

Question Paper Quantitative Methods-II (132): October 2004

- | |
|--|
| <ul style="list-style-type: none">• • Answer all questions.• • Marks are indicated against each question. |
|--|

1. The following details are available with regard to a hypothesis test on difference between means of two populations: $\leq$
Answer
 $\geq$

$n_1 = 16$ $n_2 = 9$

The samples are independently collected and it is assumed that the two populations have the same variance. The pooled estimate of the underlying common population variance is 217.04. What is the estimated standard error of difference between means?

- (a) 6.138 (b) 37.675 (c) 12.276 (d) 3.069 (e) 8.682.

(1 mark)

2. If a fair die is thrown twenty times then, what is the probability of obtaining a minimum of 3 in a maximum of seventeen throws? $\leq$
Answer
 $\geq$

- (a) 0.01759 (b) 0.01428 (c) 0.98572 (d) 0.98241 (e) 0.9997.

(2 marks)

3. In which of the following ways can binomial trials be performed on a finite population with specific numbers of successes and failures? $\leq$
Answer
 $\geq$

- (a) By sampling without replacement
(b) By sampling with replacement
(c) By adding a constant to the number of successes obtained in a given number of trials
(d) By deducting a constant from the number of successes obtained in a given number of trials
(e) By multiplying a constant to the number of successes obtained in a given number of trials.

(1 mark)

4. The probability that a continuous random variable will assume a particular value is $\leq$
Answer
 $\geq$

- (a) Zero (b) Between zero and 0.50 (c) Between 0.50 and

1.00

(d) Equal to 0.50

(e) Unknown.

(1 mark)

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

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

[Answer](#)
≤
≥

Twenty values of X have been observed. What is the likelihood that exactly twelve out of the twenty observations are at least 60?

- (a) 0.60 (b) 0.5442 (c) 0.00139 (d) 0.17971 (e) 0.82029.

(1 mark)

6. Y is a binomial random variable. For 20 trials the expected value of Y is 4. What will be the standard deviation of Y for 25 trials?

- (a) 2 (b) 4 (c) 5 (d) 20 (e) 25.

[Answer](#)
≤
≥

(1 mark)

7. A restaurant serves eight items of fish, twelve items of mutton and ten items of chicken. If customers select these items randomly then, what is the probability that two of the next four customers order either mutton or chicken items?

- (a) 0.7705 (b) 0.733 (c) 0.267 (d) 0.2295
(e) 0.5378.

[Answer](#)
≤
≥

(1 mark)

8. Given a probability distribution for a random variable, the covariance of the variable with itself is equal to

- (a) 0 (b) 1
(c) The variance of the variable (d) The standard deviation of the variable
(e) The expected value of the variable.

[Answer](#)
≤
≥

(1 mark)

9. In which of the following conditions will a binomial distribution be symmetrical?

- (a) Number of trials is 30 or more
(b) Number of trials is 100 or more

[Answer](#)
≤
≥

- (c) Probability of success in any trial is less than 0.5
- (d) Probability of failure in any trial is less than the probability of success
- (e) Probability of failure in any trial is equal to 0.5.

(1 mark)

10. Which of the following is **false** with regard to a hypergeometric distribution? ≤
Answer
≥

- (a) The trials are not independent
- (b) The probability of success varies from trial to trial
- (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)

11. A continuous uniform probability distribution is described by ≤
Answer
≥

- (a) The expected value of the distribution
- (b) The variance of the distribution
- (c) The upper and lower limits of range of values of the distribution
- (d) The mid-point of the range of values of the distribution
- (e) Only the lower limit of the range of values of the distribution.

(1 mark)

12. The probability density function of a continuous uniform random variable ≤
Answer
≥

- (a) Has the same value for the entire range of values of the random variable
- (b) Has different values for subintervals of different width within the entire range of values of the random variable
- (c) Has increasing values starting from the beginning to the end of the entire range of values of the random variable
- (d) Has decreasing values starting from the beginning to the end of the entire range of values of the random variable
- (e) Always has a value greater than 1 for the entire range of values of the random variable.

(1 mark)

13. **X** and **Y** are two dependent random variables. The following details are available: ≤
Answer
≥

Standard deviation of **X** = 6

Standard deviation of $Y = 8$

Let $Z = 5X - 10Y$

It is known that the standard deviation of Z is 70.

What is the covariance between the variables X and Y ?

- (a) 36 (b) 24 (c) 64 (d) 240 (e) 2400.

(2 marks)

14. A box contains 60 ballpoint pens out of which 10 pens are defective. 8 pens are randomly picked from the box. The probability distribution of the number of pens found to be defective out of the pens which are randomly picked, will be $\leq$
Answer
 $\geq$

- (a) A discrete uniform distribution (b) A binomial distribution
(c) A hypergeometric distribution (d) A chi-square distribution
(e) An unknown distribution.

(1 mark)

15. 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 two of the absent students were girls? $\leq$
Answer
 $\geq$

- (a) 0.3768 (b) 0.6232 (c) 0.00104 (d) 0.00685 (e) 0.99315.

(1 mark)

16. What is the expected value of the number of heads when two fair coins are tossed? $\leq$
Answer
 $\geq$

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

(2 marks)

17. For which of the following distributions, the z-values and the observed values are same? $\leq$
Answer
 $\geq$

- (a) Binomial distribution
(b) Standard normal distribution
(c) Normal distribution

- (d) Lognormal distribution
- (e) Hypergeometric distribution.

(1 mark)

18. In a binomial distribution the probability of getting zero or more number of successes is equal to Answer $\leq$
 $\geq$
- (a) 0
 - (b) The probability of getting zero success
 - (c) The probability of getting successes in all the trials
 - (d) 1 minus the probability of getting successes in all the trials
 - (e) 1.

(1 mark)

19. Which of the following statements is **false** regarding **F** distribution? Answer $\leq$
 $\geq$
- (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)

20. A normally distributed population has a known standard deviation of 1.0. What is the width of a 95% confidence interval for the population mean? Answer $\leq$
 $\geq$
- (a) 1.00
 - (b) 1.96
 - (c) 0.98
 - (d) 3.92
 - (e) Cannot be determined from the given information.

(1 mark)

21. A sample has been taken from a population and the sample mean has been found to be 100. The upper limit of a 90 percent confidence interval is 112. What is the lower limit of this confidence interval? Answer $\leq$
 $\geq$
- (a) 100
 - (b) 92
 - (c) 68.26
 - (d) 88
 - (e) 144.76.

(1 mark)

22. A small sample has been taken from a normally distributed population and the sample mean has been found to be 62. The upper limit of a 95 Answer $\leq$
 $\geq$

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.

(1 mark)

23. A sample of size 96 has been taken from a population and the sample proportion has been found to be 0.40. The lower limit of a 95% confidence interval for population proportion is 0.302. What is the upper limit of the confidence interval? ≤
Answer
≥

- (a) 0.196 (b) 0.204 (c) 0.098 (d) 0.40
(e) 0.498.

(1 mark)

24. The time that it takes to find a taxi when leaving a restaurant follows a left skewed distribution with a mean of 12 minutes and a standard deviation of 3 minutes. If 100 restaurant patrons are randomly sampled and the average time that it takes for them to find a taxi is calculated, then what type of sampling distribution will be obtained? ≤
Answer
≥

- (a) Approximately normal distribution with mean = 12 minutes and standard deviation = 0.3 minutes
(b) Left skewed distribution with mean = 12 minutes and standard deviation = 0.3 minutes
(c) Left skewed distribution with mean = 12 minutes and standard deviation = 3 minutes
(d) Exactly normal distribution with mean = 12 minutes and standard deviation = 3 minutes
(e) t distribution with unknown number of degrees of freedom.

(1 mark)

25. The following information are available with regard to a sampling distribution of mean: ≤
Answer
≥

Probability that the sample mean is more than 51 = 0.1587

Population mean = 50

Population variance = 36

It is assumed that the Central Limit Theorem will be applicable.

