

- 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 numbers of new cars sold at a dealership during a 20-day period – 4

|   |   |    |    |   |    |   |   |   |    |
|---|---|----|----|---|----|---|---|---|----|
| 8 | 5 | 12 | 3  | 9 | 10 | 6 | 3 | 8 | 8  |
| 4 | 6 | 10 | 11 | 7 | 7  | 3 | 5 | 9 | 11 |

Make a box-and-whisker plot. Comment on the skewness of these data.

- (b) A study of the residents of a region showed that 20% were smokers. The probability of death due to lung cancer, given that a person smoked, was ten times the probability of death due to lung cancer, given that the person did not smoke. If the probability of death due to lung cancer in the region is 0.006 what is the probability of death due to lung cancer given that the person is a smoker ? 3
- (c) Bag A contains 2 white and 3 red balls and bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and is found to be red. Find the probability that it was drawn from bag B. 3

2. (a) Suppose the r.v.  $X$  has a following p.d.f. 4

$$f(x) = Ke^{-2x} \quad x \geq 0$$

$$= 0 \quad \text{O.W}$$

- (i) Find  $K$
- (ii) Find mean of  $X$
- (iii) Find  $P(X < 5.27)$
- (b) (i) Among 120 applicants for a job in an insurance company only 80 are actually qualified. If five of the applicants are randomly selected (from all applies) for an indepth interview, find the probability that only two of the five will be qualified. 3
- (ii) At a certain large restaurant in a city it takes an average 10 minutes to receive the order after placing. If the service time is exponentially distributed, find the probability that the customer waiting time is more than 10 minutes.
- (c) If  $X$  has a Poisson(5) distribution, find the PGF of  $y = 2x + 3$ . 3

[TURN OVER

3. (a) If the number of minutes it takes for a mechanic to check a tyre is a random variable having an exponential distribution with mean 5. What is the probability that the mechanic will take 4

- (i) More than eight minutes to check two tyres  
(ii) At least fifteen minutes to check three tyres ?

- (b) The number of claims arising in a month under a home insurance policy, follows a Poisson distribution with mean 0.075. Calculate the approximate probability that atleast 50 claims in total arise in a month under a group of 500 such independent policies. 3

- (c) Let X and Y be two random variables having the joint probability distribution 3

| Y  | X   |     |     |
|----|-----|-----|-----|
|    | -1  | 0   | 1   |
| -1 | 0   | 0.1 | 0.1 |
| 0  | 0.2 | 0.2 | 0.2 |
| 1  | 0   | 0.1 | 0.1 |

Find  $E[y/x = 1]$  and  $V[y/x = 1]$

## Section II

4. (a) Let  $X_1, \dots, X_n$  be a random sample from a population with density function 5

$$f(x/\theta) = \frac{2\theta^2}{x^3}, \theta < x < \infty, \\ = 0, \text{ otherwise.}$$

- (i) Obtain MLE of  $\theta$ .  
(ii) Obtain moment estimator of  $\theta$ . Is it unbiased ? Find its variance.  
(b) Let  $X_1, \dots, X_n$  be a random sample from geometric distribution with p.m.f. as follows— 5

$$P(X = x) = P(1-p)^x, x = 0, 1, \dots \\ = 0, \text{ otherwise.}$$

- (i) Obtain MLE of P and CRLB of variance of any unbiased estimator of P.  
(ii) Obtain moment estimator of P. Is it unbiased ? Obtain its variance.



