9, 11, 8, 19, 17, 9, 10, 14, 16, 12.
- What is a point estimate of σ^2 ? Construct 95% confidence interval for σ^2 .

N.B. : (1) All questions are compulsory.

(2) Use of simple calculators 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

Q 1. **Questionnaire design** to complete our product satisfaction survey. Your responses will help us to address any issues that you may have as well as to better target our products and services to meet your needs. Customer responses will be kept confidential.

Q 2. **Network diagram** – Furniture maker to produce new wooden framed settee with cloth-covered foam cushions.

Activity	Description	Time in days
A	Make wooden arms and legs	3
B	Make wooden back	1
C	Make wooden base	2
D	cut foam for back and base	1
E	make covers	3
F	fit covers	1
G	put everything together	1

P 3. Plot cumulative frequencies for "Price Sensitivity Meter"

Q 4. Draw "SWOT analysis" for –

- Delta Airlines Inc. chose to invest in multibillion –dollar customer service system that addresses flight delay problems experienced by 20% of passengers. Although some companies might think the move was excessive considering 80% of customers have no problems. Delta believed customer service was an important area for increasing market share and that competitors could pose a threat if Delta didn't address a problem

[TURN OVER]

Section II

Q. No. 5

- a) Describe principal components analysis and common factor analysis and the differences between the two methods of factor analysis. (10)
- b) Calculate eigenvalues and communalities from following table and interpret them (10)

Rotated Component Matrix^a

	Component			
	1	2	3	4
Recognition/rewards/growth opportunities for your performance is fair and not dependant on your gender.	.751	.051	.074	-.118
You are treated with respect as an individual by others in the organisation	.576	.013	-.027	.269
Decision making process is slow	-.534	-.113	.143	.452
In the Company, Teamwork is top priority and team work is given more importance than individual work	.522	.240	-.448	.116
Working late (inspite of no work pressure) is important to portray a positive image in the company	-.438	.722	-.154	-.082
You have adequate authority to make independent decisions necessary to do an effective and timely job, in your work area.	.301	.627	-.053	.123
In the company, you can voice your observations/opinions fearlessly, even if they are unpleasant.	.350	.554	.440	-.148
In the Company most decisions are taken top down	-.031	-.026	.876	.070
Being occasionally late to work acts as a negative impression in the company	.032	.040	-.018	.853

Q. No. 6

- a) Describe the use of hierarchical method and K –means method in cluster analysis. Also explain difference between both methods (10)
- b) Following is the output of cluster analysis of employee satisfaction survey. What conclusions company should draw based on it? (10)

ANOVA						
	Cluster		Error		F	Sig
	Mean Square	df	Mean Square	df		
You are satisfied with the responsiveness and assistance provided by Medical	53.633	1	.760	156	70.594	.000
You are satisfied with the responsiveness and assistance provided by Supply chain/logistics	5.194	1	.953	156	5.449	.021
You are satisfied with the responsiveness and assistance provided by Personnel	2.236	1	1.121	156	1.995	.160
You are satisfied with the responsiveness and assistance provided by Administration	5.153	1	.711	156	7.247	.008
You are satisfied with the responsiveness and assistance provided by HR	2.542	1	.991	156	2.564	.111
You are satisfied with the responsiveness and assistance provided by Finance	2.236	1	1.108	156	2.018	.157
You are satisfied with the responsiveness and assistance provided by IT	3.278	1	1.449	156	2.263	.135

Number of Cases in each

Cluster	
Cluster 1	2.000
Cluster 2	156.000
Valid	158.000
Missing	18.000

III Regression and Linear Model

January 2013

BB-8946

(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) Assume regression model $Y = \beta_0 + \beta_1 X + e$ where $e \sim N(0, \sigma^2)$ and n observations. 3
 (a) Prove that regression $s.s = \hat{\beta}_1^2 \sum (x_i - \bar{x})^2$ 7
 (b) Data on 10 observations gives -

$$\sum (y_i - \hat{y}_i)^2 = 3.267, \quad \sum y_i = 63.3, \quad \sum y_i^2 = 423.49$$

