

Question Paper

Quantitative Methods-II (132): October 2006

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

1. If two fair dice are thrown simultaneously ten times then, what is the likelihood that a minimum of 3 can be [< Answer >](#) simultaneously observed in both the dice, on at least nine throws?

- (a) 0.00376
- (b) 0.00030
- (c) 0.00406
- (d) 0.99594
- (e) 0.44444.

(2 marks)

2. The academic council of Ohio University has decided to bring some changes to its existing grading system. The [< Answer >](#) university wants to determine what proportion of the students is in favor of new grading system. The total number of students enrolled in the university is 60,000 and the university feels that at least 36,000 of them support the changes. What should be the sample size that will enable the university to be 89.9 percent certain of estimating the true proportion in the favor of new system within plus and minus 0.02?

- (a) 1,075
- (b) 1,411
- (c) 1,613
- (d) 1,680
- (e) 1,711.

(3 marks)

3. Which of the following is **false** with regard to a Hypergeometric distribution? [< Answer >](#)

- (a) The trials are not independent
- (b) The probability of success is variable
- (c) The outcomes can be labeled as success or failure
- (d) The composition of the population remains unchanged
- (e) The population is finite.

(1 mark)

4. There are 25 students in a class, which consists of 14 boys and 11 girls. 5 students of the class were absent on a particular day. What is the probability that maximum two girls were absent? [< Answer >](#)

- (a) 0.2072
- (b) 0.3768
- (c) 0.416
- (d) 0.584
- (e) 0.6232.

(2 marks)

5. A random variable Z , has the following probability distribution:

[< Answer >](#)

Z	40	60	80	100	120
Probability	0.15	0.25	0.30	0.20	0.10

If twenty values of Z are observed then how many observations are expected to be less than 80?

- (a) 2
- (b) 4
- (c) 6
- (d) 8
- (e) 10.

(1 mark)

6. If the sampling distribution of a sample statistic has a mean equal to the population parameter the statistic is intended to estimate, the statistic is said to be [< Answer >](#)

- (a) An unbiased estimator of the parameter
- (b) Biased estimator of the parameter
- (c) A point estimator of the parameter
- (d) Having a low standard error
- (e) Having a high standard error.

(1 mark)

7. The probability distribution for a discrete random variable which is used to compute the probability of x successes in n independent trials, such that the probability of success remains fixed in every trial, is known as the [< Answer >](#)

- (a) Normal distribution with non-zero mean
- (b) Standard normal distribution
- (c) Binomial distribution
- (d) t -distribution
- (e) Chi-square distribution.

(1 mark)

8. Which of the following is **not** required in order to exactly describe a Hypergeometric distribution? [< Answer >](#)

- (a) The population size
- (b) The number of trials
- (c) The number of elements in the population that can be labeled as success, if the number of elements that can be labeled as failure is not known
- (d) The mean of the distribution
- (e) The number of elements in the population that can be labeled as failure, if the number of elements that can be labeled as success is not known.

(1 mark)

9. A discrete uniform random variable, M , assumes the following integer values: [< Answer >](#)

1, 2, 3, 4, 5, 6, 7, 8.

The characteristic of the discrete uniform random variable is that all possible values of the variable have equal probability.

What is the expected value of the random variable M ?

- (a) 3.00
- (b) 4.50
- (c) 6.00
- (d) 0.125
- (e) 8.00.

(1 mark)

10. Variable T follows a normal distribution. The standardized value for $T = 35$ is 2 and the standardized value for $T = 40$ is 3. Which of the following are correct with regard to T? [< Answer >](#)

- (a) Mean = 5 and variance = 5
- (b) Mean = 25 and variance = 25
- (c) Mean = 5 and variance = 25
- (d) Mean = 25 and variance = 5
- (e) Mean = 2.5 and variance = 2.5.

(2 marks)

11. When decisions are made under uncertain conditions, which of the following criteria for decision making is based [< Answer >](#) on the opportunity costs of making particular decisions?

- (a) Maximin criterion
- (b) Maximax criterion
- (c) Hurwicz criterion
- (d) Regret criterion
- (e) Expected value.

(1 mark)

12. If x_1 and x_2 are two random variables independent of each other then, the variance of $x_1 + x_2$ is equal to the [< Answer >](#)

- (a) Standard deviation of x_1 plus standard deviation of x_2
- (b) Standard deviation of x_1 minus standard deviation of x_2
- (c) Variance of x_1 plus variance of x_2
- (d) Variance of x_1 minus variance of x_2
- (e) Covariance of x_1 and x_2

(1 mark)

13. The variance of the sampling distribution of mean is [< Answer >](#)

- (a) Always equal to the variance of the underlying population
- (b) Less than the variance of the underlying population if the sample size is more than one
- (c) More than the variance of the underlying population if the sample size is more than one
- (d) Equal to the standard deviation of the underlying population if the sample size is more than one
- (e) Not dependent on the sample size.

(1 mark)

14. A small sample has been taken from a population and the sample mean has been found to be 62. The upper limit of a [< Answer >](#) 95 percent confidence interval for population mean is 81.60. The population variance is known to be 2,400. What is the sample size?

- (a) 30
- (b) 24
- (c) 36
- (d) 64
- (e) 16.

(2 marks)

15. The mean and variance of a simple random sample are found to be 12 and 9.6 respectively. The sample size is 6 and [< Answer >](#) it is less than 5 percent of the population size. It is known that the population is normally distributed. What is the 95 percent confidence interval for the population mean, on the basis of the given information?

- (a) (9.432, 14.571)
- (b) (8.748, 15.252)
- (c) (8.905, 15.095)
- (d) (2.400, 21.600)
- (e) (8.252, 15.748).

(2 marks)

16. A wholesale distributor of fertilizer products finds that the annual demand for one type of fertilizer is normally [< Answer >](#) distributed with a mean of 120 tonnes and standard deviation of 16 tonnes. If he orders only once a year, what quantity should be ordered to ensure that there is only a 5 percent chance of running short?

- (a) 110
- (b) 123
- (c) 135
- (d) 147
- (e) 159.

(2 marks)

17. The following details are available with regard to a hypothesis test on difference between means of two populations: [< Answer >](#)

$$\begin{array}{ll} n_1 = 16 & s_1 = 16 \\ n_2 = 9 & s_2 = 12 \end{array}$$

The samples are independently collected and it is assumed that the two populations have the same variance. What is the estimated standard error of difference between means?

- (a) 37.68
- (b) 6.138
- (c) 13.565
- (d) 3.683
- (e) 4.911.

(1 mark)

18. The following details are available with regard to a hypothesis test on means of two populations: [< Answer >](#)

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

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

$$n_1 = 64 \quad \sigma_1^2 = 256 \quad \sum x_1 = 256$$

$$n_2 = 36 \quad \sigma_2^2 = 144 \quad \sum x_2 = 360$$

The samples collected from the two populations are independent. If the significance level is 0.05 then

- (a) The critical value is -1.96 and the null hypothesis is accepted
- (b) The critical value is 1.96 and the null hypothesis is accepted
- (c) The critical values are ± 1.96 and the null hypothesis is rejected
- (d) The critical values are ± 1.96 and the null hypothesis is accepted
- (e) The critical values are ± 1.645 and the null hypothesis is accepted.

(2 marks)

19. A researcher is interested in determining the average urban household income of the state Uttar Pradesh. He divides the state 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 [< Answer >](#)

- (a) Systematic sampling
- (b) Cluster sampling
- (c) Simple random sampling
- (d) Stratified random sampling
- (e) Judgmental sampling.

(1 mark)

20. The sample mean is an unbiased estimator of population mean because the [< Answer >](#)

- (a) Expected value of sample mean is equal to the population mean
- (b) Expected value of sample mean is less than the population mean
- (c) Expected value of sample mean is more than the population mean
- (d) Sample mean is always equal to the population mean
- (e) Sample mean is always exceeds the population mean by a constant value.

(1 mark)

21. The following details are available with regard to a test of hypothesis for the population mean. [< Answer >](#)

Finite population correction factor = 0.952

Standard error of mean = 8 (approx)

$$\sigma^2 = 4519.45$$

What is the sample size? (Round off your answer to the nearest integer).

- (a) 8
- (b) 16
- (c) 64
- (d) 71
- (e) 32.

(1 mark)

22. A local newspaper claims that 15% of the residents of the community play the state lottery. However it is felt that the percentage is less than 15%. If you plan to test the claim by taking a random sample from the community, the appropriate set of hypothesis will be [< Answer >](#)

- (a) $H_0 : p = 0.15$ and $H_1 : p = 0.15$
- (b) $H_0 : p = 0.15$ and $H_1 : p < 0.15$
- (c) $H_0 : p = 0.15$ and $H_1 : p > 0.15$
- (d) $H_0 : p < 0.15$ and $H_1 : p = 0.15$
- (e) $H_0 : p < 0.15$ and $H_1 : p > 0.15$.

