Model Question Paper
Quantitative Methods - II (MSF1B2)

- Answer all 76 questions.
- Marks are indicated against each question.

Total Marks : 100

1. A random variable, Y, has the following probability distribution:

Y	20	40	60	80	100
Probability	0.15	0.20	0.30	0.25	0.10

15 values of Y have been observed. What is the probability that less than 12 observations are such that $40 \le Y \le 80$?

(a) 0.8441
(b) 0.1559
(c) 0.0668
(d) 0.4613
(e) 0.5387. **(2marks)**

2. The t distribution is

(a) Not a symmetrical distribution
(b) A discrete probability distribution
(c) A multimodal distribution
(d) Different for different number of degrees of freedom
(e) Positively skewed. **(1 mark)**

3. Modern Books and Magazines sells a variety of books, magazines and newspapers. The manager of the shop is concerned about one particular newspaper, The City Times. This newspaper costs Rs.1.20 each to the shop and can be sold for Rs.1.50 each. Any copy of the newspaper not sold by the end of the day can be disposed off for Re.0.20 each to the paper packet manufacturers. The manager of the shop has recorded the daily sales of The City Times over the past 200 days and his observations are given below:

Number of copies sold in a day	Number of days sold
200	40
300	60
400	80
500	20
Total	200

It is assumed that the number of copies of the newspaper sold in a day is a discrete random variable which will assume only the numbers listed above, and the shop will stock only the same number of copies.

What is the optimal number of copies of the newspaper that should be stocked?

(a) 200
(b) 300
(c) 400
(d) 500
(e) 1,400. **(2marks)**

4. Consider the following joint probability distribution.

X	–2	–1	0	–1	0
Y	1	0	1	2	0
P(X,Y)	0.1	0.2	0.1	0.3	0.3

If Z = (X + Y), then the value of V(Z) is

(a) 0.79
(b) 0.69
(c) 0.59
(d) 0.49
(e) 0.39. **(2marks)**

5. Which of the following statements is/are **true** regarding assumptions of decision theory?

I. The decision maker can define all decision alternatives or strategies which are being considered.
II. The decision maker can define the various states of nature for the decision setting which are not under control.
III. The decision maker can estimate quantitatively the consequences (benefits or costs) of selecting any decision alternative and having any state of nature occurrence.

(a) Only (I) above
(b) Only (II) above
(c) Only (III) above
(d) Both (II) and (III) above
(e) All (I), (II) and (III) above. **(1 mark)**

6. Which of the following statements is/are **true** regarding normal distribution?

I. It is a unimodal distribution.
II. The mean of a normally distributed population lies at the center of its normal curve.
III. The mean, median and mode of the distribution are never coincide.
IV. The two tails of the distribution touches the horizontal axis.

(a) Only (I) above
(b) Only (III) above
(c) Both (I) and (II) above
(d) Both (III) and (IV) above
(e) All (I), (II), (III) and (IV) above. **(1 mark)**

7. A taxi cab company has 6 Ambassadors and 4 Fiats. If 3 of these taxi cabs are in the shop for repairs and ambassador is as likely to be in for repairs as a Fiat. Then the probability that two of them are Ambassadors is

(a) $\frac{1}{8}$

(b) $\frac{1}{6}$

(c) $\frac{1}{4}$

(d) $\frac{1}{2}$

(e) $\frac{1}{10}$. **(1 mark)**

8. The mean lifetime of a sample of 100 light bulbs produced by a company is found to be 1,580 hours with a standard deviation of 90 hours. We want to test the hypothesis that the mean lifetime of the bulbs produced by the company is 1,600 hours.

Which of the following conclusions can be drawn on the basis of the above test at a significance level of 5%?

(a) The mean lifetime of the light bulbs is equal to 1,600 hours
(b) The mean lifetime of the light bulbs is not equal to 1,600 hours
(c) The mean lifetime of the light bulbs is significantly more than 1,580 hours
(d) The mean lifetime of the light bulbs is equal to 1,580 hours.
(e) The mean life time of the light bulbs is significantly less than 1,580 hours. **(2marks)**

9. The following details are available with regard to a hypothesis test on population mean:

H_0: $\mu = 9$

H_1: $\mu < 9$

$n = 25$

$s^2 = 256$

$\bar{X} = 3.60$

Significance level = 0.05

The population is normally distributed. It is later known that the true population mean is 5.
Which of the following is **true** with regard to the test?

(a) There is insufficient information for doing the test
(b) The normal distribution should be used
(c) The test does not lead to either type I or type II error
(d) The test leads to a type I error
(e) The test leads to a type II error. **(2marks)**

10. A local credit card agency reported to the government that the unpaid balance for their customers is normally distributed with an average of Rs. 5,500 with a standard deviation of Rs. 500. If an auditor randomly sampled 10 of the credit card company's customers, what is the probability that the mean of the sample would be between Rs. 5,300 and Rs. 5,400?

(a) 0.1608
(b) 0.6608
(c) 0.3971
(d) 0.2363
(e) 0.6334. **(2marks)**

11. If we accept $H_0 : \mu = 20$ against $H_1 : \mu \neq 20$ at a given level of significance with a positive value of the test statistic, then a test with $H_0 : \mu = 20$ versus $H_1 : \mu > 20$ using the same sample and the same level of significance

(a) Will always accept H_0
(b) Will always reject H_0
(c) May or may not accept H_0
(d) Will certainly lead to a Type I error
(e) Will certainly lead to a Type II error. **(1 mark)**

12. The following details are available with regard to a hypothesis test on a population mean:

$H_0 : \mu = 20$

$H_1 : \mu \neq 20$

$\sigma^2 = 81$

$n = 36$

The null hypothesis is rejected if $\bar{X} \leq 17.06$ or $\bar{X} \geq 22.94$. What is the probability of committing a type I error?

(a) 0.5
(b) 0.05
(c) 0.025
(d) 0.005
(e) 1.00. **(1 mark)**

13. The following details are available with regard to a hypothesis test on means of two populations:

H_0: $\mu_1 - \mu_2 = -1$

H_1: $\mu_1 - \mu_2 \neq -1$

$n_1 = 64$ $\sum x_1^2 = 1852$ $\sum x_1 = 320$

$n_2 = 36$ $\sum x_2^2 = 891$ $\sum x_2 = 144$

The samples collected from the two populations are independent.
What is the value of the test statistic?

(a) –3.578
(b) 3.578
(c) 0

(d) –0.354
(e) 1.789. **(1 mark)**

14. A researcher is interested in determining the average urban household income for the state Uttar Pradesh (U.P) of the country. He divides U.P. into homogeneous sectors and then randomly chooses 3 cities from each sector to sample. From each city, he randomly chooses 1000 families to survey. This type of sampling is referred to as

(a) Systematic sampling
(b) Cluster sampling
(c) Simple random sampling
(d) Stratified random sampling
(e) Judgment sampling. **(1 mark)**

15. A Ketchup manufacturer is in the process of deciding whether to produce a new extra-spicy brand. The company's marketing-research department used a national telephone survey of 6,000 households and found that the extra-spicy ketchup would be purchased by 335 of them. A much more extensive study made 2 years ago showed that 5 percent of the house holds would purchase the brand then. We want to test at a 2 percent significance level, should the company conclude that there is an increased interest in the extra-spicy flavor?

Which of the following is an appropriate null hypothesis for the above test?

(a) $H_0 : p = 0.5$
(b) $H_0 : p > 0.5$
(c) $H_1 : p > 0.05$
(d) $H_0 : p = 0.05$
(e) $H_1 : p < 0.05$. **(1 mark)**

16. As the sample size increases, the interval estimate for the population mean at a given confidence level will

(a) Increase in width
(b) Remain the same
(c) Increase initially and then decrease in width
(d) Decrease in width
(e) Double in width for increase in sample size by one observation. **(1 mark)**

17. Which of the following errors can occur at the time of observation, approximation and processing of data?

(a) Sampling errors
(b) Non-sampling errors
(c) Random occurrence errors
(d) Systematic errors
(e) Estimation errors. **(1 mark)**

18. Which of the following statements is **true** according to the Central Limit Theorem?

(a) The distribution of the sample is always symmetric
(b) The distribution of the sample is symmetric only for small sample sizes
(c) The sampling distribution of the mean approaches normality as the sample size increases
(d) The sampling distribution of the mean approaches normality as the sample size decreases
(e) The sampling distribution of the mean never approaches normality. **(1 mark)**

19. Which of the following statements is/are **true**?

I. If a right tailed test is rejected at 1% significance level, then it is always accepted at 5% significance level.
II. If a left tailed test is rejected at 5% significance level, then it is always rejected at 1% significance level.
III. If a two tailed test is accepted at 10% significance level, then it is always accepted at 5% significance level.