$$\sum x_i = 139, \quad \sum x_i^2 = 2239, \quad \sum x_i y_i = 957$$

- (i) Estimate parameters β_0 and β_1 .
 (ii) Find $\text{Var}(\hat{\beta}_0)$ and s.e. $(\hat{\beta}_1)$
 (iii) Obtain coefficient of determination
 (iv) Estimate σ^2 .
2. (a) Multiple linear regression problem has six predictors and 30 observations. 5
 Regression and total sum of squares are respectively 3147.97 and 4296.97. Find multiple correlation coefficient. Find the value of test statistic for testing the hypothesis that all predictor variables have no explanatory power.
 (b) A study of regression of Y on X_1, X_2 and X_3 gave following values of VIFs. 5
 $\text{VIF}_1 = 469.7, \text{VIF}_2 = 1.0, \text{VIF}_3 = 469.4$ comment on collinearity of the data.
3. (a) Why is transformation of variables necessary? Explain what is heteroscedasticity 5
 and how is it removed by using transformations.
 (b) Describe forward selection procedure to select variables for linear regression 2
 model.
 (c) In logistic regression let π denote probability that $Y = 1$. Give two reasons, why 3
 regression model given below can not be used to describe π .
 $\pi = \beta_0 + \beta_1 X + e.$

[TURN OVER

Section II

4. (a) Write down the ANOVA for two-way analysis with interaction and equal number 10 of observations per cell.
- (b) The following table lists the number of violent crimes reported to police on randomly selected days for this year. The data are taken from three large cities of about the same size :-

City A	City B	City C
5	2	8
9	4	12
12	1	10
3	13	3
9	7	9
7	6	14
13		

Using the 5% significance level, test the null hypothesis that the mean number of violent crimes reported per day are the same for each of these three cities.
 $(F_{(2,16)} \alpha = 0.05 = 3.63)$

5. (a) Explain the reason that it is not possible to write down the ANOVA for two-way 10 analysis with one observation per cell and interaction present. In such situation how to test that there is no interaction.
- (b) Two factors A and B are being studied for their effect on the yield of an industrial process. Three levels of each of the factors A and B are being studied and the measurements were replicated four times. Summary of the data is given below :-

In this case our Model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

$$i = 1, 2, \dots, p, \quad j = 1, 2, \dots, q, \quad k = 1, 2, \dots, r$$

$$p = 3, \quad q = 3, \quad r = 4.$$

Total sums of square = 7514.75

Sums of square due to A = 228.5

Sums of square due to B = 4565.167

Sums of square due to A * B = 575.333.

Analyze the data assuming that the levels of A and B are random samples from large populations of the levels of A and B, respectively. ($\alpha = 0.05$)

$$F_{(2,27)} \alpha = 0.05 = 3.35$$

$$F_{(4,27)} \alpha = 0.05 = 2.73.$$

6. (a) Explain 'Nested model' with examples. Further write down the SAS code for 10 this model.
- (b) Analyse the following data from a 2 way nested classification :-

Course	Section of Course	Individual
English	1	6
	2	7, 8, 9
Geology	1	7, 9
	2	5, 4, 3, 1
	3	3, 7

(3 Hours)

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

Q1: (a) Explain the relative merits and demerits of the Qualitative Methods in comparison with the Quantitative Methods available for forecasting future values of key business parameters.

(b) How will you proceed to prepare a forecast of demand for second homes in the hilly region within the Mumbai - Pune - Nashik triangle?

Q2: Which of the following five statements are correct? Mention your opinion as 'CORRECT' or 'WRONG' against each statement number. Also, give detailed reasons to support your opinion.

Statement Number 2.1: Delphi method can be used to forecast future values of a cardinal variable.

Statement Number 2.2: Moving averages of period 3 (three) should effectively smoothen the quarterly sales figures of a fairness cream.

Statement Number 2.3: Time series analysis can not detect and quantify cyclical behaviour from two-year data of a company's quarterly sales.

Statement Number 2.4: The error term in a multiplicative time series model should ideally be zero.

Statement Number 2.5: Strong correlation between two variables does not necessarily mean a cause and effect relationship between the two.

Q3: (a) Explain the difference between 'structured decisions' and 'unstructured decisions'. Give one example of each.

(b) Describe the technique of decision tree analysis with an appropriate illustrative example.

[TURN OVER]

Section II

4. Explain the term 'inventory' and give reasons for carrying inventories. Define the following :
- Set up cost
 - Holding cost
 - Shortage cost
5. A shopkeeper has a uniform demand of an item at the rate of 600 items per year. He buys from a supplier at a cost of Rs. 8 per item and the cost of ordering is Rs. 12 each time. If the stock holding cost is 20% per year of stock value, how frequently should he replenish his stock and what is the optimal order quantity ?