With what sample size is the sampling distribution of mean associated?

- (a) 30 (b) 36 (c) 18 (d) 54 (e) 60.

(2 marks)

26. The finite population multiplier need not be used for calculating the standard error if $\leq$
Answer
 $\geq$
- (a) The sampling fraction is more than 0.05
 - (b) The sampling fraction is less than 0.05
 - (c) The sample size is less than 30
 - (d) The sample size is more than 30
 - (e) The sample mean is greater than the population mean.

(1 mark)

27. Other things remaining the same, if the sample size increases then the estimated standard error of proportion $\leq$
Answer
 $\geq$
- (a) Will increase
 - (b) Will decrease
 - (c) Will remain unchanged
 - (d) Will increase by half
 - (e) Will decrease by half.

(1 mark)

28. The concept of sampling distribution applies to $\leq$
Answer
 $\geq$
- (a) Only discrete probability distributions from which random samples are obtained
 - (b) Only continuous probability distributions from which random samples are obtained
 - (c) Any probability distribution from which random samples are obtained
 - (d) Only normal probability distributions
 - (e) Only binomial probability distributions.

(1 mark)

29. The coefficient of variation of a sample is 20 and the sample mean is 15. What is the point estimate of population variance? $\leq$
Answer
 $\geq$
- (a) 3
 - (b) 9
 - (c) 45
 - (d) 300
 - (e) 0.75.

(1 mark)

30. A sample of size 96 has been taken from a population and a 95 percent confidence interval for the population proportion has been constructed. The upper and lower limits of the confidence interval are 0.4776 and 0.2424 respectively. What is the sample proportion? $\leq$
Answer
 $\geq$

- (a) 0.12 (b) 0.06 (c) 0.64 (d) 0.36 (e)
0.18.

(1 mark)

31. If repeated random samples of size 50 are taken from an infinite population, the distribution of sample means will always be ≤
Answer
≥
- (a) Normal because the population is infinite
(b) Approximately normal because of the Central Limit Theorem
(c) Normal because the sample means are always normal
(d) Normal because we do not know the sampling distribution
(e) A standard normal distribution.

(1 mark)

32. A sample of size 16 yielded a mean of 10. The population mean is known to be 12. What is the expected value of the sampling distribution of mean? ≤
Answer
≥
- (a) 10 (b) 12 (c) 2.5 (d) 3 (e)
1.732.

(1 mark)

33. In a test of ANOVA, the first estimate of the population variance based on the variance among sample means is 7.54. The second estimate of the population variance based on the variances within the samples is 15.68. The test statistic will be equal to ≤
Answer
≥
- (a) 0.18 (b) 0.28 (c) 0.38 (d) 0.48 (e)
1.58.

(1 mark)

34. ABC Co., a manufacturer of light bulbs, claims that the distribution of its light bulb life span has a mean of 600 hours and a standard deviation of 60 hours. The competitors decide to check this claim by randomly purchasing 36 light bulbs manufactured by ABC Co., and testing them to determine the life span of the light bulbs. If the manufacturer's claim is true, then which of the following is **correct** with regard to the sampling distribution of the mean for the sample of 36 light bulbs? ≤
Answer
≥
- (a) The sampling distribution of mean of 36 light bulbs will have a mean of less than 600 hours
(b) The sampling distribution will have a variance of 3600 hours

- (c) The sampling distribution will have a variance of more than 3600 hours
- (d) The sampling distribution will be approximately normal with a mean of 600 hours
- (e) The sampling distribution will be approximately the same as the population distribution which is unknown.

(1 mark)

35. A local credit card agency reported to the government that the unpaid balance for their customers is normally distributed with an average of Rs. 5500 with a standard deviation of Rs. 500. If an auditor randomly sampled 10 of the credit card company's customers, find the probability that the mean of the sample would be less than Rs. 5700.
- (a) 0.50 (b) 0.3971 (c) 0.1029 (d) 0.6029 (e) 0.8971.

(1 mark)

36. 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
- (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)

37. A(n) _____ is a numerical descriptive measure of a population, usually unknown.
- (a) Statistic (b) Point estimator
 (c) Parameter (d) Interval estimator
 (e) Test statistic.

(1 mark)

38. A researcher is interested in determining the average urban household income for the state Uttar Pradesh (U.P). 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) Judgemental sampling.

(1 mark)

39. The observations from a simple random sample are given below:

15 9 11 8 11 6

The sample size is less than 5 percent of the population size. What is the estimated standard error of mean?

- (a) 3.098
- (b) 1.265
- (c) 3.919
- (d) 0.516
- (e) 1.6.

(1 mark)

40. The arithmetic mean of the upper and lower limits of the confidence interval for population mean is equal to the

- (a) Population mean
- (b) Sample mean
- (c) Population standard deviation
- (d) Sample standard deviation
- (e) Standard error of mean.

(1 mark)

41. If the population is normally distributed, then for any given sample size,

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

the sampling distribution of the statistic (where the symbols have their usual meaning), is

- (a) A normal distribution with unknown mean and variance
- (b) The standard normal distribution
- (c) A **t** distribution
- (d) A **t** distribution only when the sample size is 30 or less
- (e) A **t** distribution only when the sample size is more than 30.

(1 mark)

42. Variable **Y** increases as variable **X** increases. Variable **X** cannot explain 36 percent of the variations in **Y**. What is the coefficient of correlation between **X** and **Y**?

- (a) 0.80
- (b) 0.20
- (c) 0.64
- (d) -0.64
- (e) -0.80.

(1 mark)

- 43.** If a deseasonalized value in a time series is 0.79 and the actual data corresponding to the deseasonalized value is 2, the seasonal index is
- (a) 0.395 (b) 2.5316 (c) 253.16 (d) 2.1356 (e) 231.56.
- Answer $\leq$
 $\geq$

(1 mark)

- [illegible]

(1 mark)

45. If the relative cyclical residual measure for a year in a time series is less than zero, it indicates that
- (a) The actual time-series value lies below the trend line and the percent of trend is more than 100
 - (b) The actual time-series value lies below the trend line and the percent of trend is equal to 100
 - (c) The actual time-series value lies above the trend line and the percent of trend is less than 100
 - (d) The actual time-series value lies above the trend line and the percent of trend is more than 100
 - (e) The actual time-series value lies below the trend line and the percent of trend is less than 100.

(1 mark)

- 46.** Zee Company has recorded the following numbers (in Rs. lakh) of total sales of its line of microwave ovens over the last 5 years.

Year	1999	2000	2001	2002	2003
Sales	3.5	3.8	4.0	3.7	3.9

Based on the above information and considering a linear trend, in which year do the sales show the largest percent of trend?

- (a) 1999 (b) 2000 (c) 2001 (d) 2002 (e) 2003.

(2 marks)

47. Which of the following is an example for cyclical variation in time series analysis? ≤
Answer

- (a) Substantial increase in the demand for greeting cards in December of every year ≥
 (b) Improvement in the quality and uses of the product due to the continuous development of the technology
 (c) Fall in the demand for products due to the depression in the economy
 (d) Increase in the growth of population due to improvement in medical facilities
 (e) Increase in the population in Andhra Pradesh resulting from the sudden migration of people from Orissa due to the occurrence of floods.

(1 mark)

48. In the linear regression model, the distribution of residuals is assumed to have ≤
Answer

- (a) Nonconstant variance (b) Constant variance ≥
 (c) Nonconstant sample size (d) Constant sample size
 (e) Decreasing means.

(1 mark)

49. The standard error of estimate ≤
Answer

- (a) Is a measure of central tendency for the distribution of the values estimated by the regression equation ≥
 (b) Measures the variability of the observed values of the dependent variable around their mean
 (c) Measures the variability of the observed values of the dependent variable around the regression line
 (d) Measures the variability of the observed values of the independent variable around their mean
 (e) Measures the variability of the observed values of the independent variable around the regression line.

