(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 frequency table gives salaries (in '000) of a sample of 42 MBAs (04) in the Mumbai city. Sketch the Box and Whisker plot for this data. Find inter quartile range.

Salary:	47	48	49	50	51	52	54	56	57	60
Frequency:	4	1	3	5	8	10	5	2	3	1

  (b) Five cards are drawn from a standard 52- card playing desk. What is the (03) probability that all 5 cards will be of the same suit?

  (c) The probability that a family owns a house is 0.76. The probability that a family (03) owns a house and is a married couple is 0.69. Find the conditional probability that a randomly selected family is a married couple given that it owns a house.

2. (a) The number of minutes it takes for a mechanic to check a tyre is a r.v. having (03) exponential distribution with mean 5. What is the probability that the mechanic will take at least 6 minutes to check a tyre?

  (b) X is a continuous r.v. having following p.d.f. (04)
$$f(x) = \theta(1-x)^{\theta-1} \quad , \quad 0 < x < 1, \ \theta > 1$$
$$= \quad 0 \qquad \qquad , \text{ otherwise}$$
Obtain mean, variance and median of this distribution. Write the name of this distribution and its parameters.

  (c) Obtain probability generating function of the binomial distribution with (03) parameters n and p. Hence find its mean.

3. (a) Following table shows a bivariate probability distribution for two discrete (04) random variables X and Y.

Y \ X	0	1	2
1	0.15	0.20	0.25
2	0.05	0.15	0.20

Obtain marginal p.m.f. of X and Y. Obtain conditional p.m.f. of Y given X=1. Check whether X and Y are independent.

  (b) Suppose the r.v. of X and Y have the following joint p.d.f. (04)
$$f(x,y) = \quad 3x, \quad 0 < y < x < 1$$
$$= \quad 0 \ , \quad \text{otherwise}$$
    (i) Find marginal p.d.f. of X and Y.
    (ii) Find conditional p.d.f. of Y given X . Hence find $E(Y|X)$.

(c) Obtain the moment generating function of a uniform r.v. U(a,b)   (02)

SECTION II

4. (a) Define unbiased estimator. Show that sample variance is an unbiased estimator   (05)
   of population variance.

   (b) $x_1, x_2, ..., x_n$ is a random sample from $f(x; \theta)$ where $f(x; \theta) = \theta\, e^{-\theta x}$ $x \geq$ (05)
   $0, \theta > 0$. Obtain moment estimator of θ.

5. (a) $x_1, x_2, ..., x_n$ is a random sample from $N(\theta, 1)$ distribution. Obtain $\hat{\theta}$, MLE of θ.   (05)
   Also obtain asymptotic distribution of $\hat{\theta}$.

   (b) The consumption of artificial sweetners in India in 2000 was equivalent to 24   (05)
   KGS of sugar per person per year. Suppose that a random sample of 150 Indians
   taken this year showed an annual average consumption of 27 KGS of sugar with
   a standard deviation of 9.0 KG. Does the sample supports the alternative
   hypothesis that the current mean consumption of artificial sweetners is more than
   24 KGS of sugar? Use $\alpha = 0.01$.

6. (a) A sample of 7 students selected from a small college produced the following data   (05)
   on their heights (inches)
   65, 74, 70, 62, 60, 67, 66
   Find sample variance and test the hypothesis $H_0: \sigma^2 > 15$ $against$ $H_1: \sigma^2 > 15$.
   Take $\alpha = 0.01$.

   (b) Violence and lack of discipline have become major problems in schools. A   (05)
   random sample of 300 adults were selected and their opinion about freedom to
   school teachers to punish students for violence and lack of discipline was noted.

	Infavour	Against
Men	93	82
Women	87	38

   Test the hypothesis that gender and opinion on the issue are independent. Take
   $\alpha = 0.05$

****

II marketing Res sem I
II m arketing Research
Jan 2015

QP Code : 18476

(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 sections should be written on separate answer-books.

SECTION I

1.	Explain classification of qualitative research procedures.	(05)
2.	Explain principles of sampling.	(05)
3.	Explain SWOT analysis in details.	(10)
4.	Design a questionnaire for new soup flavour launch in market.	(10)

SECTION II

5. (a) What is the difference between Hierarchical method of cluster formation and (05)
   K-means method of cluster formation? What is the use of each method in
   performing cluster analysis?