Section III

(10)

6. Following is the data taken from records of Air Ltd.

Particulars	Rupees
Selling Price Per Unit	20
Direct Material Cost Per Unit	5
Direct Labour Cost per Unit	3
Variable Overhead cost per Unit	2
Fixed Cost per annum	4,00,000/-

You are required to compute

- Break even point (In Rupees and In Units)
 - Break even point (In Rupees and In Units) if selling price per unit is increased by 30%
 - Break even point (In Rupees and In Units) if selling price per unit is decreased by 10%
-

I Statistical Process Control

Jan, 2013

Con. 11090-12.

BB-8973

(3 Hours)

[Total Marks : 60]

- N.B. :** (1) All questions are compulsory.
 (2) Figures to the right indicate marks.
 (3) Answers to the two sections must be submitted on separate answer books.
 (4) Calculators are allowed.

SECTION I

Instructions:

1. Use of Statistical Software, MS Excel & Excel Templates are allowed
2. Write down Hypothesis, Conclusions, Session Window output with tests used with all analysis details.

1. **Explain in detail following**

- a. CTQ, CBP & Scoping
- b. Specification Limits & Operating Limits.
- c. Value Added Process with examples
- d. Elevator Speech & your project speech

(10)

2. (a) Bank wants to understand the which service counter is better & reliable in terms of time they take for responding to Customer queries. Also help them to set the improvement Target for each Counter. (Data is in time taken in minutes)

Counter A	75	65	60	54	69	70	82	77	59	62	76	63
Counter B	65	80	52	74	79	80	72	88	68	60	82	73

- (b) Given below Wastage % of a brand. Have they achieved Target of 18 % per month. Kindly help in setting improvement Target for their Project Team.

Month	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
Yield %	15	30	2	24	29	30	22	38	18	10	32	23

(10)

3. Following 3 factors determine the mileage w.r.t kms/ltr. Of a Car. Conduct the DOE & identify the factors affecting mileage.

Speed	Driving Skills	Type of Petrol	km/ltr
-1	-1	-1	12.85
1	-1	-1	13.01
-1	1	-1	14.52
1	1	-1	14.71
-1	-1	1	12.93
1	-1	1	13.09
-1	1	1	14.61
1	1	1	14.81

(10)

[TURN OVER]

Con. 11090-BB-8973-12.

2

SECTION II

4. (a) Pin manufacturing company had complaints from the retailers about the quality. Company found that amount of metal used in the manufacturing of these pins is varying. The required metal is 10 milligram and any variation more than 20% on both the sides could cause quality issues. Which QC tool should be used by the company? Use appropriate tool and draw diagram. Find out items going outside acceptable limits. How many items are not in limits?

Weight of sample items is as follows.

10.05, 10.10, 10.10, 9.95, 9.90, 10, 9.75, 9.95, 10, 10.05

(05)

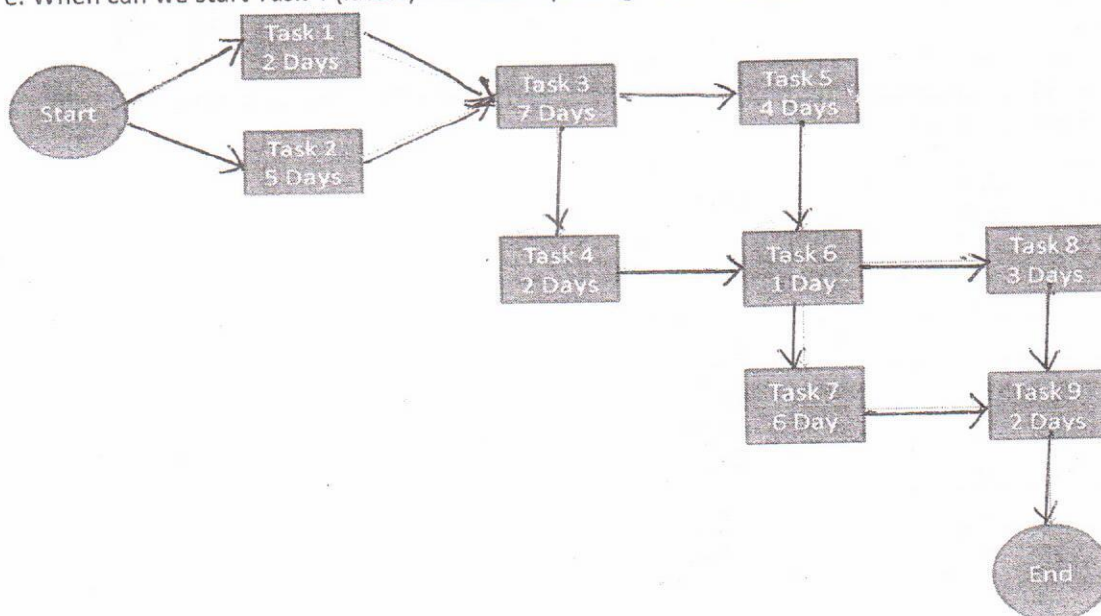
- (b) An chain restaurant owner would like to analyze and prioritise the complaints received from its customers as the perception is growing that restaurant is no more providing quality service. The complaint data is as below:

Type of complaint	Number
Food	7
Unavailability of items	15
Ambience	2
Waiting time	23
Cleanliness	3

Which QC tool can be used to represent this? Use appropriate QC tool, draw appropriate analysis and write your findings.

(05)

5. (a) Following activity diagram represents project plan of road repairs to be done before monsoon. Answer following questions.
- Calculate early start and finish time for each task
 - Calculate late start and finish time for each task
 - Identify Critical path
 - When the project must start if required to be completed on 31st Jan?
 - When can we start Task 4 (latest) without impacting overall schedule?



(05)

- (b) Three shops (Church Gate, Andheri and Borivali) of Jumbo wada-pav conducted a survey of sale of wada-pav across 3 seasons and following is the data.

There was also a data available of age wise sale. Management of Jumbo wanted to find out whether the pattern change in sale across seasons are actually due to sale patter dependent on age group.

- Which tool will you use to find out if there is any relationship between Season and Age impacting sale?
- Use the tool and present data appropriately.
- List your findings

Sale per day is as follows

	Monsoon	Winter	Summer
ChurchGate	870	770	525
Andheri	1210	1125	915
Borivali	1125	1305	1810

	Age < 20	Between 20 and 40	>40
ChurchGate	35%	45%	20%
Andheri	44%	32%	24%
Borivali	27%	23%	50%

(05)

6. (a) Suggest appropriate QC tool that can be used for following problems

Sr.	Problem
1	University is facing an issue of poor attendance in class and decides to involve professors from various departments to find out root cause of this and implement new policy which will benefit students in their learning.
2	A firm of consulting engineers wants to ensure that all eventualities and their relations are covered in an investigation report into the laying of a new cross-country gas pipeline.
3	A pharmaceutical company examines the pain-killing drugs of its subsidiaries in terms of the cost to product and general efficacy. Products which are high cost but are not of highest efficacy are dropped. Low-cost drugs of reasonable efficacy are promoted, and high-cost drugs have a project initiated to reduce production cost.
4	A production group in an electrical goods manufacturer measures and maps out the tasks required to build a toaster. They redesign and reallocate tasks to reduce the critical path time.
5	BMC studying whether there is any relationship between increase in property rates across various zones of city and increase in slums.

(05)

[TURN OVER]

- (b) Let us consider a case study that has one 2-level factor (A) and three 3-level factors (B, C and D). Calculate degrees of freedom. Which technique can be used to fit this case study with L9 orthogonal array? Explain your answer and draw a table of Experiment layout. (Refer following L9 orthogonal array)

Expt. No	Factors			
	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

(05)

VI medical statistics
Jan. 2013

Con. 11091-12.

BB-8976

(3 Hours)

[Total Marks : 60]

- N.B. :** (1) All questions are **compulsory**.
 (2) **Figures** to the **right** indicate marks.
 (3) **Calculators** are **allowed**.

1. Briefly explain the application of following tests/methods in analysis of clinical research data. [Any 4 of 6, 2.5 Marks each]
 - Non-inferiority trial and its hypothesis
 - Wilcoxon rank sum test
 - Linear Regression
 - ANCOVA
 - Binomial Test
 - Bonferroni and Turkey's adjustment

(10)
2. Answer following questions (5 marks each)
 - Type I error, type II error and power of a test, Confidence Interval
 - What information is needed by statistician in calculating Sample size for a clinical trial?