(1 mark)

50. If the coefficient of correlation between two variables lies between -1 and 0 , then the covariance between them is

$\leq$
Answer
 $\geq$

- (a) Positive
- (b) Negative
- (c) Zero
- (d) Greater in magnitude than the variance of each of the variables
- (e) Lesser in magnitude than the variance of each of the variables.

(1 mark)

51. If the standard error of estimate for a regression equation is zero then,

$\leq$
Answer
 $\geq$

- (a) The estimated values of the dependent variable will always be less than its observed values
- (b) The estimated values of the dependent variable will always be more than its observed values
- (c) The estimated values of the dependent variable will always be equal to its observed values
- (d) The coefficient of correlation between the variables involved will be zero
- (e) The coefficient of correlation between the variables involved will be positive.

(1 mark)

52. Which of the following represents the proportion of variation in the dependent variable that is explained by the regression line?

$\leq$
Answer
 $\geq$

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

(1 mark)

53. If two variables **X** and **Y** are perfectly positively correlated then their covariance will be equal to

$\leq$
Answer
 $\geq$

- (a) 0
- (b) 1
- (c) The product of the standard deviations of **X** and **Y**
- (d) The sum of the standard deviations of **X** and **Y**
- (e) The difference between the standard deviations of **X** and **Y**.

(1 mark)

54. Which of the following is **true** with regard to a given coefficient of correlation and its corresponding coefficient of determination?

$\leq$
Answer
 $\geq$

- (a) The coefficient of determination is always greater than or equal to zero, and less than or equal to 1

- (b) The coefficient of determination is always less than zero
- (c) The coefficient of determination is always equal to 1
- (d) The coefficient of determination always has the same sign as the coefficient of correlation
- (e) The magnitude of coefficient of determination is always higher than the magnitude of coefficient of correlation.

(1 mark)

55. Brilliant Finance Ltd. is a non-banking finance company. In the recent years there has been variations in the demand for loans as well as in the rate of interest. The general manager of the company wants to know whether the variations in the amount of advances explain the variations in the profit after tax of the company. ≤
Answer
≥

The following data pertain to the operations of the company in the recent years:

Year	2003	2002	2001	2000	1999	1998
Amount of advances (Rs. in crores)	7.0	5.5	5.0	4.5	4.0	4.0
Profit after tax (Rs. in lakhs)	54	48	42	27	19	14

What percentage of the variations in profit after tax is explained by the variations in the amount of advances?

- (a) 84.31% (b) 15.69% (c) 91.82% (d) 8.18% (e) 10%.

(1 mark)

56. A simple regression relationship was developed between the variables **X** and **Y**, with **X** as the independent variable. The estimated value of **Y** when **X** = 2, is 7.0; and the estimated value of **Y** when **X** = 8, is 11.8. ≤
Answer
≥

$$\Sigma Y^2 = 791.31 \quad \Sigma XY = 452.7 \quad \Sigma Y = 78.1 \quad n = 8$$

The standard error of estimate for the regression line is

- (a) 1.1113 (b) 8.4572 (c) 7.8482 (d) 7.5123 (e) 6.5893.

(2 marks)

57. Two variables, **X** and **Y**, are related by a regression equation such that **X** is the independent variable and **Y** is the dependent variable. The following details are available: ≤
Answer
≥

Coefficient of correlation between X and Y = 0.80
 Standard deviation of X = 5
 Standard deviation of Y = 8

The slope of the regression equation is

- (a) 0.80 (b) 1.28 (c) 6.4 (d) 32 (e) 64.

(1 mark)

58. For a simple regression equation the following results were obtained:

If $X = 5$, $\hat{Y} = 31$

If $X = 10$, $\hat{Y} = 47$

where X is the independent variable and $\hat{Y}$ is the estimated value of the dependent variable Y .

$\sum X = 60$

Number of data points = 6

What is the mean of the observed values of the dependent variable, Y ?

- (a) 15 (b) 32 (c) 47 (d) 18.2 (e) 25.

(2 marks)

59. The Municipal Corporation of Kolkata has gathered data on the number of traffic accidents and number of cricket matches that occur in the city over a weekend. According to it, there is a close relationship between the number of accidents and the number of cricket matches during weekends. The data collected in this regard are given below :

No. of cricket matches	20	30	10	12	15	25	34
No. of accidents	6	9	4	5	7	8	9

Assuming that there is a linear relation between these two, what would be your prediction of the number of accidents that will occur on a weekend during which 33 cricket matches take place in Kolkata?

- (a) 6.25 (b) 7.35 (c) 8.65 (d) 9.25 (e) 9.85.

(2 marks)

60. If the regression coefficient of the independent variable in a simple regression equation is negative then, which of the following statements is correct?

Answer
 $\leq$
 $\geq$

Answer
 $\leq$
 $\geq$

Answer
 $\leq$
 $\geq$

- (a) The coefficient of correlation between the variables is zero >
- (b) The coefficient of correlation between the variables is the positive square root of the coefficient of determination
- (c) The coefficient of correlation between the variables is the negative square root of the coefficient of determination
- (d) The coefficient of correlation between the variables is 1
- (e) The coefficient of correlation between the variables is between zero and 1.

(1 mark)

61. Which of the following is **not true** with regard to a regression line between **X** (independent variable) and **Y** (dependent variable)? ≤
Answer

- (a) The point $(\bar{X}, \bar{Y})$ always lies on the regression line >
- (b) The sum of the residuals is always zero
- (c) The mean of the fitted values of **Y** is the same as the mean of the observed values of **Y**
- (d) There are always as many points above the fitted line as there are below it
- (e) The regression line minimizes the sum of the squared residuals.

(1 mark)

62. If the correlation between body weight and annual income were high and positive, we could conclude that: ≤
Answer

- (a) High incomes cause people to eat more food >
- (b) Low incomes cause people to eat less food
- (c) High income people tend to spend a greater proportion of their income on food than low-income people, on average
- (d) High income people tend to be heavier than low-income people, on average
- (e) High incomes cause people to gain weight.

(1 mark)

63. A multiple regression relationship contains two independent variables. The standard error of estimate is 3.8. $\Sigma(Y - \hat{Y})^2 = 129.96$. The number of data points is ≤
Answer

- (a) 9 (b) 10 (c) 11 (d) 12 (e) 13.

(1 mark)

64. A multiple regression equation has to be developed between the ≤

variables **Y**, **X₁** and **X₂**. **X₁** and **X₂** are the independent variables, and **Y** is the dependent variable. The following details are available:

Number of data points = 10

$$\begin{array}{llll} \sum Y & = 272 & \sum X_1 Y & = 12005 \\ \sum X_1 & = 441 & \sum X_2 & = 147 \end{array} \quad \begin{array}{ll} \sum X_2 Y & = 4013 \\ \sum X_1^2 & = 19461 \end{array} \quad \begin{array}{l} \sum X_1 X_2 = 6485 \\ \sum X_2^2 = 2173 \end{array}$$

What is the intercept of the multiple regression equation? (Choose the nearest value)

- (a) -1.383 (b) 1.383 (c) -13.83 (d) 13.83 (e) 138.3

(2 mark)

65. Which of the following is **true**, if the value of adjusting constant is greater than one, in the context of analyzing seasonal variations?

- (a) Unadjusted means are more than seasonal indices
(b) Seasonal indices are less than zero
(c) Seasonal indices are less than 100
(d) Seasonal indices are greater than 100
(e) Seasonal indices are more than unadjusted means.

(1 mark)

66. Which of the following indicates the long term behaviour of a variable?

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

(1 mark)

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

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

where $x = \text{Year} - 1998$

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

The observed value of **Y** for the year 2000 is 160.

What is the relative cyclical residual for the year 2000?

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

(1 mark)

68. Which of the following is a technique for studying seasonal variations? ≤
Answer
≥
- (a) Regression analysis (b) Residual method
(c) Hypothesis testing (d) Ratio to moving average method
(e) Linear programming.