(a) Only (I) above
(b) Only (II) above

(c) Only (III) above
(d) Both (I) and (II) above
(e) All (I), (II) and (III) above. **(1 mark)**

20. The manager of Pioneer Enterprises wants to formulate a relationship between total cost and sales revenue, which may be used to estimate the total cost for a given level of sales. He collects the following information from the past records.

Sales (Rs. in lakhs)	20	23	38	28	45	40
Total cost (Rs. in lakhs)	17	22	35	22	39	36

Formulate a relationship between sales and total cost with sales as the independent variable using simple regression technique.

What is the standard error of estimate for the regression relationship?

(a) 1.858
(b) 2.747
(c) 3.636
(d) 4.525
(e) 5.414. **(2marks)**

21. The following table represents the heights of a sample of 12 fathers and their eldest sons:

Father (inches)	65	63	67	64	68	62	70	66	68	67	69	71
Son (inches)	68	66	68	65	69	66	68	65	71	67	68	70

The correlation coefficient between the heights of the fathers and sons is

(a) 0.7026
(b) 0.4946
(c) 0.5486
(d) 0.7856
(e) 0.8956. **(2marks)**

22. Mr. Ganapati Mendali, a second year MBA student, is doing a study on companies going public for the first time. He is curious to see whether there is a significant relationship between the sizes of the offering (Rs. in crores) and the price per share after the issue. The data are given below:

Size (Rs. in crores)	108.00	39.00	68.40	51.00	10.40	4.40
Price (Rs.)	12.00	13.00	19.00	12.00	6.50	4.00

The coefficient of determination for the above data set is

(a) 0.4523
(b) 0.5842
(c) 0.6753
(d) 0.7664
(e) 0.8575. **(2marks)**

23. Which of the following statements is/are **true**?

I. $F = 0 \Leftrightarrow RSS = 0 \Leftrightarrow R^2 = 0$.

II. $F = 1 \Leftrightarrow RSS = 0 \Leftrightarrow R^2 = 0$.

III. $F = 0 \Leftrightarrow RSS = 0 \Leftrightarrow R^2 = 1$.

(a) Only (I) above
(b) Only (II) above
(c) Only (III) above
(d) Both (I) and (II) above
(e) All (I), (II) and (III) above. **(2marks)**

24. The following details are available with regard to an estimated regression relationship $\hat{Y} = a + bX$

Error sum of squares = 56
n = 8

At X = 2, $\hat{Y} = 24$

What is the 90 percent prediction interval for the value of Y at X = 2?

(a) (17.56, 29.988)
(b) (17.56, 30.125)
(c) (18.064, 29.936)
(d) (18.064, 30.125)
(e) (18.064, 30.756). **(1 mark)**

25. The coefficient of correlation between variables X and Y is 0.80.

Cov (X, Y) = 20
$\sigma_X = 5$

What is the variance of Y?

(a) 5
(b) 20
(c) 25
(d) 50
(e) 100. **(1 mark)**

26. In a University, correlation coefficient between college entrance exam grades and scholastic achievement was found to be –1.08. On the basis of this you would tell the university that:

(a) The entrance exam is a good predictor of success
(b) They correlation coefficient is incorrect
(c) The exam is a poor predictor of success
(d) Students who do best on this exam will make the worst students.
(e) Students of this university are underachievers. **(1 mark)**

27. Which of the following statements is/are **true** regarding Scatter Diagram?

I. One of the easiest ways of studying the correlation between the two variables is with the help of a scatter diagram.
II. It gives an indication of the nature of the potential relationship between the variables.
III. Using scatter diagram if the variables are related, we cannot see what kind of line, or estimating equation, describes this relationship.

(a) Only (I) above
(b) Only (II) above
(c) Only (III) above
(d) Both (I) and (II) above
(e) All (I), (II) and (III) above. **(1 mark)**

28. For a data $\sum X = 725, \sum Y = 750, n = 10$. In the construction of a simple linear regression line, $Y = a + bX$, the coefficient 'b' is 0.81, then the value of coefficient 'a' is

(a) 133.725
(b) 162.750
(c) 6.561
(d) 16.275
(e) Data is insufficient. **(1 mark)**

29. The simple correlation coefficients between temperature (X_1), corn yield (X_2) and rainfall (X_3) are r_{12} (Between X_1 and X_2) = 0.59, r_{13} (Between X_1 and X_3) = 0.46 and r_{23} (Between X_2 and X_3) = 0.77.

The multiple correlation coefficient $R_{1.23}$ for the three variables is

(a) 0.5561
(b) 0.5671
(c) 0.5781
(d) 0.5901
(e) 0.6021. **(1 mark)**

30. A multiple regression relationship contains two independent variables. The standard error of estimate is 4.8. Error sum of squares = 576.

What is the number of data points?

(a) 24
(b) 25
(c) 26
(d) 27
(e) 28. **(1 mark)**

31. For a regression relationship the ratio of regression sum of squares to error sum of squares is 3 : 2. The total sum of squares is 90. What is the regression sum of squares?

(a) 54
(b) 45
(c) 36
(d) 18
(e) 27. **(1 mark)**

32. The following regression relationship between two variables, X and Y, has been obtained:

$\hat{Y} = 1{,}236 - 104X$

Where X is the independent variable and Y is the dependent variable.
The following details are also available:

$\Sigma Y^2 = 19{,}00{,}400$

$\overline{Y} = 612$

$\Sigma XY = 18{,}100$

Number of observations = 5

What is the approximate 90 percent prediction interval for Y if X = 7.50?

(a) (412.26, 409.38)
(b) (421.62, 490.38)
(c) (121.26, 499.38)
(d) (222.26, 444,38)
(e) (123.45, 543.21). **(2marks)**

33. Which of the following will be **true,** if in a regression relationship error sum of squares is equal to total sum of squares?

(a) Coefficient of determination = 1
(b) 0 < Coefficient of determination < 1
(c) Coefficient of determination = 0
(d) Coefficient of determination < 0
(e) Coefficient of determination >1. **(1 mark)**

34. If the coefficient of correlation between two variables X and Y is zero, then which of the following is **true**?

(a) $\sigma^2_x = 0$
(b) $\sigma^2_y = 0$
(c) $\overline{X} = 0$
(d) $\overline{Y} = 0$
(e) Cov (X, Y) = 0. **(1 mark)**

35. Which of the following is **true** with regard to multicollinearity in multiple regression analysis?

(a) It increases the influence of the individual variables in the model
(b) It reduces the accuracy of the model
(c) It increases the reliability of the regression coefficients
(d) It reduces the effectiveness of sensitivity analysis of the model

(e) It arises, if there is a low correlation among the independent variables. **(1 mark)**

36. If Error sum of Squares is 83.88 and Total sum of Squares is 1134.66, then the value of Coefficient of Determination is

(a) 0.926
(b) 0.074
(c) 0.740
(d) 0.092
(e) 0.0854. **(1 mark)**

37. Which of the following statements is/are **true** regarding Coefficient of Determination?

I. It is the square of the correlation coefficient.
II. It explains to what extent the variation of a dependent variable is expressed by the independent variable.
III. A low value of it shows a good linear relationship between the two variables.

(a) Only (I) above
(b) Only (II) above
(c) Only (III) above
(d) Both (I) and (II) above
(e) All (I), (II) and (III) above. **(1 mark)**

38. Which of the following is/are advantage(s) of chain base method in index numbers?

I. The link relatives calculated by using the chain base method, enable comparisons over successive years.
II. It enables the introduction of new items and the deletion of old items without altering the original series.
III. Whenever found necessary, weights can be adjusted in chain base method.
IV. Seasonal variations have minimal impact on chain index numbers.

(a) Only (I) above
(b) Only (II) above
(c) Both (III) and (IV) above
(d) (II), (III) and (IV) above
(e) All (I), (II), (III) and (IV) above. **(1 mark)**

39. A food grains merchant wants to know the overall change in prices of the commodities he deals in. He provides the following information:

Commodity	Year			
	2006		2008	
	Price (Rs./kg.)	Quantity (kg.)	Price (Rs./kg.)	Quantity (kg.)
Wheat	7	5000	10	4800
Rice	10	6000	14	5500
Pulses	25	3000	28	3500
Sugar	10	3000	14	2400
Salt	2	500	4	600

What is the Fisher's ideal price index for the year 2008 using 2006 as the base year?