(10)
3. Describe in detail, role of a statistician at various stages of execution of a clinical trial project.

(10)
4. What are ICH guidelines? Briefly Discuss ICH E9 guideline.

(10)
5. Please read the situation below and answer questions that follow.

Researchers had historical data on effect of drug A in treatment of Psoriasis. Available data from previously conducted clinical studies indicated that 80 out of 200 subjects responded (40%) by achieving PASI50 criteria after 12 week treatment with Drug A.

PASI50 criteria is assessed by calculating % reduction in PASI score (Psoriasis area and severity index). PASI score combines the assessment of the severity of psoriasis lesions and the area affected by psoriasis into a single score in the range 0 (no disease) to 72 (maximal disease).

A subject (i.e. patient) with 50% or greater reduction in PASI score at week 12 (as compared to baseline), is considered to be PASI50 responder. A subject with < 50% reduction in PASI score at week 12 is considered as non-responder.

Researchers believed that Drug B is better than Drug A in treatment of psoriasis and would help more subjects achieve PASI50 response as compared to Drug A.

Therefore they conducted a research study, in which they administered Drug A to 400 subjects and Drug B to 400 subjects for 12 weeks. Subjects were assigned to respective treatments using a simple 1:1 randomization. PASI50 response was the primary endpoint for this study. After 12 weeks of treatment with Drug B, 250 of 400 subjects achieved PASI50 response. On the other hand, after treatment with Drug A, 175 out of 400 subjects achieved PASI50 response. Please answer following questions on this study.

[TURN OVER]

- a. Please identify the design of the study conducted by researchers? (Parallel, Cross over or Single arm study).
- b. What was the data type of the primary endpoint used in this trial?
- c. Which statistical test would be useful to check if Drug B response is different from Drug A response? Please justify your choice of statistical method.
- d. How will you interpret following results? Is there a significant difference between two treatments
Relative risk of achieving PASI50 response for treatment A as compared to treatment B was 0.7 with 95% confidence interval between (0.61, 0.80)

(10)

6. Multiple Choice Question Answers – **EACH QUESTION HAS ONLY ONE CORRECT CHOICE** [2 marks each].

Select the correct choice and write the numbers against the choice in your answer book for the following questions as shown in the example below.

(a) Sample question text _____?

- | | |
|---------------|----------------|
| (i) Choice 1 | (iii) Choice 3 |
| (ii) Choice 2 | (iv) Choice 4 |

Answer : (a) (iv).

1. Which one of the following statements about SAS PROCs is NOT correct?
 - a. PROC REPORT procedure cannot produce a summary report
 - b. PROC FREQ procedure produces a visible output
 - c. PROC SORT procedure does not produce a visible output
 - d. PROC CONTENTS procedure displays structure of a dataset
2. Which of the following statement about PROC SQL is correct?
 - a. PROC SQL does not produce a visible output.
 - b. PROC SQL can be used to view structure of a table
 - c. PROC SQL must end in a RUN statement for it to execute the query
 - d. PROC SQL cannot be used to combine datasets vertically
3. PROC FREQ can be used for performing
 - a. Chi Square test and Fisher's Exact Test and Binomial test
 - b. Chi Square test and McNemar's Test but not Fisher's Exact Test
 - c. Fisher's Exact Test and McNemar's Test but not Chi Square Test
 - d. Binomial Test and Fisher's Exact Test but not McNemar's Test

4. Which of the following syntax will correctly produce a dataset ABCD, with 100 records and a single variable whose value on each record increases by 1
- Data ABCD (1 to 100);**
A+1;

RUN;
 - Data ABCD;**
A+1;

IF A < 100 THEN Output;

RUN;
 - Data ABCD (Drop= i);**
DO i = 1 to 100;

A+1;

Output;

END;

RUN;
 - Data ABCD;**
DO i = 1 to 200 until (i =100);

A+1;

END;

RUN;
5. If you submit the following program, which variables appear in the new data set Drug?

```

Data ABCD;
  Set EFGH;
    if result > 130 then resp="Fail";
    keep pat group resp ;
run;

```

- rat group resp result
- resp
- pat group resp
- result resp

(10)

Con. 11092-12.