  (b) Cingular mobile company has conducted consumer research in which (10)
   consumers' opinion on eleven statements was obtained. A 10 point scale was
   used where score 1 indicates that the statement does not describe Cingular at
   all, rating 5 shows midway in describing Cingular and score 10 means
   describes Cingular very well. Output of Factor analysis is shown below. What
   inference would you draw from the output?

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.827
Bartlett's Test of Sphericity	Approx. Chi-Square	1321.828
	df	55
	Sig.	.000

BB-Con. 12289-14.

[TURN OVER

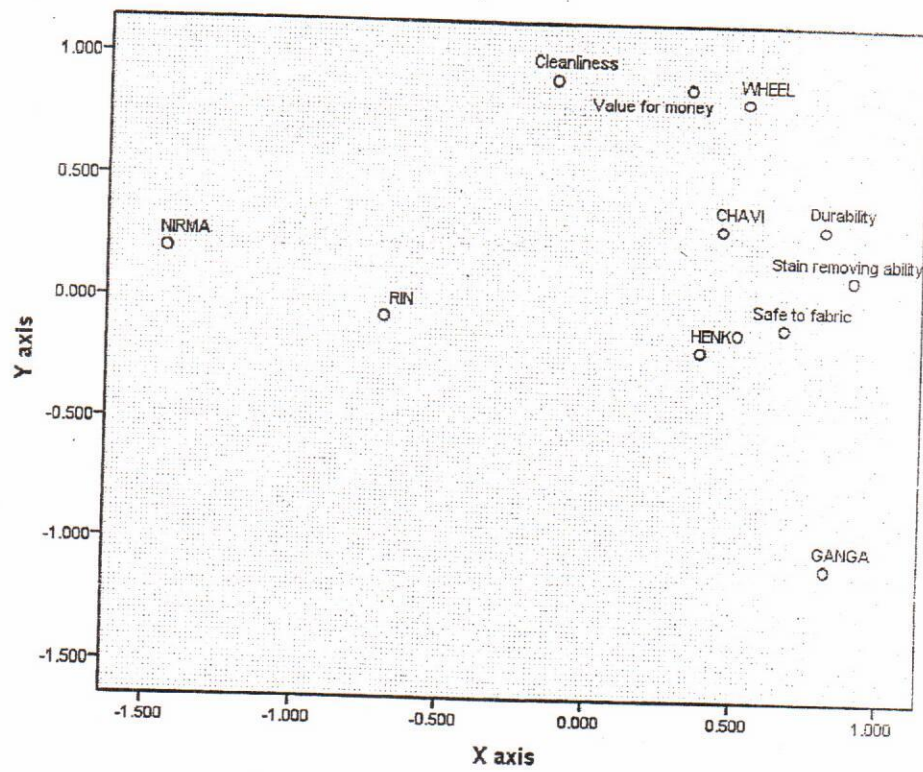| Communalities		
	Initial	Extraction
Cingular's coverage area meets your needs	1.000	.530
Consumer is able to make or receive calls where he/she lives and travels	1.000	.729
Consumer gets few dropped calls	1.000	.493
Voice Quality of the calls is very good	1.000	.639
Cingular has favorable contract requirements	1.000	.608
Cingular has calling plans that meet consumers' needs	1.000	.519
Cingular has selection of phones that meet consumers' needs	1.000	.302
Cingular has error-free billing statement	1.000	.463
Cingular provides high-quality customer service	1.000	.511
Cingular has conveniently-located stores	1.000	.253
Cingular has lower prices	1.000	.542
Extraction Method: Principal Component Analysis.		

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	4.083	37.115	37.115
2	1.508	13.705	50.819
3	.992	9.018	59.837
4	.834	7.585	67.422
5	.734	6.675	74.096
6	.684	6.217	80.313
7	.511	4.648	84.961
8	.494	4.494	89.455
9	.467	4.245	93.700
10	.390	3.548	97.248
11	.303	2.752	100.000

BB-Con. 12289-14.