(a) 115.5032
(b) 121.5032
(c) 129.5032
(d) 135.5032
(e) 139.5032. **(2marks)**

40. Estimate unweighted aggregate price index number from the following data.

Commodities (in Units)	Price in Base year (2006)	Price in Current Year (2008)
Rice (kg)	16.00	36.00
Wheat (kg)	12.25	22.10
Milk (lit.)	15.75	23.75

Egg (doz.)	15.30	28.50
Cheese (kg)	25.10	38.50

(a) 75.4826
(b) 99.8726
(c) 103.7626
(d) 125.8626
(e) 176.3626. **(1 mark)**

41. If the ratio between Laspeyre's price index number and Paasche's price index number is 28 : 27.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
X	1	10	2	5
Y	1	5	?	2

What is the price of the commodity Y in current year?

(a) 1
(b) 2
(c) 3
(d) 4
(e) 5. **(1 mark)**

42. Mr. Neelam Singh a dealer of grocery-items requires an overall comparison of the prices of the commodities, between the years 2006 and 2008. The following information is provided by him:

Commodities	Sugar	Salt	Wheat	Rice	Pulses
$P_0(Q_0 + Q_1)$	13750	9436	15000	14256	28700
$P_1(Q_0 + Q_1)$	18750	13500	25000	28435	38700

The Marshall–Edgeworth price index for the year 2008, using the year 2006 as the base year, is

(a) 123.2530
(b) 138.6530
(c) 153.2930
(d) 165.4530
(e) 175.6430. **(1 mark)**

43. Consider the following data.

Commodities	Base Year (2006)		Current Year (2008)	
	Price	Quantity	Price	Quantity
A	3	18	9	15
B	5	6	5	9
C	4	20	12	26
D	2	14	4	15

What is the weighted average of relatives price index that is to be calculated with the base values?

(a) 125.1667
(b) 175.1667
(c) 200.1667
(d) 225.1667
(e) 254.1667. **(1 mark)**

44. Which of the following statements is **false** regarding index numbers?

(a) Using index numbers we can establish trends
(b) Index numbers guide policy making
(c) Index numbers are not used to determine the purchasing power of a rupee
(d) Index numbers play a vital role in adjusting the original data to reflect reality
(e) Index number is calculated as a ratio of the current value to a base value and expressed as a percentage. **(1 mark)**

45. An aggregates price index, where the price for each item is weighted by its base period quantity is known as

(a) Paasche's price index
(b) Unweighted aggregates price index
(c) Fisher's ideal price index
(d) Laspeyre's price index
(e) Unweighted average of relatives price index. **(1 mark)**

46. If $\sum P_0Q_0 = 218$, $\sum P_1Q_0 = 391$, $\sum P_1Q_1 = 371$ and $\sum P_0Q_1 = 271$, then the value index is

(a) 130.4435
(b) 151.7735
(c) 160.5835
(d) 170.1835
(e) 195.9435. **(1 mark)**

47. Index number for the base year is always

(a) 1
(b) 10
(c) 100
(d) 1000
(e) 10000. **(1 mark)**

48. A time series for the years 1997-2008 had the following Relative Cyclical Residuals (RCR), in chronological order as follows:

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
RCR	–1%	–2%	1%	2%	–1%	–2%	1%	2%	–1%	–2%	1%	2%

What will be the RCR for 2009?

(a) +3%
(b) –1%
(c) –2%
(d) +2%
(e) Cannot be determined on the basis of the given information. **(2marks)**

49. In general, if the actual value of the dependent variable in an year in a time series is more than the value estimated by the trend equation then the Relative Cyclical Residual (RCR) is

(a) More than 0%
(b) Less than 0%
(c) Equal to 0%
(d) Always equal to 100%
(e) Always equal to 10%. **(1 mark)**

50. If the Relative Cyclical Residual (RCR) for an year is zero, then the percent of trend for that year is

(a) 100
(b) 50
(c) 10
(d) 5
(e) 1. **(2marks)**

51. The linear trend estimating equation for a time series is given below:

$\hat{Y} = 139 + 7.5x$, where $x = \text{Year} - 2006$

The observed value of Y for the year 2008 is 160.

What is the Relative Cyclical Residual (RCR) for the year 2008?

(a) –2.1698
(b) 10.8961
(c) 3.8961
(d) 96.2561

(e) 103.961. **(2marks)**

52. The seasonal index sales of a company for four quarters of a year are 200, 180, 160 and 260 respectively. If the total sales in the first quarter are worth Rs.50,000, estimate the worth of sales expected during the2nd quarter of the year.

(a) Rs. 45,000
(b) Rs. 55,000
(c) Rs. 35,000
(d) Rs. 55,500
(e) Rs. 65,000. **(2marks)**

53. SE company has recorded the following numbers (in Rs. millions) of total sales of its line of washing machines over 5 years:

Year	2004	2005	2006	2007	2008
Sales	3.5	3.8	4.0	3.7	3.9

Based on the above information, sales in which of the following year was closest to the trend line?

(a) 2004
(b) 2005
(c) 2006
(d) 2007
(e) 2008. **(3marks)**

54. The following information pertains to the actual and estimated sales on the basis of trend line, of wallets from the year 2001 to 2008.

Year	2001	2002	2003	2004	2005	2006	2007	2008
Actual Sales (Rs. in lakhs)	7.5	7.8	8.2	8.2	8.4	8.5	8.7	9.1
Estimated Sales (Rs. in lakhs)	7.6	7.8	8.0	8.2	8.4	8.6	8.8	9.0

Under Relative cyclical Residual (RCR) measure, in which year does the largest fluctuation from trend occur?

(a) 2001
(b) 2003
(c) 2005
(d) 2006
(e) 2008. **(1 mark)**

55. A business school is studying the number of enrolments to its various programs during six years. The following data are collected from past records:

Year	2003	2004	2005	2006	2007	2008
Number of enrolments in '000	185	175	190	180	185	200

What is the expected number of enrolments in the year 2009?

(a) 155.332
(b) 165.332
(c) 175.332
(d) 185.332
(e) 195.332. **(2marks)**

56. Consider the following data.

4	5	8	6	2	7	3	1	9

What is the moving average of order three associated with 7 of the following data?

(a) 5.667
(b) 6.333
(c) 5.333
(d) 5.000

(e) 4.000. **(1 mark)**

57. If deseasonalized value in time series is 2.00 and actual data is 1.58, the seasonal index is

(a) 79
(b) 61
(c) 19
(d) 12
(e) 2. **(1 mark)**

58. Consider the following data.

$$n = 16, \bar{\bar{X}} = 125.85, \bar{R} = 6.4 \text{ and } d_2 = 3.532$$

The Center Line (CL) and the Lower Control Limit (LCL) are

(a) 7.866 and 124.073
(b) 7.866 and 127.627
(c) 125.85 and 124.491
(d) 125.85 and 127.627
(e) 0.40 and 124.073. **(2marks)**

59. For a data $\bar{\bar{P}}$ = 0.45 and Upper Control Limit (UCL) = 0.737. What is the value of n (number of samples)? (Round off your answer to the nearest integer)

(a) 27
(b) 31
(c) 34
(d) 38
(e) 42. **(1 mark)**

60. Each of 20 lots of rubber belts contains 2000 rubber belts. The numbers of defective rubber belts in these lots are as follows:

405	418	324	332	292	340	282	300	320	296
397	434	294	324	220	460	258	226	460	190

Calculate the value of Center Line (CL) for the p chart for the fraction of defectives.

(a) 0.1321
(b) 0.1432
(c) 0.1643
(d) 0.1552
(e) 0.1762. **(1 mark)**

61. The following data provides the values of sample mean $(\bar{x})$, and range (R), for samples of size 5 each:

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean ($\bar{x}$)	11.2	11.8	10.8	11.6	11.0	9.6	10.4	9.6	10.6	10.0
Range (R)	7	4	8	5	7	4	8	4	7	9

The value for the Center Line (CL) of the $\bar{x}$ chart is

(a) 16.20
(b) 15.33
(c) 14.64
(d) 12.65
(e) 10.66. **(1 mark)**

62. The following data provides the values of sample range (R), for samples of size 6 each.

Sample No.	1	2	3	4	5	6	7	8	9	10
Range (R)	7	4	8	5	7	4	8	4	7	9

The value of Upper Control Limit (UCL) for the R chart is (Use the relevant D_3 or D_4 factor)

(a) 13.3182
(b) 12.6252
(c) 12.4365
(d) 11.5648
(e) 11.6428.

(1 mark)

63. The manager of a warehouse wants to construct a control chart for proportion of defectives obtained in repeated random samples of size 150 from a process, which is considered to be under control when the proportion of defectives, p, is equal to 0.15.

