



Your Personal Exams Guide



NDA



CDS



SSC CGL



CBSE UGC NET



IAS



SSC CHSL



CTET



MPSC



AFCAT



CSIR UDC NET



IBPS PO



UP POLICE



SSC MTS



SBI PO



BPSC



UP TET



IBPS RRB



IBPS CLERK



IES



UPSC CAPF



SSC Stenogr..



RRB NTPC



SSC GD



RBI GRADE B



RBI Assistant



DSSSB

SSC CGL 2021 (Tier-II) JSO Prev. Year Paper (10-Aug-2022)

Total Time: 2 Hour

Total Marks: 200

Instructions

Sl No.	Section Name	No. of Question	Maximum Marks	Negative Marks	Positive Marks
1	Quantitative Aptitude	100	200	0.5	2

- 1.) A total of 120 minutes is allotted for the examination.
- 2.) The server will set your clock for you. In the top right corner of your screen, a countdown timer will display the remaining time for you to complete the exam. Once the timer reaches zero, the examination will end automatically. The paper need not be submitted when your timer reaches zero.
- 3.) There will, however, be sectional timing for this exam. You will have to complete each section within the specified time limit. Before moving on to the next section, you must complete the current one within the time limits.

Your Personal Exams Guide

Quantitative Aptitude

1. For the production data

(+2, -0.5)

Year:	1	2	3	4	5	6
Production:	25	95	55	94	25	75

the third 3-year simple moving average is:

- a. 59
- b. 56
- c. 58
- d. 57

2. For the frequency distribution of income (in lakh) of the employees in factory

(+2, -0.5)

Class:	1.5-2.5	2.5-3.5	3.5-4.5	4.5-5.5
Frequency:	1	3	4	2

the value of mode is

- a. 3.833
- b. 3.533
- c. 3.933
- d. 3.333

3. The mean and median of the distribution is 12 and 15. Then the mode equals (+2, -0.5)
to:

- a. 21
- b. 24
- c. 15
- d. 18

4. Which option is incorrect? (+2, -0.5)

- a. For k treatments and N observations, the degree of freedom of variation between group is $k - 1$
- b. mean sum of square of total = mean sum of square of treatment + mean sum of square of error
- c. sum of square of total = sum of square of treatment + sum of square of error
- d. For k treatments and N observations, the degree of freedom of variation within group is $N - k$

5. In a two-way ANOVA table (+2, -0.5)

Source of Variation	Degree of Freedom	Sum of square	Mean sum of squares	F
Due to Level A	2	294	147	F_A
Due to Level B	2	6	3	F_B
Due to error	4	12	3	
Total	x	312		

the value of x, F_A , F_B are:

- a. (8, 49, 3)
- b. (9, 49, 3)
- c. (8, 49, 1)
- d. (9, 49, 1)

6. If the first quartile of data set 8, 10, 8, 7, 9 is 7.5, then the value of quartile deviation is (+2, -0.5)

- a. 1.0
- b. 2.5
- c. 7.5
- d. 9.5

7. For two items, tea (1 kg) and sugar (1 kg), the prices in the year 2019 were Rs. 100 and Rs. 50, respectively, whereas the prices in the year 2020 were Rs. 125 and Rs. 60, respectively. The value of Laspeyres price index is: (+2, -0.5)

- a. 123.33
- b. 122.33
- c. 125.33
- d. 120.33

8. For the frequency distribution of x , number of grammatical mistakes per line (+2, -0.5) is as follows.

x :	0	2	4
$p(x)$	0.5	0.3	0.2

The third factorial moment of X is:

- a. 4.8
- b. 2.8
- c. 3.2
- d. 3.6

9. The fair dice is rolled 15 times and face value are noted (+2, -0.5)

Face value:	1	2	3	4	5	6
# of times	3	0	4	2	5	1

The empirical probability of getting a number greater than 4 when a dice is rolled, is

- a. 0

b. $\frac{2}{5}$

c. 1

d. $\frac{1}{2}$

10. The interquartile range excludes _____ of the values. (+2, -0.5)

a. 50%

b. 100%

c. 75%

d. 25%

11. If the third quartile of the following data set 7, 10, 7, 8, 9 is 9.5, then the value of quartile deviation is: (+2, -0.5)

a. 4.25

b. 1.25

c. 2.75

d. 7.00

12. Two random variables X and Y are said to be independent if: (+2, -0.5)

a. $E(XY) = E(X) E(Y)$

b. $E(XY) = E(X) + E(Y)$

c. $E(XY) = XE(Y)$

d. $E(XY) = YE(X)$

13. For the cumulative distribution function $(F(x) = \begin{cases} 0; x < -1 \\ \frac{1}{2}\{(x+1)\}^2; -1 \leq x < 0 \\ 1 - \frac{1}{2}\{(1-x)\}^2; 0 \leq x < 1 \\ 1; 1 \leq x < \infty \end{cases})$ (+2, -0.5)

the upper quartile point is

- a. $1 - \sqrt{0.5}$
- b. $1 + \sqrt{0.25}$
- c. $1 + \sqrt{0.5}$
- d. $1 - \sqrt{0.25}$

14. The mode of the given data set is 12. The sum of the frequencies on both sides of mode are 16. The skewness: (+2, -0.5)

- a. equals to ± 1
- b. equals to -1
- c. equals to 1
- d. does not exist

15. If the first, fifth and ninth decile of frequency distribution x_i are 3, 10, 16, respectively, then Kelly's coefficient of skewness is: (+2, -0.5)

- a. $-\frac{2}{13}$
- b. $-\frac{1}{13}$

c. $-\frac{3}{13}$

d. $-\frac{4}{13}$

16. The deseasonalised time-series data will have only trend (T), cyclical (C) (+2, -0.5) and irregular (I) components and is expressed as:

a. $(T.C.I) \times 100$

b. $\frac{T.I}{C} \times 100$

c. $\frac{T.C}{I} \times 100$

d. $\frac{C.I}{T} \times 100$

17. If $P(A) = 0.4$; $P(B|A) = 0.05$; $P(C|A) = 0.04$; $P(B|A \cap C) = 0.09$; $P(C|A \cap B) = 0.07$, (+2, -0.5)

then the probability of occurrence of all events equals to:

a. 0.014

b. 0.0014

c. 0.14

d. Given information is incomplete

18. Using the method of semi-average, secular trend is measured when: (+2, -0.5)

a. time series is based on annual values

b. times series comprises even number of values

- c. trend is linear
- d. trend is symmetric about the mean

19. If the population skewness of the observations 8, 6, 3, 1, 2, 5 is 0.233, then the population skewness of 16, 12, 6, 2, 4, 10 is: (+2, -0.5)

- a. 0.233
- b. 0.466
- c. 0.932
- d. 0.1165

20. X_1 and X_2 represent number of occurrences of event A and B that follow Poisson distribution with mean rate λ_1 and λ_2 . If Y_1 and Y_2 are inter-occurrence times of event A and B, then $\min(Y_1, Y_2)$ follows (+2, -0.5)

- a. Exponential distribution with mean rate $\min(\lambda_1, \lambda_2)$
- b. Poisson distribution with mean rate $|\lambda_1 - \lambda_2|$
- c. Exponential distribution with mean rate $(\lambda_1 + \lambda_2)$
- d. Poisson distribution with mean rate $(\lambda_1 + \lambda_2)$

21. If r_p be partial correlation computed on sample $T_{AB,C}$ computed from sample of size n , the test statistic for significance testing is (+2, -0.5)

a. $t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}}$

b. $t = \frac{r_p \sqrt{n+v}}{\sqrt{1+r_p^2}}$

c. $t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}}$

d. $t = \frac{r_p \sqrt{n+v}}{\sqrt{1-r_p^2}}$

22. The mean deviation and coefficient of mean deviation of 5 observations are 1.2 and 0.4. If the sum of the first four terms 10, then the fifth term is equal to: (+2, -0.5)

a. 4.5

b. 5.5

c. 4

d. 5

23. Identify, from the following, the moment used as a measure of skewness? (+2, -0.5)

a. First moment

b. Fourth moment

c. Third moment

d. Second moment

24. If the difference between the rank of the 4 observations are 2.5, 0.5, -1.5, -1.5, then Spearman's rank correlation coefficient equals to: (+2, -0.5)

a. -0.2

- b. 0.1
- c. -0.1
- d. 0.2

25. From standard pack of 52 cards, 3 cards are drawn at random without replacement. The probability of drawing a king, a queen and a jack in order is (+2, -0.5)

- a. $\frac{4}{16575}$
- b. $\frac{16}{16575}$
- c. $\frac{8}{16575}$
- d. $\frac{32}{16575}$

26. The grouped data for the observation are (+2, -0.5)

Class:	1-3	3-5	5-7
Frequency:	2	1	2

The population skewness

- a. is zero
- b. is negative
- c. is positive
- d. cannot be computed

27. For the ANOVA, which option is wrong? (+2, -0.5)

- a. Total sum of square = Total variation in data
- b. $F = \frac{\text{Mean sum of square within group}}{\text{Mean sum of square between group}}$
- c. Total degree of freedom = between degree of freedom + within degree of freedom
- d. Mean sum of square between group = $\frac{\text{sum of square between group}}{\text{degree of freedom between group}}$

28. For the distribution (+2, -0.5)

X:	-1	0	1
p(x):	0.3	0.5	0.2

the third factorial moment is

- a. -0.3
- b. 0.3
- c. 1.8
- d. -1.8

29. Which one is not non-probability sample method (+2, -0.5)

- a. Cluster sampling
- b. Snowball sampling
- c. Quota sampling

d. Purposive sampling

30. The median of the following observations 10, 11, 9, 12, 10, 10, 12, 10, 9, 11 is: (+2, -0.5)

a. 8

b. 10

c. 9

d. 11

31. For completely randomised design for k treatments and n observation, y_{ij} = response from the j^{th} unit receiving i^{th} treatment, $\sum_i \sum_j y_{ij} = y$ and $\sum_j^{n_i} y_{ij} = y_i$. Which of the following options is correct? (+2, -0.5)

a. $\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ represents sum of squares of due to error

b. $\sum_{i=1}^k (\bar{y}_i - \bar{y})^2$ represents sum of squares due to treatments

c. $\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ represents total sum of squares

d. $\sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$ represents sum of squares due to treatments

32. Which one is not basis of classification of data? (+2, -0.5)

a. Geological classification

b. Geographical classification

c. Qualitative classification

d. Temporal classification

33. If first, second, and third moment about origin are 1, 6, and 15 respectively, (+2, -0.5)
then Karl Pearson beta coefficient of skewness β_1 is

- a. $\frac{1}{625}$
- b. $\frac{1}{25}$
- c. $\frac{1}{125}$
- d. $\frac{1}{5}$

34. The value of k so that following is probability mass function (+2, -0.5)

X:	-2	-1	0	1	2	3
P(X = x):	2k	3k	4k	3k	2k	k

is

- a. $\frac{2}{15}$
- b. $\frac{1}{14}$
- c. $\frac{1}{16}$
- d. $\frac{1}{15}$

35. A physical instructor claims that the mean weight of students in school is (+2, -0.5)
greater than 82 kg with standard deviation 20. If a sample of size 81
students is selected with mean weight of 90. The test statistic equals to

- a. $z = 3.0$
- b. $z = 3.6$

c. $z = 2.4$

d. $z = 3.2$

36. For the data set

(+2, -0.5)

x:	1	2	3	4
y:	4	5	3	2

the regression coefficient b_{yx} (y on x) equals to:

a. -0.86

b. 0.8

c. 0.86

d. -0.8

37. Which of the following options is correct when data is classified on the basis of attributes? (+2, -0.5)

a. Qualitative classification

b. Geographical classification

c. Geological classification

d. Temporal classification

38. The sixth decile (D_6) of 5, 3, 2, 6, 8, 4 is:

(+2, -0.5)

a. 5.25

- b. 5.30
- c. 5.15
- d. 5.20

39. For the two variables X and Y, the following observations are tabulated. (+2, -0.5)

X:	3	4	4
Y:	10	10	9

The Spearman's correlation coefficient is:

- a. -0.125
- b. -0.120
- c. -0.100
- d. -0.110

40. The median of following observations 8, 9, 7, 10, 8, 8, 10, 8, 7, 9 is (+2, -0.5)

- a. 8.2
- b. 8.0
- c. 8.4
- d. 8.5

41. Which of the following options is INCORRECT? (+2, -0.5)

- a. For k treatments and N observations, the degree of freedom of total variation is $N - 1$
- b. For k treatments and N observations, the degree of freedom of the ratio of mean sum of square of treatments and residual mean sum is $N - 1$ and $k - 1$
- c. For k treatments and N observations, the degree of freedom of variation within groups is $N - k$
- d. For k treatments and N observations, the degree of freedom of variation between groups is $k - 1$

42. The standard error of the given data 15, 5, 12, 10, 20, 4 is (+2, -0.5)

- a. $\frac{\sqrt{45}}{3}$
- b. $\frac{\sqrt{43}}{3}$
- c. $\frac{\sqrt{46}}{3}$
- d. $\frac{\sqrt{44}}{3}$

43. Which option is incorrect for the component of time series? (+2, -0.5)

- a. Shifts in the level of a time series that cannot be explained are referred to as seasonal cycles
- b. A trend is a gradual upward or downward shift in the level of the series
- c. A sudden, temporary shift is referred as pulse
- d. A non-seasonal cycle is a repetitive, possibly unpredictable, pattern in the series values

44. If Z follows standard normal distribution with mean 0 and variance 1, then Z^2 follows: (+2, -0.5)

- a. Gamma distribution with $\alpha = 1$ and $\beta = 1$
- b. Normal distribution with mean 0 and variance 1
- c. Chebyshev distribution with degree of freedom 1
- d. Beta distribution with $\alpha = 1$ and $\beta = 1$

45. For the price-quantity chart (+2, -0.5)

Good	Base year (2000)		Current year (2020)	
	Quantity	Price (in \$)	Quantity	Price (in \$)
1	20	20	30	50
2	40	30	50	60
3	60	40	70	70

the Laspeyres price index for current year is

- a. 184
- b. 188
- c. 180
- d. 190

46. Which of the following statements is INCORRECT?

(+2, -0.5)

- a. Collecting primary data is not quite expensive both, in the terms of time and money.
- b. Primary data are original.
- c. Primary data are those that are collected for the first time.
- d. Primary data are more reliable and suitable.