(1 mark)

23. Which of the following **can** be used for knowing the test statistic in hypothesis tests? [< Answer >](#)

- I. The sample mean can be used when the test involves the population mean.
- II. The difference between two sample means can be used when the test involves the difference between two population means.
- III. The sample proportion can be used when the test is about the population proportion.

- (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)

24. The following details are available with regard to a hypothesis test on difference between means of two populations: [< Answer >](#)

$$n_2 = 12$$

The samples are independently collected and it is assumed that the two populations are normally distributed and have the same variance. The pooled estimate of the underlying common population variance is 180.535 and the estimated standard error of difference between means is 4.365.

What is the size of the first sample?

- (a) 24
- (b) 16
- (c) 32
- (d) 64
- (e) 45.

(1 mark)

25. The null hypothesis regarding a chi-square test of independence for a contingency table with 6 rows and 4 columns [< Answer >](#) should be rejected (at 95% confidence) if the test statistic is greater than

- (a) 1.96
- (b) 1.64
- (c) 24.996
- (d) 36.415
- (e) 1.753.

(1 mark)

26. The estimated regression equation of the variable Y on X is: $\hat{Y} = 10 + 5X$. [< Answer >](#)

$$\text{Let } Z = Y + 2.$$

What is the intercept of the estimated regression equation of Z on X (X is the independent variable)?

- (a) 10
- (b) 8
- (c) 12
- (d) 20
- (e) 5.

(2 marks)

27. From the following details: [< Answer >](#)

Coefficient of correlation between variables X and Y = 0.75

$$\sigma_x^2 = 256$$

$$\sigma_y^2 = 576$$

What will be the slope of a regression relationship between X and Y with X as the independent variable?

- (a) 1.330
- (b) 1.125
- (c) 2.083
- (d) 0.500
- (e) 1.254.

(1 mark)

28. 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)

29. For the regression relationship the ratio of regression sum of squares to error sum of squares is 5 : 2. The total sum [of squares is 730](#). What is the regression sum of squares?

- (a) 452.0
- (b) 386.2
- (c) 544.3
- (d) 480.7
- (e) 521.2.

(2 marks)

30. On a population survey, we came to know that, as population increases their needs also increases and vice versa. But [the study on population cannot explain 45 percent of the variations in needs](#). What is the coefficient of correlation between population and needs?

- (a) 0.7416
- (b) 0.5250
- (c) 0.4500
- (d) - 0.4500
- (e) - 0.5500.

(1 mark)

31. Genesis Systems Ltd. produces different types of products in the field of information technology. The research and [development department of the company has been witnessing a growth in the research activities over the past few years](#). The data on research expenditure in the recent past are given below:

Year	2000	2001	2002	2003	2004	2005
Research expenditure (Rs. in lakh)	40	47	54	62	69	77

Estimate the research expenditure (in Rs. lakh) for the year 2006.

- (a) 76.37
- (b) 78.57
- (c) 80.77
- (d) 82.97
- (e) 84.07.

(2 marks)

32. An economist believes that the rent charged by the owners of apartments is determined by the number of rooms and the distance from the city center. The following data pertains to a randomly collected sample of tenants living in apartments situated at various parts of the city: [< Answer >](#)

Rent (Rs.)	Number of rooms	Distance from city center (in kms)
3300	2	2
7500	5	4
3750	3	2
4200	4	3
3300	4	10
2750	3	6

If the Y denotes rent in rupees, X_1 denotes the number of rooms and X_2 denotes the distance from the city centre in kilometers, then obtain a multiple regression relationship of the form

$$Y = a + b_1 X_1 + b_2 X_2.$$

What is the value of the coefficient b_1 ?

- (a) 1314.62
- (b) 1415.73
- (c) 1526.84
- (d) 1637.95
- (e) 1748.06.

(3 marks)

33. The manager of Pioneer Enterprises wants to formulate a relationship between total cost and sales revenue, which [< Answer >](#) 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 analysis.

What is the estimate of total cost when sales revenues are Rs.35 lakh, according to the regression relationship?

- (a) Rs.28.908 lakhs
- (b) Rs.30.897 lakhs
- (c) Rs.32.786 lakhs
- (d) Rs.34.675 lakhs
- (e) Rs.36.564 lakhs.

(2 marks)

34. Trinity Computers Ltd. manufactures and sells personal computers and accessories for office automation and the home users. The company is planning to expand its manufacturing facility if the sales of its computers increase beyond 150,000 units this year. It has compiled the sales figures of its computers across the country for the last five years. It feels that with the increase in the income level of the consumers and the advanced features coming in the personal computers, which enable it to be used not only as an educational tool but also for entertainment at home, will boost its sales in the household segment in the coming years. [< Answer >](#)

Year	2001	2002	2003	2004	2005
Number of PCs sold (in thousands)	67	71	76	82	96

Fit a curvilinear estimating equation of type $\hat{Y} = a + bx + cx^2$, where $x = X - \bar{X}$ (X represents year). What is the value of the coefficient **b** in the estimating equation?

- (a) 4.7
- (b) 5.8
- (c) 6.9
- (d) 7.5
- (e) 8.1.

(2 marks)

35. A student of statistics wants to find out the coefficient of correlation between two variables **X** and **Y**. He has obtained the following data for the deviations from assumed means of the **X** series and **Y** series. There are 8 paired data of **X** and **Y** [< Answer >](#)

d_x	9	20	30	-9	-10	10	-1	-8
d_y	13	25	44	0	-5	24	11	-4

Find the coefficient of correlation between **X** and **Y**.

- (a) 0.94
- (b) 0.95
- (c) 0.96
- (d) 0.97
- (e) 0.98.

(2 marks)

36. The following details pertain to a correlation analysis of variables **X** and **Y**: [< Answer >](#)
 $r = 0.80$, $\sum xy = 60$, $\sigma_y = 2.5$ and $\sum x^2 = 90$.

If x and y denote the deviations from the arithmetic means of the data in **X** series and **Y** series respectively, then find the number of items in each series. It is assumed that the correlation analysis is performed among the entire population of the values of **X** and **Y** respectively.

- (a) 8
- (b) 9
- (c) 10
- (d) 11
- (e) 12.

(1 mark)

37. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible: [< Answer >](#)

Variance of $X = 9$

Regression equation of:

Y on X (X is the independent variable): $8X - 10Y + 66 = 0$

X on Y (Y is the independent variable): $40X - 18Y = 214$

Find the correlation between X and Y .

- (a) 0.4
- (b) 0.5
- (c) 0.6
- (d) 0.7
- (e) 0.8.

(2 marks)

38. From the records, it is relatively easy to determine the amount of liquor consumed per capital and the number of cigarettes consumed per capital for each of the 10 provinces of Canada. These are plotted on a scatter diagram and a high positive correlation is found. Which of the following is **correct**? [< Answer >](#)

- (a) This implies that heavy smoking causes people to drink more
- (b) This implies that heavy drinking causes people to smoke more
- (c) We cannot conclude cause and effect, but this also implies that there is a high positive correlation between cigarette smoking and alcohol consumption for individuals
- (d) This could be an example of a correlation caused by a common cause because both activities are highly correlated with average family income and average income varies widely among the provinces
- (e) We cannot conclude cause and effect, but this also implies that some individuals smoke and consume liquor together.

(1 mark)

39. The intercept on Y -axis of a simple regression line is 5. If the independent variable has a value of 5 and the estimated value of the dependent variable is 20 then the slope of the regression equation is [< Answer >](#)

- (a) 3
- (b) 5
- (c) 20
- (d) 27
- (e) 28.

(1 mark)

40. Which of the following measures is used to find the correlation between the two variables when both the variables are ordinal in nature? [< Answer >](#)

- (a) Covariance
- (b) Variance
- (c) Pearson's coefficient of correlation
- (d) Coefficient of determination
- (e) Rank correlation.

(1 mark)

41. Let for a basket of goods the Laspeyres price index for the year 2003 was 235. The aggregates value of the same basket of goods using base year quantities in the year 2003 was Rs.7,250. The aggregate value of the same basket of goods using base year quantities in the year 2005 was Rs.8300. What was the Laspeyres price index for the year 2005?

[< Answer >](#)

- (a) 205.0
- (b) 269.0
- (c) 235.0
- (d) 290.0
- (e) 280.0.