Calculate the Lower Control Limit (LCL) for the p chart.

(a) 0.1625
(b) 0.1425
(c) 0.1225
(d) 0.0825
(e) 0.0625. **(1 mark)**

64. The manager of a food chain outlet wants to ensure that the variability in the service time by their bearers is in control. He does a sampling of the average time taken by the bearers in serving the customers in the outlet. He finds that the mean of the sample range to be 1.5 minutes. If he has collected five samples each of ten observations to conduct the test then what is the Upper Control Limit (UCL) of the R chart? (Given that $d_2 = 3.078$ and $d_3 = 0.797$ for sample size of 10)

(a) 0.3352
(b) 1.0052
(c) 1.5002
(d) 2.0052
(e) 2.6652. **(1 mark)**

65. Which of the following charts deal(s) with qualitative factor(s) possessed by the data?

I. R charts.
II. p charts.
III. $\bar{x}$ charts.

(a) Only (I) above
(b) Only (II) above
(c) Only (III) above
(d) Both (I) and (III) above
(e) Both (II) and (III) above. **(1 mark)**

66. For generating $\bar{X}$ chart 40 samples are taken. The sum of all the observations is 560 and the value to be used for the centre line is 2.8. The size of each sample is

(a) 5
(b) 10
(c) 18
(d) 20
(e) 36. **(1 mark)**

67. The demand for a particular spare part in a factory was found to vary from day to day. In a sample study the following information was obtained:

Days	Mon	Tue	Wed	Thurs	Fri	Sat
No. of parts demanded	1124	1125	1110	1120	1126	1115

We need to test the hypothesis that the number of parts demanded does not depend on the day of the week. What is the value of the test statistic for the above test?

(a) 0.159
(b) 0.179
(c) 0.279
(d) 0.259
(e) 0.079. **(2marks)**

68.

A study compared the effects of four 1-month point-of-purchase promotions on sales. The unit sales for five stores using all four promotions in different months follow:

Free samples	78	87	81	89	85
One-pack gift	94	91	87	90	88
Rupees off	74	78	70	82	76
Refund by mail	79	83	78	69	81

A test of ANOVA has to be performed.

What is the estimated population variance using the variance between the sample means?

(a) 194
(b) 196
(c) 198
(d) 200
(e) 202. **(1 mark)**

69. The fitness levels of the soldiers of two battalions of the Indian Army are believed to have same variance. The senior officers of the Indian Army want to compare the actual variances in the fitness levels of the two battalions. A sample of 16 soldiers from the first battalion had a variance of 3.75 in their fitness levels. A sample of 10 soldiers from the second battalion had a variance of 5.38 in their fitness levels. The officer wants to test the belief at a significance level of 10%.

What conclusion can be drawn on the basis of the above test?

(a) There is no significant difference in the variances of the fitness levels of the two battalions
(b) The two sample variances are incorrect
(c) The variance in the fitness levels of first battalion is less than the variance in the second battalion
(d) The variance in the fitness levels of first battalion is more than the variance in the second battalion
(e) No conclusion can be drawn because the population means are not known. **(3marks)**

70. The following details are available with regard to a multiple regression analysis:

Regression sum of squares = 21.6
Error sum of squares = 2.296
Number of observations = 10
Number of independent variables = 2
What is the value of the F statistic for the multiple regression model?

(a) 3.293
(b) 32.9268
(c) 10.8268
(d) 2.2968
(e) 21.668. **(1 mark)**

71. The following data shows the number of claims processed per day by four employees of an insurance company, observed for a number of days. It has to be tested whether the average number of claims processed per day by all the four employees is same.

Employee 1:	15	17	14	12		
Employee 2:	12	10	13	17		
Employee 3:	11	14	13	15	12	
Employee 4:	13	12	13	15	10	9

What is the critical value for the appropriate statistical test at 5% level of significance?

(a) 2.87
(b) 3.29
(c) 4.43
(d) 4.77
(e) 5.29. **(1 mark)**

72. In a chi-square test, the contingency table has 4 rows and 4 columns. The number of degrees of freedom is

(a) 3
(b) 4
(c) 8
(d) 9
(e) 16. **(1 mark)**

73. In a test involving analysis of variance the following are found:

Estimate of the population variance based on the variance among sample means = 32.5
Estimate of the population variance based on the variance within the samples = 14.4

The F-ratio is equal to

(a) 0.44
(b) 0.0694
(c) 0.64
(d) 1.2569
(e) 2.2569. **(1 mark)**

74. If X follows continuous uniform distribution. It's probability density function is given by

$$f(x) = \begin{cases} \frac{1}{4} & ; \quad 1 \leq X \leq 5 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

What is the variance of the distribution?

(a) $\frac{1}{2}$
(b) $\frac{2}{3}$
(c) $\frac{4}{3}$
(d) $\frac{5}{3}$
(e) 4. **(2marks)**

75. Which of the following statements is/are **true** regarding simulation?

I. If the mathematical model set up could always be optimized by the analytical approach, then, there would be no need for simulation.
II. Only when interrelationships are too complex or there is uncertainty regarding the values that could be assumed by the variables or both, we would have to resort to simulation.
III. Simulations cannot be used to break down complex systems into sub-systems and study the behavior of each of these subsystems.

(a) Only (I) above
(b) Only (II) above
(c) Only (III) above
(d) Both (I) and (II) above
(e) All (I), (II) and (III) above. **(1 mark)**

76. Consider the following probability distribution of a variable X :

X	5.0	10.0	15.0	20.0	25.0	30.0	35.0
p	0.1	0.05	0.05	0.1	0.3	0.2	0.2

In a Monte Carlo simulation process the number 49 was generated in a particular run from a uniform distribution of 0 – 99. The value of X in this run will be

(a) 35
(b) 20
(c) 25
(d) 15

(e) 30. **(1 mark)**

END OF QUESTION PAPER

Suggested Answers
Quantitative Methods - II (MSF1B2)

ANSWER **REASON**

1. E The event $40 \leq Y \leq 80$ occurs when

$Y = 40$ or 60 or 80

$\therefore P(40 \leq Y \leq 80) = 0.20 + 0.30 + 0.25 = 0.75$

Let Z denote the number of observations of Y which satisfy the condition $40 \leq Y \leq 80$.

$\therefore$ Z follows a binomial distribution with number of trials = 15 and

Probability of success = 0.75.

$\therefore P(Z < 12) = P(Z \leq 11) = 1 - P(Z \geq 12)$

$= 1 - [P(Z = 12) + P(Z = 13) + P(Z = 14) + P(Z = 15)]$

$= 1 - [0.2252 + 0.1559 + 0.0668 + 0.0134]$

$= 1 - 0.4613$

$= 0.5387.$

2. D The t distribution is different for different degrees of freedom.

3. B Profit per newspaper sold = 1.50 – 1.20 = Re.0.30

Loss per unsold newspaper = 1.20 – 0.20 = Re.1.00

Deriving the probabilities:

Daily sales	200	300	400	500	
Number of days sold (f)	40	60	80	20	$\Sigma f = 200$
Probability $\left(p = \frac{f}{\Sigma f}\right)$	$\frac{40}{200} = 0.20$	$\frac{60}{200} = 0.30$	$\frac{80}{200} = 0.40$	$\frac{20}{200} = 0.10$	$\Sigma p = 1.00$

Conditional profit table

Stock level	Daily demand (Probability)				Expected daily profit (Rs.)
	200 (0.20)	300 (0.30)	400 (0.40)	500 (0.10)	
200	60	60	60	60	60
300	–40	90	90	90	64
400	–140	–10	120	120	29
500	–240	–110	20	150	–58

From the above table we can see that the maximum expected daily profit (Rs.64) is associated with the stock level of 300 newspapers. Hence the optimal number of the newspaper, The City Times, that should be stocked is 300.

Working notes:

Conditional profit for any stock level = Number of newspapers demanded and sold x 0.30 – Number of unsold newspapers x 1.00.

4. B Given that $Z = X+Y$.

Notice that Z takes on the values –1, 1 and 0 with probabilities 0.3, 0.4 and 0.3 respectively.