47. If the first, second, and third moment about the origin are 2, 8, and 18 respectively, then third moment about mean is

(+2, -0.5)

- a. 12
- b. 14
- c. -14
- d. -12

48. For the following frequency distribution

(+2, -0.5)

Class:	3-5	5-7	7-9	9-11
Frequency:	1	4	2	1

the value of mode is:

- a. 6.25
- b. 6.20
- c. 6.00

d. 6.40

49. For frequency distribution presentation, which option is wrong? (+2, -0.5)

- a. In a histogram, a bar is created above each score (or class interval) so that the height of the bar corresponds to the frequency
 - b. Bar graph presents score categories that are measured from a nominal or an ordinal scale
 - c. The smooth curve emphasizes the fact that the distribution is not showing the exact frequency for each category.
 - d. In a polygon, an additional line is not drawn at each to bring the graph back to a zero frequency
-

50. If the population kurtosis of the observations 16,12,6,2,4,10 is 1.7414, then population kurtosis of the 8,6,3,1,2,5 is (+2, -0.5)

- a. 0.8707
 - b. 1.7414
 - c. 3.4828
 - d. 0.43535
-

51. The mean and median of the distribution are 10 and 12 respectively, then the mode equals to (+2, -0.5)

- a. 20
- b. 16

c. 14

d. 18

52. If mean and mode of the distribution is 32 and 21, then the distribution: (+2, -0.5)

a. is negatively skewed

b. is not skewed

c. cannot be determined

d. is positively skewed

53. The probability of getting a total of 7 on two dice thrown together is: (+2, -0.5)

a. $\frac{8}{36}$

b. $\frac{7}{36}$

c. $\frac{6}{36}$

d. $\frac{5}{36}$

54. Recession in industry is associated with the: (+2, -0.5)

a. trend

b. seasonal component

c. irregular component

d. cyclical component

55. The data sets of sizes 6 and 9 have standard deviation 3 and 4, respectively, and arithmetic means 4 and 4, respectively. The standard deviation of combined data set of size 15 is: (+2, -0.5)

a. $\sqrt{\frac{68}{5}}$

b. $\sqrt{\frac{72}{5}}$

c. $\sqrt{\frac{66}{5}}$

d. $\sqrt{\frac{70}{5}}$

56. If each observation in a data set for number of employees in different divisions is doubled then the coefficient of quartile deviation: (+2, -0.5)

a. is also doubled

b. remains same

c. is fourtimes of the original

d. is halved

57. If each observation is halved then the coefficient of quartile deviation (+2, -0.5)

a. is one-fourth of original

b. is also halved

c. is doubled

d. remains same

58. If the sum of lower and upper quartiles is 6 and quartile deviation is 1.5, then the value of coefficient of quartile deviation is (+2, -0.5)

- a. 0.6
- b. 0.5
- c. 0.7
- d. 0.4

59. If mean and median of the distribution are 12 and 21, then the distribution (+2, -0.5)

- a. is not skewed
- b. is negatively skewed
- c. can't be determined for its skewness
- d. is positively skewed

60. Fundamental principles of design of experiment are (+2, -0.5)

- (I) Randomization
- (II) Replication
- (III) Local control

Which option is correct?

- a. Only (I) and (III)
- b. All (I), (II), and (III)
- c. Only (I) and (II)

d. Only (II) and (III)

61. Which option is WRONG?

(+2, -0.5)

- a. Secondary data requires less time and money than primary data
 - b. Secondary data is less reliable and less suitable than primary data
 - c. Secondary data refer to those data that have already been collected by some other person.
 - d. Secondary data is original
-

62. If moment generating function of continuous random variable X is $\frac{\lambda}{\lambda - t}$ $t < \lambda$, then $E(X^3)$ equals to:

(+2, -0.5)

- a. $\frac{8}{\lambda^3}$
 - b. $\frac{6}{\lambda^3}$
 - c. $\frac{4}{\lambda^3}$
 - d. $\frac{2}{\lambda^3}$
-

63. Which of the following correctly completes the given statement?

(+2, -0.5)

Index number helps in:

- (I) determining the cost of living
- (II) fixing the dearness allowances
- (III) reflecting the real income

a. Only II and III

- b. Only I and II
- c. All I, II and III
- d. Only I and III

64. If A, B and C are arbitrary events, then $P(A \cap B \cap C)$ equals to: (+2, -0.5)

- a. $P(A) P(B|A) P(C|A \cap B)$
- b. $P(A) P(B) P(C)$
- c. $P(A) P(A|B) P(A \cap B|C)$
- d. $P(A) + P(B) + P(C)$

65. The standard deviation of Y is double of standard deviation of x. The correlation coefficient between X and Y is 0.5. (+2, -0.5)

The acute angle between lines of regression is

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

66. If moment generating function of discrete random variable X is $(q + pe^t)^n$, (+2, -0.5)
then $E(X^2)$ equal to

- a. $nq(np + q)$

- b. $np(p + nq)$
- c. $nq(p + nq)$
- d. $np(np + q)$

67. For an experiment we have the following data set: $n = 4$, $\sum x = a$, $\sum y = 10$, $\sum xy = 21$, $\sum x^2 = 30$, $\sum y^2 = 30$. If the correlation coefficient is -0.8 then the value of a is: (+2, -0.5)

- a. 7
- b. 10
- c. 8
- d. 9

68. Two data set of sizes 9 and 6 have standard deviation 3 and 4 respectively and arithmetic means 3 respectively. The standard deviation of combined data set of size 15 is (+2, -0.5)

- a. $\sqrt{\frac{177}{15}}$
- b. $\sqrt{\frac{176}{15}}$
- c. $\sqrt{\frac{178}{15}}$
- d. $\sqrt{\frac{175}{15}}$

69. To estimate the average work experience of MBA students at a management institute, five students are selected at random from each (+2, -0.5)

type of background, say commerce, science and engineering. This type of sampling is called:

- a. systematic sampling
- b. stratified sampling
- c. simple random sampling
- d. cluster sampling

70. The prices (in Rs.) for the commodity ABC, XYZ, MNO, and IJK in base year (2020) are 20, 18, 12, 24 and in current year (2022) are 25, 22, 15, 28 respectively. The value of price index by simple aggregative method is (+2, -0.5)

- a. 111.62
- b. 125.62
- c. 121.62
- d. 115.62

Your Personal Exams Guide

71. If the second and third moment about the origin are 8 and 18 and the third moment about mean is -14, then the first moment about the origin is: (+2, -0.5)

- a. 3
- b. 1
- c. 1.5
- d. 2

72. For the ANOVA table

(+2, -0.5)

Source of variations	Sum of squares	Degree of freedom
Between treatment	75	3
Error	48	16
Total	123	19

the F - statistics is

- a. 8.99
- b. 8.33
- c. 8.60
- d. 7.33

73. Which is not relative measures of skewness?

(+2, -0.5)

- a. $\frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$
- b. $\frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$
- c. $\frac{D_9 - 2D_5 + D_1}{D_9 - D_1}$
- d. mean - mode

74. In a 3 races, 2 genders and 5 in each treatment group for two-way ANOVA, the degree of freedom for source of variation due to interaction, error and total respective are

(+2, -0.5)

- a. (6,24,29)

b. (6,30,30)

c. (2,24,29)

d. (2,24,30)

75. The grouped data for the observation are as follows.

(+2, -0.5)

Class:	2-4	4-6	6-8
Frequency:	2	1	2

The population skewness:

a. is negative

b. Data is insufficient

c. is positive

d. is zero

76. The arithmetic mean of the following frequency distribution of number of members in family in the society

(+2, -0.5)

X:	3	5	6	8	9	10
Frequency:	1	4	2	1	3	2

is

a. 4.923

b. 6.923

c. 7.923

d. 5.923

77. The Pearson's correlation coefficient between following observation (+2, -0.5)

X:	1	2	3	4
Y:	3	4	2	1

is -0.8. If each observation of X is halved and of Y is doubled, then Pearson's correlation coefficient equals to

a. -0.82

b. -0.79

c. -0.80

d. -0.81

78. If the odds in favour of any random event A are 5 : 6, then the odds against the event are: (+2, -0.5)

a. 6 : 5

b. 5 : 11

c. 6 : 11

d. 11 : 6

79. The purchasing power of money is equal to: (+2, -0.5)

a. $\frac{1}{(\text{price index number})^2}$

b. $\frac{1}{\text{price index number}}$

c. $\sqrt{\text{price index number}}$

d. $\frac{1}{\sqrt{\text{price index number}}}$

80. The arithmetic mean of the following frequency distribution of number of accidents X on week working days is: (+2, -0.5)

X:	2	4	6	8	10	12
Frequency:	3	4	2	1	4	2

- a. 4.625
b. 6.625
c. 7.625
d. 5.625

81. The 95% confidence interval of average age of accidents in any city during last year for a sample of size 100 with mean age 34.25 from population of standard deviation 10 is (+2, -0.5)

- a. [32.29, 36.21]
b. [32.29, 36.58]
c. [32.605, 35.895]
d. [31.92, 36.58]

82. If r and R denote correlation and multiple correlation coefficient for the data set for X_1, X_2 and X_3 . Which option is correct? (+2, -0.5)

- a. $r_{12} = 0.24, r_{13} = 0.22, r_{23} = 0.23, R_{1.23} = 0.21$
- b. $r_{12} = 0.69, r_{13} = 0.22, r_{23} = 0.23, R_{1.23} = 0.69$
- c. $r_{12} = 0.21, r_{13} = 0.22, r_{23} = 0.23, R_{1.23} = 0.20$
- d. $r_{12} = 0.69, r_{13} = 0.22, r_{23} = 0.23, R_{1.23} = 0.21$

83. If X and Y represent waiting time and service time of customers is shopping mall, have joint density $f(x,y) = kx; 0 \leq y \leq x \leq 1$, then the value of k is (+2, -0.5)

- a. 1
- b. 3
- c. 4
- d. 2

84. The value of a and b so that the following is probability mass function (+2, -0.5)

$X:$	0	1	2
$P(X = x):$	$3a$	$3b$	$4b$

with mean 1.1, is:

- a. (0.2, 0.2)
- b. (0.1, 0.1)
- c. (0.1, 0.2)

d. (0.2, 0.1)

85. For the variables X and Y, we collect 4 observations with $\sum x = 10$, $\sum y = 14$, $\sum x^2 = 30$, $\sum y^2 = 54$, $\sum xy = 31$. The regression line y on x is (+2, -0.5)

a. $y = -0.8x - 5.5$

b. $y = 0.8x - 5.5$

c. $y = -0.8x + 5.5$

d. $y = 0.8x + 5.5$

86. For the random variable X with probability density function $f(x) = \frac{(x-3)^2}{5}$; $x = 3, 4, 5$, the variance of X is: (+2, -0.5)

a. $\frac{2}{25}$

b. $\frac{2}{5}$

c. $\frac{4}{5}$

d. $\frac{4}{25}$

87. If A and B are mutually exclusive, the general addition rule is: (+2, -0.5)

a. $P(A \cup B) = P(A) + P(B)$

b. $P(A + B) = P(A) + P(B)$

c. $P(A \cup B) = P(A) + P(B) + P(A \cap B)$

d. $P(A \cap B) = 0$

88. If random variable X follows binomial distribution with parameter n and p with mean 15 and variance 10, then the value of mode is (+2, -0.5)

- a. $\frac{47}{3}$
- b. $\frac{48}{3}$
- c. $\frac{49}{3}$
- d. $\frac{46}{3}$

89. If the mean and variance of a binomial distribution are 5 and 4, respectively, then the value of n is: (+2, -0.5)

- a. 25
- b. 10
- c. 15
- d. 20

90. If the first, second and third moment about origin are 2, 8 and 14, respectively, then Karl Pearson gamma coefficient of skewness γ_1 is: (+2, -0.5)

- a. -2.25
- b. 0.25
- c. -0.25
- d. -1.25

91. If A and B are mutually exclusive events such that $P(A) P(B) > 0$, then which option is correct? (+2, -0.5)

- a. A and B are independent
- b. A and B are not independent
- c. $A \subset B$
- d. $B \subset A$

92. For the ANOVA table (+2, -0.5)

Source of variations	Sum of squares	Degrees of freedom
Between treatment	45	3
Error	32	16
Total	99	19

the F - statistics is:

- a. 7.4
- b. 7.3
- c. 7.5
- d. 7.2

93. A random sample of size 225 is drawn from the population of mean μ and standard deviation σ . The sample mean follows the distribution with mean 100 and standard distribution $4/3$. The value of μ and σ are: (+2, -0.5)

a. (100, 18)

b. (100, 15)

c. (100, 24)

d. (100, 20)