5. (a) Eleven school boys were given a test in drawing. They were given a month's further tuition and second test of equal difficulty was held at the end of it. Do the marks give evidence that students have benefited by the extra coaching? (use  $\alpha = .05$ ). 5

| Roll No.                      | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 |
|-------------------------------|----|----|----|----|----|----|----|----|----|----|----|
| Marks of 1 <sup>st</sup> test | 23 | 20 | 19 | 21 | 18 | 20 | 18 | 17 | 23 | 16 | 19 |
| Marks of 2 <sup>nd</sup> test | 24 | 19 | 22 | 18 | 20 | 22 | 20 | 20 | 23 | 20 | 17 |

- (b) The ages of a random sample of 5 university professors are 39, 54, 61, 72 and 59. Using this information, find a 99% confidence interval for the population standard deviation of the ages of all professors at the university, assuming that the ages of university professors are normally distributed. 5
6. (a) The Mendelian theory states that the number of a certain type of peas falling into the classifications round and yellow, wrinkled and yellow, round and green and wrinkled and green should be in the ratio 9 : 3 : 3 : 1. Suppose that 100 such peas revealed 56, 19, 17 and 8 in the respective categories. Are these data consistent with the model? Use  $\alpha = .05$ . 4
- (b) Organic chemists often purify organic compounds by a method known as fractional crystallization. An experimenter wanted to prepare and purify 4.85 g of aniline. Ten 4.85-g specimens of aniline were prepared and purified to produce acetanilide. The following dry yields were obtained 3.85, 3.88, 3.9, 3.62, 3.72, 3.8, 3.85, 3.36, 4.01, 3.82. Construct a 95% confidence interval for the mean number of grams of acetanilide that can be recovered from 4.85 g of aniline. 6

- 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. **Questionnaire design** for the customer who recently contacted the customer service department. Short survey to tell us about customer experience. Customer feedback will be used to improve our service. Customer responses will be kept confidential.
2. **Network diagram** – To identify critical path and figure out how much time the whole project will take.

| Activity | Description         | Required Predecessor | Duration |
|----------|---------------------|----------------------|----------|
| A        | Product design      | (None)               | 5 months |
| B        | Market research     | (None)               | 1        |
| C        | Production analysis | A                    | 2        |
| D        | Product model       | A                    | 3        |
| E        | Sales brochure      | A                    | 2        |
| F        | Cost analysis       | C                    | 3        |
| G        | Product testing     | D                    | 4        |
| H        | Sales training      | B, E                 | 2        |
| I        | Pricing             | H                    | 1        |
| J        | Project report      | F, G, I              | 1        |

3. Explain entire process in **"Report preparation and Presentation"**
4. Draw **"SWOT analysis"** of a market position of a small management consultancy with specialism in Human Resource Management.

**[ TURN OVER**



## Section II

5. a) How do you find the reliability of factors derived by using factor analysis technique? (5)  
 b) Calculate eigenvalues and communalities from following table and interpret them. Name the factors. (5)

Rotated Component Matrix<sup>a</sup>

|   | Component |      |
|---|-----------|------|
|   | 1         | 2    |
| Q9.Highly ethical company                       | .855      | .090 |
| Q8.Company cares about the society              | .830      | .234 |
| Q17.Company does its fair share to help society | .692      | .332 |
| Q14.Company I can trust                         | .689      | .373 |
| Q16.Known for treating its employees well       | .651      | .366 |
| Q7.Value offered                                | .642      | .343 |
| Q15.Company has advertising I really like       | .624      | .290 |
| Q13.Financially sound company                   | .169      | .826 |
| Q10.Leader in the IT industry                   | .235      | .801 |
| Q12.Company has strong capable senior leaders   | .424      | .717 |
| Q11.Innovative company                          | .511      | .586 |

6. a) Explain Ward's method of calculating distance between two clusters in cluster analysis. (5)  
 b) Using following output of cluster analysis, what suggestions would you like to give to company? (5)

Final Cluster Centers

|   | Cluster |   |   |
|---|---------|---|---|
|   | 1       | 2 | 3 |
| Q23.Overall quality of products or services purchased | 2       | 4 | 5 |
| Q24.Overall quality of non-technical customer service | 2       | 3 | 4 |
| Q25.Overall quality of training                       | 2       | 3 | 4 |
| Q26.Overall quality of technical support              | 2       | 3 | 5 |