Z	–1	0	1
P(Z)	0.3	0.3	0.4

$E(Z) = -1\times0.3 + 0\times0.3 + 1\times0.4 = 0.1$

$V(Z) = (-1-0.1)^2(0.3) + (0-0.1)^2(0.3) + (1-0.1)^2(0.4) = 0.69$

Or

X	P(X)	XP(X)	$(X-(-0.7))^2$ P	Y	P(Y)	YP(Y)	$(Y-0.8)^2$ P

			(X)				(Y)
–2	0.1	–0.2	0.169	0	0.5	0	0.32
–1	0.5	–0.5	0.045	1	0.2	0.2	0.008
0	0.4	0	0.196	2	0.3	0.6	0.432
Total		**-0.7**	**0.41**	**Total**		**0.8**	**0.76**

To compute Cov(X,Y), we must examine the joint distribution

$$\text{COV}(X,Y) = (-2+0.7)(1-0.8)(0.1) + (-1+0.7)(-0.8)(0.2) + (0.7)(1-0.8)(0.1)$$
$$+(-1+0.7)(2-0.8)(0.3) + (0.7)(-0.8)(0.3)$$
$$= (-0.24)$$

V(Z) = V(X+Y)=V(X)+V(Y)+2Cov(X,Y) = 0.41+0.76+2×(-0.24) = 0.69.

5. E The following statements are **true** regarding assumptions of decision theory.

I. The decision maker can define all decision alternatives or strategies which are being considered.

II. The decision maker can define the various states of nature for the decision setting which are not under control.

III. The decision maker can estimate quantitatively the consequences (benefits or costs) of selecting any decision alternative and having any state of nature occurrence.

So option (e) is correct.

6. C The following statements are **true** regarding normal distribution.

I. It is a unimodal distribution.

II. The mean of a normally distributed population lies at the center of its normal curve.

Given statements (III) and (IV) are false. Since correct statements are

The mean, median and mode of the distribution are **coincide**.

The two tails of the distribution extend indefinitely and **never** touches the horizontal axis.

Hence Option (c) is correct.

7. D Given that N (total number of taxi cabs) = 10,

M(number of Ambassador cars) = 6,

n (sample size) = 3,

x (number of Ambassadors) = 2

$$P(X=x) = \frac{{}^{M}C_{x}\ {}^{(N-M)}C_{(n-x)}}{{}^{N}C_{n}}$$

$$= \frac{{}^{6}C_{2}\ {}^{(10-6)}C_{(3-2)}}{{}^{10}C_{3}} = \frac{15\times 4}{120} = \frac{1}{2}.$$

8. B The appropriate null hypothesis for the given test of hypothesis is given by, H0 : μ = 1,600.

The appropriate alternative hypothesis for the given test of hypothesis is given by, H1 : $\mu \neq$ 1,600.

Given that μ = 1,600, $\bar{x}$ = 1,580, s = 90, n = 100.

The standard error of mean = $\frac{s}{\sqrt{n}} = \frac{90}{\sqrt{100}} = 9.$

The test statistic for this test is, $z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{1580-1600}{9} = -2.22$

The critical value of z = ± 1.96 for a two tailed test at a 5% level of significance.

Since the test statistic z = – 2.22 falls in the rejection region, we reject the null hypothesis and conclude at 5% level of significance that the mean lifetime of the light bulbs is not equal to 1,600 hours.

9. E

$H_0 \; : \; \mu \; = \; 9$

$H_1 \; : \; \mu \; < \; 9$

The sample is small and the population variance is not known. The sample variance is specified. The population is normally distributed. Hence we should use the t distribution with 25 – 1 = 24 d.o.f.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{256}{25}} = 3.20$$

$$t = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{3.60-9}{3.20} = -1.688$$

At $\alpha = 0.05$, the critical value is –1.711.

The test statistic is more than the left tail critical value. So it falls in the acceptance region.

$\therefore$ We accept H_0. The true mean is 5. So H_0 is false. Hence the test leads to a type II error.

10. A Because the sample size is small we cannot apply the CLT. However, since the population is normally distributed the sampling distribution of mean will also be normally distributed with mean = population mean and standard deviation $= \frac{\sigma}{\sqrt{n}}$

$$E(\bar{x}) = \mu = 5500$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{10}} = 158.11$$

$$\therefore P(5300 < \bar{x} < 5400) = P\left(\frac{5300-5500}{158.11} < z < \frac{5400-5500}{158.11}\right)$$

$= P(-1.265 < z < -0.632)$

$= P(-1.265 < z < 0) - P(-0.632 < z < 0)$

$= P(0 < z < 1.265) - P(0 < z < 0.632)$

$$= \left[0.3962 + \frac{(0.3980-0.3962)}{(1.27-1.26)}(1.265-1.26)\right] - \left[0.2357 + \frac{(0.2389-0.2357)}{(0.64-0.63)}(0.632-0.63)\right]$$

$= 0.3971 - 0.2363 = 0.1608.$

11. C If we accept H_0:μ = 20 against of H_1:$\mu \neq$ 20 at a given level of significance with a positive value of the test statistic, then a test with H_0:μ = 20 versus H_1:μ > 20 using the same sample and the same level of significance may or may not accept H_0 because in a one tailed test the critical region on the relevant tail is larger than a two tailed test for the same level of significance.

12. B Probability of committing a type I error = Probability of rejecting the null hypothesis when it is true

By the central limit theorem $\bar{x}$ is normally distributed with mean = μ and variance = $\frac{\sigma^2}{n}$

Hence Probability of rejecting the null hypothesis when it is true =

$$P(\bar{x} \le 17.06 \text{ or } \bar{x} \ge 22.94) = P(\bar{x} \le 17.06) + P(\bar{x} \ge 22.94)$$

$$= P\left(\frac{17.06-20}{\sqrt{\frac{\sigma^2}{n}}} \le z\right) + P\left(\frac{22.94-20}{\sqrt{\frac{\sigma^2}{n}}} \ge z\right)$$

$$= P\left(\frac{-2.94}{\sqrt{\frac{81}{36}}} \le z\right) + P\left(\frac{2.94}{\sqrt{\frac{81}{36}}} \ge z\right)$$

$$= P(z \le -1.96) + P(z \ge 1.96)$$

$$= 0.025 + 0.025 = 0.05$$

13. B Estimated standard error of difference between means:

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{s_1^2}{n_1}+\frac{s_2^2}{n_2} = \frac{4}{64}+\frac{9}{36} = 0.3125$$

$$s_1^2 = \frac{1}{n_1-1}\left[\sum x_1^2 - n_1\bar{x}_1^2\right] = \frac{1}{63}\left[1852-64\left(\frac{320}{64}\right)^2\right] = 4$$

$$s_2^2 = \frac{1}{n_2-1}\left[\sum x_2^2 - n_2\bar{x}_2^2\right] = \frac{1}{35}\left[891-36\left(\frac{144}{36}\right)^2\right] = 9$$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{0.3125} = 0.559$$

Value of the test statistic =

$$\frac{(\bar{x}_1-\bar{x}_2)-(\bar{\mu}_1-\bar{\mu}_2)}{\text{-}\ \sigma_{\bar{x}_1-\bar{x}_2}} = \frac{\left(\frac{320}{64}-\frac{144}{36}\right)-(-1)}{0.559} = \frac{2}{0.559} = 3.578$$

14. D This type of sampling is referred to as stratified random sampling because the entire population is divided in to classes (sectors) or strata which are homogeneous in nature and samples a fixed number of observations from each stratum.

15. D The appropriate null hypothesis for the given test of hypothesis is given by, H_0: p= 0.05.

The appropriate alternative hypothesis for the given test of hypothesis is given by, H_1: $p > 0.05$.

16. D As the sample size increases, the interval estimate for the population mean at a given confidence level will decrease in width

17. B Non sampling errors occur at the time of observation, approximation and processing of data.

18. C According to the CLT the sampling distribution of mean approaches

Normality as the sample size increases.

19. C Given statement (III) is true.

If a two tailed test is accepted at 10% significance level, then it is always accepted at 5% significance level.

Given statements (I) and (II) are false. Since correct statements are

If a right tailed test is rejected at 1% significance level, then it is always **rejected** at 5% significance level.

If a left tailed test is rejected at 5% significance level, then it **may or may not** rejected at 1% significance level.

Hence Option (c) is correct.