94. Which of the following is an example of using a sample to make inference about a population? (+2, -0.5)

a. Assembly elections

b. Statistics of a cricket player in one-day matches

c. Census

d. Pre-election poll by media

95. The seventh decile (D_7) of data set 4, 3, 7, 10, 9, 1 is (+2, -0.5)

a. 8.9

b. 8.8

c. 8.6

d. 8.7

96. For the ANOVA, which of the following options is INCORRECT? (+2, -0.5)

a. Null hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_n$

b. F - ratio belongs to $[-\infty, \infty]$

- c. Alternative hypothesis H_1 : At least one population mean is different from one another
- d. Variance are compared in F ratio to determine mean differences are significantly bigger than chance

97. For the variables X, Y and Z, $r_{xy} = 0.80$, $r_{xz} = 0.64$, and $r_{yz} = 0.79$, the square of multiple correlation coefficient R^2_{xyz} is: (+2, -0.5)

- a. 0.43
- b. 0.33
- c. 0.53
- d. 0.64

98. Which of the following is the most relevant for deriving a point estimate? (+2, -0.5)

- a. Population size
- b. Sample size
- c. Variability in the populaiton
- d. Confidence desired

99. If price-quantity are related for base year (0) and current year (1) are (+2, -0.5)

$\sum p_0 q_0 = 260$, $\sum p_1 q_0 = 395$, $\sum p_0 q_1 = 264$, $\sum p_1 q_1 = 422$, then Marshall Edgeworth price index equals to

- a. 145.92

- b. 155.92
- c. 165.92
- d. 175.92

100. For the distribution with unknown θ

(+2, -0.5)

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

We set the testing of hypothesis $H_0: \theta = 1$ vs $H_1: \theta = 2$. When the critical region $X \geq 0.4$, the value of probability of type-II error is:

- a. 0.30
- b. 0.24
- c. 0.25
- d. 0.20

Your Personal Exams Guide

Answers

1. Answer: b

Explanation:

The correct answer is **56**. Find the third 3 - year by calculating the moving average for each set of three consecutive years.

★ Key Points

Simple Moving Average:

- In time series moving average is a method of calculating the average of a certain number of items which move through the time series.
- The movement of the items happens in the series such that **the top data of the previous averaged group drops and adds into the next successive average.**
- It is calculated by adding all the **the data points during a specific period and dividing the sum by the number of time periods.**
- To find out the grand average, one must calculate the percentage of the derived data.
- The investors use this method to track the directions in the trend series.

Calculation:

The first three year (1 - 3), the average is :

$$MA = \frac{25+95+55}{3}$$

$$MA = 55$$

The second three year (2 - 4), the average is :

$$MA = \frac{95+55+94}{3}$$

$$MA = 84$$

The third three year (3 - 5), the average is :

$$MA = \frac{55+94+25}{3}$$

$$MA = 56$$

Hence, the 3rd year simple moving average is 56.

2. Answer: a

Explanation:

The correct answer is **3.833**.

★ Key Points

Frequency Distribution:

- A tabulation or chart of frequencies of the given data that shows the actual number of observations falling in each range.
- When the data in the chart is represented in terms of percentage of occurrence it is known as relative frequency distribution.

Example:

- Height of ten students of a class are given (in cm): 139, 145, 150, 145, 136, 150, 152, 144, 138, 138
- Frequency distribution of the same data would look like:

Height	Frequency
139	1
145	2
150	2
136	1
152	1
144	1
138	2

★ Additional Information

Types of frequency distribution:

- Grouped frequency distribution: Often to avoid any human error of calculating a large set of data it is organized into more manageable groups.
- Simple or ungrouped frequency distribution: The frequency distribution explained above is known as ungrouped or simple frequency distribution as the data is not categorized into groups.
- Cumulative frequency distribution: Cumulative frequencies are calculated for each group of data that are arranged in a frequency table.
- Relative frequency distribution: A relative frequency distribution shows the proportion of the total number of observations associated with each value or class of values and is related to a probability distribution, which is extensively used in statistics.
- Relative cumulative frequency distribution: The distribution shows relative cumulative frequency that can be found by dividing the frequency of each interval by the total number of observations.

Calculation:

The data that has been organized into groups is grouped data. To calculate the mode for the grouped data use the below formula:

$$Mode = L + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} * h$$

where,

L = lower boundary of the mode class interval

f_0 = frequency of the class interval before the mode class

f_1 = frequency of the mode class

f_2 = frequency of the class interval after the mode class

h = size of the class interval (width)

⇒ The class interval 3.5 – 4.5 has the highest frequency is of 4

Substituting the values, we get,

$$Mode = L + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} * h$$

$$Mode = 3.5 + \frac{(4 - 3)}{2 * 4 - 3 - 2} * 1$$

$$Mode = 3.833$$

Hence, the mode is equal to 3.833

3. Answer: a Your Personal Exams Guide

Explanation:

The correct answer is **21**.

★ Key Points

Mode:

- The mode is the measure of central tendency that identifies the category or score that occurs the most frequently within the distribution of data.
- In other words, mode represents the data that occurs the highest number of times.

Types of Mode:

Four types of mode are seen. These are –

- **Unimodal:** The data has one mode. For example, the Mode of data set {14, 15, 16, 17, 20, 15, 31, 42, 15, 69, 15, 18, 15, 19} is 15 as there is only one value repeating itself. Hence, it is a unimodal data set.
- **Bimodal:** The dataset contains exactly two modes. This implies that exactly two data values are having the highest frequencies in the given data. For example, the mode of data set {1, 0, 20, 8, 13, 3, 13, 14, 15, 13, 17, 17, 19, 13, 7, 19} is 13 and 17 because both 13 and 17 are repeating themselves in the given set. Hence, it is a bimodal data set.
- **Trimodal:** The data set having exactly three modes. For instance, the Mode of data set {2, 2, 2, 3, 5, 5, 1, 0, 9, 4, 4, 5, 2, 6, 5, 4, 7, 5, 8} is 2, 4, and 5 because all the three values are repeating in the given set. Hence, it is a trimodal data set.
- **Multimodal:** A set of data with four or more than four modes is known as a Multimodal Mode. For example, the mode of data set {100, 80, 80, 95, 95, 80, 99, 102, 66, 85, 90, 110, 100, 90, 90, 100, 95} is 80, 90, 95, and 100 because all the four values are repeated multiple times in the given set. Hence, it is a multimodal data set.

★ Additional Information

- **Mean:** The average of all the values in the given data set.
- **Median:** The middle most value of the given data set.
- **Relationship between mean, median and mode:**
- The relationship between mean median and mode can be expressed by using Karl Pearson's Formula as:
 - $(\text{Mean} - \text{Median}) = \frac{1}{3} (\text{Mean} - \text{Mode})$
 - $3 (\text{Mean} - \text{Median}) = (\text{Mean} - \text{Mode})$
 - $\text{Mode} = \text{Mean} - 3(\text{Mean} - \text{Median})$
 - $\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$
- Mean, median and mode are the **three measures** of central tendency in Statistics.

Calculation:

The mode can be calculated using the below formula:

$$\text{Mean} - \text{Mode} = 3 * (\text{Mean} - \text{Mode})$$

$$12 - \text{Mode} = 3 * (12 - 15)$$

$$\text{Mode} = 3 * 12 - 3 * 15$$

$$\text{Mode} = 21$$

Hence, mode is 21.

4. Answer: b

Explanation:

The option 2) represents the wrong relationship between the mean sum of squares. The correct relationship is:

Mean sum of squares of total = Mean sum of square treatment + Mean sum of square of error + residual mean square.

Thus, the right answer is **option 2)**

Note -

1. The option 1) is correct because the degrees of freedom of variation between group is $k - 1$
2. The option 3) is correct because the it represents the formula
3. The option 4) is correct because the degrees of freedom within group is $N - k$

5. Answer: c

Explanation:

X is degrees of freedom for the total source of variation. To find the degrees of freedom use the formula,

$$N - 1 = (\text{number of levels in Factor A}) * (\text{number of levels of Factor B}) - 1$$

$$N = 3 * 3 - 1$$

$$N = 8$$

Thus, the value of **8**

The F_A is calculated using the below formula,

$$F_a = \frac{MSS}{MSW}$$

$$F_a = \frac{147}{3}$$

$$F_a = 49$$

The F_B is calculated using the below formula,

$$F_b = \frac{MS_b}{MS_w}$$

$$F_b = \frac{3}{3}$$

$$F_b = 1$$

The correct option is **option 3) (8, 49, 1)**

6. Answer: a

Explanation:

The correct answer is **1.0**.

★ Key Points

- The **quartile deviation**, also known as semi-interquartile range (IQR), is a measure of dispersion in a data set and is calculated as the difference between the third quartile (Q3) and the first quartile (Q1).
 - It is a robust statistic, meaning it is not affected by outliers in the data set as much as the range or standard deviation.
 - It is a useful summary statistic when comparing the spread of data sets with different sample sizes, as it is not influenced by sample size in the same way

that the standard deviation is.

★ Additional Information

- Given the first quartile of the data set is 7.5, we can calculate the IQR as follows:
 - $IQR = Q3 - Q1$ Since the data set has 5 observations, Q3 can be calculated as the median of the upper half of the data set, which is $(9 + 10) / 2 = 9.5$.
Therefore, $IQR = 9.5 - 7.5 = 2.0$
 - Finally, the quartile deviation can be calculated as $IQR / 2 = 2.0 / 2 = 1.0$
 - So, the value of quartile deviation is 1.0.
- The **quartile deviation** (also known as semi-interquartile range) is a measure of the dispersion of a data set and is defined as the difference between the first quartile (Q1) and the third quartile (Q3) divided by 2.
- In this case, the first quartile (Q1) is 7.5. To find the third quartile (Q3), we first sort the data set in ascending order: 7, 8, 8, 9, 10.
- The third quartile (Q3) is the median of the upper half of the data set, so $Q3 = 9$. The quartile deviation is then calculated as $(9 - 7.5) / 2 = 1.0$.

7. Answer: a

Explanation:

The correct answer is **123.33**.

★ Key Points

- Laspeyres price index:**
 - a price index used in statistics for measuring the price development of a certain basket of goods and services consumed in the base period.
 - Laspeyres Price Index = $\frac{\sum(P_1Q_0)}{\sum(P_0Q_0)} \times 100$

According to the given data, let's assume that base period = 2019 and current period = 2020

Quantity at base period = quantity at current period = Q

Commodity	P_0 (Price at 2019)	Q_0 (Quantity at 2019)	P_1 (Price at 2020)	Q_1 (Quantity at 2020)	P_0Q_0	P_1Q_1
Tea	Rs. 100	Q	Rs. 125	Q	100Q	125Q
Sugar	Rs. 50	Q	Rs. 60	Q	50Q	60Q
Total	-	-	-	-	150Q	185Q

According to Laspeyres Price Index

$$= \frac{185Q}{150Q} \times 100$$

$$= 123.33$$

Hence, the correct answer is 123.33

8. Answer: a

Explanation:

The correct answer is 4.8

★ Key Points

- The **third factorial moment** is a measure of the skewness of the probability distribution of a random variable. It measures how much the distribution deviates from a perfectly symmetric distribution.
- The formula for the third factorial moment involves the expected value of the product of three powers of the random variable, and it can be derived using the properties of the expected value operator and combinatorics.

- In general, if the third factorial moment is positive, then the distribution is said to be skewed to the right, which means that it has a longer tail on the right side of the mean.
- If the third factorial moment is negative, then the distribution is skewed to the left, which means that it has a longer tail on the left side of the mean.
- If the third factorial moment is zero, then the distribution is symmetric.

★ Additional Information

- To find the third factorial moment of X , we can use the following formula:
 - $E[X(X-1)(X-2)] = \sum (x^3 - 3x^2 + 2x)p(x)$
- where $E[X]$ is the expected value of X , and $p(x)$ is the probability of X taking the value x .
- Using the values given in the problem, we can calculate the third factorial moment as:
 - $E[X(X-1)(X-2)] = (0^3 - 3(0)^2 + 2(0))(0.5) + (2^3 - 3(2)^2 + 2(2))(0.3) + (4^3 - 3(4)^2 + 2(4))(0.2) = 0 + 1.2 + 3.6 = 4.8$
- Therefore, the third factorial moment of X is 4.8.

★ Important Points

- It's important to note that the third factorial moment is just one measure of skewness, and there are other measures, such as the coefficient of skewness, that can be used to quantify skewness as well.
- Finally, it's worth noting that the third factorial moment is only defined for random variables that have a third moment. Some distributions, such as the Cauchy distribution, do not have a defined third moment, and therefore the third factorial moment is not applicable to them.

9. Answer: b

Explanation:

The correct answer is $2/5$

★ Key Points

- **Empirical probability** is based on actual observations or experiments.
 - It is calculated by dividing the number of favorable outcomes by the total number of outcomes.
- The more observations or experiments are conducted, the more reliable the empirical probability becomes.
- **Probability**
 - It is a measure of the likelihood or chance of an event occurring.
 - Its values range from 0 (impossible) to 1 (certain).
 - It can be expressed as a fraction, decimal, or percentage.