ANOVA

|   | Cluster     |    | Error       |     | F       | Sig. |
|---|-------------|----|-------------|-----|---------|------|
|   | Mean Square | df | Mean Square | df  |         |      |
| Q23.Overall quality of products or services purchased | 103.410     | 2  | .409        | 426 | 252.552 | .000 |
| Q24.Overall quality of non-technical customer service | 121.200     | 2  | .405        | 426 | 299.246 | .000 |
| Q25.Overall quality of training                       | 125.143     | 2  | .428        | 426 | 292.516 | .000 |
| Q26.Overall quality of technical support              | 172.625     | 2  | .348        | 426 | 495.914 | .000 |

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each  
Cluster

|         |   |         |
|---------|---|---------|
| Cluster | 1 | 51.000  |
|         | 2 | 202.000 |
|         | 3 | 176.000 |
| Valid   |   | 429.000 |
| Missing |   | 0.000   |

# III Regression & Linear model

July 2013.

AGJ 1st half con-code (j+) 63

Con. 10393-13.

(3 Hours)

BB-2019

[ 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  $Y = \beta_0 + \beta_1 X + e$  and  $n$  observations :— 3  
(i) Find :—  
(1)  $\text{var.}(\hat{\beta}_0)$   
(2)  $\text{s.e.}(\hat{\beta}_1)$   
(3) expression for  $R^2$ .  
(ii) Suppose a regression analysis provides the following results :—  
 $\hat{\beta}_0 = 1, \hat{\beta}_1 = 2, \text{s.e.}(\hat{\beta}_0) = 0.5, \text{s.e.}(\hat{\beta}_1) = 0.25,$   
 $\text{SST} = 117.29, \text{SSE} = 30.0, n = 24$  and  $t_{22, .025} = 2.07$   
(1) Test the hypothesis  $H_0 : \beta_1 = 0$  against  $H_1 : \beta_1 \neq 0$  using 5% level of significance.  
(2) Compute  $R^2$ .  
(b) (i) If  $\hat{Y}$  denotes estimated value of  $y$ , give range of value of  $\text{cor}(Y, \hat{Y})$ , assume simple regression. 5  
(ii) In (i) find standard error of  $\hat{Y}$  when  $x = x_0$  2
2. (a) Consider multiple regression model with four predictor variables. State the procedure to test the hypothesis that all predictor variables have no explanatory power. You are given that regression s.s. = 23665352, residual s.s. (with 88 d.f.) = 22657938. Compute F statistic to test the hypothesis you have stated and find  $R^2$  value. 5  
(b) What do you understand by multicollinearity ? In a regression model with three predictor variables, values of VIF's and Eigen values of correlation matrix of predictor variables are given below. Comment on the collinearity of the variables using the given information :— 5  
 $\text{VIF}_1 = 469.7, \quad \text{VIF}_2 = 1.0, \quad \text{VIF}_3 = 469.4$   
 $\lambda_1 = 1.999, \quad \lambda_2 = 0.998, \quad \lambda_3 = 0.003$

[ TURN OVER



3. (a) Output of logistic regression using  $X_1$ ,  $X_2$  is given below :—

4

| Variable | Coefficient | s.e.  | Z - test | p - value | odds ratio |
|----------|-------------|-------|----------|-----------|------------|
| constant | -0.550      | 0.951 | -0.58    | .563      |            |
| $X_1$    | 0.157       | 0.075 | 2.10     | 0.036     | 1.17       |
| $X_2$    | 0.195       | 0.122 | 1.59     | 0.112     | 1.21       |

Classify the following observations to Group 1 ( $Y = 1$ ), Group 2 ( $Y = 0$ ).

| $X_1$ | $X_2$ |
|-------|-------|
| 3.3   | -3.5  |
| -27.9 | 6.3   |
| 59.0  | 23.1  |

- (b) Why is transformation of variables necessary ? Transform the following models to bring them to linear form :—

3

(i)  $Y = \alpha X^\beta$

(ii)  $Y = \alpha e^{X^\beta}$

(iii)  $Y = \frac{X}{\alpha X - \beta}$

- (c) Describe backward selection procedure to select variables for linear regression model.