20. A The equation of the regression line is $\hat{Y} = a + bX$

where, X = Sales (Rs. in lakhs)

Y = Total costs (Rs. in lakhs)

$$b = \frac{n\sum XY - \sum X\sum Y}{n\sum X^2 - \left(\sum X\right)^2}\ ;\quad a = \bar{Y} - b\bar{X}.$$

$\sum X = 194$; $\sum Y = 171$; $n = 6$;

$\sum XY = 5987$; $\sum X^2 = 6782$; $\bar{Y} = 28.5$.

$$b = \frac{6(5987)-(194)(171)}{6(6782)-(194)(194)} = \frac{2748}{3056} = 0.899.$$

$$a = \frac{171}{6} - (0.899)\left(\frac{194}{6}\right) = -0.568$$

$\therefore \hat{Y} = -0.568 + 0.899\,x$

Standard error of estimate (S_e) = $\sqrt{\dfrac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$

$\sum Y^2 = 5299$

$S_e = \sqrt{\dfrac{5299-(-0.568)(171)-(0.899)(5987)}{6-2}}$

$= 1.858.$

21. A

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2}}$$

	X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})\times(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
	65	68	-1.67	0.42	-0.7014	2.7889	0.1764
	63	66	-3.67	-1.58	5.7986	13.4689	2.4964
	67	68	0.33	0.42	0.1386	0.1089	0.1764
	64	65	-2.67	-2.58	6.8886	7.1289	6.6564
	68	69	1.33	1.42	1.8886	1.7689	2.0164
	62	66	-4.67	-1.58	7.3786	21.8089	2.4964
	70	68	3.33	0.42	1.3986	11.0889	0.1764
	66	65	-0.67	-2.58	1.7286	0.4489	6.6564
	68	71	1.33	3.42	4.5486	1.7689	11.6964
	67	67	0.33	-0.58	-0.1914	0.1089	0.3364
	69	68	2.33	0.42	0.9786	5.4289	0.1764
	71	70	4.33	2.42	10.4786	18.7489	5.8564
Total	800	811			40.3332	84.6668	38.9168

$$r = \frac{40.3332}{\sqrt{84.6668 \times 38.9168}} = 0.7026.$$

22. A To compute the coefficient of correlation between the two variables we tabulate them as below:

Size (X)	Price (Y)	$x = X-\bar{X}$	$y = Y-\bar{Y}$	x^2	y^2	xy
108	12	61.13	0.92	3736.8769	0.8464	56.2396
39	13	-7.87	1.92	61.9369	3.6864	–15.1104
68.4	19	21.53	7.92	463.5409	62.7264	170.5176
51.0	12	4.13	0.92	17.0569	0.8464	3.7996
10.4	6.5	-36.47	-4.58	1330.0609	20.9764	167.0326
4.40	4.0	-42.47	-7.08	1803.7009	50.1264	300.6876
Total				7413.1734	139.2084	683.1666

The mean values

$$\bar{X} = \frac{\sum X}{n} = \frac{281.2}{6} = 46.87$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{66.5}{6} = 11.08$$

We know that the coefficient of correlation

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$\therefore \quad r = \frac{683.1666}{\sqrt{7413.1734 \times 139.2084}} = 0.6725$$

The coefficient of determination = $r^2 = (0.6725)^2 = 0.4523$.

23. A Given statement (I) is correct.

$$F = 0 \Leftrightarrow RSS = 0 \Leftrightarrow R^2 = 0$$

Hence option (a) is correct.

24. C $S_e = \sqrt{\frac{ESS}{n-2}} = \sqrt{\frac{56}{8-2}} = 3.055$

90% prediction interval = $\hat{Y} \pm t \times S_e = 24 \pm t_{6,0.10} S_e$

$$24 \pm (1.943 \times 3.055) = (18.064,\ 29.936).$$

25. C Given that,

$$r_{xy} = 0.80$$

$$0.80 = \frac{\text{cov}(x,y)}{\sigma_X \sigma_y}$$

$$\sigma_y = \frac{20}{5 \times 0.80}$$

$$\sigma_y = 5 \quad \Rightarrow \sigma_y^2 = 25.$$

26. B As correlation coefficient should always fall between +1 to –1 the value calculated by the statistician is not correct.

27. D The following statements are **true** regarding Scatter Diagram.

I. One of the easiest ways of studying the correlation between the two variables is with the help of a scatter diagram.

II. It gives an indication of the nature of the potential relationship between the variables.

Given Statement (III) is false. Since correct statement is

Using scatter diagram if the variables are related, we **can** see what kind of line, or estimating equation, describes this relationship.

So option (d) is correct.

28. D Given that $\sum X = 725, \sum Y = 750, n = 10$.

$$\bar{X} = \frac{725}{10} = 72.5 \quad \text{and} \quad \bar{Y} = \frac{750}{10} = 75.$$

In the construction of a simple linear regression line, $Y = a + bX$, the coefficient 'b' is 0.81 then the coefficient 'a' is given by $a = \bar{Y} - b\bar{X}$

$$a = 75 - (0.81) \times 72.5 = 16.275.$$

29. D We have to calculate the multiple correlation coefficient treating X_1 as dependent and X_2 and X_3 as independent variables, i.e., we have to find

$$R_{1.23} = \sqrt{\frac{r^2_{12} + r^2_{13} - 2\, r_{12}\, r_{13}\, r_{23}}{1 - r^2_{23}}}$$

Substituting the given values

$$R_{1.23} = \sqrt{\frac{(0.59)^2 + (0.46)^2 - 2(0.59 \times 0.46 \times 0.77)}{1-(0.77)^2}} = 0.5901.$$

30. E In multiple regression relationship:

Standard error of estimate, $s_e = \sqrt{\frac{\sum\left(Y-\hat{Y}\right)^2}{n-k-1}} = \sqrt{\frac{ESS}{n-k-1}}$

Given : ESS = 576

n = ?

k = 2

s_e = 4.8

$$\therefore 4.8 = \sqrt{\frac{ESS}{n-2-1}}$$

$$\therefore \frac{ESS}{n-3} = 23.04$$

$$\text{or } \frac{576}{n-3} = 23.04$$

$$\text{or } n = \frac{576}{23.04} + 3 = 28$$

31. A

Coefficient of determination $= \frac{RSS}{TSS} = \frac{RSS}{ESS+RSS}$

$$= \frac{\frac{RSS}{ESS}}{\frac{ESS}{ESS} + \frac{RSS}{ESS}} = \frac{\frac{3}{2}}{1+\frac{3}{2}} = 0.60$$

$\therefore \frac{RSS}{TSS} = 0.60$ or RSS = 0.60 × TSS = 0.60 × 90 = 54.

32. B

Standard error of estimate, $s_e = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n-2}}$

ΣY^2 = 19,00,400 (given)

ΣY = $\overline{Y}$ x n = 612 x 5 = 3060

ΣXY = 18,100 (given)

$\hat{Y}$ = 1,236 – 104X

\ a = 1,236 and b = -104

$$\therefore s_e = \sqrt{\frac{19{,}00{,}400 - (1236)(3060) - (-104)(18{,}100)}{5-2}} = 14.61.$$

The sample size is less than 30 and the standard deviation of the population is not known. Hence the appropriate distribution to be used for the prediction interval is the t-distribution.

Degrees of freedom = n – 2 = 5 – 2 = 3.

The appropriate t–value is 2.353 (from the t–table)

The limits of the prediction interval are given below:

For, x = 7.50, $\hat{Y}$ = 1236 – 104 (7.50) = 456

The prediction interval is:

$\hat{Y} - t\,(s_e)$ = 456 ± 2.353 (14.61)

= 456 ± 34.38

= (421.62, 490.38).

33. C RSS + ESS = TSS

$\therefore$ ESS = TSS $\Rightarrow$ RSS = 0

If RSS = 0 then , coefficient of determination = $\frac{RSS}{TSS}$ =0.

34. E

$$r = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

$\therefore r = 0 \Rightarrow Cov(X, Y) = 0.$

35. D Multicollinearity reduces the effects of the individual variables in the model; however the variables may together explain the dependent variable. It does not reduce the accuracy of the model. It reduces the effectiveness of sensitivity analysis of the model. Hence (a), (b), (c) and (e) are not true, (d) is true.

36. A

Coefficient of determination $= 1 - \frac{ESS}{TSS}$

$= 1 - \frac{83.88}{1134.66}$

$= 0.926.$

37. D The following statements are **true** regarding Coefficient of Determination.

I. It is the square of the correlation coefficient.

II. It explains to what extent the variation of a dependent variable is expressed by the independent variable.

Given Statement (III) is false. Since correct statement is

A **high** value of it shows a good linear relationship between the two variables.

So option (d) is correct.

38. E The following are advantages of chain base method in index numbers.

I. The link relatives calculated by using the chain base method, enable comparisons over successive years.

II. It enables the introduction of new items and the deletion of old items without altering the original series.

III. Whenever found necessary, weights can be adjusted in chain base method.

IV. Seasonal variations have minimal impact on chain index numbers.

Hence Option(e) is correct.