(1 mark)

69. The movements of the dependent variable above and below the secular trend line over periods longer than one year can be captured by which of the following? ≤
Answer
≥
- (a) Coefficient of correlation (b) Regression coefficient
(c) Interval estimate (d) Relative cyclical residual
(e) Index number.

(1 mark)

70. Seasonal variations have minimal impact on ≤
Answer
≥
- (a) Price index (b) Quantity index
(c) Weighted index (d) Unweighted index
(e) Chain index.

(1 mark)

71. Which of the following index numbers does **not** satisfy the time reversal test? ≤
Answer
≥
- (a) Fisher's ideal index
(b) Marshall-Edgeworth index
(c) Fixed weights aggregate index
(d) Laspeyres index
(e) Time reversal test is not applicable to index numbers.

(1 mark)

72. A dealer of grocery-items requires an overall comparison of the prices of the commodities he deals in, between the years 2000 and 2003. The following information is provided by him: ≤
Answer
≥

Commodities	Sugar	Salt	Wheat	Rice	Pulses
$P_0(Q_0 + Q_1)$	15750	1512.50	10500	18450	28600
$P_1(Q_0 + Q_1)$	20650	2337.50	13125	22500	40950

The Marshall–Edgeworth price index for the year 2003, using the year

2000 as the base year, is

- (a) 123.25 (b) 128.65 (c) 133.08 (d) 145.45 (e) 155.64.

(1 mark)

73. Consider the following information:

Commodity	2000 (Base year)		2003	
	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

Fisher's ideal price index for the year 2003 is equal to

- (a) 112.82 (b) 113.55 (c) 127.35 (d) 128.12 (e) 133.58.

(2 marks)

74.

Commodity	$\frac{P_1}{P_0} \times 100$	$P_n Q_n$
Rice	118.42	10,000
Wheat	130.77	4,050
Salt	160	1,300
Sugar	145	9,600
Pulses	137.78	13250

Consider the following information for a given year :

The weighted average of relatives price index for the year is

- (a) 118.42 (b) 130.77 (c) 134.54 (d) 137.78 (e) 139.20.

(1 mark)

75. Which of the following patterns of $\bar{x}$ charts indicates that two distinct populations are being observed?

- (a) Increasing trend levels (b) Jumps in process
(c) Cycles in process level (d) Hugging the control limits

(e) Hugging the center line.

(1 mark)

76. Which of the following is used to monitor the variability in a process?

(a) $\bar{x}$ charts

(b) **R** chart

(c) **p** chart

(d) Scatter diagram

(e)

Frequency polygon.

(1 mark)

77. The Lower Control Limit in an **R** chart if **n** = 7, and $\bar{R} = 7.1$, is

(a) 0.338

(b) 0.438

(c) 0.538

(d) 0.658

(e)

0.748.

(1 mark)

78. In case of **p** chart, for **n** = 55, $\bar{p} = 0.33$, the value of centre line will be

(a) 0.67

(b) 0.35

(c) 0.33

(d) 0.83

(e)

0.73.

(1 mark)

79. The mean and standard deviation of a sampling distribution are 15.00 and 0.03 respectively. The Upper Control Limit for the corresponding $\bar{x}$ chart will be

(a) 0.15

(b) 0.45

(c) 15.09

(d) 16.09

(e)

18.09.

(1 mark)

80. In a control chart there are some observations that lie outside the upper and lower control limits. These points are referred to as the

(a) Outliers

(b) Attribute

(c) Inherent variation

(d) Assignable variation

(e)

Common variation.

(1 mark)

81. The following details are available for a test involving ANOVA:

Number of samples = 4

Size of the first sample = 5

Size of the second sample = 5

Size of the third sample = 6

Size of the fourth sample = 4

The number of degrees of freedom in the numerator and denominator of the **F** statistic are

- (a) 4 and 6 respectively
- (b) 4 and 4 respectively
- (c) 5 and 4 respectively
- (d) 3 and 16 respectively
- (e) 4 and 20 respectively.

(1 mark)

82. Which one of the following probability distributions is used for determining whether or not a population has a binomial probability distribution? Answer $\leq$
 $\geq$

- (a) **F** distribution
- (b) Normal distribution with $\mu \neq 0$
- (c) Binomial distribution
- (d) Chi-square distribution
- (e) Standard normal distribution.

(1 mark)

83. The value of chi-square statistic **cannot** be negative, because Answer $\leq$
 $\geq$

- (a) The observed frequencies are always greater than the expected frequencies
- (b) The differences between the observed and expected frequencies are squared
- (c) The chi-square statistic is based upon the absolute value of the differences between the observed frequencies and expected frequencies
- (d) The chi-square statistic is the summation of squares of observed frequencies
- (e) The chi-square statistic is the summation of squares of expected frequencies.

(1 mark)

84. Calculate the test statistic for a hypothesis test on population standard deviation, from the following data: Answer $\leq$
 $\geq$

Sample size = 30

Hypothesized value of the population standard deviation = 1

Sample standard deviation = 0.73

- (a) 15.45
- (b) 16.50
- (c) 17.65
- (d) 18.60
- (e) 19.00.

(1 mark)

85. Shanta Biotech is a renowned company in providing bio-chemistry and bio-physics products. It provides clones of popular genes such as DNA, RNA and Oops. The number of pairs sold for each brand on several different days is as follows:

DNA	RNA	Oops
8	11	9
4	10	7
2	4	5
10	10	9

For a test of ANOVA the estimated common population variance, using the variance within the samples, is

- (a) 3.0833 (b) 5.4853 (c) 7.6453 (d) 9.0833 (e) 12.4533.

(2 marks)

86. 60 account executives were sampled from all account executives in a specific industry and were classified into three age groups, labeled A, B, and C. It is believed that there is an equal likelihood that any account executive selected at random will fall in any of the three age groups. There were 26 account executives in group A, 19 in group B and 15 in group C. The belief is to be tested at a significance level of 0.05, using a suitable technique.

What is the value of the test statistic?

- (a) 3.10 (b) 12.00 (c) 62 (d) 50 (e) 10.25.

(2 marks)

87. The null hypothesis regarding a contingency table with 6 rows and 4 columns should be rejected (at 5% significance level) if the test statistic is

- (a) ≥ 1.96 (b) ≤ 1.64 (c) ≥ 24.996 (d) ≥ 36.415 (e) ≤ 24 .

(1 mark)

88. Determine the critical value of the test statistic in a test of ANOVA, when there are two samples and a total of twenty observations. Assume 0.05 level of significance.

- (a) 2.41 (b) 3.41 (c) 4.41 (d) 7.41 (e) 9.41.

(1 mark)

Suggested Answers
Quantitative Methods-II (132): October 2004

(a)

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

(d)

Let X be the number obtained on any throw. The desired event is $X = 3$ or 4 or 5 or 6 i.e. $X \geq 3$.

This can be considered as success. $\therefore P(\text{Success}) = P(X \geq 3) = \frac{4}{6} = \frac{2}{3}$

Let Y denote the number of throws in which success ($X \geq 3$) happens. Y follows a binomial

distribution with $p = \frac{2}{3}$ and number of trials = 20

$$\begin{aligned} P(Y \leq 17) &= 1 - [P(Y = 18) + P(Y = 19) + P(Y = 20)] \\ &= 1 - [{}^{20}C_{18}p^{18}(1-p)^2 + {}^{20}C_{19}p^{19}(1-p) + {}^{20}C_{20}p^{20}] \\ &= 1 - [190 \times \left(\frac{2}{3}\right)^{18} \left(\frac{1}{3}\right)^2 + 20 \times \left(\frac{2}{3}\right)^{19} \left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^{20}] \\ &= 1 - [0.01428 + 0.00301 + 0.0003] \\ &= 0.98241 \end{aligned}$$

(b)

By sampling with replacement binomial trials can be performed on a finite population with specific number of successes and failures.