[TURN OVER

Rotated Component Matrix[a]		
	Component	
	1	2
Cingular has favorable contract requirements	.775	.085
Cingular has lower prices	.730	.094
Cingular has calling plans that meet consumers' needs	.719	.048
Cingular provides high-quality customer service	.668	.254
Cingular has error-free billing statement	.619	.282
Cingular has selection of phones that meet consumers' needs	.479	.270
Cingular has conveniently-located stores	.432	.257
Consumer is able to make or receive calls where he/she lives and travels	.086	.850
Voice Quality of the calls is very good	.221	.768
Cingular's coverage area meets your needs	.223	.693
Consumer gets few dropped calls	.167	.682
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.		

6. Following is the output of factor analysis showing factor scores and map (15)
    showing attribute points and brand points for drawing Attribute based
    perceptual map of Detergent brands. Explain the strengths and weaknesses of
    each brand. Is there any scope for launching new brand?

FACTOR SCORES

	X AXIS	Y AXIS
Cleanliness ability	-.122	.914
Safe to fabric	.666	-.102
Stain removing ability	.902	.105
Value for money	.335	.883
Durability	.802	.310
NIRMA	-1.440	.200
WHEEL	.530	.830
GANGA	.820	-1.100
CHAVI	.450	.300
RIN	-.700	-.070
HENKO	.380	-.200

Perceptual map (X axis vs Y axis) showing brands NIRMA, RIN, HENKO, CHAVI, WHEEL, GANGA and attributes Cleanliness, Value for money, Durability, Stain removing ability, Safe to fabric.

****

QP Code : 18477

(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) Assume regression model $Y = \beta_0 + \beta_1 X + e$ where $e \sim N(0, \sigma^2)$ with n observations.
    (I) prove that regression s.s.$= \hat{\beta}_1^2 \Sigma (x_i - \bar{x})^2$ (03)
    (II) Data on 10 observations gives, (05)
    $\Sigma (y_i - \bar{y})^2 = 3.267$, $\Sigma (x_i - \bar{x})^2 = 306.9$, $\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 77.33$;
    $\Sigma x_i = 139$, $\Sigma y_i = 63.3$
        (i) Estimate β_0, β_1.
        (ii) Find coefficient of determination.
        (iii) Estimate y when X=15. Find 95% confidence interval for estimated value (give expression only).
  (b) Why is transformation of variables necessary? (02)
2. (a) Data on 10 observations with two predictor variables X_1, X_2 and a response variable Y (06) gives rise to the following ANOVA table, where some entries are missing. Assume linear regression model.

ANOVA TABLE

Source	d.f.	s.s.	M.SS.	F	p
Regression	–	9.0009	–	21.36	0.001
Error	–	–	–		
Total	–	–			

        (i) Fill in the blanks and write completed table.
        (ii) Find the value of R^2.
        (iii) Could you conclude that both regression coefficients are zero? Justify your answer.

  (b) A study of regression of Y on X_1, X_2, X_3 gives following VIF values. Comment on (02) collinearity of data.
    $VIF_1 = 469.7$, $VIF_2 = 1.0$, $VIF_3 = 469.4$
  (c) What is heteroscedasticity ? How do you remove it, if $V(e) = K^2 \sigma^2$ (02)

3. (a) Relationship between age(X) and presence or absence of disease (D) is studied using (06) 100 subjects. Dependent variable Y is defined as,
$$Y = \begin{cases} 1 \ if \ D \ is \ present \\ 0 \ if \ D \ is \ absent \end{cases}$$
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? What is the rate of correct classification?

BB-Con: 12284-14. **[TURN OVER**

X	Y
26	0
55	0
75	1

(b) What is hat matrix? How do you express fitted values in terms of observed responses? (02)
Find hat matrix in case of simple regression model.

(c) Define forward selection procedure to select variables for linear regression model. (02)

SECTION II

4. Following is the output of SPSS of two-way classified data with two factors (10)
contractibility and tensile strength with covariate resilience. Answer the following
questions.

   i.   Write the mean sum of squares for covariate effect. Is it significant?
   ii. Write total sum of squares.
   iii. Write degrees of freedom for error.
   iv. How many levels are there for factor tensile strength?
   v.   Write hypothesis and conclusion.
   vi. Write the model.