★ Additional Information

- The empirical probability of getting a number greater than 4 when a dice is rolled can be calculated by adding up the number of times the dice lands on a 5 or 6, and dividing by the total number of rolls:
 - Empirical probability = $(\text{number of 5s} + \text{number of 6s}) / \text{total number of rolls} = (5 + 1) / 15 = 6 / 15 = 2 / 5$
 - Therefore, the correct answer is $2/5$.
- The theoretical probability of rolling a number greater than 4 on a fair dice is $2/6$ or $1/3$, since there are two favorable outcomes (rolling a 5 or 6) out of six possible outcomes (rolling a number from 1 to 6).
- Empirical probability can be used to estimate theoretical probability, but the two may not always be equal due to chance or sampling error.

10. Answer: a

Explanation:

The correct answer is **50%**.

★ Key Points

- The **interquartile range** is a useful measure of variability that is less affected by outliers than the range, which is the difference between the maximum and minimum values in a dataset.

- The IQR is sometimes used to identify outliers, which are values that fall more than 1.5 times the IQR below the first quartile or above the third quartile.
- The IQR can be used to define the box in a box plot, which is a graphical representation of the distribution of a dataset.
- The whiskers of the box plot extend to the minimum and maximum values within 1.5 times the IQR of the first and third quartiles, respectively.
- The IQR can be calculated using the formula $IQR = Q3 - Q1$, where $Q1$ and $Q3$ are the first and third quartiles, respectively.

★ Additional Information

- The **interquartile range**
 - It is a robust measure of variability that is useful for datasets with skewed distributions or outliers.
 - It is a standard measure of variability that is taught in many introductory statistics courses and is commonly used in research and data analysis.
 - It is a measure of variability that describes the spread of a dataset by dividing it into quarters.
 - Specifically, the IQR is the difference between the upper quartile ($Q3$) and the lower quartile ($Q1$).
- $Q1$ represents the 25th percentile, which is the value below which 25% of the data falls, while $Q3$ represents the 75th percentile, which is the value below which 75% of the data falls.
 - Therefore, the IQR includes the middle 50% of the data, and excludes both the lowest 25% and the highest 25% of the values.
- In summary, the correct answer is that the interquartile range excludes 50% of the values, with the remaining 50% falling within the range.

11. Answer: b

Explanation:

The quartile deviation formula is given by :

$$Q.D = \frac{Q_3 - Q_1}{2}$$

where, Q3 is given as 9.5. In order to calculate the 1st quartile use following steps:

- Arrange the data in ascending order that is, 7,7,8,9,10
- The upper half of the data is 7,7 where the 25% of data falls
- Thus, the Q1 is:

$$Q_1 = \frac{7+7}{2}$$

$$Q_1 = 7$$

Substituting the resultant value in quartile deviation formula we get,

$$Q.D = \frac{9.5-7}{2}$$

$$Q.D = 1.25$$

The correct answer is option 2) **1.25**

12. Answer: a

Explanation:

Two random variables X and Y are independent if the covariance between X and Y is zero, which is expressed mathematically as

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = 0.$$

This is equivalent to the expectation of their product being equal to the product of their individual expectations, or $E(XY) = E(X) E(Y)$.

13. Answer: a

Explanation:

The correct answer is $1 - \sqrt{0.5}$

★ Key Points

- The **cumulative distribution function (CDF)** is a function that describes the probability distribution of a random variable.
 - In this case, the CDF is defined by the function $F(x)$, which has different values for different ranges of x .
- The **upper quartile point** of a data set is the value that corresponds to the 75th percentile.
- To calculate the upper quartile point using the CDF, we need to find the value of x for which $F(x)$ is equal to 0.75.
 - To understand this, we can look at the formula for $F(x)$ for the range $0 \leq x < 1$.
 - We can see that this part of the function is a quadratic function with a maximum value of 1 at $x = 0$.
 - In this case, we can see that $F(x)$ is equal to 0.75 when x is equal to $1 - \sqrt{1/2}$.

★ **Additional Information**

- The upper quartile corresponds to the value of x at which $F(x)$ is equal to 0.75, or the point where 75% of the data is below this value.
- Since the maximum value of the quadratic function is at $x = 0$, we can calculate the upper quartile by finding the root of the quadratic equation $1 - (1 - x)^2/2 = 0.75$, which simplifies to $(1 - x)^2 = 0.5$.
- Solving for x , we get $x = 1 - \sqrt{1/2}$, which is the upper quartile point.
- Overall, the CDF can be a useful tool for analyzing and modeling probability distributions, and the upper quartile point is a useful measure of variability that can provide insight into the spread and shape of a data set.

14. Answer: d

Explanation:

The information provided is not enough to calculate skewness. Skewness requires information on the distribution's mean and standard deviation, which are not provided.

Thus, the correct answer is **does not exist**.

15. Answer: b

Explanation:

The calculation for Kelly's coefficient of skewness can be performed as follows:

$$D_9 = 16(9thdecile)$$

$$D_5 = 10(5thdecile)$$

$$D_1 = 3(1stdecile)$$

$$SK_k = \frac{(D_9 - 2D_5 + D_1)}{(D_9 - D_1)}$$

$$SK_k = \frac{(16 - 2 \cdot 10 + 3)}{(16 - 3)}$$

$$SK_k = \frac{(-1)}{13}$$

Thus, the correct answer is $\frac{-1}{3}$

16. Answer: a

Explanation:

The de-seasonalized time-series data is a transformation of original time-series data to remove the impact of seasonality. The resulting time-series will only

have the trend (T), cyclical (C), and irregular (I) components. The equation $T.C.I \times 100$ represents the decomposition of the time-series

data into these three components, where the multiplication by 100 is used to express the values as percentages.

17. Answer: b

Explanation:

Consider,

$$P(A \cap B \cap C) = P(A) * P(C|A) * P(C|A \cap B)$$

$$P(A \cap B \cap C) = 0.4 * 0.04 * 0.07$$

$$P(A \cap B \cap C) = 0.0014$$

The LHS and RHS are same. Therefore, the correct answer is **0.0014**

18. Answer: c**Explanation:**

The method of semi-average is used to measure linear secular trend. It involves dividing the time period into two equal parts, calculating the

average value of the variable for each part, and then subtracting one average from the other. The difference between the two averages

represents the linear secular trend, as it measures the rate of change in the variable over time.

19. Answer: a**Explanation:**

Skewness of a data set can be calculated using the formula:

$$S_k = \frac{\sum (x_i - \bar{x})^3}{\sigma^3}$$

where deviations from the mean are calculated as (observation - mean). This formula takes into account the asymmetry of

the data set by considering the deviations raised to the power of 3, and then divided by the cube of the standard deviation to normalize the result.

Multiplying every value in the data set by a constant factor will result in the same skewness, as the mean, standard deviation, and

deviations from the mean will all be multiplied by the same constant factor.

20. Answer: c

Explanation:

The inter-occurrence time between two events follows an exponential distribution with a mean rate equal

to the inverse of the mean number of occurrences of that event.

For events A and B, the inter-occurrence time Y_1 and Y_2 follow exponential distributions with mean rates $1/\lambda_1$ and $1/\lambda_2$ respectively.

The minimum of these two times, $\min(Y_1, Y_2)$, is the time until the next occurrence of either event A or event B. Since both events

are occurring simultaneously and independently, the total rate of occurrences for both events is the sum of the individual rates: $\lambda_1 + \lambda_2$.

The mean rate for the exponential distribution of the minimum inter-occurrence time is therefore $(\lambda_1 + \lambda_2)$.

21. Answer: c

Explanation:

The correct answer is $t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}}$

★ Key Points

- **Partial correlation** is a statistical measure that represents the strength of the relationship between two variables, after controlling for the effect of one or more additional variables.
- The **test statistic** for significance testing of partial correlation, r_p , is used to determine if the partial correlation between two variables is significant and non-zero.
- The formula for the test statistic for partial correlation is given as: $t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}}$, where n is the sample size and v is the number of control variables. The numerator of the formula represents the strength of the partial correlation, and the denominator represents the standard error of the estimate.
- The value of t obtained from this formula can then be compared to the critical value of t from a t -distribution table with $(n - v - 2)$ degrees of freedom, to determine the significance of the partial correlation. If the calculated t -value is greater than the critical value, the partial correlation is considered significant and non-zero.

★ Additional Information

- **Partial correlation** is used to study the relationship between two variables after controlling for the effects of other variables.
- The test statistic for partial correlation is a ratio of the strength of the partial correlation to the standard error of the estimate.
- The calculated t -value is compared to a critical value from a t -distribution table to determine the significance of the partial correlation.
- A significant and non-zero partial correlation indicates that there is a relationship between the two variables being studied, after controlling for the effects of the control variables.

22. Answer: d

Explanation:

The correct answer is 5.

★ Key Points

- Let's denote the observations as X_1, X_2, X_3, X_4 , and X_5 . The mean of the observations can be calculated as:
- $\text{mean} = (X_1 + X_2 + X_3 + X_4 + X_5)/5$
- The **mean deviation** for each observation is calculated as:
 - Mean Deviation (X_1) = $X_1 - \text{mean}$
 - Mean Deviation (X_2) = $X_2 - \text{mean}$
 - Mean Deviation (X_3) = $X_3 - \text{mean}$
 - Mean Deviation (X_4) = $X_4 - \text{mean}$
 - Mean Deviation (X_5) = $X_5 - \text{mean}$
- The sum of the mean deviations of all observations should be zero as every deviation above the mean should have a corresponding deviation below the mean. This can be expressed as:
 - **Sum of Mean Deviations** = Mean Deviation (X_1) + Mean Deviation (X_2) + Mean Deviation (X_3) + Mean Deviation (X_4) + Mean Deviation (X_5) = 0
- The coefficient of mean deviation is calculated as the ratio of the mean deviation to the mean and is expressed as a percentage. In this case, it is given as 0.4. This can be expressed as:
 - **Coefficient of Mean Deviation** = $\text{Mean Deviation} / \text{mean} * 100$

★ Additional Information

- Given that the mean deviation and coefficient of mean deviation of 5 observations are 1.2 and 0.4, we can use the following formula to calculate the sum of the observations:
 - Sum of observations = (Number of observations) * (Mean)
- Since the coefficient of mean deviation is 0.4, it follows that:
 - Mean deviation / Mean = 0.4 Mean deviation = $0.4 * \text{Mean}$
- Since the mean deviation is 1.2, it follows that:
 - $0.4 * \text{Mean} = 1.2$
 - $\text{Mean} = 1.2 / 0.4 = 3$
- Therefore, the sum of the observations is:
 - Sum of observations = $5 * 3 = 15$
- Since the sum of the first four terms is 10, the fifth term must be equal to:

- Fifth term = Sum of observations – Sum of first four terms
Fifth term = $15 - 10 = 5$
- In statistics, the mean deviation and coefficient of mean deviation are measures of dispersion in a data set.
- The mean deviation is the average of the absolute deviations from the mean, while the coefficient of mean deviation is the ratio of the mean deviation to the mean.
- The fifth term in the set of 5 observations is equal to 5.
- The mean deviation and coefficient of mean deviation provide valuable information about the dispersion in a data set, and can be used to calculate the mean, as well as individual terms in the data set.

23. Answer: c

Explanation:

The correct answer is Third moment

★ Key Points

- The **first moment** is called the **MEAN** which describes the center of the distribution.
- The **second moment** is the **VARIANCE** which describes the spread of the observations around the center.
- The **third moment** describes other aspects of a distribution such as how the distribution is skewed from its mean or peaked.
- **Skewness** is a measure of the asymmetry of a probability distribution about its mean.
- The third moment is often used as a measure of skewness, as it indicates the degree of asymmetry of a distribution.
- A **positive third moment** indicates that the distribution is skewed to the right, while a negative third moment indicates that the distribution is skewed to the left.

★ Additional Information

- Skewness is a measure of the asymmetry of a distribution
- Third moment is used as a measure of skewness
- **Positive third moment** indicates a right-skewed distribution
- **Negative third moment** indicates a left-skewed distribution
- Other measures of skewness include the mean deviation and the coefficient of mean deviation.
- The third moment should be used with caution as it can be sensitive to outliers, and may not always accurately reflect the shape of a distribution.
 - It is important to consider the overall shape of a distribution and use multiple measures of skewness when analyzing data.

24. Answer: c

Explanation:

The correct answer is **-0.1**.

★ Key Points

- **Spearman's rank correlation coefficient** is used to measure the association between two variables when the data are ordinal or ranked.
- The formula for calculating the Spearman's rank correlation coefficient (ρ) is given by:
 - $\rho = 1 - (6 * \text{sum of differences}^2) / (n^3 - n)$, where n is the sample size and sum of differences is the sum of squared differences between the ranks of the two variables.
- In this case, given that the difference between the rank of the 4 observations are 2.5, 0.5, -1.5, -1.5, the sum of squared differences would be $2.5^2 + 0.5^2 + (-1.5)^2 + (-1.5)^2 = 12.5$.
- Plugging in $n = 4$, we have: $\rho = 1 - (6 * 12.5) / (4^3 - 4) = 1 - 37.5 / 8 = -0.1$.
- The Spearman's rank correlation coefficient measures the association between two variables when the data are ordinal or ranked.
- The formula for calculating the Spearman's rank correlation coefficient involves finding the sum of squared differences between the ranks of the two variables and using that value in the formula.

★ Additional Information

- A **positive value of the Spearman's rank correlation coefficient** indicates a positive association between the variables, a negative value indicates a negative association, and a value of zero indicates no association.
- Spearman's rank correlation coefficient is often used when Pearson's correlation coefficient cannot be applied, for example, when the data are not normally distributed or are not continuous.
- Spearman's rank correlation coefficient ranges from -1 to 1 , with -1 indicating a perfect negative association, 1 indicating a perfect positive association, and 0 indicating no association.
 - It is important to choose the appropriate correlation coefficient based on the nature of the data and the research question being studied.
- The interpretation of the correlation coefficient and its statistical significance depends on the sample size and the strength of the association between the variables.