(a)

The probability that a continuous random variable will assume a particular value is zero.

(d)

The event that X assumes a value at least 60 occurs when $X = 60$ or 80 or 100

$$\begin{aligned} \therefore P(X \text{ assumes a value at least } 60) &= P(X \geq 60) \\ &= P(X = 60) + P(X = 80) + P(X = 100) \\ &= 0.30 + 0.20 + 0.10 \\ &= 0.60 \end{aligned}$$

Let Y be the no. of observations of the variable X, that are at least 60. Then Y will follow a binomial distribution with number of trials = 20 and probability of success (i.e. getting at least 60) = 0.60

$$\begin{aligned} P(Y = 12) &= {}^{20}C_{12} (0.60)^{12} (1-0.60)^{20-12} \\ &= 0.17971. \end{aligned}$$

(a)

$$\text{For 20 trials } E(Y) = n \times p = 20 \times p = 4$$

$$\therefore p = \frac{4}{20} = 0.20$$

$$\begin{aligned}\therefore \text{ For 25 trials } V(Y) &= n \times p \times q \\ &= 25 \times 0.20 \times (1 - 0.20) \\ &= 4\end{aligned}$$

$$\begin{aligned}\therefore \text{ For 25 trials standard deviation of } Y &= \sqrt{V} \\ &= \sqrt{4} = 2\end{aligned}$$

(d)

The customers will select either chicken or mutton.

The number of such items is $12 + 10 = 22$

Total number of items = $8 + 12 + 10 = 30$

$$\therefore P(\text{chicken or mutton}) = \frac{22}{22 + 8} = \frac{11}{15}$$

Let X represent the number of chicken or mutton items ordered by the customers.

$\therefore$ X follows a binomial distribution with $n = 4$ and $p = \frac{11}{15}$

$$\begin{aligned}P(X = 2) &= {}^4C_2 \left(\frac{11}{15}\right)^2 \left(1 - \frac{11}{15}\right)^{4-2} \\ &= 6 \times \left(\frac{11}{15}\right)^2 \left(\frac{4}{15}\right)^2 \\ &= 0.2295\end{aligned}$$

(c)

The covariance of any random variable with itself is equal to the variance of the variable.

(e)

A binomial distribution is symmetrical if the probability of success and the probability of failure are equal. If $p + q = 1$, this means that the probability of failure is equal to 0.5.

(d)

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

(c)

A continuous uniform probability distribution is described by the upper and lower limits of the interval.

(a)

The probability density function of a continuous uniform random variable has the same value for the entire range of values of the random variable.

(b)

$$V(Z) = V(5X - 10Y) = 5^2 V(X) + 10^2 V(Y) - 2 \times 5 \times 10 \times \text{Cov}(X, Y)$$

$$\text{Or } 70^2 = 25 \times 6^2 + 100 \times 8^2 - 100 \text{ Cov}(X, Y)$$

$$\text{Or } 4900 = 900 + 6400 - 100 \text{ Cov}(X, Y)$$

Or $100 \text{ Cov}(X,Y) = 900 + 6400 - 4900$

Or $100 \text{ Cov}(X,Y) = 2400$

Or $\text{Cov}(X,Y) = \frac{2400}{100} = 24$.

(c)

The number of defective pens out of the pens randomly picked from the box follows a hypergeometric distribution because the population of ballpoint pens is finite and it contains a specific number of successes (defective pens).

(a)

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

= No. of elements in the population = 25

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

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

= Desired number of successes = 2

$P(X = 2) = \frac{{}^r C_x \cdot {}^{(N-r)} C_{(n-x)}}{{}^N C_n} = \frac{{}^{11} C_2 \cdot {}^{(25-11)} C_{(5-2)}}{{}^{25} C_5} = \frac{55 \times {}^{14} C_3}{{}^{25} C_5} = 0.3768$

(c)

When two fair coins are tossed the probability distribution of the number of heads will be as follows:

Number of heads	Probability
0	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
1	$\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
2	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$(\text{Number of heads}) = \left(0 \times \frac{1}{4}\right) + \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{4}\right) = 1$

(b)

In a standard normal distribution the Z-values and observed values are same.

(e)

In a binomial distribution the probability of getting zero or more number of successes

$= P(x = 0) + P(x = 1) + \dots P(x = n) = 1.00$.

(b)

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

(e)

Width of confidence interval for mean = UCL – LCL

$$= (\bar{x} + z\sigma_{\bar{x}}) - (\bar{x} - z\sigma_{\bar{x}})$$

$$= 2z\sigma_{\bar{x}}$$

$\sigma_{\bar{x}}$ cannot be known without knowing the sample size, which is not given. Hence the width can not be determined on the basis of the given information.

(d)

$$\begin{aligned} \text{UCL} &= \bar{x} + z\sigma_{\bar{x}} \\ \text{or } z\sigma_{\bar{x}} &= \text{UCL} - \bar{x} = 112 - 100 = 12 \\ \text{LCL} &= \bar{x} - z\sigma_{\bar{x}} = 100 - 12 = 88. \end{aligned}$$

(b)

Since the population variance is known, the distribution to be used is the $N(0, 1)$ distribution notwithstanding the sample size is small.

$$\begin{aligned} \text{UCL} &= \bar{x} + 1.96\sigma_{\bar{x}} \\ \text{or } 1.96\sigma_{\bar{x}} &= \text{UCL} - \bar{x} = 81.60 - 62 = 19.60 \\ \sigma_{\bar{x}} &= \frac{19.60}{1.96} = 10 \end{aligned}$$

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

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

(e)

$$\begin{aligned} \text{LCL} &= \hat{p} - z\sqrt{\frac{\hat{p}\hat{q}}{n}} \text{ or } 0.302 = 0.40 - z\hat{\sigma}_{\hat{p}} \text{ or} \\ z\hat{\sigma}_{\hat{p}} &= 0.40 - 0.302 = 0.098 \\ \text{UCL} &= \hat{p} + z\hat{\sigma}_{\hat{p}} = 0.40 + 0.098 = 0.498. \end{aligned}$$

(a)

By Central Limit Theorem for large samples the sample mean is approximately normally distributed with

$$\text{population mean and standard deviation} = \frac{\sigma}{\sqrt{n}}$$

$$E(\bar{x}) = \mu = 12 \text{ min}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{100}} = 0.30 \text{ min}$$

Approximately normal distribution with $E(\bar{x}) = 12 \text{ min}$ and $\sigma_{\bar{x}} = 0.30 \text{ min}$.

(b)

By Central Limit Theorem for large samples the sample mean is approximately normally distributed with

$$\text{population mean and standard deviation} = \frac{\sigma}{\sqrt{n}}$$

$$P(\bar{x} > 51) = 0.1587 = P(z > k)$$

$$P(0 < z < k) = 0.50 - 0.1587 = 0.3413$$

From standard normal table we can see that $k = 1.00$

$$z = 1 = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \text{ or } 1 = \frac{51 - 50}{\sigma_{\bar{x}}} \text{ or } \sigma_{\bar{x}} = \frac{1}{1} = 1$$

$$\frac{\sigma}{\sqrt{n}} = 1 \text{ or } n = \left(\frac{\sigma}{1}\right)^2 = \sigma^2 = 36.$$

(b)

The finite population multiplier need not be used for calculating the standard error if the sampling fraction is less than 0.05.

(b)

Estimated standard error of proportion $\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

As 'n' increases $\hat{\sigma}_{\bar{p}}$ decreases.

(c)

The concept of sampling distribution applies to any probability distribution from which random samples are obtained.