3

## Section II

4. (a) Write down the ANOVA for one-way analysis with SAS commands.
- (b) A farmer wants to test three brands of weight-gain diets for chickens to determine if the mean weight gains for each of these brands are the same. He selected 15 chickens and randomly put each of them on one of these three brands of diet. The following table lists the weights (in pounds) gained by these chickens after a period of one month :—

10

| Brand A | Brand B | Brand C |
|---------|---------|---------|
| .8      | .6      | 1.2     |
| 1.3     | 1.3     | .8      |
| 1.7     | .6      | .7      |
| .9      | .4      | 1.5     |
| .6      | .7      | .9      |

At the 1% significance level, can you conclude that the mean weight gains by all chickens are the same for each of these three diets, ( $F_{(2, 12)} \alpha = 0.01 = 6.92$ ).



5. (a) Explain two-way analysis with one observation per cell, without interaction and 10 give example.

- (b) In an experiment to study the effect of two factors, A and B, a two-way layout was used and the measurements were replicated three times. Summary of the data is given below :—

Total sums of square = 5872.33

In this case our Model is —

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

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

$$p = 5, q = 4, r = 3$$

Total sums of square = 5872.33

Sums of Square due to A = 1843.167

Sums of Square due to B = 2258.867

Sums of Square due to A \* B = 1398.967

Analyse the data for  $\alpha = 0.05$ .

$$F_{(3, 40) \alpha = 0.05} = 2.84$$

$$F_{(4, 40) \alpha = 0.05} = 2.61$$

$$F_{(12, 40) \alpha = 0.05} = 2.00$$

6. (a) Explain 'Nested Model' with examples. Further write down the SAS code for this 10 model.

- (b) The purity of the drug was determined, three times by each of the three technicians in each of the four laboratories. The responses from the experiment are as follows :

| Laboratories |    |     |             |    |    |     |      |    |    |    |     |
|--------------|----|-----|-------------|----|----|-----|------|----|----|----|-----|
| I            |    |     | II          |    |    | III |      |    | IV |    |     |
|              |    |     | Technicians |    |    |     |      |    |    |    |     |
| I            | II | III | IV          | V  | VI | VII | VIII | IX | X  | XI | XII |
| 88           | 82 | 83  | 81          | 89 | 91 | 94  | 90   | 90 | 87 | 80 | 94  |
| 89           | 83 | 80  | 78          | 84 | 81 | 90  | 89   | 85 | 80 | 81 | 89  |
| 91           | 86 | 84  | 79          | 80 | 87 | 88  | 87   | 87 | 82 | 84 | 88  |

Total Sum of Squares = 666.75

Sum of squares due to Laboratories = 148.97

Sum of Squares due to Technicians in Laboratories = 308.44

Analyze the data. ( $\alpha = .01$ )

$$F_{(3, 24, 0.01)} = 4.72$$

$$F_{(8, 24, 0.01)} = 3.36.$$

- 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 in detail the relative merits and demerits of the 'Regression Model' and the 'Time Series Decomposition Model' from the viewpoint of forecasting and decision making.  
(b) A 20-year old shoe company needs a forecast of its business. How will you proceed to forecast its monthly sales over the next six months?

- 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: Mean Square Error provides an assessment of the goodness of the forecasting model.

Statement Number 2.2: Leading indicator model of regression is the most appropriate technique to forecast the rank of India in international competitiveness next year from its current year's rank in the world exports.

Statement Number 2.3: Strong correlation between two variables does not necessarily mean a cause and effect relationship between the two.

Statement Number 2.4: It is always safer to use the additive model:  $Y = T + S + C + E$  when we are not confident that the four components Trend (T), Seasonal (S), Cyclical (C) & Error (E) of our time series are mutually independent.