25. Answer: c

Explanation:

The correct answer is $\frac{8}{16575}$

★ Key Points

- The **probability** of drawing a king, a queen and a jack in order is the product of the probabilities of drawing each card.
- There are 4 kings, 4 queens, and 4 jacks in a standard deck of 52 cards, so the probability of drawing any one of them is $\frac{4}{52}$.
 - The first card can be any of the 3 cards in order. The probability of drawing a king first is $\frac{4}{52}$.
 - The second card can be either a queen or a jack, but it cannot be the same as the first card.
- There are 3 queens or jacks left, and 51 cards in total, so the probability of drawing the second card is $\frac{3}{51}$.

- The same logic applies for the third card, and the probability of drawing a queen or a jack is $\frac{2}{50}$.
 - Therefore, the probability of drawing a king, a queen, and a jack in order is $\frac{4}{52} * \frac{3}{51} * \frac{2}{50} = \frac{8}{16575}$ }

★ Additional Information

- The number of each card in the deck affects the probability of drawing them.
- The order of drawing the cards affects the final probability.
- This problem assumes that the cards are drawn without replacement, which means that once a card is drawn, it is not returned to the deck.
- The problem calculates the probability of drawing the cards in a specific order, but there are other ways to calculate the probability of drawing a specific combination of cards, regardless of order.

26. Answer: a

Explanation:

The correct answer is **zero**.

★ Key Points

- **Population skewness** measures the asymmetry of a population distribution.
- A value of zero indicates a symmetric distribution.
- The **formula for population skewness** is given as: $\text{Skewness} = 3 * (\text{Mean} - \text{Median}) / \text{Standard Deviation}$
- **Skewness** is one of the measures of shape and central tendency.
- Other measures such as kurtosis and standard deviation are also important in understanding the overall distribution of data.

★ Additional Information

- The population skewness measures the asymmetry of a population distribution, with a value of zero indicating a symmetric distribution. The formula for population skewness is given as:

- $Skewness = 3 * (Mean - Median) / StandardDeviation$
- In a grouped data set, the standard deviation cannot be calculated, hence the skewness is only estimated.
 - However, in this case, the observation of equal frequencies in two of the classes suggests that the distribution is symmetric and the skewness is close to zero.
- It's important to note that skewness is one of the measures of shape and central tendency and can help us understand the distribution of data
 - however it's not the only one and other measures such as kurtosis and standard deviation are also important in understanding the overall distribution of data.

27. Answer: b

Explanation:

The correct answer is $F = \frac{\text{Mean sum of square within group}}{\text{Mean sum of square between group}}$

★ Key Points

- **ANOVA** is used to test the equality of means among more than two groups
- The **F-statistic** is the ratio of the mean sum of squares between groups to the mean sum of squares within groups
- A larger F-statistic indicates a significant difference in the means among the groups and supports rejection of the null hypothesis
- The **critical value** of the F-statistic is determined based on the degrees of freedom and significance level.
- In ANOVA, the test statistic F is used to test the null hypothesis that all population means are equal.
 - F is calculated as the ratio of the mean sum of squares between groups (MSB) to the mean sum of squares within groups (MSW).

★ Additional Information

- The formula for F is given by:
 - $F = MSB / MSW$

- MSB is the average variance among the group means, and MSW is the average variance within the groups.
- If the null hypothesis is true, then the group means should be similar and the variance within the groups should be larger than the variance between the groups. As a result, the F-statistic should be close to 1.
- However, if the null hypothesis is false, and the group means are significantly different, then MSB will be larger than MSW, resulting in a larger F-statistic.
- In such cases, the F-statistic is used to determine the significance of the results, and the null hypothesis is rejected if the F-statistic is larger than the critical value.

28. Answer: d

Explanation:

The correct answer is **-1.8**

★ Key Points

- **r^{th} factorial moment about origin (μ'_r)**
 - In probability theory of Statistics, the r^{th} factorial moment is a mathematical quantity defined as the expectation of the falling factorial (here, r^{th}) of a random variable.
 - Let, X assumes the values $x_1, x_2, x_3, \dots, x_n$

$X: = X_i$	$p(x) = P_i$	$P_i X_i$	$X_i^2 P_i$	$X_i^3 P_i$
-1	0.3	-0.3	0.3	-0.3
0	0.5	0	0	0
1	0.2	0.2	0.2	0.2
Total :	-	-0.1	0.5	0.1

- third factorial moment of the given series
 $= -0.1 - 3 \cdot 0.5 + 2 \cdot (-0.1)$
 $= -0.1 - 1.5 - 0.2$
 $= -1.8$

Hence, the correct answer is **(-1.8)**

Your Personal Exams Guide

29. Answer: a

Explanation:

The correct answer is **cluster sampling**.

★ Key Points

- **Non-probability sampling methods** are often used when the population is difficult to define or when the researcher has a specific purpose for selecting a certain group of individuals.
- **Probability sampling methods** are considered to be more representative of the population and are therefore often considered to be more scientifically rigorous

than non-probability sampling methods.

- **Cluster sampling**
 - It is a probability sampling method, not a non-probability sampling method.
 - In cluster sampling, the population is divided into smaller groups called clusters, and a random sample of these clusters is selected.
 - The individuals within the selected clusters are then surveyed to gather data.
- Cluster sampling is a probability sampling method
- Snowball sampling, quota sampling, and purposive sampling are non-probability sampling methods
- Non-probability sampling methods do not involve random selection of individuals, while probability sampling methods do.

★ Additional Information

- **Snowball sampling** is a non-probability sampling method in which individuals are selected based on referrals from other individuals.
- **Quota sampling** is a type of non-probability sampling method in which individuals are selected based on specific characteristics such as age, gender, or race.
 - The sample is created by dividing the population into subgroups based on these characteristics, and then selecting a set number of individuals from each subgroup.
- **Purposive sampling** is a type of non-probability sampling method in which the researcher selects individuals based on specific criteria that are deemed important for the study.
 - This method is often used when the population is small or when the researcher has a specific purpose for selecting a certain group of individuals.
- The choice between probability and non-probability sampling methods will depend on the research question, the resources available, and the specific needs of the study.
- It is important for researchers to be transparent about their sampling methods and to discuss the limitations of their sample when presenting their results.

30. Answer: b

Explanation:

The correct answer is **10**.

★ Key Points

- The **median** of a set of observations is the value that separates the upper half from the lower half of the data.
 - To find the median of a set of observations, the data must be arranged in numerical order.
- In this case, the observations are: 9, 9, 10, 10, 10, 10, 11, 11, 12, 12.
- Since there are an even number of observations, the median is the average of the two middle values, which in this case are 10 and 10.
- The median is therefore 10.
- The median is the value that separates the upper half from the lower half of the data
- To find the median, the data must be arranged in numerical order
- If the number of observations is odd, the median is the **middle value**
- If the number of observations is even, the median is the average of the two middle values.

★ Additional Information

- The median is one of several measures of central tendency, which describe where the data is centered.
 - Other measures of central tendency include the mean (average) and the mode (most common value).
- The median is considered to be a more robust measure of central tendency than the mean, as it is not influenced by outliers or extreme values.
- The **choice of measure of central tendency** depends on the specific needs of the study and the type of data being analyzed
 - It is important to understand the strengths and limitations of each measure of central tendency
 - to choose the most appropriate measure for the data being analyzed.

31. Answer: b

Explanation:

The correct answer is: $\sum_{i=1}^k (\bar{y}_{i.} - \bar{y})^2$ represents the sum of squares due to treatments.

★ Key Points

- In a completely randomized design, y_{ij} represents the response from the j^{th} unit receiving the i^{th} treatment.
- The symbol " Σ " represents the sum of all observations.
- The expression $\sum_j y_{ij} = y_{i.}$ represents the sum of the responses for a specific treatment (i) and is equal to the mean response for that treatment ($\bar{y}_{i.}$).
- The expression $\sum_{i=1}^k (\bar{y}_{i.} - \bar{y})^2$ represents the sum of squares due to treatments, which is a measure of the variation in the response variable that can be attributed to the different treatments. This expression calculates the difference between the mean response for each treatment ($\bar{y}_{i.}$) and the overall mean response ($\bar{y}$), squares these differences, and then sums these squared differences across all treatments.

★ Additional Information

- The sum of squares due to treatments is used in the **analysis of variance (ANOVA)** to determine if there are significant differences in the response variable between the different treatments.
- The sum of squares due to treatments is just one component of the total sum of squares, which represents the total variation in the response variable.
- Understanding the concept of sum of squares and its use in ANOVA is important for conducting and interpreting experiments involving treatment comparisons.
- It is important to have a clear understanding of the experimental design and the data collection process in order to accurately interpret the results of an ANOVA.

32. Answer: a

Explanation:

The correct answer is **Geological classification**.

★ Key Points

- The basis of classification of data can be broadly categorized into four categories, namely, geographical, temporal, qualitative, and quantitative classification.
- **Geological classification** is not one of the commonly used basis of data classification.
- **Geographical classification** is based on location or region.
- **Temporal classification** is based on time or date.
- **Qualitative classification** is based on non-numeric attributes such as color, shape, and size.
- **Quantitative classification** is based on numeric attributes such as height, weight, and temperature.

★ Additional Information

- The basis of classification of data is an important aspect of data analysis and helps in organizing and summarizing data.
- The choice of the basis of classification depends on the type of data and the objective of the analysis.
 - For example, if the objective is to analyze the sales of a company based on different regions, then geographical classification would be an appropriate basis.
- The basis of classification should be chosen based on the type of data and the objective of the analysis.
- A clear and appropriate basis of classification can make the data analysis process easier and more effective.

33. Answer: c

Explanation:

The correct answer is 1/125

Given:

★ Key Points

$$\mu_1' = 1; \mu_2' = 6; \mu_3' = 15$$

Calculation:

$$B_1 = \frac{(\mu_3')^2}{(\mu_2')^3} - \frac{(\mu_1')^2}{(\mu_2')^3} \quad (1)$$

$$\text{Now, } \mu_2 = \mu_2' - (\mu_1')^2$$

$$\Rightarrow 6 - (1)^2$$

$$\Rightarrow 5$$

$$\mu_3 = \mu_3' - 3(\mu_2')(\mu_1') + 2(\mu_1')^3$$

$$\Rightarrow 15 - 3(6)(1) + 2(1)^3$$

$$\Rightarrow 15 - 18 + 2$$

$$\Rightarrow -1$$

By putting the above two values in (1)

$$\Rightarrow B_1 = \frac{(-1)^2}{(5)^3}$$

$$\Rightarrow B_1 = 1/125$$

★ Additional Information

- The **Karl Pearson's coefficient** of skewness is a measure of the degree of skewness or asymmetry of a probability distribution.

- It is calculated as the third standardized moment about the mean, which is the ratio of the third central moment and the cube of the standard deviation.
- The **coefficient of skewness** has the following properties:
 - If $\beta_1 = 0$, then the distribution is symmetric and has no skewness.
 - If $\beta_1 > 0$, then the distribution is positively skewed and has a long right tail.
 - If $\beta_1 < 0$, then the distribution is negatively skewed and has a long left tail.
- The coefficient of skewness is an important descriptive statistic that provides information about the shape of a distribution.
- The coefficient of skewness should be used in conjunction with other measures such as the mean, median, and standard deviation to describe the distribution and understand its properties

34. Answer: d

Explanation:

The correct answer is 1/15

★ Key Points

- The sum of all the probabilities is always equal to 1.
- For a **discrete random variable** X with a probability mass function, the probabilities must sum to 1, which is a requirement for a valid probability distribution.
 - Hence, the value of k can be determined by setting the sum of all the probabilities equal to 1 and solving for k .
- In this case, the sum of all the probabilities is given by
 - Hence, $2k + 3k + 4k + 3k + 2k + k = 1$
 - $\Rightarrow 15k = 1$
 - $\Rightarrow k = 1/15$
- A **probability mass function** is a function that assigns a probability to each value of a discrete random variable.
- The probabilities assigned to the values of a discrete random variable must sum to 1.

- The value of k can be determined by solving for the sum of all the probabilities equal to 1.

★ Additional Information

- Once the value of k is determined, the probabilities can be found by multiplying each probability by k . In this case, the probabilities are
- $P(X = -2) = 2k = 2/15$
 $P(X = -1) = 3k = 3/15$
 $P(X = 0) = 4k = 4/15$
 $P(X = 1) = 3k = 3/15$
 $P(X = 2) = 2k = 2/15$
 $P(X = 3) = k = 1/15$
- A probability mass function must be a valid probability distribution, meaning that all probabilities must be non-negative and the sum of all probabilities must be equal to 1.
- The value of k can be determined by solving for the sum of all the probabilities equal to 1.
- The probabilities assigned to the values of a discrete random variable provide information about the likelihood of each value of X .

35. Answer: b

Explanation:

The correct answer is $z = 3.6$

★ Key Points

- The **test statistic** is a measure of the difference between the sample mean and the population mean in terms of standard deviations.
- In this case, the test statistic can be calculated as follows:
 - $z = (\text{sample mean} - \text{population mean}) / (\text{standard deviation} / \sqrt{(\text{sample size})})$
 - $z = (90 - 82) / (20 / \sqrt{(81)})$
 - $z = 3.6$

★ Additional Information

- The **test statistic** is a measure of the difference between the sample mean and the population mean in terms of standard deviations.
- The test statistic is calculated by dividing the difference between the sample mean and the population mean by the standard deviation divided by the square root of the sample size.
- The test statistic can be used to determine the significance of the difference between the sample mean and the population mean.
- If the test statistic is large enough, it can be concluded that the sample mean is significantly different from the population mean.