Tests of Between-Subjects Effects

Dependent Variable: Electric_current

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	99.268a	7	14.181	6.883	.007
Intercept	35.979	1	35.979	17.463	.003
Contractibility	3.187	3	1.062	.516	.683
Tensile_strength	2.599	3	.866	.420	.743
Resilience	55.268	1	55.268	26.825	.001
Error	16.482	8	2.060		
Total	88028.000	16			
Corrected Total	115.750	15			

a. R Squared = .858 (Adjusted R Squared = .733)

5. Explian the Turkey's test procedure for non-additivity? (10)

6. A quality manager wishes to test the effect of five different testing methods for in- (10)
process quality check on the percentage rejection of times at the final assembly.
Perform the analysis and draw proper conclusions. $F_{4,8,0.05} = 3.84$.

M₁	M₂	M₃	M₄	M₅
15	02	10	07	05
20	10	19	10	12
06		02		14

****

QP Code : 18482

(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 sections should be written on separate answer-books.

SECTION I

1. (a) Explain the relative strengths and drawbacks of the Qualitative Methods and the Quantitative Methods used for forecasting future values of key business parameters. (10)

   (b) How will you proceed to prepare a forecast of demand for low cost smartphones?

2. Which of the following five statements are correct? Mention your opinion as 'CORRECT' or 'WRONG' against each statement and give detailed reasons to support your opinion. (10)

   Statement Number 2.1: Ideally, the error term in a multiplicative time series model should be one.
   Statement Number 2.2: Forecasting future demand for e-books is not possible unless past sales data are made available.
   Statement Number 2.3: Time series analysis is not powerful enough to detect a cyclical component in a two year data of a company's monthly travelling expenses.
   Statement Number 2.4: Strong correlation between two variables confirms a cause and effect relationship between the two conclusively.
   Statement Number 2.5: In a SWOT analysis of a business organization, the opportunities and threats emanate from external sources while the strengths and weaknesses are controllable by the organization.

3. (a) Describe the Six Thinking Hats technique for making decision in complex situations. (10)

   (b) Apply this technique to a TV Channel dedicated to general entertainment which wants to revamp its business strategy for future.

SECTION II

4. Explain in detail the reasons for carrying inventories. (10)

5. Explain the meaning of buffer stock (or safety stock). Also explain normal lead time consumption and re-order level taking suitable example. (10)

BB-Con. 12292-14. [TURNOVER

SECTION III

6. From the following particulars, you are required to calculate: (10)

    (i)   Profit Volume Ration (P/V Ratio)

    (ii)   Break Even Point for Sales (BEP)

    (iii)   Margin of Safety

    (iv)   Profit when sales are Rs. 2,00,000/-

    (v)   Sales required to earn a profit of Rs. 40,000/-

Year	Sales	Profit
I	2,40,000	18,000
II	2,80,000	26,000

****

BB-Con. 12292-14.

QP Code : 18484

(3 Hours)   [Total Marks:60

N.B. (1) All question 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 & Excel along with six sigma excel template is
   allowed.
2. Write down Null & Alternative hypothesis for the test with name/s of the test/s
   used, basis of test and conclusion along with output of session window.

1. Customer Service Dept. collected data of complaint resolution time in   (10)
   Hours. Kindly advise whether Bank performed better in 2013-14. Which year Bank
   performed consistently month over month? What improvement Target can be set
   for the year 2014-15 & for consistency?

Year	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
2012-13	140	156	132	150	153	165	149	166	155	145	165	166
2013-14	131	146	118	140	145	146	138	154	134	126	148	139

2. Data on Carpet area of 3 flats measured by 3 diff. surveyors is given below &   (05)
   kindly advise Society on measurement analysis. Also explain in detail the output
   session window & graph window.