BB-8979

(3 Hours)

[Total Marks : 60]

- N.B. :** (1) All questions are compulsory.
 (2) Figures to the right indicate marks.
 (3) Answers to the two sections must be submitted on separate answer books.
 (4) Calculators are allowed.

SECTION I

1. (a) The distance from point $P = (x_1, x_2, x_3)$ to origin $O = (0, 0, 0)$ is defined as $d(0, P) = \max(|x_1|, |x_2|, |x_3|)$. Compute distance from $P = (-3, 4, -7)$ to the origin using above formula. Define statistical distance from an arbitrary point $P(x_1, x_2, \dots, x_p)$ to a fixed point $Q(y_1, y_2, \dots, y_p)$ (04)
- (b) Let $X = \begin{bmatrix} 2 & 5 & 5 \\ 3 & 2 & 1 \\ 6 & 7 & 5 \end{bmatrix}$
 Obtain (a) mean vector
 (b) Variance Covariance matrix (divisor 3)
 (c) Generalized sample variance and its interpretation (06)
2. (a) State p.d.f. of a bivariate normal distribution having $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$, $i=1,2$ and correlation coefficient between X_1 and X_2 as ρ . Obtain the conditional distribution of X_2 given X_1 (06)
- (b) Random vector $X \sim N_3(\mu, I_3)$ where $\mu' = [\mu_1 \ \mu_2 \ \mu_3]$. Obtain mean vector and variance-covariance matrix of the distribution

$$\begin{bmatrix} X_1 & - & X_3 \\ X_3 & - & X_2 \end{bmatrix}$$
 (04)
3. (a) Explain the method of χ^2 plots to test normality of data (04)
- (b) The p.d.f. of a 5-variate normal distribution is given by

$$\frac{1}{(2\pi)^{\frac{5}{2}} 6} e^{-\frac{1}{2} \left[\frac{x_1^2}{2} + (x_2 - 1)^2 + \frac{(x_3 - 5)^2}{3} + \frac{(x_4 - 8)^2}{3} + \frac{x_5^2}{2} \right]}$$

 Find mean vector and variance-covariance matrix. (03)
- (c) Obtain maximum likelihood estimators of mean vector and variance covariance matrix using the following sample from bivariate population.

$$X = \begin{bmatrix} 5 & 7 \\ 4 & 7 \\ 3 & 6 \\ 4 & 4 \end{bmatrix}$$
 (03)

[TURN OVER]

4. (a) A random sample of size n_1 $X_{11}, X_{12}, X_{13}, \dots, X_{1n_1}$ is selected from $N_p(\mu_1, \Sigma)$. Another random sample of size n_2 $X_{21}, X_{22}, \dots, X_{2n_2}$ is selected from $N_p(\mu_2, \Sigma)$ which is independent of the 1st random sample. Discuss the steps to test the hypothesis $H_0: \mu_1 = \mu_2$. Hence state the $100(1-\alpha)\%$ simultaneous confidence intervals for the components of vector $\mu_1 - \mu_2$ (06)
- (b) Describe multivariate multiple regression model. Indicate method of estimating parameters. (04)

SECTION II

5. (a) Define canonical correlations and canonical variables. (04)
- (b) If $\rho = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$ where
- $$\rho_{11} = \begin{bmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{bmatrix} \quad \rho_{22} = \begin{bmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix} \quad \rho_{12} = \begin{bmatrix} 0.5 & 0.6 \\ 0.3 & 0.4 \end{bmatrix}$$
- And first pair of canonical variables is $U_1 = 0.86 z_1^{(1)} + 0.28 z_2^{(1)}$,
 $V_1 = 0.54 z_1^{(2)} + 0.73 z_2^{(2)}$ obtain correlation coefficients between $(U_1, Z_1^{(1)})$ and
 between $(V_1, Z_1^{(2)})$ (04)
- (c) Give a test to decide whether it is worthwhile to proceed for canonical correlation analysis. (02)
6. (a) Give a procedure to obtain principal components from correlation matrix $\rho_{p \times p}$. Also obtain variance of first principal component. (04)
- (b) Write down orthogonal factor model. Explain the terms (i) communality (ii) specific variance. (04)
- (c) Explain how discriminant analysis helps in data reduction. State Fisher's discriminant rule for two normal populations. State the assumption. (02)