39. C

Fisher's ideal price index $= \sqrt{\frac{\sum P_1Q_0 \times \sum P_1Q_1}{\sum P_0Q_0 \times \sum P_0Q_1}} \times 100$

P_0	Q_0	P_1	Q_1	P_0Q_0	P_1Q_1	P_0Q_1	P_1Q_0
7	5000	10	4800	35000	48000	33600	50000
10	6000	14	5500	60000	77000	55000	84000
25	3000	28	3500	75000	98000	87500	84000
10	3000	14	2400	30000	33600	24000	42000
2	500	4	600	1000	2400	1200	2000
Total				201000	259000	201300	262000

$$=\sqrt{\frac{262000\times 259000}{201000\times 201300}}\times 100=129.5032.$$

40. E Here base year is 2006 and current year is 2008

commodities	P_0	P_1
Rice	16	36
Wheat	12.25	22.1
Milk	15.75	23.75
Egg	15.3	28.5
Cheese	25.1	38.5
Total	84.4	148.85

Unweighted aggregate price index = $\frac{\sum P_1}{\sum P_0}\times 100=\frac{148.85}{84.4}\times 100=176.3626.$

41. D Let the misting value be 'x'

Given that

$$\frac{L}{P}=\frac{28}{27}$$

$$or,\quad \frac{\frac{\sum P_1Q_0}{\sum P_0Q_0}\times 100}{\frac{\sum P_1Q_1}{\sum P_0Q_1}\times 100}=\frac{28}{27}$$

$$or,\quad \frac{\frac{20+5x}{10+5}}{\frac{10+2x}{5+2}}=\frac{28}{27}$$

$$or,\quad \frac{20+5x}{15}\times\frac{7}{10+2x}=\frac{28}{27}$$

$$or,\quad \frac{20+5x}{10+2x}=\frac{20}{9}$$

$$or,\quad x=4.$$

42. C

Marshall –Edgeworth price index = $\frac{\sum P_1(Q_0+Q_1)}{\sum P_0(Q_0+Q_1)}$ x 100

$$=\frac{124385}{81142}\times 100=153.2930.$$

43. E Weighted average of relatives price index

$$=\frac{\sum\left[\left(\frac{P_1}{P_0}\times 100\right)(P_0Q_0)\right]}{\sum P_0Q_0}$$

P_0	Q_0	P_1	Q_1	P_1/P_0	$(P_1/P_0)*100$	P_0Q_0	$((P_1/P_0)*100)P_0Q_0$
3	18	9	15	3	300	54	16200
5	6	5	9	1	100	30	3000
4	20	12	26	3	300	80	24000
2	14	4	15	2	200	28	5600
Total						192	48800

$$=\frac{48800}{192}=254.1667.$$

44. C The following statements **true** regarding index numbers.

Using Index numbers we can establish trends

Index numbers guide policy making

Index numbers play a vital role in adjusting the original data to reflect reality

Index number is calculated as a ratio of the current value to a base value and expressed as a percentage.

Given option(c) is false. Since correct statement is

Index numbers are **used** to determine the purchasing power of the rupee.

Hence option (c) is correct.

45. D A weighted aggregate price index where the weight for each item is its base period quantity is known as Laspeyres price index.

46. D Value Index is $\frac{\Sigma P_1 Q_1}{\Sigma P_0 Q_0} \times 100 = \frac{371}{218} \times 100 = 170.1835.$

47. C Index number for the base year is always 100.

48. E Based on the given information, it is not possible to determine the value, as it is difficult to forecast the cyclical fluctuations. It is because of the fact that time period of the cyclical variations varies from time to time.

49. A In general, if the actual value in a year in a time series is more than the estimated value then the relative cyclical residual is more than 0%.

50. A Relative cyclical residual $= \frac{Y - \hat{Y}}{\hat{Y}} \times 100$

$= \left(\frac{Y}{\hat{Y}} \times 100\right) - 100$

= Percent of trend – 100

$\therefore$ 0 = Percent of trend – 100

$\Rightarrow$ Percent of trend = 100.

51. C Year = 2008

$\therefore$ x = 2008 – 2006 = 2

$\hat{Y} = 139 + (7.5 \times 2) = 154$

Relative cyclical residual $= \left(\frac{Y - \hat{Y}}{\hat{Y}}\right) \times 100$

Y = 160

$\therefore$ Relative cyclical residual for the year 2008 = $\left(\frac{160 - 154}{154}\right) \times 100$ = 3.8961.

52. A Estimated sales for the 2nd quarter

$= \frac{\text{sales for the 1st quarter}}{\text{sales index for the 1st quarter}} \times \text{sales index for the 2nd quarter}$

$= \frac{50000}{200} \times 180$ = Rs. 45000.

53. E Percent of trend measure $= \frac{\text{Actual sales}}{\text{Estimated sales}} \times 100$

Equation of secular trend line:

$\hat{y}$ = a + bx

Where, x = X – 2006

$$b = \frac{\Sigma xY}{\Sigma x^2} = \frac{(-2\times3.5)+(-1\times3.8)+(0\times4)+(1\times3.7)+(2\times3.9)}{4+1+0+1+4} = \frac{0.7}{10} = 0.07$$

$$a = \overline{y} = \frac{3.5+3.8+4.0+3.7+3.9}{5} = 3.78$$

$\therefore\ \hat{y} = 3.78 + 0.07x$

x value 2004 is −2; $\hat{Y}$ =3.78+0.07x = 3.78+0.07(−2) = 3.64

x value 2005 is −1; $\hat{Y}$ =3.78+0.07x = 3.78+0.07(−1) = 3.71

x value 2006 is 0; $\hat{Y}$ =3.78+0.07x = 3.78+0.07(0) = 3.78

x value 2007 is 1; $\hat{Y}$ =3.78+0.07x = 3.78+0.07(1) = 3.85

x value 2008 is 2; $\hat{Y}$ =3.78+0.07x = 3.78+0.07(2) = 3.92

Year	2004	2005	2006	2007	2008
Actual sales	3.5	3.8	4.0	3.7	3.9
Forecasted sales	3.64	3.71	3.78	3.85	3.92
Percent of trend (A)	96.15%	102.43%	105.82%	96.1%	99.49%
Relative cyclical residual = A – 100	–3.85%	2.43%	5.82%	–3.9%	–0.51%

Sales was closest to the trend line in year 2008, because the relative cyclical residual had the least magnitude in that year.

54. B

Relative Cyclical Residual Measure = $\frac{Y-\hat{Y}}{\hat{Y}} \times 100$

Year	Y	$\hat{Y}$	$\frac{Y-\hat{Y}}{\hat{Y}}\times100$
2001	7.5	7.6	-1.31579
2002	7.8	7.8	0
2003	8.2	8	2.5
2004	8.2	8.2	0
2005	8.4	8.4	0
2006	8.5	8.6	-1.16279
2007	8.7	8.8	-1.13636
2008	9.1	9	1.111111

The largest fluctuation from trend has taken place in the year 2003, i.e., 2.5 %

55. E The estimating equation is : $\hat{y} = a + bx$

Year (X)	Y	$x = (X-\overline{X})\times2$	xY	x^2
2003	185	–5	-925	25
2004	175	–3	-525	9
2005	190	–1	-190	1
2006	180	1	180	1
2007	185	3	555	9
2008	200	5	1000	25
Total	1115	0	95	70

$$b = \frac{\Sigma xY}{\Sigma x^2} = \frac{95}{70} = 1.357$$

$$a = \overline{Y} = \frac{1115}{6} = 185.833$$

$\therefore$ The estimating equation is : $\hat{Y} = 185.833 + 1.357x$

When X = 2009, x = 7 and expected number of enrolments is $\hat{Y} = 185.833 + 1.357(7) = 195.332$.

56. E

Data Points.	Moving total of order 3.	Moving averages
4		
5	17	5.667
8	19	6.333
6	16	5.333
2	15	5.000
7	12	4.000
3	11	3.667
1	13	4.333
9		

Here moving total 17 represents 4 + 5 + 8; moving total 19 represents 5 + 8 + 6 etc,.
Here moving average 5.667 is associated to 5,
Here moving average 6.333 is associated to 8,
Similarly,
Moving average 4.000 is associated to 7.