(1 mark)

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

[< Answer >](#)

Commodity	Year			
	1998		2005	
	Price (Rs./Kg.)	Quantity (Kg.)	Price (Rs./Kg.)	Quantity (Kg.)
Wheat	7.50	5000	10.25	4800
Rice	10.00	6000	14.00	5500
Pulses	25.00	3000	28.00	3500
Sugar	10.60	3000	14.50	2400
Salt	2.60	500	4.50	600

What is Fishers ideal price index for the year 2005 using 1998 as the base year?

- (a) 130.36
- (b) 129.24
- (c) 128.12
- (d) 127.48
- (e) 126.72.

(2 marks)

43. The prices and quantity of goods consumed by a hostel in the years 2002 and 2005 are as follows:

[< Answer >](#)

Goods	Year 2002		Year 2005	
	Price (Rs. per kg.)	Quantity (Kgs.)	Price (Rs. per kg.)	Quantity (Kgs.)
Pulses	15	30	25	26
Oil	45	40	65	30
Rice	60	25	65	25
Wheat	8	50	12	45
Maize	18	30	25	28

Calculate weighted average of relatives price index for the year 2005 with the year 2002 as the base year and the year 2005 for weighting.

- (a) 136.18
- (b) 137.29
- (c) 138.31
- (d) 139.42
- (e) 140.53.

(1 mark)

44. Which of the following is **true** with regard to index numbers?

[< Answer >](#)

- (a) The fixed weight aggregates price index does not allow the flexibility to select the base period for comparison of prices
- (b) Paasches price index tends to underestimate the rise in prices
- (c) Unweighted aggregates price index numbers link the price changes to the usage or consumption levels
- (d) Laspeyers price index has a downward bias
- (e) The Laspeyers and Paasches methods tend to produce similar extremes in index values of the same data.

(1 mark)

45. A fruit merchant in the fruit market of the city is worried about the rising prices of the fruits and its effect on the quantity consumed by the consumers. He wants to understand this effect by finding a suitable index. The following table produces the prices and quantities of some fruits in the year 2003 and 2005:

[< Answer >](#)

Fruits	2003		2005	
	Price (Rs/kg)	Quantity ('000 kgs)	Price (Rs/kg)	Quantity ('000 kgs)
Mangoes	16	32	26	30
Grapes	46	42	66	32
Guava	62	26	68	26
Dates	9	52	14	46
Cherry	20	32	26	30

Find out Paasche's ideal price index number for the year 2005, taking 2003 as the base year.

- (a) 130.98
- (b) 132.90
- (c) 134.70
- (d) 136.61
- (e) 138.52.

(1 mark)

46. Which of the following is **true** with regard to Marshall-Edgeworth price index?

[< Answer >](#)

- (a) It is the arithmetic mean of Laspeyers and Paasches price indices
- (b) It is the geometric mean of Laspeyers and Paasches price indices
- (c) It is the product of Laspeyers and Paasches price indices
- (d) It can be calculated by using the ratio of base year quantities to current year quantities as weights
- (e) It can be calculated by using the sum of base year and current year quantities as weights.

(1 mark)

47. A composite price index where the prices of the items in the composite are weighted by their relative importance is known as the

[< Answer >](#)

- (a) Price relative
- (b) Unweighted aggregates price index
- (c) Weighted aggregates price index
- (d) Value index
- (e) Quantity index.

(1 mark)

48. Which of the following tests is/are **not** used for testing the consistency of index numbers?

[< Answer >](#)

- I. Chi-square test.
- II. Time reversal test.
- III. Circular test.
- IV. Factor reversal test.

- (a) Only (I) above
- (b) Only (II) above
- (c) Only (III) above
- (d) Both (II) and (IV) above
- (e) Both (III) and (IV) above.

(1 mark)

49. There are four commodities in a basket. The average price of the commodities in the basket is Rs.40 in the current year and their average price is Rs.25 in the base year.

[< Answer >](#)

What is the unweighted aggregates price index for the current year?

- (a) 162.5
- (b) 260
- (c) 362.5
- (d) 160
- (e) 200.

(1 mark)

50. One of the leading liquor companies holds the information regarding the time series analysis of an actual profit of the year 2004 as Rs.256.75 lakh and relative cyclical residual measure as – 3.56 %. Then what can be the estimated profit of the company in the year 2005?

[< Answer >](#)

- (a) Rs.275.69 lakh
- (b) Rs.247.60 lakh
- (c) Rs.266.22 lakh
- (d) Rs.265.89 lakh
- (e) Rs.914.03 lakh.

(1 mark)

51. Which of the following are common reasons for studying both secular trends and seasonal variations?

[< Answer >](#)

- I. To allow the elimination of the seasonal component from the series.
- II. To describe past patterns.
- III. To project past patterns into the future.

- (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)

52. Which of the following is/are **true** regarding the relationship between percent of trend measure and relative cyclical residual measure? [< Answer >](#)

- I. When the value under percent of trend is 100, the value under relative cyclical residual measure is 0.
 - II. When the value under percent of trend is more than 100, the value under relative cyclical residual measure is positive.
 - III. When the value under percent of trend is less than 100, the value under relative cyclical residual measure is negative.
- (a) Only (I) above
 - (b) Only (II) above
 - (c) Both (I) and (II) above
 - (d) Both (II) and (III) above
 - (e) All (I), (II) and (III) above.

(1 mark)

53. Which of the following statement(s) is/are **true**? [< Answer >](#)

- (a) The secular trend allows us to describe the historical pattern of the data
- (b) Secular trend represents the short to medium term variation in time series data
- (c) The component of a time series that oscillates predictably above and below the trend line for a period less than one year is known as irregular variation
- (d) Seasonal variations are periodic and repetitive in nature
- (e) Both (a) and (d) above.

(1 mark)

54. Which of the following represents the proportion of variation in the dependent variable that is explained by the regression line? [< Answer >](#)

- (a) Coefficient of determination
- (b) Coefficient of correlation
- (c) Coefficient of variation
- (d) Standard error of estimate
- (e) Standard deviation.

(1 mark)

55. The sale of Mehta Industries Ltd. in the year 2005 is Rs.7.50 crores. The estimated sale for the same year as per the secular trend equation is Rs.6.25 crores. The relative cyclical residual for the year 2005 is [< Answer >](#)

- (a) 15.0
- (b) 20.0
- (c) 45.0
- (d) 83.3
- (e) 120.0.

(1 mark)

56. Which of the following can be identified using the percent of trend measure? [< Answer >](#)

- (a) Secular trend
- (b) Cyclical variation
- (c) Seasonal variation
- (d) Irregular variation
- (e) Coefficient of variation.

(1 mark)

[< Answer >](#)

57. If the pattern of $\bar{X}$ chart of a process is 'Hugging the center line', then it indicates that the

- (a) Process mean may be drifting
- (b) Process mean may have shifted
- (c) Two distinct populations are being observed
- (d) Variation has been reduced
- (e) Few individual outliers are present.

(1 mark)

[< Answer >](#)

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

$$\hat{Y} = 139 + 7.5x$$

where $x = \text{Year} - 2003$

$\hat{Y}$ is the estimated value of the dependent variable Y .

The observed value of Y for the year 2005 is 160.

What is the relative cyclical residual for the year 2005?

- (a) -3.9
- (b) 10.8
- (c) 3.9
- (d) 96.25
- (e) 103.9.

(1 mark)

[< Answer >](#)

59. The movement of the dependent variable above and below the secular trend line over periods longer than one year is known as

- (a) Seasonal variation
- (b) Secular variation
- (c) Irregular variation
- (d) Cyclical variation
- (e) Regular variation.

(1 mark)

[< Answer >](#)

60. A new shearing machine is set to cut off a piece of steel from a long bar. For various reasons the machine at some times cuts off a piece that is too long or too short. These unacceptable pieces are automatically dropped in a box and the operator of the shearing machine must count these defectives after every 100 pieces sheared off. The record after the first day of operation is:

Number of sheared off pieces	100	100	100	100	100
Number of defectives	5	6	7	4	8

Set out the lower control limit for the p chart.

- (a) 0.012
- (b) 0.011
- (c) 0.051
- (d) 0.000
- (e) -0.012.

(1 mark)

61. A daily sample of 30 items was taken over a period of 14 days in order to establish control limits. If 21 defectives were found, what should be the upper control limit of the proportion of defectives? [< Answer >](#)

- (a) 1.1720
- (b) 0.2110
- (c) 1.1010
- (d) 0.1500
- (e) 0.1694.

(2 marks)

62. A hole formed by a drilling machine has the mean diameter of 0.5230 cm and the standard deviation of 0.0032 cm. [< Answer >](#)
Calculate the lower control limit for means with samples of size 4 each.