(b)

$$\text{Coefficient of variation} = \frac{s}{\bar{x}} \times 100$$

$$\text{or } \frac{s}{15} \times 100 = 20$$

$$\text{or } s = 20 \times \frac{15}{100} = 3$$

$$\therefore \text{Point estimate of population variance} = s^2 = 9$$

(d)

$$\begin{aligned} \text{UCL} &= \hat{p} + z\hat{\sigma}_{\bar{p}} = 0.4776 \\ \text{CL} &= \hat{p} - z\hat{\sigma}_{\bar{p}} = 0.2424 \\ \hat{p} &= \frac{\text{UCL} + \text{LCL}}{2} = \frac{0.4776 + 0.2424}{2} = 0.36. \end{aligned}$$

(b)

If repeated random samples of size 50 are taken from an infinite population, the distribution of sample means will always be approximately normal because of the Central Limit Theorem. This is because the CLT is applicable whenever the sample size is large (>30).

(b)

The expected value of sampling distribution of mean = Population mean = 12.

(d)

The first estimate of the population variance based on the variance among sample means is 7.54. The second estimate of the population variance based on the variances within the samples is 15.68. The F statistic will be 7.54/15.68, i.e., 0.48.

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

(d)

By Central Limit Theorem for large samples the sample mean is approximately normally distributed with

population mean and standard deviation $= \frac{\sigma}{\sqrt{n}}$

$$(\bar{x}) = \mu = 600$$

$$(\bar{x}) = (\sigma_{\bar{x}})^2 = \frac{\sigma^2}{n} = \frac{3600}{36} = 100$$

ence statement (d) is correct.

(e)

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 $= \frac{\sigma}{\sqrt{n}}$

$$(\bar{x}) = \mu = 5,500$$

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

$$P(\bar{x} < 5700) = P\left(z < \frac{5700 - 5500}{158.11}\right) = P(z < 1.265)$$

$$= 0.5000 + \left[0.3962 + \frac{(0.3980 - 0.3962)}{(1.27 - 1.26)}(1.265 - 1.26)\right] = 0.8971.$$

(a)

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

(c)

A Parameter is a numerical descriptive measure of a population, usually unknown.

(d)

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.

(b)

When the sample size is less than 5 percent of the population size the estimated standard error of mean =

$$= \left[\frac{1}{(n-1)} \left\{ \sum_{i=1}^n X^2 - \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} \right\} \right]^{1/2}$$
$$= \left[\frac{1}{(6-1)} \left\{ 648 - \frac{60^2}{6} \right\} \right]^{1/2} = \left[\frac{1}{5} \times 48 \right]^{1/2} = 3.098$$

$$\text{estimated standard error of mean} = \frac{s}{\sqrt{n}} = \frac{3.098}{\sqrt{6}} = 1.265$$

(b)

The arithmetic mean of the upper and lower limits of the confidence interval for population mean is equal to sample mean.

(c)

$\bar{x}$ = sample mean, μ = population mean, s = sample standard deviation, n = sample size. If the population

is normally distributed, then for any given sample size, the sampling distribution of $\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ is a t distribution irrespective of the sample size being large or small.

(a)

Since Y increases as X increases, they are positively correlated. Therefore the coefficient of correlation between X and Y is the positive square root of the coefficient of determination.

$$\begin{aligned} \text{Coefficient of correlation} &= \sqrt{\text{Coefficient of determination}} \\ &= \sqrt{1 - \text{Proportion of unexplained variations}} \\ &= \sqrt{1 - 0.36} \\ &= 0.80 \end{aligned}$$

(c)

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

$$\begin{aligned} \therefore \text{Seasonal Index} &= \frac{\text{Actual data}}{\text{Deseasonalized value}} \\ &= \frac{2}{0.79} \times 100 = 253.16 \end{aligned}$$

(e)

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

$$\therefore \text{Estimated sales value} = \frac{\text{Actual sales value}}{\text{Percent of trend measure}}$$

$$= \frac{3.4}{98.7} \times 100 = \text{Rs.3.445 lakh}$$

(e)

If the relative cyclical residual measure for a year in a time series is less than zero, it indicates that the actual time series value lies below the trend line and the percent of trend is less than 100.

(c)

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

where, $x = X - \bar{X}$

$$b = \frac{\sum xY}{\sum x^2} \quad a = \bar{Y}$$

$$\therefore b = \frac{\sum xY}{\sum x^2} = \frac{0.70}{10} = 0.07$$

$$a = \bar{Y} = \frac{\sum Y}{n} = \frac{18.9}{5} = 3.78$$

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

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

x value 1999 is -2; $\hat{Y} = 3.78 + 0.07x$

x value for 1999 is -2; $\hat{Y} = 3.78 + 0.07(-2) = 3.64$

x value for 2000 is -1; $\hat{Y} = 3.78 + 0.07(-1) = 3.71$

x value for 2001 is 0; $\hat{Y} = 3.78 + 0.07(0) = 3.78$

x value for 2002 is 1; $\hat{Y} = 3.78 + 0.07(1) = 3.85$

x value for 2003 is 2; $\hat{Y} = 3.78 + 0.07(2) = 3.92$

	Year	1999	2000	2001	2002
$x = X - \bar{X}$		-2	-1	0	1
Actual sales		3.5	3.8	4.0	3.7
Forecasted sales		3.64	3.71	3.78	3.85
Percent of trend		96.15	102.43	105.82	96.1

Largest percent of trend was prevailed in 2001.

(c)

Substantial increase in the demand for greeting cards in December of every year is an example for seasonal variation. Improvement in the quality and uses of the product due to the continuous development of the technology is an example for secular trend. Fall in the demand for products due to the depression in the economy is an example for cyclical variation. Increase in the growth of population due to improvement in medical facilities is an example for secular trend. Increase in the population in Andhra Pradesh resulting from the sudden migration of people from Orissa due to the occurrence of floods is an example for irregular variation.

(b)

In the linear model, the distribution of residuals is assumed to have constant variance

(c)

The standard error of estimate measures the variability of the observed values of the dependent variable around the regression line.

(b)

If the coefficient of correlation between two variables lies between -1 and 0 , then the covariance between them will be negative.

(c)

If the standard error of estimate for a regression equation is zero then, the estimated values of the dependent variable will always be equal to its observed values.

(a)

The proportion of variations that is explained by the regression line is represented by the coefficient of determination.

(c)

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

$$r = +1 \Rightarrow \text{Cov}(x, y) = \sigma_x \cdot \sigma_y.$$

(a)

a. The coefficient of determination is

coefficient of correlation (r). Hence it will always be ≥ 0 and ≤ 1 .

b. From above we can see that the coefficient of determination cannot be less than zero.

c. The coefficient of determination is will be equal to 1 , only if the coefficient of correlation is equal to ± 1 ; in other cases it will be > 0 and < 1 .

d. The coefficient of determination is always positive; The coefficient of correlation may be negative.

e. Since $-1 \leq r \leq 1$, the coefficient of determination which is the square of coefficient of correlation (r^2) will always have magnitude less than or equal to r .

(a)

Let the following notations be used:

X : Amount of advances (Rs. in crores)

Y : Profit after tax (Rs. in lakhs)

Coefficient of correlation is given by:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_x \cdot \sigma_y} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \cdot \Sigma(Y - \bar{Y})^2}}$$

X	7.0	5.5	5.0	4.5	4.0	4.0	$\Sigma X = 30$
Y	54	48	42	27	19	14	$\Sigma Y = 204$
$(X - \bar{X})$	2	0.5	0	-0.5	-1	-1	

$(Y - \bar{Y})$	20	14	8	-7	-15	-20	
$(X - \bar{X})^2$	4	0.25	0	0.25	1	1	$\Sigma(X - \bar{X})^2 = 6.5$
$(Y - \bar{Y})^2$	400	196	64	49	225	400	$\Sigma(Y - \bar{Y})^2 = 1334$
$(X - \bar{X})(Y - \bar{Y})$	40	7	0	3.5	15	20	$\Sigma(X - \bar{X})(Y - \bar{Y}) = 85.5$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{30}{6} = 5.0$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{204}{6} = 34$$

$$\therefore r = \frac{85.5}{\sqrt{6.5 \times 1334}} = 0.9182$$

Proportion of variations in profit after tax that is explained by the variations in the amount of advances
 $r^2 = (0.9182)^2 = 0.8431$

$\therefore$ Percentage of variations in profit after tax that is explained by the variations in the amount of advances
 $0.8431 \times 100 = 84.31\%$.