Statement Number 2.5: Forecasting future values of a quantitative variable is possible even if past quantitative data on it are not available.

- Q3: (a) Describe the Six Thinking Hats technique for making decision in complex situations.  
(b) Apply this technique to a toy manufacturing company which wants to make long term decisions about entering new lines of business.

[TURN OVER



## Section II

4. Define problem of Economic Order Quantity with known demand. The shortages are allowed. Rate of replenishment of inventory is not finite. Determine optimum production quantity for each production run stating the conditions and diagram.

5. For a fixed order quantity system find out the E.O.Q, optimum buffer stock, normal lead time consumption and reorder level with the following data;

Annual consumption in units = 5,000

Cost in Rs. Of one unit = 1.5

Cs ( set up cost ) in Rs. per Production Run = 13

C1 ( holding cost)in Rs.per unit = 0.22

Past lead times = 15 days, 25 days, 13 days, 14 days, 30 days & 17 days.

## Section III

(10)

6. Two competing companies Dot X Ltd and Line X Ltd produced and sell the same type of products in the market. Following is the data available

| Particulars          | Dot X Ltd | Line X Ltd |
|----------------------|-----------|------------|
| Sales                | 2,50,000  | 2,50,000   |
| Less:- Variable Cost | 1,50,000  | 2,00,000   |
| Less: Fixed Cost     | 75,000    | 25,000     |
| Net Profit           | 25,000    | 25,000     |

You are required to compute for both the companies

- P/V Ratio
- Break Even Sales (In Rupees)

You are also required to state which company is likely to earn greater profit in conditions of (i) Low demand and (ii) High demand

- N.B. :** (1) All questions are compulsory.  
 (2) Figures to the right indicates 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

**Q1. Explain in detail following :-**

(10)

- a) Method of selecting Six Sigma Project  
 b. SPC Rules  
 c) Filter Score  
 d. SIPOC & Scoping

**Q2.**

(10)

- a) Mfg. Company wants to understand which Brand is better & reliable w.r.t. quality in order to launch their premium brand. (Higher the number – Bad the quality)

|         |     |     |     |     |     |     |     |     |     |     |     |     |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Brand A | 175 | 165 | 160 | 154 | 169 | 170 | 182 | 177 | 159 | 162 | 176 | 163 |
| Brand B | 165 | 180 | 152 | 174 | 179 | 180 | 172 | 188 | 168 | 160 | 182 | 173 |

- b) Given below Yield % of a brand. Kindly help in setting improvement Target for their Project Team.

|         |      |      |      |      |      |      |      |      |      |      |      |      |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|
| Month   | Apr  | May  | Jun  | Jul  | Aug  | Sep  | Oct  | Nov  | Dec  | Jan  | Feb  | Mar  |
| Yield % | 83.4 | 72.1 | 82.9 | 93.1 | 72.4 | 81.9 | 92.2 | 82.8 | 78.6 | 88.0 | 76.5 | 79.0 |

- Q 3.** Given below is the data describing amount of precipitation (settling dirt) in gms (10)  
 using 3 factors. Conduct the DOE & write down the regression equation based on p-value. In case of further Optimization which factors will you change & in what direction?

| W  | R  | C  | Precipitation in gms |
|----|----|----|----------------------|
| -1 | -1 | -1 | 20.85                |
| 1  | -1 | -1 | 21.01                |
| -1 | 1  | -1 | 22.52                |
| 1  | 1  | -1 | 22.71                |
| -1 | -1 | 1  | 20.93                |
| 1  | -1 | 1  | 21.09                |
| -1 | 1  | 1  | 22.61                |
| 1  | 1  | 1  | 22.81                |

## SECTION II

**4.a)**

(05)

A manufacturing company outsourced it's after sales service activity to another company. They wanted to get a feel of type of complaints and it's volume in the last 6 months and which type of complaints are maximum and need immediate attention. . They asked the vendor to submit graphical representation of data, which can be presented to the management. Which quality tool you think the vendor should use to represent the data? Draw appropriate diagram to represent the data given below.