- (a) 0.5174 cm
- (b) 0.5176 cm
- (c) 0.5178 cm
- (d) 0.5180 cm
- (e) 0.5182 cm.

(1 mark)

63. The following data provides the values of sample range **R**, for samples of size 5 each. [< Answer >](#)

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

Calculate the value for the center line of the **R** chart.

- (a) 6.1
- (b) 6.2
- (c) 6.3
- (d) 6.5
- (e) 6.6.

(1 mark)

64. Which of the following patterns of $\bar{X}$ chart is generally desirable to any quality control manager? [< Answer >](#)

- (a) Increasing trend
- (b) Decreasing trend
- (c) Cycles
- (d) Hugging the center line
- (e) Hugging the control limits.

(1 mark)

65. Which of the following is a control chart plotted on the basis of qualitative factors possessed by the output? [< Answer >](#)

- (a) R chart
- (b) p chart
- (c) $\bar{X}$ chart
- (d) Scatter diagram
- (e) Histogram.

(1 mark)

[< Answer >](#)

66. The following data pertain a $\bar{X}$ chart:
- | | |
|---------------------------------------|---------|
| Grand mean | = 25.0 |
| Average of sample ranges | = 7.5 |
| Number of observations in each sample | = 9.0 |
| Control chart factor, d_2 | = 2.534 |
- The upper control limit is approximately

- (a) 22.04
- (b) 24.01
- (c) 25.00
- (d) 26.00
- (e) 27.96.

(1 mark)

[< Answer >](#)

67. Which of the following is **true** in the context of quality control?

- (a) Variations in quality of output occur only due to the existence of random or inherent variations in the process
- (b) Variations in quality of output occur only due to the existence of non-random or special cause variations in the process
- (c) All non-random variations must be identified and brought under control before redesigning the process to reduce the random or inherent variations
- (d) The process must be redesigned to eliminate all random variations before identifying the non-random variations and eliminating the same
- (e) If the process is out of control, then random variation will be observed.

(1 mark)

[< Answer >](#)

68. When we carry out a chi-square test of independence which of the following can be stated in the null hypothesis?

- (a) The population mean is equal to a specified value
- (b) The population proportion is equal to a specified value
- (c) The population distribution follows a specified probability distribution
- (d) The population can be classified on the basis of two criteria which are not associated
- (e) The population can be classified on the basis of two criteria which are associated.

(1 mark)

[< Answer >](#)

69. Suppose we are conducting a chi-square test of goodness of fit to determine whether a set of observations could have come from a normal population. The population mean and standard deviation are known and specified. We select a sample of 100 observations and organize them into a frequency distribution with 6 categories. How many degrees of freedom are there?

- (a) 5
- (b) 9
- (c) 3
- (d) 8
- (e) 10.

(1 mark)

70. Which of the following is **false** regarding the chi-square test?

[< Answer >](#)

- (a) It is used for a test of significance for difference between means of two or more populations
- (b) It is used to test the deviations between observed and theoretical frequencies
- (c) It is used to test whether there is any association between two attributes or not
- (d) It is used to test the goodness of fit
- (e) It is used to test the population variance.

(1 mark)

71. Sintel Corporation, a manufacturer of silicon chips, requires so-called "clean rooms", where the air is specially filtered to keep the number of dust particles at a minimum. The company wants to make sure that each of its five clean rooms has the same average number of dust particles. Five observations on air quality measured in terms of dust scores have been taken from each room for performing an ANOVA. From this data, the between-column variance has been found to be 18.48 and within-column variance has been found to be 14.65.

The significance level for the test is 0.05

Which of the following statements is **false**?

- (a) The probability distribution applicable for the test is the F-distribution
- (b) The value of the test statistic is 1.26
- (c) The critical value for the test is 2.87
- (d) The critical value for the test is 2.60
- (e) It can be concluded at a 0.05 significance level that the rooms do not have significantly different dust particles.

(2 marks)

72. In performing a chi-square test of goodness of fit, as the difference between the respective observed and expected frequencies decreases, the likelihood of concluding that the population follows the assumed probability distribution

[< Answer >](#)

- (a) Decreases
- (b) Increases
- (c) Increases or decreases depending upon the assumed probability distribution
- (d) Remains the same
- (e) Tends to become equal to the likelihood of rejecting the assumption.

(1 mark)

73. Which of the following is **false** regarding a chi-square distribution?

[< Answer >](#)

- (a) Chi-square distribution has only one parameter
- (b) The distribution becomes sharply positively skewed when the number of degrees of freedom is large
- (c) For large number of degrees of freedom the distribution approaches symmetry
- (d) The distribution is continuous
- (e) The distribution is unimodal.

(1 mark)

74. The supervisor of security at a large department store would like to know if the store apprehends different numbers of shoplifters during the Christmas holiday season than in the weeks before or after the holiday. He gathered data on the number of shoplifters apprehended in the store during the months of November, December and January over the past 6 years. The information is as below: [< Answer >](#)

November	43	38	59	55	39	48
December	54	40	48	35	50	49
January	35	27	34	40	30	32

A test of ANOVA has to be performed

Find the estimate of the population variance based on the variance within the samples.

- (a) 47
- (b) 48
- (c) 49
- (d) 50
- (e) 51.

(1 mark)

75. Which of the following is **not** a step used in Analysis of variance (ANOVA)? [< Answer >](#)

- (a) We determine an estimate of the population variance from the variance that exists among the sample means
- (b) We determine an estimate of the population variance from the variance that exists within the samples
- (c) We prepare a table of observed and estimated frequencies
- (d) We accept the null hypothesis if the two estimates of the population variance are not significantly different
- (e) We reject the null hypothesis if the F ratio is too high and falls in the critical region.

(1 mark)

76. 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: [< Answer >](#)

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.

(2 marks)

77. Which of the following statements is **false** regarding **F**-distribution?

[< Answer >](#)

- (a) Similar to chi-square distribution, it is a family of distributions
- (b) F-distribution is bimodal in nature
- (c) As the number of degrees of freedom varies, so does the shape of distribution
- (d) It has two numbers of degrees of freedom
- (e) As the number of degrees of freedom increases, the distribution tends to become symmetrical.

(1 mark)

78. Which of the following is **true** with regard to the **p** chart in quality control?

[< Answer >](#)

- (a) The center line is drawn on the basis of overall sample mean
- (b) The lower control limit can only be greater than or equal to zero
- (c) The upper control limit is equal to the overall sample proportion
- (d) The lower control limit is equal to the overall sample proportion
- (e) The upper and the lower control limits are based on the overall sample mean.

(1 mark)

Suggested Answers

Quantitative Methods-II (132): October 2006

1. Answer : (c)
Reason :

[< TOP >](#)

Let success be defined as obtaining at least three from both the dice simultaneously.

$$\text{On any die } P(\text{At least 3}) = P(3 \text{ or } 4 \text{ or } 5 \text{ or } 6) = \frac{4}{6} = \frac{2}{3}$$

$$\therefore P(\text{Success}) = P(\text{At least 3 on both the dice}) = \left(\frac{2}{3}\right)^2 = \frac{4}{9}$$

Let X denote the number of successes.

$$\therefore X \text{ follows a binomial distribution with } n = 10 \text{ and } p = \frac{4}{9}$$

$$\begin{aligned} P(X \geq 9) &= P(X = 9) + P(X = 10) \\ &= {}^{10}C_9 \left(\frac{4}{9}\right)^9 \left(1 - \frac{4}{9}\right)^{10-9} + {}^{10}C_{10} \left(\frac{4}{9}\right)^{10} \left(1 - \frac{4}{9}\right)^{10-10} \\ &= 0.00376 + 0.00030 \\ &= 0.00406 \end{aligned}$$

2. Answer : (c)

[< TOP >](#)

Reason : The z value in the normal table for 89.9 percent confidence level is 1.64. We want our estimate to be within ± 0.02 , so,

$$Z \sigma_{\bar{p}} = 0.02$$

$$\text{And } z = 1.64$$

$$\text{Then } 1.64 \sigma_{\bar{p}} = 0.02$$

Substituting the value of $\sigma_{\bar{p}}$ in the equation,

$$1.64 \sqrt{\frac{pq}{n}} = 0.02$$

$$\text{or, } \sqrt{\frac{pq}{n}} = 0.0122$$

As the university feels that the proportion of students favoring the new system is $\frac{36000}{60000} = 0.60$. We can take this value as the estimate of the population parameter.

$$\text{Therefore, } \sqrt{\frac{0.6 \times 0.4}{n}} = 0.0122$$

$$\text{or, } n = \frac{0.6 \times 0.4}{(0.0122)^2} = 1612.4697$$

Therefore a sample size of **1613** would be appropriate to estimate the proportion of the students favouring the new system in an interval of ± 0.02 at a confidence level of 89.9%.