★ Important Points

- The test statistic
 - It is only meaningful if the sample is a random sample.
 - It assumes that the population is normally distributed.
 - It is only one piece of information used to make inferences about the population mean.
 - Other factors, such as the sample size and the variability of the population, must also be taken into account.

36. Answer: d

Explanation:

The correct answer is – 0.8 .

★ Key Points

- The **regression coefficient** is a measure of the strength of the linear relationship between two variables.
- It is calculated as the covariance between the two variables divided by the variance of the independent variable. In this case, the regression coefficient can be calculated as follows:
 - $\text{Cov}(x, y) / \text{Var}(x) = -0.8$

- **The regression coefficient**

- It is a measure of the strength of the linear relationship between two variables.
- It is calculated as the covariance between the two variables divided by the variance of the independent variable.
- It can be used to make predictions about the dependent variable based on the value of the independent variable. A negative regression coefficient indicates a negative linear relationship between the two variables, while a positive regression coefficient indicates a positive linear relationship.
- It only measures linear relationships between the two variables.
- It assumes that the relationship between the two variables is linear.
- It does not indicate causality between the two variables.

★ **Additional Information**

- To calculate the regression coefficient (b_1) between x and y , we can use the formula:
 - $b_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$
- Where $\bar{x}$ and $\bar{y}$ are the means of x and y , respectively.
- First, let's calculate the means:
 - $\bar{x} = (1 + 2 + 3 + 4)/4 = 2.5$ $\bar{y} = (4 + 5 + 3 + 2)/4 = 3.5$
- Next, let's calculate the numerator:
 - $\sum(x - \bar{x})(y - \bar{y}) = (1 - 2.5)(4 - 3.5) + (2 - 2.5)(5 - 3.5) + (3 - 2.5)(3 - 3.5) + (4 - 2.5)(2 - 3.5)$
 - $= -1.5 * 0.5 + -0.5 * 1.5 + 0.5 * -0.5 + 1.5 * -0.5 = -1.5 * 0.5 + -0.5 * 1.5 + 0.5 * -0.5 + 1.5 * -0.5$
 - $= -0.75 + -0.75 + 0.25 - 0.75$
 - $= -2.5$
- Next, let's calculate the denominator:
 - $\sum(x - \bar{x})^2 = (1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2$
 - $= 1.25 + 0.25 + 0.25 + 1.25$
 - $= 3$
- Finally, let's calculate b_1 :
- $b_1 = -2.5/3 = -0.83333...$
 - Therefore, the regression coefficient (b_1) between x and y is approximately -0.83 .

37. Answer: a

Explanation:

The correct answer is **Qualitative classification**.

★ Key Points

- **Qualitative classification** is a method of grouping data based on attributes or qualities that are inherent in the data.
- The data is classified into categories based on its characteristics, such as color, shape, size, or texture.
- **Qualitative classification** groups data based on attributes or qualities inherent in the data.
- The data is classified into categories based on its characteristics.

★ Additional Information

- Qualitative classification is often used in fields such as biology, sociology, and psychology to categorize individuals or objects based on specific traits.
- Qualitative classification should be used when the data being classified cannot be easily quantified.
- The categories used in qualitative classification should be mutually exclusive and exhaustive, meaning that each item should only belong to one category and all items should belong to at least one category.
- The number of categories used in qualitative classification should be kept to a minimum to ensure the accuracy and usefulness of the classification.

38. Answer: d

Explanation:

The correct answer is **5.2**.

★ Key Points

- **Deciles** divide a data set into ten equal parts.
- The n th decile separates the lowest $(10n - 90)\%$ of the data from the highest $(90 - 10n)\%$ of the data.
- To calculate the decile, the data must be sorted and the formula $D_n = X[(n/10) * (N + 1)]$ must be used.
- If the result is not a whole number, interpolation must be used to estimate the decile value.
- Deciles are often used in statistics to summarize a data set and to identify the values that separate the data into different percentiles.
- Deciles are just one type of quantile. Other quantiles, such as quartiles and percentiles, can also be calculated.
- The value of the decile can be affected by outliers and extreme values in the data set. It is important to identify and handle outliers before calculating deciles to ensure accurate results.

★ Additional Information

- The deciles are a set of quantiles that divide a data set into ten equal parts.
- The 6th decile (D_6) is the value that separates the lowest 60% of the data from the highest 40% of the data.
 - In other words, 60% of the data values are below D_6 and 40% of the data values are above D_6 .
 - To calculate the 6th decile for the given data set, we first need to sort the data in ascending order: 2, 3, 4, 5, 6, 8.
- The formula to calculate the decile is as follows:
 - $D_n = X[(n/10) * (N + 1)]$, where X is the sorted data set, n is the decile (in this case, 6), and N is the number of data points (in this case, 6).
 - $D_6 = X[(6/10) * (6 + 1)] = X[(6/10) * 7] = X[4.2]$
- Since 4.2 is not a whole number, we use interpolation to estimate the 6th decile:
 - $D_6 = (X[4] + X[5])/2 = (5 + 6)/2 = 5.2$

39. Answer: a

Explanation:

The correct answer is **-0.125**.

★ Key Points

- **Spearman's correlation coefficient** is a measure of the monotonic association between two variables.
 - It measures how well the rank order of one variable corresponds to the rank order of the other variable.
- A positive correlation means that as the rank of one variable increases, the rank of the other variable also increases.
- A negative correlation means that as the rank of one variable increases, the rank of the other variable decreases.
- To **calculate** the Spearman's correlation coefficient, we first need to calculate the ranks of the data in X and Y.
- The ranks are assigned based on the order of the data, with the smallest value receiving the rank of 1 and the largest value receiving the rank of the number of data points.
- If there are ties, the average rank is assigned.
- The ranks for X are: 1, 2, 2
- The ranks for Y are: 3, 3, 1
- Next, we calculate the difference between the ranks for each data point, square the differences, and sum them.
- The formula is as follows:
 - $\rho = 1 - \frac{6 * \sum(d^2)}{N * (N^2 - 1)}$, where N is the number of data points and d is the difference between the ranks of X and Y for each data point.
 - $\rho = 1 - \frac{6 * ((1 - 3)^2 + (2 - 3)^2 + (2 - 1)^2)}{3 * (3^2 - 1)} = 1 - \frac{6 * (4 + 1 + 1)}{3 * 8} = 1 - \frac{6 * 6}{24} = 1 - 1 = 0$
- The result of -0.125 means that there is a weak negative correlation between the two variables.

★ Additional Information

- **Spearman's correlation coefficient** measures the monotonic association between two variables.

- It measures the correspondence between the rank orders of the variables.
- A **positive correlation** means that as the rank of one variable increases, the rank of the other variable also increases.
- A **negative correlation** means that as the rank of one variable increases, the rank of the other variable decreases.

★ Important Points

- Spearman's correlation coefficient is a non-parametric measure, meaning that it does not make assumptions about the underlying distribution of the data.
 - It can be used with both continuous and ordinal data.
- Spearman's correlation coefficient is only appropriate to use when there is a monotonic relationship between the variables.
 - If the relationship is not monotonic, other measures, such as Pearson's correlation coefficient, may be more appropriate.
- The strength of the correlation is determined by the magnitude of the coefficient.
 - A value close to +1 or -1 indicates a strong correlation, while a value close to 0 indicates a weak correlation.

40. Answer: b

Explanation:

The correct answer is **8.0**.

★ Key Points

- The **median** is a measure of central tendency that gives us the middle value in a dataset.
- The median is useful when the dataset has outliers or extreme values that can distort the mean.
- To find the median, we need to arrange the data in order and choose the middle value or average of the two middle values.
- It's important to note that the median is not always the same as the mean, which is another measure of central tendency that gives us the average value

in a dataset.

- Unlike the median, the mean is sensitive to outliers and extreme values, which can have a significant impact on the result.

★ Additional Information

- The median is a measure of central tendency that gives us the middle value in a dataset when it is arranged in order.
 - To find the median of a dataset, we first need to arrange the data in ascending or descending order, and then choose the middle value.
- If the dataset has an odd number of observations, the median is simply the middle value.
- If the dataset has an even number of observations, the median is the average of the two middle values.
- In this case, the dataset has 10 observations, which is an even number.
 - After arranging the data in ascending order, the dataset becomes: 7, 7, 8, 8, 8, 8, 9, 9, 10, 10
 - The two middle values are 8 and 8, and the average of these two values is 8.0, which is the median of the dataset.

★ Important Points

- The median is a useful measure of central tendency for datasets with outliers or extreme values.
- It's essential to understand the difference between the median and the mean and how to calculate them.
- In practice, different measures of central tendency may give different results, and it's important to choose the appropriate one based on the characteristics of the dataset.

41. Answer: b

Explanation:

The correct answer is for k treatments and N observations, the degree of freedom of the ratio of the mean sum of the square of treatments and the residual mean sum is

$N - 1$ and $k - 1$.

★ Key Points

- The number of degrees of freedom is $k-1$, $N-K$, where m is the number of samples and N is the total size of all the samples together.
- The sampling distribution of the ratio of the variances of two samples taken from a normal population is known as the **F-distribution**. It is directly applied in determining whether two samples originate from populations with the same variance.
- The more significant usage is in **variance analysis**, even if you may infrequently see it employed to assess equality of variances (ANOVA). ANOVA is used to determine if three or more samples are from populations with the same mean, at least in its most basic form as described in this chapter.
- **ANOVA** determines whether the samples are from populations with equal means or not by determining whether the variance of the observations is more closely related to the variance of each observation from the mean of its sample or to the variance of the means of the samples from the overall mean.

★ Additional Information

- The divergence of data points from the mean value is measured by the sum of squares.
- While a smaller result shows little deviation from the mean, a bigger sum of squares suggests higher variability.
- Subtract the data points from the mean, square the differences, and then add the resulting amount to determine the sum of squares.
- **Total, residual, and regressive sums of squares are the three types.**
- The **sum of squares** can be used by investors to assist in improving their investing decisions.

42. Answer: c

Explanation:

is the right solution $SE = \frac{\sqrt{46}}{3}$



key points

- The formula for the standard error is: $SE = \frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation and n is the total number of observations.
- Standard deviation is calculated using the following table:

X	$\bar{X}$	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
15	11	4	16
5	11	-6	36
12	11	1	1
10	11	-1	1
20	11	9	81
4	11	-7	49

$$\text{mean} = \bar{X} = \frac{15+5+12+10+20+4}{6} = 11$$

$$\text{standard deviation} = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}} = \sqrt{\frac{92}{3}}$$

$$\text{standard error} : \frac{\sqrt{\frac{92}{3}}}{6} = \frac{\sqrt{46}}{3}$$

Thus the correct answer is- *The standard error of the given data 15, 5, 12, 10, 20 and 4 is $\frac{\sqrt{46}}{3}$.*



Additional Information

- Standard error (SE) is the estimated standard deviation of a statistical sample population.
- The difference between the computed mean of a population and one that is widely accepted to be true is described by the standard error.
- The standard error decreases when more data points are used in mean computations.

43. Answer: a

Explanation:

The correct option is **Shifts in the level of a time series that cannot be explained are referred to as seasonal cycles.**



Key Points

- When a time series is impacted by seasonal elements like the season or day of the week, a seasonal pattern develops. Seasonality always has a set and recognizable periodicity. The seasonality in the monthly sales of anti-diabetic medications seen above is caused in part by a change in the price of the medications at the end of the year.
- The act of estimating and then eliminating systematic and calendar-related influences from a time series is known as seasonal adjustment. Observed data must be seasonally adjusted because seasonal influences might mask both the series' true underlying movement and some non-seasonal traits that analysts may find interesting.



Additional Information

- Although they are commonly mistaken, seasonal and cyclical behaviour is extremely different. If there is no definite frequency for the fluctuations, they are cyclical; if there is a fixed frequency that is seasonal and linked to a certain aspect of the calendar.
- The average length of a cycle is often longer than the average length of a seasonal pattern, and the magnitudes of cycles are typically more variable than those of seasonal patterns.
- Many time series include trends, cycles, and seasonality. We must first identify the time series patterns present in the data before selecting a forecasting approach that can accurately capture the patterns.

Hence in the level of a time series that cannot be explained are referred to as seasonal cycles **are not true in the case of time series.**

44. Answer: c

Explanation:

The correct answer is **Chebyshev distribution with degree of freedom 1**.

★ Key Points

- If Z follows a **standard normal distribution**, then Z^2 follows a chi-square distribution with 1 degree of freedom.
- The chi-square distribution is a family of distributions that is used to model the sum of the squares of independent standard normal random variables.
- The distribution is characterized by the degree of freedom, which is the number of standard normal variables that are squared and summed.
- For a chi-square distribution with 1 degree of freedom, the distribution is known as the Chebyshev distribution.
- The Chebyshev distribution has a mean of 1 and a variance of 1. It is a right-skewed distribution with a peak near 1 and an exponential tail.