Flat No	1	1	1	1	1	1	2	2	2
Surveyor No	1	2	3	1	2	3	1	2	3
Trial No	1	1	1	2	2	2	1	1	1
Carpet Area	515	482	529	519	502	503	409	366	342

Flat No	2	2	2	3	3	3	3	3	3
Surveyor No	1	2	3	1	2	3	1	2	3
Trial No	2	2	2	1	1	1	2	2	2
Carpet Area	408	368	358	438	445	450	455	442	461

BB-Con. : 12287-14.   **[TURN OVER**

3. Find below to set the base-line   (05)
   a. Sample Size to be used with SD of 10 & tolerance of 2 is allowed at 95% confidence.
   b. What will be DPMO & Zst for having 3 defects out of 35 Units with OFE of 3 ?
   c. What will be DPMO & Zst for USL-320 ml & LSL of 270 ml with current variance of 64 ml?

4. Following 3 factors determine the Production output of a machine.   (10)
   Conduct DOE & identify the factors affecting Production.

Local	In-house Trained	Round Edge	-1
Imported	OEM trained	Square Edge	+1
Material Quality	Technical Skills	Type of Product	Production In Mil.
-1	-1	-1	14.35
1	-1	-1	14.51
-1	1	-1	16.02
1	1	-1	16.21
-1	-1	1	14.43
1	-1	1	14.59
-1	1	1	16.11
1	1	1	16.31

SECTION II

5. Draw a flowchart for a process of selling at a counter of McDonald, which   (05) involves

   1. Look for next customer
   2. If customer waiting in queue then call for next customer, if no customer then wait for customer to arrive at store following activities.
   3. Take down the order
   4. Enter it in the order entry system
   5. Collect cash/credit card/debit card
   6. If cash then count it and put it in drawer
   7. If card then swipe it
   8. If card swipe fails then request for another card or cash
   9. If card swipe successful then ask for customer's signature
   10. Confirm order

[TURN OVER

BB-Con. : 12287-14.

11. Request customer to wait for the delivery.

6. Refer following excel and answer following questions   (05)

	A	B	C	D	E	F
1	Emp No	Name	Gender	Salary	Bonus	Department
2	1	A1	Male	151760	30350	Sales
3	2	A2	Female	126139	18920	Accounts
4	3	A3	Male	37818	5670	Admin
5	4	A4	Male	31656	6330	Marketing
6	5	A5	Male	150957	15100	Production
7	6	A6	Female	197914	29690	Export
8	7	A7	Male	24869	3730	HR
9	8	A8	Male	113073	22610	Sales
10	9	A9	Male	57160	5720	Production
11	10	A10	Female	154987	23250	Finance
12	11	A11	Male	24985	2500	Production
13	12	A12	Male	102137	20430	Sales
14	13	A13	Male	162869	24430	Accounts
15	14	A14	Female	147878	22180	Admin
16	15	A15	Male	36250	7250	Marketing
17	16	A16	Male	62627	6260	Production
18	17	A17	Male	32679	4900	Export
19	18	A18	Female	198271	29740	HR
20	19	A19	Male	103469	20690	Sales
21	20	A20	Male	185912	18590	Production

1. Write formulae for calculating average salary of employees
2. Write formulae to calculate number of Female employees
3. Write formulae to calculate total expenditure on bonus by Sales department
4. Write formulae to calculate maximum bonus and minimum bonus
5. Write formulae to find whether total bonus is greater than 15% of total salary. Answer should be Yes or No