[ TURN OVER



| Complaint Type | Number of complaints |
|----------------|----------------------|
| A              | 27                   |
| B              | 95                   |
| C              | 6                    |
| D              | 15                   |
| E              | 46                   |
| F              | 7                    |

4.b)

(05)

Students have to do a project as part of their final semester. A group of 6 students decided to visit a manufacturing unit and collect their quality data for analysis and use it for the project. It involved following activities.

| Srl | Activity  | Time required in Days |
|-----|---|-----------------------|
| 1   | Permission of the managing director of the manufacturing unit | 8                     |
| 2   | Booking of ticket for travel                                  | 10                    |
| 3   | Collection of data at the unit                                | 2                     |
| 4   | Analysis of data set 1 with 2 students                        | 3                     |
| 5   | Analysis of data set 2 with 2 students                        | 5                     |
| 6   | Analysis of dataset 3 with 2 students                         | 4                     |
| 7   | Final analysis based on analysis of data set 1, 2 and 3       | 2                     |
| 8   | Writing of project Report                                     | 3                     |
| 9   | Printing 3 copies of the report by professional printer       | 2                     |
| 10  | Submission of the project                                     | 1                     |

1. Draw activity diagram for these activities.
2. Find our critical path
3. If the project submission date is 1<sup>st</sup> July then when should travel tickets be booked latest?

4.c)

(05)

Cold drink manufacturing company is assessing risks involved in it's launch of new product, which involves significant investment and would like to ensure that top risks are dealt with appropriate counter measures that will minimise the loss in case of risk occurrence.

- i. Following are the risks identified with it's corresponding probability of risk occurrence. Calculate risk exposure and find out the most serious risk.

| Risk                          | Probability of occurring | Loss if risk occurs |
|-------------------------------|--------------------------|---------------------|
| Product recall situation      | 2%                       | 80,000              |
| Significant product rejection | 0.1%                     | 1,000,000           |
| Competitive strike            | 10%                      | 25,000              |



- ii. Following are the countermeasure against the most serious risk with it's corresponding costs and new probabilities of reduced risks. Find out new risk exposure and risk reduction leverage.

| Countermeasure       | Total Cost | New Risk Probability | New Total Loss |
|----------------------|------------|----------------------|----------------|
| Advertising Campaign | 40,000     | 3%                   | 5,000          |
| Price Promotions     | 30,000     | 5%                   | 10,000         |
| Simultaneous Launch  | 10,000     | 8%                   | 15,000         |

- iii. Find out the cost effective countermeasure

(05)

5.a)

Meteorological department wanted to find out whether there is any relationship between amount of rain fall during day on rainy day and minimum temperature at night on that day during rainy season. They collected data as follows. Which QC tool can be used for this? What is your conclusion based on this data?

| Rain fall in mm | Min Temperature in degree cc |
|-----------------|------------------------------|
| 2.2             | 27                           |
| 3.2             | 26.5                         |
| 1.2             | 29                           |
| 0.8             | 30.5                         |
| 3.6             | 25.5                         |
| 4.8             | 22                           |
| 2.6             | 26                           |
| 4.5             | 21                           |
| 2.4             | 25                           |
| 0.3             | 28                           |
| 0.6             | 29                           |
| 3.5             | 26                           |
| 4.4             | 23.5                         |
| 0.5             | 29                           |
| 1.8             | 27                           |

5.b)

(05)

Suggest appropriate QC tool from 7 new QC tools that can be used for following problems.

|   |  |
|---|--|
| 1 | 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.   |
| 2 | The design department and the production department of a company had continuous issues of bad changes to design by production department and designs provided by design department to be impractical to build. They felt there could be common causes to these issues. |