3. Answer : (d)

[< TOP >](#)

Reason : In a hypergeometric distribution the composition of the population changes from trial to trial.

4. Answer : (d)

[<TOP>](#)

Reason : The number of absentees who are girls follows a hypergeometric distribution. The following details are available:

N = No. of elements in the population = 25

r = No. of elements in the population labelled success = No. of girls = 11

n = No. of trials = No. of absent students = 5

x = Desired number of successes

$$\begin{aligned} \text{In a hypergeometric distribution } P(x) &= \frac{{}^r C_x {}^{(N-r)} C_{(n-x)}}{{}^N C_n} \\ P(x \leq 2) &= P(x = 1) + P(x = 2) = \frac{{}^{11} C_1 {}^{(25-11)} C_{(5-1)}}{{}^{25} C_5} + \frac{{}^{11} C_2 {}^{(25-11)} C_{(5-2)}}{{}^{25} C_5} \\ &= \frac{11 \times {}^{14} C_4}{{}^{25} C_5} + \frac{55 \times {}^{14} C_3}{{}^{25} C_5} \\ &= 0.2072 + 0.3768 \\ &= 0.584 \end{aligned}$$

5. Answer : (d)

[<TOP>](#)

Reason : Our desired event is $Z < 80$. $P(Z < 80) = 0.15 + 0.25 = 0.40$

Let X be the number of times Z assumes values less than 80. Then X is binomially distributed with number of trials = 20 and probability of success = 0.40

The expected number of observations less than 80 = $E(Z) = n.p = 20 \times 0.40 = 8$

6. Answer : (a)

[<TOP>](#)

Reason : If the sampling distribution of a sample statistic has a mean equal to the population parameter the statistic is intended to estimate, the statistic is said to be an unbiased estimator of the parameter.

7. Answer : (c)

[<TOP>](#)

Reason : The probability distribution for a discrete random variable which is used to compute the probability of x successes in n independent trials is known as the binomial probability distribution.

8. Answer : (d)

[<TOP>](#)

Reason : The mean of the distribution need not be known in order to exactly describe a hypergeometric Distribution.

9. Answer : (b)

[<TOP>](#)

Reason : The random variable M is discrete uniform. So each of its 8 possible values has equal probability =

$$\frac{1}{8} \text{ i.e., } \frac{1}{8}$$

$$E(M) = \frac{1}{8} \{1 + 2 + 3 + \dots + 8\}$$

$$E(M) = \frac{k+1}{2} \text{ (Where } k = 8 \text{ by the formula for expected value)}$$

$$= \frac{8+1}{2} = 4.5$$

10. Answer : (b)

[<TOP>](#)

Reason :

For T = 35 standardised value = 2.

$$2 = \frac{35 - \mu}{\sigma} \text{ or } 2\sigma = 35 - \mu \dots \dots (A)$$

For T = 40 standardised value = 3.

$$3 = \frac{40 - \mu}{\sigma} \text{ or } 3\sigma = 40 - \mu \dots \dots (B)$$

Subtracting equation (A) from (B) we get $\sigma = 5$

Putting $\sigma = 5$ in (A) we get $\mu = 35 - 2\sigma = 35 - (2 \times 5)$
or $\mu = 25$

Mean = 25

Variance = $\sigma^2 = 5 \times 5 = 25$

11. Answer : (d)

[<TOP>](#)

Reason : In decision making under uncertainty the regret criterion is based on opportunity costs of decisions.

12. Answer : (c)

[<TOP>](#)

Reason : $\text{Var}(x_1 + x_2) = \text{Var}(x_1) + \text{Var}(x_2) + 2\text{Cov}(x_1, x_2)$

If x_1 and x_2 are independent then, $\text{Cov}(x_1, x_2) = 0$

$\therefore \text{Var}(x_1 + x_2) = \text{Var}(x_1) + \text{Var}(x_2)$.

13. Answer : (b)

[<TOP>](#)

Reason : The variance of the sampling distribution of mean is less than the variance of the underlying population if the sample size is more than one.

14. Answer : (b)

[<TOP>](#)

Reason : Since the population variance is known, the distribution to be used is the $N(0, 1)$ distribution not with standing that the sample size is small.

$$\therefore \text{UCL} = \bar{x} + 1.96\sigma_{\bar{x}}$$

$$\text{or } 1.96\sigma_{\bar{x}} = \text{UCL} - \bar{x} = 81.60 - 62 = 19.60$$

$$\text{or } \sigma_{\bar{x}} = \frac{19.60}{1.96} = 10$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{ or } n = \left(\frac{\sigma}{\sigma_{\bar{x}}} \right)^2 = \frac{\sigma^2}{\sigma_{\bar{x}}^2} = \frac{2,400}{100} = 24$$

Since the population variance (σ^2) is known the standard error of mean will be based on σ .

15. Answer : (b)

[<TOP>](#)

Reason : The sample size is small; so the appropriate distribution to apply for the confidence interval is t distribution with $6 - 1 = 5$ degrees of freedom.

Sample standard deviation, $s = \sqrt{9.6}$

Estimated standard error of mean $= \frac{s}{\sqrt{n}} = \frac{\sqrt{9.6}}{\sqrt{6}} = 1.265$

$\therefore$ 95% confidence interval for population mean =

$$\begin{aligned}\bar{X} \pm 2.571 \frac{s}{\sqrt{n}} &= 12 \pm 2.571 \times 1.265 \\ &= (8.748, 15.252)\end{aligned}$$

16. Answer : (d)

Reason : Let the annual demand (in tonnes) be denoted by the random variable X.

$$\text{Therefore } z = \frac{x-120}{16}$$

The area in the right tail of the normal curve is 0.05. Since the area between the mean and the given value of X is 0.45, therefore, from the table we get this area of 0.45 corresponding to $z = 1.645$ (with interpolation).

Substituting this value of $z = 1.645$ in standard normal variate, we get

$$1.645 = \frac{x-120}{16} \text{ or } X = 146.32.$$

Therefore the wholesale distributor should order 147 tonnes (Note that the value has been rounded off to the higher value because otherwise probability of shortfall will be more than 5%).

17. Answer : (b)

Reason : Pooled estimate of the underlying common population variance, $s_p =$

$$\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{(16-1)16^2 + (9-1)12^2}{16+9-2} = 217.04$$

$$\text{Estimated standard error of difference between means} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{217.04 \left(\frac{1}{16} + \frac{1}{9} \right)} = 6.138$$

18. Answer : (c)

Reason : $H_0: \mu_1 = \mu_2 + 1 \Rightarrow \mu_1 - \mu_2 = 1$

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

It is a two-tailed test at a significance level of 0.05. Both the samples are large. Hence the normal distribution should be used to approximate the sampling distribution of the difference between sample means.

Value of the test statistic =

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\left(\frac{256}{64} - \frac{360}{36} \right) - 1}{\sqrt{\frac{256}{64} + \frac{144}{36}}} = \frac{-6-1}{2.828} = -2.475$$

The critical values under the left and right tails obtained from the standard normal distribution are ± 1.96 . The observed value of the test statistic is less than the left tail critical value; so it falls in the critical region. So we reject the null hypothesis.

19. Answer : (d)

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

[<TOP>](#)

[<TOP>](#)

[<TOP>](#)

[<TOP>](#)

20. Answer : (a) [<TOP>](#)

Reason : The sample mean is an unbiased estimator of population mean because the expected value of sample mean is equal to the population mean.

21. Answer : (c) [<TOP>](#)

Reason :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \text{Finite population correction factor}$$

$$\text{or } 8 = \sqrt{\frac{4519.45}{n}} \times 0.952$$

$$\text{or } n = \left[\frac{\sqrt{4519.45} \times 0.952}{8} \right]^2 = 63.9999 \approx 64$$

22. Answer : (b) [<TOP>](#)

Reason : $H_0 : p = 0.15$ and $H_1 : p < 0.15$

23. Answer : (e) [<TOP>](#)

Reason : The sample mean is used when the test involves the population mean. The difference between two sample means is used when the test involves the difference between two population means. The sample proportion is used when the test is about the population proportion.