7. (a) Following activity diagram represents project plan of road repairs to be done before monsoon. Answer following questions.   (05)
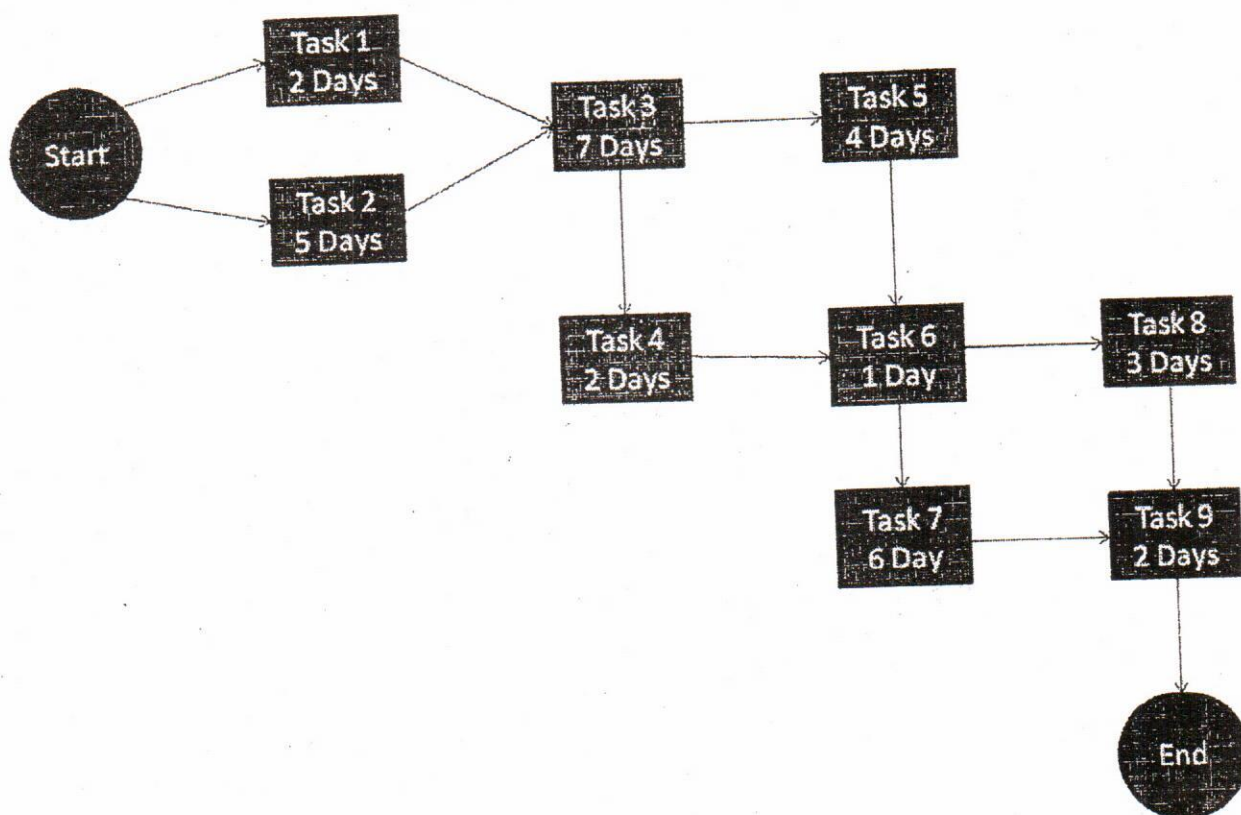
a. Calculate early start and finish time for each task

b. Calculate late start and finish time for each task

c. Identify Critical path

d. When the project must start if required to be completed on 31st May?

e. When can we start Task 4 (latest) without impacting overall schedule?

[TURN OVER

(b) Suggest appropriate QC tool that can be used for following problems   (05)

Sr.	Problem
1	A washing powder has different efficiencies at achieving 'softness' and 'stain removal' in garments made of acrylic, polyester, wool and various fiber mixtures. If similar affects are found in a group of fibers, then changing the powder ingredients may affect the whole group in a similar way.
2	NGO working on environmental issues visit a village where there are many environmental issues. They want villagers to be part of their environment improvement programme and collect suggestions from all.
3	A funeral director maps out the tasks involved in a funeral, paying attention to how multiple funerals might overlap. By rearranging tasks, he enables one more funeral to be fitted into a day, thus increasing customer responsiveness along with business volume.
4	BMC studding whether there is any relationship between increase in property rates across various zones and increase in slums
5	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.

8. (a) An airlines company would like to analyze and prioritise the quality   (05) complaints received from its customers. The complaint data is as below:

Type of complaint Number

Baggage delay   23

Missed connections   15

Lost baggage   7

Poor cabin service 3

Ticketing error   2

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

(b) In a case study there are four 2-level factors A, B, C and D. We want to (05) estimate their main effects and also the interaction A x B, B x C and B x D. Calculate degrees of Freedom. Use following L8 orthogonal array and

interaction table and draw a table of experiment layout. Show columns representing interaction between A x B, B x C and B x D.