[ TURN OVER



|   |   |
|---|---|
| 3 | A toy store was aiming to increase sales while improving the satisfaction of its customers with the toys that it sold. It would like to measure both the initial appeal (which related to actual purchase) and the longer term satisfaction (which related to company image) of a range of toys for boys aged 5 to 10, both being scored on a one-to-ten scale. |
| 4 | BMC studding whether there is any relationship between increase in property rates across various zones of city and increase in slums.   |
| 5 | 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.   |

5.c)

(05)

In a case study there are four 2-level factors A, B, C and D. We want to 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

|   | 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) |

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

1. Briefly explain the application of following tests/methods in analysis of clinical research data. [ Any 4 of 5, 2.5 Marks each ] (10)

- ANOVA and Dunnett's test
- Wilcoxon rank sum test
- Linear regression
- Fisher's Exact Test
- Relative Risk

2. Describe process flow of a single Clinical trial project and discuss the role of a statistician in a clinical trial project (10)

3. Answer following questions (5 marks each) (10)

- Explain Type I error, type II error and power of a test
- What are analysis populations? What are important analysis populations in a clinical trial and how are they determined?

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

5. Please read the situation described below and answer the questions mentioned below. (10)

You are planning a study comparing two topical medications (eye drops) that lower the intra-ocular pressure. Indication for use of eye drops is open angle Glaucoma. Glaucoma represents a condition where there is increase in intra-ocular pressure (Normal intra-ocular pressure is between 10 to 20 mmHg). . This will be a randomized double blind controlled clinical trial.

You want to prove that after 12 weeks of treatment, Test eye drops produce more reduction in intra-ocular pressure as compared to standard eye drops.

- a. Describe the null and alternative hypothesis for this study?
- b. What information will you need to calculate sample size for this study?
- c. Please discuss statistical methodology to test hypothesis in this study.

[ TURN OVER ]



6. Multiple Choice Question Answers – EACH QUESTION HAS ONE CORRECT CHOICE [ 2 marks each] (10)

### GENERAL INSTRUCTIONS TO ANSWER QUESTIONS BY SELECTING CHOICES

To answer a question, mention Question number followed by the alphabet representing the correct choice in your answer paper.

Example:

1. Sample question text: \_\_\_\_\_?
- |             |             |
|-------------|-------------|
| a. Choice 1 | b. Choice 2 |
| c. Choice 3 | d. Choice 4 |

If your answer is "Choice 4", then write following in your Answer paper.

Q. 1: d

- 1) All of the following statements about Proc Freq procedure are True, except.
  - a. Tables statement determines, for which variables, the frequency will be calculated.
  - b. Norow, nocol and nopercnt are valid options in Tables statement.
  - c. 'Fisher' option in tables statement will calculate Fisher's exact test
  - d. This procedure does not produce any visible output.
  
- 2) Which of the following statement is correct?
  - a. Proc Sort procedure is used to Transpose a SAS dataset.
  - b. Proc SQL must end with a RUN statement for it to execute.
  - c. Proc Plan procedure can be used to generate 2:2:1 randomization schedule
  - d. Proc Format procedure is used to produce formatted RTF output.
  
- 3) PROC FREQ procedure is commonly used for all of the following except
  - a. Produce summary statistics for continuous data.
  - b. Produce confidence interval for difference in success proportions for two treatment groups.
  - c. Generate exact binomial confidence intervals for success proportion in single sample
  - d. Generate CMH Test statistic.

4) Which of the following statement about the code below is correct?

```
Proc SQL;
```

```
Select * From Work.ABCD;
```

```
QUIT;
```

- a. Summary of All continuous variables from ABCD dataset in work library will be displayed.
- b. ABCD dataset will be created in work library.
- c. Rows from ABCD dataset in work library will be displayed and will be ordered based upon first variable on the dataset
- d. All rows from ABCD dataset will be displayed.