24. Answer : (e) [<TOP>](#)

Reason : Estimated standard error of difference between means =

$$s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{180.535 \left(\frac{1}{n_1} + \frac{1}{12} \right)} = 4.365$$

Squaring both sides:

$$180.535 \left(\frac{1}{n_1} + \frac{1}{12} \right) = 19.0532$$

$$\text{or } 15.0446 + \frac{180.535}{n_1} = 19.0532$$

$$\text{or } \frac{180.535}{n_1} = 19.0532 - 15.0446 = 4.0086$$

$$\text{or } n_1 = \frac{180.535}{4.0086} = 45.03 \approx 45$$

25. Answer : (c) [<TOP>](#)

Reason : 95% confidence level implies 5% significance level. The null hypothesis regarding chi-square test of independence for a contingency table with 6 rows and 4 columns should be rejected (at 95% confidence level i.e., 5% significance level) if the test statistic is greater than a chi-square value (for $\text{dof} = (6 - 1)(4 - 1) = 15$) = 24.996.

26. Answer : (c) [<TOP>](#)

Reason : Slope of the estimated regression equation of Z on X =

$$\begin{aligned}
\frac{n \sum XZ - \sum X \sum Z}{n \sum X^2 - (\sum X)^2} &= \frac{n \sum X(Y+2) - \sum X \sum (Y+2)}{n \sum X^2 - (\sum X)^2} \\
&= \frac{n[\sum XY + 2 \sum X] - \sum X(\sum Y + \sum 2)}{n \sum X^2 - (\sum X)^2} \\
&= \frac{n \sum XY + 2n \sum X - \sum X \sum Y - \sum X \sum 2}{n \sum X^2 - (\sum X)^2} \\
&= \frac{n \sum XY + 2n \sum X - \sum X \sum Y - 2n \sum X}{n \sum X^2 - (\sum X)^2} \\
&= \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} \\
&= \text{Slope of estimated regression equation of Y on X} \\
&= 5
\end{aligned}$$

Intercept of the estimated regression equation of Z on X =

$$\begin{aligned}
\bar{Z} - b\bar{X} &= \frac{\sum Z}{n} - b\bar{X} \\
&\quad (\text{Where 'b' is slope of estimated Regression equation of Z on X}) \\
&= \frac{\sum (Y+2)}{n} - b\bar{X} = (\bar{Y} + 2) - b\bar{X} = 2 + (\bar{Y} - b\bar{X}) \\
&= 2 + \text{Intercept of estimated Regression equation of Y on X} \\
&= 2 + 10 = 12
\end{aligned}$$

27. Answer : (b)

[<TOP>](#)

$$\text{Reason : } r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 0.75 \sqrt{256 \times 576} = 288$$

$$\text{Slope of regression equation of X and Y} = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \frac{288}{256} = 1.125$$

28. Answer : (e)

[<TOP>](#)

Reason : In multiple regression relationship:

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

Given : ESS = 576

$$n = ?$$

$$k = 2$$

$$s_e = 4.8$$

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

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

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

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

29. Answer : (e)

[<TOP>](#)

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

$$\frac{\frac{RSS}{ESS}}{\frac{ESS}{ESS} + \frac{RSS}{ESS}} = \frac{\frac{5}{2}}{1 + \frac{5}{2}} = 0.714$$

$$\therefore \frac{RSS}{TSS} = 0.714 \quad \text{or } RSS = 0.714 \times TSS = 0.714 \times 730 = 521.2.$$

30. Answer : (a)

[<TOP>](#)

Reason :

$$\begin{aligned} \text{Coefficient of correlation} &= +\sqrt{\text{Coefficient of determination}} \\ &= +\sqrt{1 - \text{Proportion of unexplained variations}} \\ &= \sqrt{1 - 0.45} \\ &= 0.7416 \end{aligned}$$

(Positive square root because population increases as needs increases and vice versa)

31. Answer : (e)

[<TOP>](#)

Reason : Let the following notations be used:

X : Year

x : Coded value of year = $2(X - \bar{X})$

Y : Research expenditure (in Rs.)

Year X	$x = 2(X - \bar{X})$	Y	xY	x^2
2000	-5	40	-200	25
2001	-3	47	-141	9
2002	-1	54	-54	1
2003	1	62	62	1
2004	3	69	207	9
2005	5	77	385	25
$\Sigma X = 12015$		$\Sigma Y = 349$	$\Sigma xY = 259$	$\Sigma x^2 = 70$

$$\bar{X} = \frac{12015}{6} = 2002.5$$

$$\hat{Y} = a + bx$$

$$b = \frac{\Sigma xY}{\Sigma x^2} = \frac{259}{70} = 3.7$$

$$a = \bar{Y} = \frac{349}{6} = 58.17$$

$\therefore$ The linear trend equation is $\hat{Y} = 58.17 + 3.7x$

$$\text{Year (X)} = 2006$$

$$x = 2 (2006 - 2002.5) = 7$$

$$Y = 58.17 + 3.7 (7) = 84.07$$

∴ The research and development expenditure for the year 2006 is Rs.84.07 lakh.

32. Answer : (c)

[<TOP>](#)

Reason : Let the following notations be used:

Y : Rent (Rs.)

X₁ : Number of rooms

X₂ : Distance (kms) from the city center

X ₁	2	5	3	4	4	3	ΣX ₁ =	21
X ₂	2	4	2	3	10	6	ΣX ₂ =	27
Y	3300	7500	3750	4200	3300	2750	ΣY =	24800
X ₁ Y	6600	37500	11250	16800	13200	8250	ΣX ₁ Y =	93600
X ₂ Y	6600	30000	7500	12600	33000	16500	ΣX ₂ Y =	106200
X ₁ ²	4	25	9	16	16	9	ΣX ₁ ² =	79
X ₂ ²	4	16	4	9	100	36	ΣX ₂ ² =	169
X ₁ X ₂	4	20	6	12	40	18	ΣX ₁ X ₂ =	100

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

$$\Sigma Y = na + b_1 \Sigma X_1 + b_2 \Sigma X_2$$

$$\Sigma X_1 Y = a \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2$$

$$\Sigma X_2 Y = a \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2$$

Putting the values we get–

$$24800 = 6a + 21b_1 + 27b_2 \quad \text{--} \quad (A)$$

$$93600 = 21a + 79b_1 + 100b_2 \quad \text{--} \quad (B)$$

$$106200 = 27a + 100b_1 + 169b_2 \quad \text{--} \quad (C)$$

Solving the above three simultaneous equations we get the values,

$$a = 96.54, b_1 = 1526.84 \text{ and } b_2 = -290.476.$$

33. Answer : (b)

[<TOP>](#)

Reason : The equation of the regression line is $\hat{Y} = a + bX$

where, X = Sales (Rs. in lakhs)

Y = Total costs (Rs. in lakhs)

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

$$\Sigma X = 194; \quad \Sigma Y = 171; \quad n = 6;$$

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

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

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

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

Sales (x) = Rs.35 lakhs (given)

$$\therefore \hat{Y} = -0.568 + (0.899)(35) = 30.897$$

$\therefore$ Total costs = Rs.**30.897 lakhs**.

34. Answer : (c)

[< TOP >](#)

Reason : For a curvilinear trend we fit the equation of type $Y = a + bX + cX^2$.

The corresponding estimating equation is $\hat{Y} = a + bx + cx^2$. Where x represents the coded time variable. The calculation is shown below:

Let X represents the Year and Y represents the number of PCs sold in thousands.

Data	X	Y	$x = X - \bar{X}$	x^2	x^4	xY	x^2Y
1	2001	67	-2	4	16	-134	268
2	2002	71	-1	1	1	-71	71
3	2003	76	0	0	0	0	0
4	2004	82	1	1	1	82	82
5	2005	96	2	4	16	192	384
Total	10015	392	0	10	34	69	805

$$\text{Mean of X, } \bar{X} = \frac{10015}{5} = 2003$$

$$\text{We have } \sum Y = an + c \sum x^2,$$

$$\sum x^2 Y = a \sum x^2 + c \sum x^4,$$

$$\text{And } b = \frac{\sum xY}{\sum x^2}$$

Substituting the values in these equations, we have

$$392 = 5a + 10c \quad \dots\dots\dots (i)$$

$$805 = 10a + 34c \quad \dots\dots\dots (ii)$$

$$b = \frac{69}{10} = \mathbf{6.9} \quad \dots\dots\dots (iii)$$

35. Answer : (d)

[< TOP >](#)

Reason : The coefficient of regression can be found out as follows:

d_x	9	20	30	-9	-10	10	-1	-8	$\sum d_x = 41$
d_y	13	25	44	0	-5	24	11	-4	$\sum d_y = 108$
d_x^2	81	400	900	81	100	100	1	64	$\sum d_x^2 = 1727$

d_y^2	169	625	1936	0	25	576	121	16	$\Sigma d_y^2 = 3468$
$d_x d_y$	117	500	1320	0	50	240	-11	32	$\Sigma d_x d_y = 2248$

The coefficient of correlation,

$$\begin{aligned}
 r &= \frac{N \Sigma d_x d_y - \Sigma d_x \Sigma d_y}{\sqrt{N \Sigma d_x^2 - (\Sigma d_x)^2} \sqrt{N \Sigma d_y^2 - (\Sigma d_y)^2}} \\
 &= \frac{8 \times 2248 - 41 \times 108}{\sqrt{8 \times 1727 - (41)^2} \sqrt{8 \times 3468 - (108)^2}} \\
 &= 0.97 \text{ (approx.)}
 \end{aligned}$$

36. Answer : (c)

[<TOP>](#)

Reason : Given that: $r = 0.8$, $\Sigma xy = 60$, $\sigma_y = 2.5$ and $\Sigma x^2 = 90$.