L8 – Orthogonal Array

	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
	A	B	C	D	E	F	G

Interaction table for L8

Column	Column						
	1	2	3	4	5	6	7
1	(1)	3	2	5	4	7	6
2		(2)	1	6	7	4	5
3			(3)	7	6	5	4
4				(4)	1	2	3
5					(5)	3	2
6						(6)	1
7							(7)

****

BB-Con. : 12287-14.

(3 Hours)   Total Marks : 60

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

1. Describe application of following tests/statistical methods in clinical research studies (any 4
   of 5 – 2.5 marks each)   (10)
   a. Wilcoxon signed rank test
   b. Binomial Test
   c. ANCOVA
   d. Relative risk
   e. Bonferroni Adjustment

2. Describe Contents of a Statistical Analysis Plan for a Clinical Trial.   (10)

3. Answer following questions (5 marks each)   (10)
   a. What are different types of data collected in a clinical trial?
   b. Briefly describe steps involved in clinical data management activities for a clinical
      trial project.

4. Describe the information needed to calculate sample size for a clinical trial   (10)

5. Please read the following scenario and answer questions mentioned below.   (10)

   A group of investigators felt that combination of Rifampicin and Vancomycin for treatment
   of Methicillin Resistant Staphylococcus Aureus (MRSA) pneumonia would result in better
   cure rate than using Vancomycin alone. Cure is defined as complete eradication of the
   infectious organism within 14 days. If organism persists, treatment would be considered to
   have failed for respective subject.

   Hence they planned a clinical trial comparing the Combination treatment (rifampicin +
   Vancomycin) against vancomycin alone for treatment of MRSA pneumonia. In proposed
   study, patients with MRSA pneumonia are planned to be randomly allocated, in 1:1
   proportion, to either combination therapy or to valcomycin alone.

   They have sought your help in providing statistical plan and sample size calculation.

   A. What information will you need to calculate sample size for this study.[3 marks]
   B. How will you write the statistical hypothesis for this study .[3 marks]
   C. Discuss statistical methods that will be used for testing investigators' objectives in this
      study. [4 marks]

[TURN OVER

BB-Con. : 12288-14.

6.   Please answer following multiple choice questions (by indicating on your answer sheet) question number followed by option that represents the correct answer (e.g. Mention 6a : (i) for question 6a if option numbered (i) is the correct answer in your opinion) [2 Marks Each-]   (10)

a.   Which of the following statement about PROC SQL is True?

(i) Proc SQL never produces a visible output.

(ii) Proc SQL must end in a RUN statement to work correctly.

(iii) Proc SQL can be used to query information from multiple tables

(iv) Proc SQL cannot produce summary statistics by groups

b.   Which of the following is the correct syntax to create a dataset with 100 records numbering from 1 to 100?

(i)   Data ABCD;

            A= 1 to 100;

            Output;

            Run;

(ii)   Data ABCD;

            A= ranuni(1234)*100;

            Output;

            Run;

(iii)   Data ABCD (drop=i);

            Do i = 1 to 100;

            A+1;

            Output;

            End;

            Run;

(iv) Proc SQL;

            Create table abcd as 1 to100 from NULL;

            Quit;

c.   Which of the following is a correct PROC SORT statement?
(i)   PROC SORT DATA = A; BY X Y; RUN;

(ii)   PROC SORT DATA = A; BY X, Y; RUN;

    (iii)   PROC SORT DATA = A; BY; X Y; RUN;

    (iv)   PROC SORT DATA = A; BY X, BY Y; RUN;

d. Which of the following is a correct statement about SAS procedures?

    (i)   Proc means can provide mean, median and mode, but not quartiles

    (ii)   Proc freq can provide frequencies but not column percentages

    (iii)   Data step cannot be used to create block randomization.

    (iv)   Proc univariate can be used to perform tests of normality.

e. Which one of the following SAS procedure will display output by default?

    (i) PROC POWER

    (ii) PROC COPY

    (iii) PROC TRANSPOSE

    (iv) PROC SORT

****

QP Code : 18489

(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 **both** section should be written in **separate** answer-books.