(a)

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

Given: $7 = a + b(2)$ (For $X=2$)

Or $7 = a + 2b$(A)

$11.8 = a + b(8)$ (For $X=8$)

or $11.8 = a + 8b$(B)

$$\therefore b = \frac{(11.8-7)}{(8-2)} = 0.80$$

Putting $b = 0.80$ in (A)

$$7 = a + 2(0.80)$$

or $a = 7 - 1.6 = 5.4$

$\therefore$ The regression equation is

$$\hat{Y} = 5.4 + 0.80X$$

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

$$\Sigma Y^2 = 791.31 \quad \Sigma XY = 452.7$$

$$\Sigma Y = 78.1 \quad n = 8$$

$$\begin{aligned} \therefore S_e &= \sqrt{\frac{791.31 - (5.4 \times 78.1) - (0.8 \times 452.7)}{8 - 2}} \\ &= 1.1113 \end{aligned}$$

(b)

$$\text{Slope of a regression equation, } b = \frac{\text{Cov}(X, Y)}{\sigma_x^2}$$

$$\begin{aligned} &= \frac{r \sigma_x \sigma_y}{\sigma_x^2} \\ &= \frac{0.80 \times 5 \times 8}{25} = 1.28 \end{aligned}$$

(c)

A simple regression equation is of the form:

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

Given:

$$\text{If } X = 5, \quad \hat{Y} = 31 \therefore 31 = a + 5b \dots \dots \dots (A)$$

$$\text{If } X = 10, \quad \hat{Y} = 47 \therefore 47 = a + 10b \dots \dots \dots (B)$$

Subtracting (A) from (B) we get:

$$47 - 31 = (10 - 5)b$$

$$16 = 5b$$

$$\text{or } b = \frac{16}{5} = 3.2$$

Putting the values of b in (B):

$$47 = a + 10 \times 3.2 = a + 32$$

$$\text{or } a = 47 - 32 = 15$$

$$\therefore \hat{Y} = 15 + 3.2X$$

$$\text{Now, } a = \bar{Y} - b\bar{X} = \bar{Y} - b\left(\frac{\Sigma X}{n}\right)$$

$$\therefore \bar{Y} = a + b\left(\frac{\Sigma X}{n}\right) = 15 + 3.2\left(\frac{60}{6}\right) = 47$$

Mean of the observed values of Y, i.e. $\bar{Y} = 47$

(d)

Let X be the number of cricket matches played and Y be the number of accidents during weekends.

$$X = 146, \quad \Sigma Y = 48, \quad n = 7$$

$$\begin{aligned} \sum XY &= 1101, \\ \sum X^2 &= 3550, \\ \bar{X} &= 20.86, \bar{Y} = 6.86 \end{aligned}$$

$$\text{The coefficient of regression } b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{1101 - 7 \times 20.86 \times 6.86}{3550 - 7 \times (20.86)^2} = 0.197$$

$$\text{The Y intercept } a = \bar{Y} - b \bar{X} = 6.86 - 0.197 \times 20.86 = 2.751$$

$$\text{Therefore the estimating equation is } \hat{Y} = 2.751 + 0.197 X$$

$$\text{The estimate of Y for } X = 33, \text{ is } \hat{Y} = 2.751 + 0.197 \times 33 = 9.251 \approx 9.25$$

(c)

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

' is the regression coefficient of the independent variable. $b < 0$ implies that the variables x and y are negatively correlated.

$$\therefore \text{Coefficient of correlation} = \text{Negative}$$

Not of coefficient of determination.

(d)

The statement that there are always as many points above the fitted line as there are below it is wrong.

(d)

If the correlation between body weight and annual income were high and positive, we could conclude that high-income people tend to be heavier than low-income people, on average.

(d)

$$\begin{aligned} \text{The number of data points can be obtained as } & \frac{\sum (y - \hat{y})^2}{s_e^2} + (k + 1) \\ &= \frac{129.96}{3.8 \times 3.8} + 2 + 1 = 12 \end{aligned}$$

(c)

A multiple regression equation is obtained by solving the following equations:

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

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

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

Putting in the relevant values in the equation :

$$272 = 10a + 441 b_1 + 147 b_2 \dots \dots \dots (A)$$

$$12005 = 441a + 19461 b_1 + 6485 b_2 \dots \dots \dots (B)$$

$$4013 = 147a + 6485 b_1 + 2173 b_2 \dots \dots \dots (C)$$

Multiplying (A) by 441 and (B) by 10, and subtracting (B) from (A):

$$\begin{array}{rclcl}
 119952 & & 4410a & & 194481b_1 & & 64827b_2 \\
 = & & + & & + & & \\
 120050 & & 4410a + & & 194610b_1 & & 64850b_2 \\
 = & & & & + & & \\
 - & & - & & - & & \\
 \therefore 98 = 129 & & - & & & & b_1 + \\
 \hline
 -98 & & & & -129b_1 & & -23b_2 \\
 = & & & & & & \\
 \hline
 \end{array}$$

.....(D)

Multiplying (B) by 147 and (C) by 441, and subtracting (C) from (B):

$$\begin{array}{rclcl}
 1764735 & & 64827a & & 2860767b_1 & + & 953295 \\
 = & & + & & & & b_2 \\
 \\
 1769733 & & 64827a & & 2859885b_1 & + & 958293b_2 \\
 = & & + & & & & \\
 - & & & & - & & \\
 = & & - & & - & & \\
 882b_1 & + & & & & & 4998b_2 \quad \dots(E) \\
 \hline
 -4998 & & & & 882b_1 - & & 4998b_2 \\
 = & & & & & & \\
 \hline
 \end{array}$$

Multiplying (D) by 882 and (E) by 129, and subtracting (E) from (D):

$$\begin{array}{rclcl}
 86436 & = & 113778b_1 & & 20286b_2 \\
 & & + & & \\
 644742 & & -113778b_1 & & 644742b_2 \\
 = & & + & & \\
 \hline
 731178 & & & & 665028b_2 \\
 = & & & & \\
 \hline
 \end{array}$$

$$\therefore b_2 = \frac{731178}{665028} = 1.099$$

Putting the values of b_2 in (D) :

$$98 = 129b_1 + (23 \times 1.099)$$

$$\text{or } b_1 = \frac{98 - 25.277}{129} = 0.564$$

Putting the values of b_1 and b_2 (A):

$$72272 = 10a + (441 \times 0.564) + (147 \times 1.099)$$

$$\text{or } a = \frac{272 - 248.724 - 161.553}{10} = -13.828$$

∴ The multiple regression equation

$$\hat{Y} = -13.828 + 0.564X_1 + 1.099X_2$$

The value of 'a', can also be found out by first eliminating 'b₁' and 'b₂'. The value of 'a' will be the same. However, if we attempt to eliminate 'b₁' and 'b₂' first then the system of equations will involve unmanageably large numbers. Hence we have eliminated 'a' and 'b₁' first, and then we have derived the value of 'a' in the end from the values of 'b₁' and 'b₂'. The sequence of steps shown above gives the solution for the entire equation, for the complete understanding of the student).

(e)

Seasonal index = Unadjusted means × Adjusting constant

From the above equation, if the value of adjusting constant is greater than one, it can be understood that seasonal index is more than unadjusted mean.

(a)

The secular trend shows the long term behaviour of a variable.

(c)

$$\text{Year} = 2000$$

$$\therefore x = 2000 - 1998 = 2$$

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

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

$$Y = 160$$

$$\text{Relative cyclical residual for the year 2000} = \left(\frac{160 - 154}{154} \right) \times 100 = 3.9.$$

(d)

Ratio-to-moving-average method is a technique for studying seasonal variations.