★ Additional Information

- The **Chebyshev distribution** has important applications in the fields of engineering, mathematics, and statistics.
 - For example, it is used to model the distribution of the maximum absolute value of independent standard normal random variables.
 - It is not the same as the standard normal distribution.
- The standard normal distribution is symmetrical and bell-shaped, while the Chebyshev distribution is right-skewed and has a peak near 1.
- The Chebyshev distribution is a special case of the chi-square distribution.
- The **chi-square distribution** has more degrees of freedom and is more flexible, but it also requires more information to specify.

- The Chebyshev distribution is a useful distribution for modeling the behavior of extreme values in a data set.

45. Answer: d

Explanation:

The correct answer is **190**.

★ Key Points

- The **Laspeyres price index** is a measure of the average change in prices between two periods, where the prices in the base year are used to weigh the prices in the current year.
- To calculate the Laspeyres price index, we first need to calculate the weighted average of the prices in the current year, using the quantities in the base year as weights.
- Then, we divide this weighted average by the weighted average of the prices in the base year, and multiply the result by 100 to obtain the index.

★ Additional Information

- The calculation is as follows:
 - Weighted average of prices in the current year:

$$(30 * 20 + 50 * 40 + 70 * 60) / (20 + 40 + 60) = (600 + 2000 + 4200) / (120) = 6800 / 120 = 56.67$$
 - Weighted average of prices in the base year: $(20 * 20 + 30 * 40 + 40 * 60) / (20 + 40 + 60) = (400 + 1200 + 2400) / (120) = 3600 / 120 = 30$
- Laspeyres price index = $(\text{Weighted average of prices in the current year}) / (\text{Weighted average of prices in the base year}) * 100 = 56.67 / 30 * 100 = 189$
- So, the Laspeyres price index for the current year is 190.

46. Answer: a

Explanation:

The correct answer is **collecting primary data is not quite expensive both, in the terms of time and money.**

★ Key Points

- The process of acquiring data through surveys, interviews, or experiments is known as primary data collection. Household surveys are a common type of primary data.
- It may be more expensive to obtain primary data than secondary data. This is so that volunteers may be found, or a vendor or research firm can be hired. Sometimes secondary data is even free.
- In so far, as it was gathered with the intention of solving a specific study issue, primary data is frequently trustworthy, real, and unbiased. It is noteworthy that primary data collection is uncommon due to the high implementation costs.

★ Additional Information

- The Obama Administration determined a few years ago that any research conducted with federal public dollars should be made freely available to the general public. Additionally, data management plans should be in place to keep the data safe and secure for virtually all time. These data sets are available for download and manipulation for study and are published as Supplementary Materials in the scientific literature.
- Your choice of data for your analysis or research is influenced by things like:
 - Time constraints
 - Expertise in the field of interest
 - Existing (secondary) data availability
 - Ability and experience of your business to gather primary data.

Hence, **collecting primary data is not quite expensive both, in the terms of time and money is not a correct statement.**

47. Answer: c

Explanation:

The correct answer is -14

★ Key Points

- The third moment about the **mean** is $E((X - \mu)^3)$.
- Here the mean, symbolized by μ , is the first point near the origin of a random variable.
- Expanding $E((X - \mu)^3)$ we get: $E((X - \mu)^3) = E(X^3) - 3E(X)E(X^2) + 2E(X)^3$
- Here $E(X^3)$, $E(X^2)$, $E(X) (= \mu)$ are the **third, second, and first moments** about the origin respectively.

Substituting the values respectively in the third moment about the mean expression we get:

$$E((X - \mu)^3) = 18 - (3 * 2 * 8) + (2 * 2^3)$$

$$E((X - \mu)^3) = -14$$

★ Additional Information

- The method of moments is a technique used in statistics to estimate population parameters. To derive higher moments like skewness and kurtosis, the same method is applied. First, functions of the pertinent parameters are used to express the population moments. The sample moments are then put equal to those expressions.
- These dimensionless, normalized central moments depict the distribution without dependence on any linear scale change.

Hence, if the first, second, and third moments about the origin are 2, 8, and 18 respectively, then the third moment about the mean is -14

48. Answer: b

Explanation:

The correct answer is **6.20**

★ Key Points

The formula of the mode used in a **frequency distribution** is:

$$M_0 = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) * h$$

where , M_0 = Mode, l = lower limit of the **modal class** , h = size of the class interval, f_1 = frequency of modal class. f_1 = frequency of the class preceding, the modal class and f_2 = The frequency of the class succeeding the modal class.

According to the problem, class 5–7 is the modal class, since it has the highest frequency.

Substituting the respective values in the mode equation gives us:

$$\begin{aligned} M_0 &= 5 + \left(\frac{4-1}{(2*4)-1-2} \right) * 2 \\ \implies M_0 &= 5 + 1.20 \\ \implies M_0 &= 6.20 \end{aligned}$$

★ Additional Information

- The value of h here is symmetrical which is $7 - 5 = 2$.
- The real **upper limit** is produced by adding 0.5 to the highest number, if it is expressed as a whole number, or by adding 0.05 to the highest number if it is represented as a decimal. The upper limit is the highest value of the class interval.
- Similarly to the upper limit, the lower limit is the smallest value in the class interval. To determine the actual lower limit, take the smallest number and either deduct 0.5 from it if it's a whole number or 0.05 from it if it's a decimal.

Hence for the given frequency distribution the value of mode is **6.20**

49. Answer: d

Explanation:

The correct answer is **in a polygon**, an additional line is not drawn at each to bring the graph back to a zero frequency.

★ Key Points

- A frequency polygon is a line graph that plots class frequency against the class midpoint. You can get it by connecting the midpoints of the rectangles' tops in the histogram.
- One must include the first and last class with zero frequencies in order to confirm that our graph is genuinely a polygon (i.e., the graph is closed-shape). Draw attention to the frequency column and the column with the midpoints for each class.

★ Additional Information

- A histogram is a bar graph with rectangle-shaped bars displaying the data, but a frequency polygon is a line graph with a curved line showing the data.
- A polygon has flat surfaces throughout. A polygon must have a minimum of three sides. Each side of a line segment must intersect another line segment only at its terminal. By counting the sides of a polygon, we may quickly ascertain its geometry.

Hence, **for frequency distribution presentation, the wrong option is in a polygon**, an additional line is not drawn at each to bring the graph back to a zero frequency.

50. Answer: b

Explanation:

The correct answer is **1.7414**

★ Key Points

- **Kurtosis** = Fourth Moment / (Second Moment)²

	Variable (x)	Deviation(x - $\bar{x}$)	Deviation ²	Deviation ⁴
	8	3.8333	14.6944	215.9267
	6	1.8333	3.3611	11.2971
	3	-1.1667	1.3611	1.8526
	1	-3.1667	10.0278	100.5563
	2	-2.1667	4.6944	22.0378
	5	0.8333	0.6944	0.4823
SUM	25	0	34.8333	352.1528
SUM/n	$\bar{x} = 4.1667$	0	5.8056	58.6921

$$\text{Kurtosis} = 58.6921 / (5.8056)^2$$

$$\text{Kurtosis} = 1.7414$$

★ Important Points

- An attribute of a dataset can be described using the statistical measure kurtosis. On a graph, regularly distributed data typically resembles an upside-down bell when plotted. The bell curve is a name for this. The tails on either side of the curve are often formed by the plotted data that are most dispersed from the data mean. Kurtosis reveals how much information is included in the tails.
- Leptokurtic (below normal), platykurtic (above normal), and mesokurtic (normal) are the three categories of kurtosis (more than normal).
- Only when combined with standard deviation is kurtosis beneficial. Even if an investment has a high kurtosis (poor), the overall standard deviation is likely to be modest (good). On the other hand, a low kurtosis (excellent) investment could have a large total standard deviation (bad).

Hence, if the population kurtosis of the observations 16,12,6,2,4,10 is 1.7414, then the population kurtosis of the 8,6,3,1,2,5 is 1.7414

51. Answer: b

Explanation:

The correct answer is 16 .

★ Key Points

- The formula used to derive this result is Mean - Mode = 3 (Mean - Median).
The solution to this problem is:
 $10 - \text{Mode} = 3(10 - 12) \implies (10 + 6) = \text{Mode} \implies \text{Mode} = 16$
- The Mean in a frequency distribution is the average of all the observations. The average of all the observations is done by adding them up and dividing them by the total number of observations in a frequency distribution.
- The Median is the middlemost value of a distribution when all the values of the observation are arranged from ascending to descending order.
- Mode is the value that has the highest frequency.

★ Additional Information

- The relation between Mean, Median and Mode determines the nature of the frequency distribution curve.
- If Mean=Median=Mode then the distribution curve is **symmetrical**.
- If Mean>Median>Mode then the distribution curve is **positively skewed**.
- The Mode>Median>Mean then the distribution curve is **negatively skewed**.

Hence, the mean and median of the distribution are 10 and 12 respectively, then the mode equals to the mean and median of the distribution are 10 and 12 respectively, then the mode is 16.

52. Answer: d

Explanation:

The correct answer is **positively skewed**.

★ Key Points

- The first step is to find out the value of the **median**. The media can be found out from the relation, $\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$. The values of mean and mode are 32 and 21. Therefore the value of the median is, $32 - 21 = 3(32 - \text{Median}) \implies 11 = 96 - 3\text{Median} \implies \text{Median} = 28.33$.
- In this case The value of $\text{Mean} > \text{Mode} > \text{Median}$. This, according to the nature of the curve of frequency distribution, refers that the distribution is positively skewed.

Thus the correct answer is- The given values of **mean** and **mode**, the given distribution is positively skewed.

★ Additional Information

- If the $\text{Mean} < \text{Median} < \text{Mode}$, the distribution would have been **negatively skewed**.
- If $\text{Mean} = \text{Median} = \text{Mode}$, the distribution would have been **symmetrically distributed**.
- The values of mean and mode can also be computed using the given relation.

Hence, If mean and mode of the distribution is 32 and 21, then the distribution is positively skewed.

53. Answer: c

Explanation:

The correct answer is $\frac{6}{36}$

★ Key Points

- The **probability** of a particular event E occurring is given by $P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes or sample space.}}$
- The sample space of this problem, where two dice are thrown together is 36
- The sum of values on the dice resulting in a total of 7 is given by the following table:

Serial number	Value of observation in dice 1	Value of observation in dice 2
1	1	6
2	6	1
3	3	4
4	4	3
5	2	5
6	5	2

- The number of favourable outcomes, in this case, is 6.
- Let F be the **event** of getting a 7 when two dice are thrown together.
- Therefore $P(F) = \frac{6}{36}$

★ Important Points

- Sample space is the set of all the possible outcomes in the given experiment. In this case, there are 36 outcomes possible when 2 dice are thrown together.
- Event denotes the probability of getting a single outcome of an experiment. In this problem, the event is to get the value of 7.
- The outcome is the result of the experiment.

Hence, the probability of getting a total of 7 on two dice thrown together is $\frac{6}{36}$

54. Answer: d

Explanation:

The correct answer is a **cyclical component** .

★ Key Points

- A cyclical industry is one that responds to the **business cycle** , with revenues typically being higher during periods of economic expansion and prosperity and lower during periods of recession and contraction.
- Businesses in cyclical industries can deal with this type of instability by implementing staff cutbacks and layoffs to make up for lost income during lean **economic times** , and by giving bonuses and hiring en masse during booming economic times.
- In reality, a **recession** is a particular kind of vicious cycle with cascading decreases in output, employment, income, and sales that feedback into a further decline in output, quickly spreading from industry to industry and region to region.

★ Additional Information

- A recession, according to the NBER, lasts anywhere from a few months to more than a year and occurs between the peaks and troughs of a business cycle when there is a significant decline in economic activity across the board.
- Cyclical industries produce or market goods that we can live without or put off purchasing during difficult economic times. Luxury products, leisure vacations, and new construction are a few examples.

Hence, **recession in the industry is associated with the cyclical component.**

55. Answer: c

Explanation:

The correct answer is $\sqrt{\frac{66}{5}}$

★ Key Points

- The formula used to derive this formula is $s = \sqrt{\frac{n_1*s_1^2 + n_2*s_2^2 + n_1*d_1^2 + n_2*d_2^2}{n_1 + n_2}}$, where $d_1 = \bar{x}_1 - \bar{x}$ and $d_2 = \bar{x}_2 - \bar{x}$ also, $\bar{x} = \frac{n_1*\bar{x}_1 + n_2*\bar{x}_2}{n_1 + n_2}$ where n_1 is the total observation of the first data set, n_2 is the total observation of the second data set. Also, s_1, x_1 and s_2, x_2 are the standard deviation and mean for data set 1 and data set 2 respectively.
- Here $n_1 = 6, n_2 = 9$ and $s_1 = 3, s_2 = 4$ and $\bar{x}_1 = 4, \bar{x}_2 = 4$. The following calculations are done:

$$\bar{x} = \frac{6*4 + 9*4}{6+9}$$

$$\bar{x} = 4$$

$$d_1 = 4 - 4 = 0$$

$$d_2 = 4 - 4 = 0$$

- Substituting the values in the expression s, we get:

$$s = \sqrt{\frac{6*9 + 9*16 + 6*0 + 9*0}{15}}$$

$$s = \sqrt{\frac{198}{15}}$$

$$s = \sqrt{\frac{66}{5}}$$

★ Additional Information

- Combined Standard Deviation** is the result of calculating the standard deviation of two or more series.
- The number of series that can be included in the Combined Standard Deviation calculation is N.
- If we know the mean and quantity of items in each group of the data, we can calculate **the combined arithmetic mean**.