1. (a) Explain the use of statistical distance in multivariate analysis and its advantage over the Enclidean distance. **5**

(b) Four students gave three tests and their marks are as follows :— **5**

Student	Marks		
1	8	9	7
2	6	8	6
3	8	4	6
4	6	7	5

Obtain the mean vector, variance-covariance matrix (unbiased) and the generalized sample variance.

2. (a) Discuss the use of dot plots and scatter plots in summarizing statistical data with multivariate observations. **4**

(b) A random sample of 10 observations is selected from a trivariate normal distribution with mean vector μ and variance covariance matrix Σ. **4**

State the distribution of (i) $5\overline{X}$ (ii) $\sum_{i=1}^{10} i\,X_i$.

(c) Let $\underline{X} \sim N_p(\mu, \Sigma)$. **2**
State distribution of $(X - \mu)'\,\Sigma^{-1}(X - \mu)$.
State the maximum likelihood estimator of Σ.

3. (a) Let $X \,!\, N_3\,(\mu, \Sigma)$ **4**

where $\mu = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ $\Sigma = \begin{bmatrix} 9 & 2 & -1 \\ 2 & 5 & 0 \\ -1 & 0 & 1 \end{bmatrix}$

Obtain the conditional distribution of X_1 given $\begin{bmatrix} X_2 = 3 \\ X_3 = 1 \end{bmatrix}$

(b) Explain in brief the method of χ^2 plots to test normality of data. **3**
(c) Show that the T^2 statistic is invariant under changes in the units of measurements for X of the form $\underset{p \times 1}{Y} = CX + d;$ (C is non – sin gular) **3**

[TURN OVER

BB-Con. : 12289-14.

4. (a) Let $X_1, X_2 \ldots X_n$ be a random sample from $Np(\mu, \Sigma)$. Obtain the likelihood ratio   5
test to test the hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. State its relation to the
Hotelling's T^2 statistic.

(b) Explain the regression method of modelling relationship between 'm' response   5
variables $Y_1, Y_2 \ldots Y_m$ and single set of predictor variables $X_1, X_2, \ldots X_r$.

Section II

5. (a) The distances between pairs of five items are given below :—   5

$$
\begin{array}{c c c c c c}
 & 1 & 2 & 3 & 4 & 5 \\
1 & 0 & & & & \\
2 & 4 & 0 & & & \\
3 & 6 & 9 & 0 & & \\
4 & 1 & 7 & 10 & 0 & \\
5 & 6 & 3 & 5 & 8 & 0
\end{array}
$$

Cluster the five items using complete linkage. Draw the dendrogram.

OR

(a) If $\varrho = \begin{pmatrix} \varrho_{11} & \varrho_{12} \\ \varrho_{21} & \varrho_{22} \end{pmatrix}$ where $\varrho_{11} = \begin{pmatrix} 1 & \cdot 4 \\ \cdot 4 & 1 \end{pmatrix}$, $\varrho_{22} = \begin{pmatrix} 1 & \cdot 2 \\ \cdot 2 & 1 \end{pmatrix}$ and $\varrho_{12} = \begin{pmatrix} \cdot 5 & \cdot 6 \\ \cdot 3 & \cdot 4 \end{pmatrix}$ and   5

first pair of canonical variables is :

$$U_1 = \cdot 86 \, Z_1^{(1)} + \cdot 28 \, Z_2^{(1)}$$
$$V_1 = \cdot 54 \, Z_1^{(2)} + \cdot 73 \, Z_2^{(2)}$$

Obtain first canonical correlation.

Obtain correlation coefficients between U_1 and $Z_1^{(1)}$ and between V_1 and $Z_1^{(2)}$.

(b) Write down the orthogonal fator model. Explain the terms cammunality and specific   5
variance. Explain the principal component method to extract the initial fctor loading.

6. (a) State the minimum ECM rule for classifying two multivariate normal populations.   5
State your assumptions. Give the ECM rule for special cases.

(b) Define first two principal components Y_1, Y_2 of Σ matrix. Also obtain their variances.   5
Obtain first two principal components of the following ϱ matrix.

$$\varrho = \begin{pmatrix} 1 & \cdot 75 \\ \cdot 75 & 1 \end{pmatrix}$$