Now we know that

$$\begin{aligned}
 \sigma_x^2 &= \frac{\Sigma(X - \bar{X})^2}{N} = \frac{\Sigma x}{N} = \frac{90}{N} \\
 r &= \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{N \sigma_x \sigma_y} = \frac{\Sigma xy}{N \sigma_x \sigma_y}
 \end{aligned}$$

Squaring we get,

$$\begin{aligned}
 r^2 &= \frac{(\Sigma xy)^2}{N^2 \sigma_x^2 \sigma_y^2} = \frac{(\Sigma xy)^2}{N^2 \frac{\Sigma x^2}{N} \sigma_y^2} \\
 \text{or, } (0.8)^2 &= \frac{(60)^2}{N^2 \times \frac{90}{N} \times 2.5^2} = \frac{3600}{90N \times 6.25}
 \end{aligned}$$

or, $N = 10$.

37. Answer : (c)

[<TOP>](#)

Reason : The **b** coefficient of the regression line X on Y : $40X - 18Y = 214$

$$\Rightarrow X = \frac{214}{40} + \frac{18}{40} Y \quad \Rightarrow b_{xy} = \frac{18}{40}$$

The **b** coefficient of the regression line Y on X : $8X - 10Y + 66 = 0$

$$\Rightarrow Y = \frac{66}{10} + \frac{8}{10} X \quad \Rightarrow b_{yx} = \frac{8}{10}$$

Since the regression coefficients are positive the correlation is also positive (hence it will be the +ve square root of the product of the two b coefficients)

$$\text{Therefore coefficient of correlation } r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{\frac{8}{10} \times \frac{18}{40}} = \sqrt{0.36} = 0.6$$

38. Answer : (d)

[<TOP>](#)

Reason : This could be an example of a correlation caused by a common cause because both activities are highly correlated with average family income and average income varies widely among the provinces. (c) is incorrect because that is an inference with regard to individuals whereas the study covers average consumption of liquor and cigarettes.

39. Answer : (a)

[<TOP>](#)

Reason : Linear regression equation $Y = a + bX$

$$20 = 5 + 5 \times b$$

$$b = \frac{20-5}{5} = 3.$$

40. Answer : (e)

[<TOP>](#)

Reason : Rank correlation is used to find the correlation coefficient between the two variables when both the underlying variables are ordinal or when the data are available in the ordinal form irrespective of the type of variable

41. Answer : (b)

[<TOP>](#)

Reason : Laspeyres price index for 2003: $\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100 = \frac{7250}{\sum P_0 Q_0} \times 100 = 235$

$$\text{Or } \sum P_0 Q_0 = \frac{7250 \times 100}{235} = 3085.10$$

$$\therefore \text{Laspeyres price index for 2005} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

$$= \frac{8300}{3085.10} \times 100 = 269.0$$

42. Answer : (c)

[<TOP>](#)

Reason : Fisher's ideal price index = $\sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \times 100$

$$\sum P_1 Q_0 = (10.25 \times 5000) + (14.00 \times 6000) + (28.00 \times 3000) + (14.50 \times 3000) + (4.50 \times 500) = 265000$$

$$\sum P_1 Q_1 = (10.25 \times 4800) + (14.00 \times 5500) + (28.00 \times 3500) + (14.50 \times 2400) + (4.50 \times 600) = 261700$$

$$\sum P_0 Q_0 = (7.5 \times 5000) + (10.00 \times 6000) + (25.00 \times 3000) + (10.60 \times 3000) + (2.60 \times 500) = 205600$$

$$\sum P_0 Q_1 = (7.5 \times 4800) + (10.00 \times 5500) + (25.00 \times 3500) + (10.60 \times 2400) + (2.60 \times 600) = 205500$$

$$\therefore \text{Fisher's ideal price index} = \sqrt{\frac{265000}{205600} \times \frac{261700}{205500}} \times 100 = 128.12.$$

43. Answer : (a)

[<TOP>](#)

Reason : For weighted average of relatives price index we do the following calculations:

$$\sum \left[\left(\frac{P_1}{P_0} \times 100 \right) (P_1 Q_1) \right] = 744239.25$$

$$\sum P_1 Q_1 = 5465$$

$$\text{Weighted average of relative price index} = \frac{\sum \left[\left(\frac{P_1}{P_0} \times 100 \right) (P_1 Q_1) \right]}{\sum P_1 Q_1}$$

$$= \frac{744239.25}{5465} \approx 136.18$$

Therefore the value of weighted average of relatives price index is **136.18**.

44. Answer : (b)

[< TOP >](#)

Reason : The fixed weight aggregates price index allows the flexibility to select the base period for comparison of prices, because it uses the weights from a representative period.

Paasche's price index tends to underestimate the rise in prices.

Unweighted aggregates price index does not reflect the reality since the price changes to the consumption levels.

Laspeyres price index has an upward bias.

Laspeyres and Paasche methods tend to produce opposite extremes in index values of the same data.

45. Answer : (b)

[< TOP >](#)

Reason : We calculate the following to find the Paasche's Index:

Now the values of $\sum P_0 Q_1 = 4578$ and $\sum P_1 Q_1 = 6084$

$$\text{We know that Paasche's index} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100 = \frac{6084}{4578} \times 100 = 132.90.$$

46. Answer : (e)

[< TOP >](#)

Reason : The Marshall-Edgeworth price index can be calculated using the sum of base year and current year quantities as weights.

47. Answer : (c)

[< TOP >](#)

Reason : A composite price index where the prices of the items in the composite are weighted by their relative importance is known as the weighted aggregate price index.

48. Answer : (a)

[< TOP >](#)

Reason : Chi-square test is not a test of consistency for index numbers.

49. Answer : (d)

[< TOP >](#)

$$\text{Reason : Unweighted aggregates price index} = \frac{\sum P_1}{\sum P_0} \times 100 = \frac{(n\bar{P}_1)}{(n\bar{P}_0)} \times 100 = \frac{(4 \times 40)}{(4 \times 25)} \times 100 = 160.$$

50. Answer : (c)

[< TOP >](#)

Reason : Relative cyclical residual measure = -3.56 %

Actual profit of the year 2004 = Rs.256.75 lakh

Percent of trend measure = 100 + Relative cyclical residual

$$= \frac{\text{Actual value}}{\text{Estimated value}} \times 100 = 100 + (-3.56) = 96.44$$

$$\therefore \text{Estimated profit} = \text{Actual profit} \times \frac{100}{96.44} = 256.75 \times \frac{100}{96.44} = \text{Rs.266.22 lakh.}$$

51. Answer : (e)

[< TOP >](#)

Reason : All the given alternatives (a), (b) and (c) are reasons for studying both secular trends and seasonal variation. Hence, alternative (e) is the answer.

52. Answer : (e)

[< TOP >](#)

Reason : When the value as per percent of trend is 100, the value under relative cyclical residual measure is 0. When the value under percent of trend is more than 100, the value under relative cyclical residual measure is positive. When the value under percent of trend is less than 100, the value under relative cyclical residual measure is negative.

53. Answer : (e) [< TOP >](#)

Reason : Both statements (a) and (d) are true. Secular trend represents long term behaviour of a variable; hence statement (b) is false. Seasonal variations are repetitive and predictable movements above and below the trend line occurring within one year or less; hence statement (c) is false.

54. Answer : (a) [< TOP >](#)

Reason : Coefficient of determination represents the proportion of variation in the dependent variable that is explained by the regression line.

55. Answer : (b) [< TOP >](#)

$$\text{Reason : Relative cyclical residual} = \frac{y - \hat{y}}{\hat{y}} \times 100 = \frac{7.50 - 6.25}{6.25} \times 100 = 20$$

56. Answer : (b) [< TOP >](#)

Reason : Cyclical variation can be identified using the percent of trend measure.