57. A

$$\text{Deseasonalized value} = \frac{\text{Actual data}}{\text{Seasonal index}} \times 100$$

$$\therefore \text{Seasonal Index} = \frac{\text{Actual data}}{\text{Deseasonalized value}} \times 100$$

$$= \frac{1.58}{2.00} \times 100 = 79.$$

58. C We know that the value of $\overline{\overline{X}}$ is used to draw the center line. Therefore, from the given data the center line is drawn at 125.85. The lower control limits is calculated as follows:

$$\text{Lower control limit} = \overline{\overline{X}} - \frac{3\overline{R}}{d_2\sqrt{n}}$$

$$= 125.85 - \frac{3(6.4)}{3.532(\sqrt{16})}$$

$$= 125.85 - \frac{19.2}{14.128}$$

$$= 125.85 - 1.359$$

$$= 124.491.$$

59. A Given that $\overline{\overline{p}} = CL = 0.45 = p$.

$$\therefore q = 1 - p = 1 - 0.45 = 0.55.$$

$$UCL = p + 3\sqrt{\frac{pq}{n}} = 0.737$$

$$\Rightarrow 0.737 = 0.45 + 3\sqrt{\frac{0.45 \times 0.55}{n}}$$

$$\Rightarrow 0.737 - 0.45 = 3\sqrt{\frac{0.2475}{n}}$$

$$\Rightarrow 0.287 = 3\sqrt{\frac{0.2475}{n}}$$

$$\Rightarrow 0.0957 = \sqrt{\frac{0.2475}{n}}$$

$$\Rightarrow (0.0957)^2 = \frac{0.2475}{n}$$

$$\Rightarrow 0.0092 = \frac{0.2475}{n}$$

$$\Rightarrow n = \frac{0.2475}{0.0092} = 26.9022 \approx 27.$$

60. C Total number of defectives out of 40,000 items in 20 samples is:
405 + 420 + 324 + … + 460 + 190 = 6572

$$\bar{p} = \frac{6572}{40000} = 0.1643$$

Therefore the central line is set at 0.1643.

61. E The central line of the $\bar{x}$ chart is = Mean of the sample means ($\bar{\bar{x}}$)

Number of samples, n = 10

Sum of the means, $\sum \bar{x}$ = 106.6

Mean of the sample means or grand mean, $\bar{\bar{x}}$ = 106.6 /10 = 10.66.

62. B The central line of for the R chart = Mean of the sample range ($\bar{R}$)

$\bar{R}$ = ∑R / n = 63 /10 = 6.3.

The value of the upper control limit for the R chart = $\bar{R}\,D_4$

= 6.3 × 2.004 = 12.6252.

63. E Lower Control Limit for p chart = $p - 3\sqrt{\frac{pq}{n}}$

$$= 0.15 - 3\sqrt{\frac{0.15 \times 0.85}{150}}$$

$$= 0.15 - 0.0875$$

$$= 0.0625.$$

64. E The Upper Control Limit for the R chart is given as,

$$UCL = \bar{R} + \frac{3d_3}{d_2}\bar{R}$$

$$= 1.5 + \frac{3 \times 0.797}{3.078} \times 1.5$$

= 1.5 + 1.1652 = 2.6652.

65. B The p charts deal with the attributes that are qualitative factors possessed by the data. Whereas the $\bar{x}$ and R charts deal with quantitative aspects

66. A Let the size of each sample be x.

The value to be used for the centre line is the grand mean. In this case the grand mean $(\bar{\bar{x}})$ = 2.8.

Sum of all the observations = Σx = 560 (given)

Now, $\bar{\bar{x}} = \frac{\Sigma x}{n \times k}$

Where,

n = Sample size =?

k = number of samples = 40 (given)

$\therefore n = \frac{\Sigma x}{\bar{\bar{x}}.k}$

$n = \frac{560}{40 \times 2.8} = 5$

67. B Here we set up the null hypothesis, H_0 that the number of parts demanded does not depend on the day of week.

Under the null hypothesis, the expected frequencies of the spare part demanded on each of the six days would be:

$$\frac{1}{6}(1124+1125+1110+1120+1126+1115) = \frac{6720}{6} = 1120$$

this is a chi-square test of independence and the contingency table is given as:

Days	Frequency		$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
	f_o	f_e		
Mon.	1124	1120	16	0.014
Tues.	1125	1120	25	0.022
Wed.	1110	1120	100	0.089
Thurs.	1120	1120	0	0
Fri.	1126	1120	36	0.032
Sat.	1115	1120	25	0.022
Total	6720	6720		0.179

Then the value of chi-square is given by:

i.e. $\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i} = 0.179$

68. D The value of the grand mean $\bar{\bar{x}} = \frac{\sum x_{ji}}{n} = \frac{420+450+380+390}{20} = \frac{1640}{20} = 82$

Variance between the samples can be computed as below:

n_j	$\bar{x}_j$	$(\bar{x}_j - \bar{\bar{x}})^2$	$n_j(\bar{x}_j - \bar{\bar{x}})^2$
5	84	4	20
5	90	64	320
5	76	36	180
5	78	16	80
Total			600

The variance between the sample means = $s^2_{\bar{x}} = \frac{\sum n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1} = \frac{600}{4-1} = 200$

69. A We set the null and alternative hypothesis as follows:

$H_0 : \sigma_1^2 = \sigma_2^2$ (Null Hypothesis: The variances in the fitness levels of two battalions are same)

$H_1 : \sigma_1^2 \neq \sigma_2^2$ (Alternative Hypothesis: The two battalions do not have same variance in their fitness levels)

We are given that $n_1 = 16$ and $s_1^2 = 3.75$ and $n_2 = 10$ and $s_2^2 = 5.38$

We calculate the F statistic, $F = \frac{s_1^2}{s_2^2} = \frac{3.75}{5.38} = 0.697$

The number of degrees of freedom in the numerator is 16 −1 = 15 and in the denominator it is 10 − 1 = 9. Since we require both the limits, the limit F (15, 9, 0.05) is directly obtained from the tables. Its value is 3.01. Therefore the upper limit value is 3.01.

The critical value in the left tail is given by F (15, 9, 0.95).

We get this value as follows:

$$F(n, d, \alpha) = \frac{1}{F(d, n, 1-\alpha)}$$

Where, n is the degree of freedom in the numerator

d is the degree of freedom in the denominator

α is the significance level.

$$\Rightarrow F(15, 9, 0.95) = \frac{1}{F(9, 15, 0.05)} = \frac{1}{2.59} = 0.386$$

The lower limit value is 0.386.

Since the value of F statistic for the sample falls in the acceptance region, we accept the null hypothesis. We conclude that the samples belonging to the two battalions have the same variance in fitness levels

70. B

$$\text{F statistic} = \frac{RSS/k}{ESS/(n-(k+1))}$$

Given: RSS = 21.6

ESS = 2.296

Number of observations (n) = 10

Number of independent variables (k) = 2

$$\therefore \quad F = \frac{(21.6/2)}{[(2.296/(10-2-1)]} = \frac{10.8}{(2.296/7)} = 32.9268.$$

71. B This is an analysis of variance. The critical value would be found from the table of F-distribution.

The number of degrees of freedom in numerator = (k – 1) = 4 – 1 = 3

The number of degrees of freedom in denominator = $(n_T - k)$ =19–4= 15.

The value for F statistic from the table at 5% level of significance is given as:

F(3, 15, 0.05) = 3.29.

72. D Number of degrees of freedom = (Number of rows – 1) (Number of columns – 1)

= (4–1)(4–1)

= 9.

73.

E

$$\text{F ratio} = \frac{\text{Estimate of the population variance based on the variance among the sample means}}{\text{Estimate of the population variance based on the variance within the samples}}$$

$$= \frac{32.5}{14.4} = 2.2569.$$

74. C The probability density function of a typical random variable is given by

$$f(x)=\begin{cases}\dfrac{1}{(b-a)} & ; \ a \le X \le b \\ 0 & ; \ \text{otherwise}\end{cases}$$

Mean= E(X)= $\dfrac{(a+b)}{2}$

Variance= V(X)= $\dfrac{(b-a)^2}{12}$

Given that a=1 and b=5.

$\therefore$ Variance = $\dfrac{(5-1)^2}{12}=\dfrac{16}{12}=\dfrac{4}{3}$.

75. D The following statements are **true** regarding Simulation.

I. If the mathematical model set up could always be optimized by the analytical approach, then, there would be no need for simulation.

II. Only when interrelationships are too complex or there is uncertainty regarding the values that could be assumed by the variables or both, we would have to resort to simulation.

Given statement III is not true. Since Simulations **can** be used to break down complex systems into sub-systems and study the behavior of each of these subsystems.

76. C

X	Probability	Cumulative Probability	Random numbers allotted
5	0.10	0.10	00 - 09
10	0.05	0.15	10 - 14
15	0.05	0.20	15 - 19
20	0.10	0.30	20 - 29
25	0.30	0.60	30 - 59
30	0.20	0.80	60 - 79
35	0.20	1.00	80 - 99

Since the random number generated was 49(which lies in between 30 – 59), the value of X in this run is 25.