(d)

- (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.

(e)

In Chain Index, the base year is not fixed but changes from year to year. Here, new relatives are constructed for each year with the previous year as base. Thus, seasonal variations have minimal impact on Chain Index.

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

(d)

Laspeyres index does not satisfy the time reversal test

(c)

$$\text{Marshall –Edgeworth price index} = \frac{\sum P_1(Q_0 + Q_1)}{\sum P_0(Q_0 + Q_1)} \times 100 = \frac{99562.50}{74812.50} \times 100 = 133.08$$

(d)

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

$$\begin{aligned} \sum P_1 Q_0 &= (10.25 \times 5000) + (14.00 \times 6000) + (28.00 \times 3000) + (14.50 \times 3000) + (4.50 \times 500) \\ &= 265000 \end{aligned}$$

$$\begin{aligned} \sum P_1 Q_1 &= (10.25 \times 4800) + (14.00 \times 5500) + (28.00 \times 3500) + (14.50 \times 2400) + (4.50 \times 600) \\ &= 261700 \end{aligned}$$

$$\sum P_0 Q_0 = (7.50 \times 5000) + (10 \times 6000) + (25 \times 3000) + (10.60 \times 3000) + (2.60 \times 500) = 205600$$

$$\sum P_0 Q_1 = (7.50 \times 4800) + (10 \times 5500) + (25 \times 3500) + (10.60 \times 2400) + (2.60 \times 600) = 205500$$

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

(c)

$$\text{Unweighted average of relatives price index} = \frac{\sum \left[\left(\frac{P_1}{P_0} \times 100 \right) P_n Q_n \right]}{\sum P_n Q_n}$$

Commodity	$\frac{P_1}{P_0} \times 100$	$P_n Q_n$	$\left(\frac{P_1}{P_0} \times 100 \right) P_n Q_n$
Rice	118.42	10,000	11,84,200
Wheat	130.77	4,050	52,96,18.50
Salt	160	1,300	20,80,000
Sugar	145	9,600	13,92,000
Pulses	137.78	13,250	18,25,585
		$\Sigma = 38,200$	$\Sigma = 51,39,403.50$

$$\frac{\sum \left[\left(\frac{P_1}{P_0} \times 100 \right) P_n Q_n \right]}{\sum P_n Q_n} = \frac{51,39,403.50}{38,200} = 134.54$$

Weighted average of relatives price index for year 2001.

$$\frac{51,39,403.50}{38,200} = 134.54$$

(d)

Hugging the control limits' pattern indicates that two distinct populations are being observed.

(b)

(a) This is a wrong answer. The $\bar{x}$ charts are constructed using the mean value as the central limit and three standard deviations from the mean as upper and lower control limits.

(b) This is the right answer. **R** charts are used to monitor the variability in the individual characteristics of the product. The control limit is the mean range and the upper and lower control limits are constructed with the standard deviation in the range values.

(c) This is the wrong answer. **p** charts deal with the attributes that are qualitative in nature. The measure that is applied here is the proportion of the data points satisfying the given criteria and the proportion of the data points that does not satisfy the criteria.

(d) This is the wrong answer.

(e) This is the wrong answer.

(c)

$$\text{The Lower Control Limit in an R Chart} = \bar{R} - \frac{3d_3 \bar{R}}{d_2}$$

$$= 7.1 - \frac{3 \times 0.833 \times 7.1}{2.704}$$

$$= 0.538.$$

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

(c)

In case of p-chart, for $n = 55$, $\bar{p} = 0.33$, the value of centre line will be the value of $\bar{p}$.i.e., 0.33.

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

(c)

The Upper Control Limit of $\bar{x}$ Chart = $\mu_{\bar{x}} + 3\sigma_{\bar{x}}$

$$15.00 + 3 \times 0.03$$

$$15.09.$$

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

(a)

(a) This is the **right** answer. In a control chart there are some observations that lie outside the upper and lower control limits. These points are referred to as the Outliers.

(b) This is the wrong answer. Qualitative variables with only two categories are termed as attributes. Monitored with the help of p chart.

(c) This is the wrong answer. This is another name of random variation.

(d) This is the wrong answer. this is the nonrandom, systematic variability in a process. It usually can be corrected without redesigning the entire process.

(e) This is the wrong answer. This is the random variability inherent in the system. It usually cannot be corrected without redesigning the entire process.

(d)

In ANOVA:

Number of degrees of freedom in the numerator of F ratio = Number of samples – 1

$$= 4 - 1 = 3$$

Number of degree of freedom in the denominator of F ratio

= Total sample size – Number of samples

$$= (5 + 5 + 6 + 4) - 4$$

16

(d)

Chi-square test is used to test the goodness of fit of a distribution. To test whether the observed frequencies follow Poisson distribution frequencies or not chi-square is used.

(b)

a. The observed frequencies may be less than the expected frequencies. Hence this is not the correct reason.

b. The differences between the observed and expected frequencies are squared. Hence the chi-square cannot be negative.

c. The chi-square statistic is not based upon the absolute value of the difference between the observed frequencies and expected frequencies.

d & e. The chi-square statistic is not based upon the summation of the squares of the observed or expected frequencies

(a)

The test statistic is the χ^2 statistic.

$$s^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$29 \times (0.73)^2 / 1$$

$$15.45.$$

(d)

The Means of the samples are $\bar{x}_1 = 6, \bar{x}_2 = 8.75, \bar{x}_3 = 7.5$.
 The grand mean, $\bar{\bar{x}} = [(8 + 4 + 2 + 10) + (11 + 10 + 4 + 10) + (9 + 7 + 5 + 9)] / 12 = 7.42$.
 We calculate Within column variance as below:

$$s^2 = \sum \left(\frac{n_j - 1}{n_T - k} \right) \times s_j^2$$

$$s^2 = \frac{(8-6)^2 + (4-6)^2 + (2-6)^2 + (10-6)^2}{4-1}$$

$$= 13.3333.$$

$$s^2 = \frac{(11-8.75)^2 + (10-8.75)^2 + (4-8.75)^2 + (10-8.75)^2}{4-1}$$

$$= 10.25.$$

$$s^2 = \frac{(9-7.5)^2 + (7-7.5)^2 + (5-7.5)^2 + (9-7.5)^2}{4-1}$$

$$= 3.6667.$$

$$s_o, \hat{\sigma}^2 = \left(\frac{4-1}{12-3} \right) \times (13.3333 + 10.25 + 3.6667)$$

$$9.0833.$$

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

(a)

The appropriate test is a chi-square test of goodness of fit. Since there is an equal likelihood for any executive to be in any of the three age groups, the probability of observing any of the age groups is $\frac{1}{3}$.
 Hence the expected frequencies for the three age groups will be as follows:

A	B	C
---	---	---

$\frac{1}{3} \times 60 = 20$	$\frac{1}{3} \times 60 = 20$	$\frac{1}{3} \times 60 = 20$
------------------------------	------------------------------	------------------------------

he observed frequencies are :

A	B	C
26	19	15

he test statistic is the chi-square statistic, which is calculated below :

$$\chi^2 = \frac{(26-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(15-20)^2}{20} = 3.10$$

(c)

The test for a contingency table is a chi-square test of independence. A contingency table with 6 rows and 4 columns i.e. at $(6-1)(4-1) = 15$ degrees of freedom should be rejected (at 95% confidence) if the test statistic is greater than or equal to 24.996 (from chi-square tables).

(c)

Degree of freedom for numerator = Number of samples – 1

$$2 - 1 = 1.$$

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

$$20 - 2 = 18.$$

From F distribution table, at 0.05 level of significance, and above degrees of freedom, F ratio comes out to be 2.10. Hence from above discussion, we can infer that option (c) is correct.

[< TOP OF THE DOCUMENT >](#)