57. Answer : (d) [< TOP >](#)

Reason : When the pattern of $\bar{X}$ chart of a process is 'Hugging the center line', then it indicates that the variation has been reduced to a great extent.

58. Answer : (c) [< TOP >](#)

Reason : Year = 2005

$$\therefore x = 2005 - 2003 = 2$$

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

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

$$Y = 160$$

$$\therefore \text{Relative cyclical residual for the year 2005} = \left(\frac{160 - 154}{154} \right) \times 100 = 3.9.$$

59. Answer : (d) [< TOP >](#)

Reason : (a) This is a wrong answer. The movement of dependent variable above and below the secular trend line over periods longer than one year is not known as Seasonal variation.

(b) This is a wrong answer. The movement of dependent variable above and below the secular trend line over periods longer than one year is not known as Secular variation.

(c) This is a wrong answer. The movement of dependent variable above and below the secular trend line over periods longer than one year is not known as Irregular variation.

(d) This is the right answer. The movement of dependent variable above and below the secular trend line over periods longer than one year is known as Cyclical variation.

(e) This is a wrong answer. The movement of dependent variable above and below the secular trend line over periods longer than one year is not known as Regular variation.

60. Answer : (d) [< TOP >](#)

Reason : Given that

Number of sheared off pieces	100	100	100	100	100
Number of defectives	5	6	7	4	8
Proportion defective	0.05	0.06	0.07	0.04	0.08

$$\text{The mean of proportions} = 0.30 / 5 = 0.06$$

$$\begin{aligned}\text{Lower Control Limit for } \mathbf{p} \text{ chart} &= \bar{p} - 3\sqrt{\frac{\bar{p}\bar{q}}{n}} \\ &= 0.06 - 3\sqrt{\frac{0.06 \times 0.94}{100}} \\ &= 0.06 - 3 \times 0.024 = -0.012\end{aligned}$$

Since the lower control limit can not be negative, we set the lower control limit as **zero**.

61. Answer : (e)

[<TOP>](#)

Reason : Given that, total number of samples = 14

Number of items per sample = 30

Number of defectives, **d** = 21

$$\begin{aligned}\bar{p} &= \text{Estimate of the process fraction defectives} = \frac{21}{14 \times 30} = 0.05 \\ \text{Central line for } \mathbf{p} \text{ chart} &= 0.05\end{aligned}$$

$$\begin{aligned}\text{Upper Control Limit for } \mathbf{p} \text{ chart} &= \bar{p} + 3\sqrt{\frac{\bar{p}\bar{q}}{n}} \\ &= 0.05 + 3\sqrt{\frac{0.05 \times 0.95}{30}} = 0.1694\end{aligned}$$

The value of UCL is given by 0.1694.

62. Answer : (e)

[<TOP>](#)

Reason : Lower Control Limit for $\bar{x}$ chart = $\mu_x - 3\sigma_y$

$$\begin{aligned}&= \mu_x - 3\left(\frac{\sigma}{\sqrt{n}}\right) \\ &= 0.5230 - 3\left(\frac{0.0032}{\sqrt{4}}\right) \\ &= 0.5230 - 3(0.0016) \\ &= 0.5182 \text{ cm.}\end{aligned}$$

63. Answer : (c)

[<TOP>](#)

Reason : The central line of the **R** chart = Mean of the sample range ($\bar{R}$)

$$\bar{R} = \Sigma R / n = 63 / 10 = \mathbf{6.3}.$$

64. Answer : (d)

[<TOP>](#)

Reason : The pattern “Hugging the center line” indicates that the deviations are minor and uniform in nature. This indicates that the variation has been reduced to a great extent. So for any quality control manager this pattern is generally desirable.

65. Answer : (b)

[<TOP>](#)

Reason : R chart is plotted using process variability. $\bar{x}$ chart is plotted using process mean. Scatter diagram is the plot of points in regression analysis. Histogram is the graphical representation of frequency distribution. p chart is plotted using qualitative factors of the output. Hence (b) is the answer.

66. Answer : (e)

[<TOP>](#)

$$\begin{aligned}\text{Reason : Upper control limit} &= \bar{x} + \frac{3\bar{R}}{d_2\sqrt{n}} = 25 + \frac{3(7.5)}{2.534\sqrt{9}} = 27.96\end{aligned}$$

67. Answer : (c) [< TOP >](#)

Reason : As long as non-random variations are present the process remains out-of-control. Once the non-random variations are eliminated the process is brought in-control, and the entire process can be redesigned to reduce the incidence of random variation.

68. Answer : (d) [< TOP >](#)

Reason : When carrying out a chi-square test of independence it can be stated in the null hypothesis that the population can be classified on the basis of two criteria which are independent (i.e., not associated)

69. Answer : (a) [< TOP >](#)

Reason : The degree of freedom is equal to the number of categories minus the number of parameters estimated minus one. In this case the population mean and the standard deviation are known. So no parameters are estimated.

$$\therefore \text{DoF} = 6 - 0 - 1 = 5$$

70. Answer : (a) [< TOP >](#)

Reason : Chi square test is used for: Test the deviations between observed and theoretical values, Can be used to test whether there is any association or relationship between two or more attributes or not. It can be used to test the goodness of fit and as a test of population variance. But it can not be used for a test of significance for difference between means of two or more populations.

71. Answer : (d) [< TOP >](#)

Reason : F ratio = Between-column variance/ Within-column variance
= 18.48/14.65 = 1.26.

Degree of freedom for numerator = Number of samples – 1
= 5 – 1 = 4.

Degree of freedom for denominator = Total number of elements present in all the samples – Number of samples

$$= 25 - 5 = 20.$$

From F distribution table, at 0.05 level of significance, and above degrees of freedom, the critical value comes out to be 2.87.

So, the value of calculated F ratio of 1.26 is within acceptable region.

Therefore, we can conclude at a significance level of 0.05 that the rooms do not have significantly different average dust scores.

Hence from above discussion, we can infer that option (d) is correct.

72. Answer : (b) [< TOP >](#)

Reason : As the difference between the observed and expected frequencies decreases, the probability of concluding that the population follows the assumed probability distribution increases.

73. Answer : (b) [< TOP >](#)

Reason : The distribution is positively skewed when the number of degrees of freedom is small.

74. Answer : (b) [< TOP >](#)

Reason : We find the variance within samples with the help of following table:

	x_{1i}	$(x_{1i} - \bar{x}_1)^2$	x_{2i}	$(x_{2i} - \bar{x}_2)^2$	x_{3i}	$(x_{3i} - \bar{x}_3)^2$
	43	16	54	64	35	4
	38	81	40	36	27	36
	59	144	48	4	34	1
	55	64	35	121	40	49
	39	64	50	16	30	9

	48	1	49	9	32	1
Sum	282	370	276	250	198	100
Mean	47		46		33	

Therefore variance within the samples =

$$\frac{\sum (n_j - 1) s_j^2}{n_T - k} = \frac{\sum \sum (x_j - \bar{x}_j)^2}{n_T - k} = \frac{370 + 250 + 100}{18 - 3} = \frac{720}{15} = 48.$$

75. Answer : (c)

[<TOP>](#)

Reason : (a) This is a step followed in the ANOVA. We determine an estimate of the population variance from the variance that exists among the sample means.

(b) This is a step followed in the ANOVA. We determine an estimate of the population variance from the variance that exists within the samples.

(c) This is **not** a step followed in ANOVA. The contingency table consisting observed and estimated frequency is prepared for testing equality of proportion from more than two samples.

(d) This is a step followed in the ANOVA. We accept the null hypothesis if the two estimates of the population variance are not significantly different.

(e) This is a step followed in the ANOVA. We reject the null hypothesis if the F ratio is too high and falls in the critical region.

76. Answer : (d)

[<TOP>](#)

$$\bar{\bar{x}} = \frac{\sum x_{ji}}{n} = \frac{420 + 450 + 380 + 390}{20} = \frac{1640}{20} = 82$$

Reason : The value of the grand mean

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

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

77. Answer : (b)

[<TOP>](#)

Reason : F distribution is unimodal in nature but not bimodal. Hence from above discussion, we can infer that option (b) is **false** regarding F distribution. Options (a), (c), (d) and (e) are all true regarding F distribution.

78. Answer : (b)

[<TOP>](#)

Reason : a. In a p chart, the center line is drawn on the basis of overall sample proportion.

b. In a p chart, the lower control limit can only be greater than or equal to zero.

c, d & e. The upper and lower control limits are fixed by adding and subtracting the estimated standard error of proportion from the overall sample proportion

[<TOP OF THE DOCUMENT>](#)