5) Identify the missing.

```
Data exam;
```

```
Number=roll/10;
```

```
Run;
```

```
Proc sort data=exam;
```

```
_____ number;
```

```
Run;
```

- a. Set
  - b. Merge
  - c. Freq
  - d. By
-



**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.

1. (a) State the statistical distance from an arbitrary point  $P(x_1, x_2, \dots, x_p)$  to a fixed point  $Q(y_1, y_2, \dots, y_p)$ . Under what condition will the statistical distance be equal to the Euclidean distance. Explain the need for statistical distance over Euclidean distance. 5
- (b) The following are 5 measurements on the variables  $x_1, x_2$  and  $x_3$ . 5

|       |    |   |   |   |    |
|-------|----|---|---|---|----|
| $x_1$ | 9  | 2 | 6 | 5 | 8  |
| $x_2$ | 12 | 8 | 6 | 4 | 10 |
| $x_3$ | 3  | 4 | 0 | 2 | 1  |

Obtain (i)  $\bar{x}$  (ii)  $S_n$  (iii) R (iv) generalised sample variance and (v) total variance.

2. (a) Consider the data given in the following matrix :— 5

$$X = \begin{bmatrix} 2 & 8 & 8 \\ 4 & 2 & 0 \\ 10 & 12 & 8 \end{bmatrix} \text{ and } \begin{matrix} b = [2 & 2 & -1]' \\ c = [1 & -1 & 3]' \end{matrix}$$

Obtain sample mean and sample variance of  $b'X$  and  $c'X$ . Also obtain covariance between  $b'X$  and  $c'X$ .

- (b) A random sample of size 6 is selected from a trivariate normal distribution with mean vector  $\mu = (2 \ 0 \ 3)'$  and variance covariance matrix 5

$$\Sigma = \begin{bmatrix} 9 & 0 & 2 \\ 0 & 4 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

Obtain distribution of (i)  $Y_1 = \bar{X}$  and (ii)  $Y_2 = \sum_{i=1}^6 iX_i$

Hence state distribution of  $Z = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$

[ TURN OVER

3. (a)  $\underline{X} \sim N_p(\mu, \Sigma)$  where

2

$$\underline{X} = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix} \quad \underline{\mu} = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and  $|\Sigma_{22}| > 0$ . State the conditional p.d.f. of  $X_{(1)}$  given  $X_{(2)} = x_{(2)}$

- (b) Explain the method of using Q – Q plots to test normality of given data. 4  
 (c) State properties of multivariate normal distribution. 4
4. (a) A random sample of size  $n$ ,  $X_1, X_2, \dots, X_n$ , is selected from  $N_p(\mu, \Sigma)$ . Explain the likelihood ratio test procedure to test  $H_0: \mu = \mu_0$  against  $H_1: \mu \neq \mu_0$ . State its relation to the Hotelling's  $T^2$  statistic. State the  $100(1 - \alpha)\%$  simultaneous confidence interval for  $\mu_i$ ,  $i = 1, 2, \dots, p$ . 5  
 (b) Explain the multivariate multiple regression model. Indicate the method of estimating the parameters. 5
5. (a) Suppose we measure two variables  $X_1$  and  $X_2$  for four items A, B, C and D. The data are as follows : 10

| Item | Observations |       |
|------|--------------|-------|
|      | $X_1$        | $X_2$ |
| A    | 5            | 4     |
| B    | 1            | -2    |
| C    | -1           | 1     |
| D    | 3            | 1     |

Use the K means clustering technique to divide the items into  $K = 2$  clusters. Start with initial groups (AB) and (CD).

- (b) Define first two principal components  $Y_1, Y_2$  of  $\Sigma$  matrix. Also obtain their variances. Obtain first two principal components of the following  $\rho$  matrix :

$$\rho = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}$$

6. (a) Write down orthogonal factor model and explain the terms — 10  
 (i) Communality  
 (ii) Specific variance.  
 (b) Show that communalities remain unchanged with factor rotation.  
 (c) Given the following data :—

$$\bar{x}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \bar{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad S_{\text{pooled}} = \begin{bmatrix} 7.3 & -1.1 \\ -1.1 & 4.8 \end{bmatrix}$$

Compute sample discriminant function.