Hence, *the standard deviation of data sets of sizes 6 and 9 have a standard deviation 3 and 4, respectively, and arithmetic means 4 and 4, respectively. The standard deviation of the combined data set of size 15 is $\sqrt{\frac{66}{5}}$*

56. Answer: b

Explanation:

The correct answer is that it remains same .

★ Key Points

The formulae we need for solving the problem are as follows:

$$\text{Quartile formula for } Q_1, \text{grouped data} = Q_1 = I_1 + \frac{(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Quartile formula for } Q_3, \text{grouped data} = Q_3 = I_1 + \frac{3*(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Here, since all the observations are doubled the value for Quartiles becomes:

$$\text{Quartile formula for } Q_1, \text{grouped data} = Q_1 = I_1 + \frac{(2N/4)-2c}{2f} * (I_2 - I_1)$$

$$\implies \text{Quartile formula for } Q_1, \text{grouped data} = Q_1 = I_1 + \frac{(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Quartile formula for } Q_3, \text{grouped data} = Q_3 = I_1 + \frac{3*(2N/4)-2c}{2f} * (I_2 - I_1)$$

$$\implies \text{Quartile formula for } Q_3, \text{grouped data} = Q_3 = I_1 + \frac{3*(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

This is the same value as the coefficient of quartile deviation when the values of the observation of the data set were unchanged.

Thus the correct answer is- *If each observation in a data set for the number of employees in different divisions is doubled then the coefficient of quartile deviation remains same.*

★ Additional Information

- The formula in this case used is for **grouped data** as the question specifically says observation of different employees for different departments.
- The same result can be derived for **ungrouped data** as well.

57. Answer: d

Explanation:

The correct answer is that it remains same .

★ Key Points

Calculation: The formulae we need for solving the problem are as follows:

$$\text{Quartile formula for } Q_1, \text{grouped data} = Q_1 = I_1 + \frac{(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Quartile formula for } Q_3, \text{grouped data} = Q_3 = I_1 + \frac{3*(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Here, since all the observations are doubled the value for Quartiles becomes:

$$\text{Quartile formula for } Q_1, \text{grouped data} = Q_1 = I_1 + \frac{(\frac{1}{2}N/4)-\frac{1}{2}c}{\frac{1}{2}f} * (I_2 - I_1)$$

$$\implies \text{Quartile formula for } Q_1, \text{grouped data} = Q_1 = I_1 + \frac{(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Quartile formula for } Q_3, \text{grouped data} = Q_3 = I_1 + \frac{3*(\frac{1}{2}N/4)-\frac{1}{2}c}{\frac{1}{2}f} * (I_2 - I_1)$$

$$\implies \text{Quartile formula for } Q_3, \text{grouped data} = Q_3 = I_1 + \frac{3*(N/4)-c}{f} * (I_2 - I_1)$$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

This is the same value as the coefficient of quartile deviation when the values of the observation of the data set were unchanged.

★ Additional Information

- The formula in this case used is for grouped data as well.
- The same result can be derived for ungrouped data as well.

Hence, if each observation is halved then the coefficient of quartile deviation remains same.

58. Answer: b

Explanation:

The correct answer is **0.5**.

★ Key Points

- The **quartile deviation** is given by $Q = \frac{(Q_3 - Q_1)}{2}$, In this sum $Q = 1.5$, therefore $\frac{(Q_3 - Q_1)}{2} = 1.5 \implies (Q_3 - Q_1) = 3$

- The sum of quartile deviation is given by 6 which means $(Q3 + Q1) = 6$
- The coefficient of quartile deviation is given by $(Q3 - Q1)/(Q3 + Q1)$. Therefore the **coefficient** of quartile deviation in this sum is $\frac{3}{6} = 0.5$.

★ Additional Information

- Mathematically speaking, the quartile deviation is equal to half of the difference between the upper and lower quartile. In this case, the quartile deviation can be represented by the symbol QD, where Q3 stands for the upper quartile and Q1 for the lower quartile. The Semi Interquartile range is another name for quartile deviation.
- The Quartile Deviation is an easy technique to calculate a distribution's spread relative to a measure of its central tendency (usually the mean). As a result, it provides you with a general sense of the range in which the middle 50% of your sample data is located.

Hence, *the coefficient of quartile deviation is 0.5.*

59. Answer: b

Explanation:

The correct answer is **negatively skewed**.

★ Key Points

- The formula used to derive this result is $\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$. The solution to this problem is then,

$$12 - \text{Mode} = 3(12 - 21)$$

$$\implies (12 + 27) = \text{Mode}$$

$$\implies \text{Mode} = 39$$

- The relation between Mean, Median and Mode determines the **nature** of the frequency distribution curve.
- The **Mode > Median > Mean** then the distribution curve is negatively skewed.

★ Additional Information

- The **mean** in a frequency distribution is the average of all the observations. The average of all the observations is done by adding them up and dividing them by the total number of observations in a frequency distribution.
- The **median** is the middlemost value of a distribution when all the values of the observation are arranged from ascending to descending order.
- mode is the value that has the highest frequency.
- If Mean = Median = Mode then the distribution curve is **symmetrical**.
- If Mean > Median > Mode then the distribution curve is **positively skewed**.

Hence, if the mean and median of the distribution are 12 and 21, then the distribution is *negatively skewed*.

60. Answer: b

Explanation:

The correct answer is **Randomization, Replication and local control**.

★ Key Points

- Randomization is the stage where statistical methods are used to detail the accuracy of the experiment.
- Replication is when the basic experiments are repeated to minimise errors.
- Local control is used it has been noted that randomization and replication do not entirely eliminate all extraneous sources of variation, leaving them uncontrollable.

★ Additional Information

- The process of randomly distributing treatments to experimental units is known as randomization. Every conceivable treatment allocation has the same chance under the random process, it is implied.
- As a result, the local control technique improves the experiment. In other words, one needs to pick a design that will allow us to control any unneeded sources of

variation. One employs local control, which refers to the degree of balancing, blocking, and grouping of experimental units, for this aim.

Hence, randomization Replication and Local control are the primary and most essential steps for designing a successful experiment.

61. Answer: d

Explanation:

The correct answer is: **"Secondary data is original".**

★ Key Points

- **Secondary data** refers to data that have already been collected by someone else, not original data.
- Secondary data can be faster and more cost-effective to access than primary data.
- Secondary data's reliability depends on the source, quality and relevance of the data.

★ Additional Information

- Secondary data can come from various sources, including government agencies, market research firms, trade associations, and online databases.
- Secondary data can provide a broad overview and context for a particular topic.

★ Important Points

- The suitability of secondary data for a specific research project depends on the research objectives, data quality and relevance, and the research methods used.
 - In some cases, secondary data may not be sufficient or relevant to answer a specific research question, requiring the collection of primary data.
-

62. Answer: b

Explanation:

The correct answer is $6/\lambda^3$

★ Key Points

- The **moment generating function (MGF)** of a random variable is a function that represents the expected values of its powers.
 - The MGF of X is defined as $M(t) = E(e^{tx})$, where E represents the expected value.
 - The n th derivative of the MGF evaluated at $t = 0$ provides the expected value of X^n .

★ Additional Information

- The **condition** $-t < \lambda/\lambda$ ensures that the MGF of X exists and is finite.
- The MGF is a useful tool to characterize the distribution of a random variable, and it provides a way to calculate its moments.

★ Important Points

- To calculate $E(X^3)$ from the MGF of X , you need to take the third derivative of $M(t)$ with respect to t , and then evaluate it at $t = 0$.
- The MGF method assumes that the random variable X has a moment generating function, which may not be the case for all distributions.

63. Answer: c

Explanation:

The correct answer is **All I, II and III.**

★ Key Points

- **Index numbers** are statistical measures that calculate changes in a set of values over time.
- **Index numbers** are commonly used in economics to measure changes in the cost of living and to track changes in real income.
- **Index numbers** are not primarily used to fix dearness allowances, but they can assist in the process.

★ Additional Information

- Index numbers are calculated relative to a base period and are used to determine the rate of inflation.
- The information obtained from index numbers is important in determining the impact of inflation on an economy and the purchasing power of consumers.
- There are several types of index numbers, including the **Consumer Price Index (CPI)** and the **Producer Price Index (PPI)**, which measure changes in the prices of goods and services.

★ Important Points

- Index numbers provide valuable information about changes in the **cost of living and real income**, which is crucial for policy makers and economists.
- Understanding the concept and calculation of index numbers is important for those studying economics and finance.
- The accuracy of index numbers is important, as they are used to make decisions that can impact the economy and people's lives.

64. Answer: a

Explanation:

The correct answer is $P(A)P(B|A)P(C|A \cap B)$

★ Key Points

- The symbol $\cap$ represents the intersection of events, meaning "and".

- The probability of the intersection of events $A, B,$ and C ($P(A \cap B \cap C)$) is equal to the probability of A times the conditional probability of B given A times the conditional probability of C given the intersection of A and B.

★ Additional Information

- The **conditional probability** of B given A ($P(B|A)$) represents the probability of event B occurring given that event A has occurred.
- The **conditional probability** of C given the intersection of A and B ($P(C|A \cap B)$) represents the probability of event C occurring given that both events A and B have occurred.
- This formula represents the chain rule of probability, which states that the probability of multiple events occurring in a certain order is equal to the product of their individual probabilities.

★ Important Points

- Understanding the **chain rule of probability** and the **concept of conditional probability** is important in various fields, including statistics and computer science.
- The formula $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$ is widely used in decision making, particularly in risk analysis and game theory.
- Careful application of the chain rule is crucial in avoiding incorrect calculations and reaching incorrect conclusions.

65. Answer: b

Explanation:

The correct answer is **arctan (3/5)**

★ Key Points

- The **acute angle** between lines of regression is given by $\arctan(b)$, where b is the slope of the regression line.

- The **slope of the regression** line can be calculated using the correlation coefficient (r) and the standard deviations of X and Y .
- In this case, the correlation coefficient is 0.5 and the standard deviation of Y is double that of X , so the slope of the regression line can be calculated as $b = 2 * r * (SD_y / SD_x) = 2 * 0.5 * (2/1) = 2$.
- Hence, the acute angle between lines of regression can be calculated as $\arctan(b) = \arctan(2) = 3/5$.

★ Additional Information

- The relationship between X and Y can be visualized by a scatter plot and a regression line.
- The **correlation coefficient** represents the strength of the linear relationship between X and Y
 - with a value of +1 indicating a perfect **positive correlation**
 - a value of -1 indicating a perfect
 - a value of 0 indicating no correlation.
- The angle between the regression line and the x -axis is called the acute angle, which is an indicator of the slope of the regression line.
- In this case, with a correlation coefficient of 0.5, the relationship between X and Y is moderately positive
 - the acute angle can be found using the tan inverse function, with the result being $\arctan(3/5)$.
- The fact that the standard deviation of Y is double of the standard deviation of X means that Y values have more spread than X values
 - the slope of the regression line (represented by the acute angle) helps to describe this relationship.

66. Answer: d

Explanation:

The correct answer is $np(np + q)$.

★ Key Points

- A **moment generating function** is a mathematical tool that helps to calculate moments of a distribution (i.e. its mean, variance, etc.).
- In this case, the moment generating function of X is given by $(q + pet)^n$.
- To find the expected value of X^2 , we can use the second derivative of the moment generating function and evaluate it at $q = 0$.

★ Additional Information

- Moment generating functions are not always unique for a given distribution.
- The expected value of X^2 can also be found by directly calculating $E(X^2)$ from the definition of expectation.

★ Important Points

- Moment generating functions can be useful in cases where it is difficult to directly calculate moments of a distribution.
- The moment generating function can provide information about all the moments of a distribution.

67. Answer: a

Explanation:

The correct answer is 7.

★ Key Points

- The **correlation coefficient** only measures linear relationships, so it does not necessarily indicate causality.
- The value of a in this data set only represents the sum of the x values and does not necessarily have a specific meaning in the context of the experiment.
- The value of a in the data set is 7.
- The correlation coefficient of -0.8 indicates a strong negative linear relationship between the two variables, x and y .
- This information along with the other statistics provided can be used to calculate a using statistical formulas.

★ Additional Information

- The correlation coefficient (r) measures the strength and direction of a linear relationship between two variables.
- A **correlation of -1** indicates a perfect negative linear relationship, while a correlation of 1 indicates a perfect positive linear relationship.
- The formula to calculate the correlation coefficient is: $r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2) * (n\sum y^2 - (\sum y)^2)}}$
- In this case, we know the value of r (-0.8), so we can rearrange the formula and solve for a .

68. Answer: a

Explanation:

The correct answer is $\sqrt{177/15}$.

★ Key Points

- It's important to note that the standard deviation is a **measure of the dispersion** of a set of data.
- The larger the **standard deviation**, the more spread out the data is.
- In this case, the standard deviation of the combined data set of size 15 is calculated by taking into account the standard deviations and sample sizes of the two original data sets.
- By combining the two data sets, the standard deviation of the combined data set takes into account the dispersion of both original data sets and provides a more accurate representation of the spread of the combined data.

★ Additional Information

- To find the **standard deviation** of the combined data set, we can use the formula for the weighted average of the standard deviations of two data sets:
- $\sigma = \sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1 + n_2 - 2}}$
- where n_1 and n_2 are the sizes of the two data sets and σ_1 and σ_2 are their standard deviations.

- Plugging in the given values, we get:
- $\sigma = \sqrt{[(9-1)3^2 + (6-1)4^2] / (9 + 6 - 2)} = \sqrt{[72 + 72] / 11} = \sqrt{144 / 11} = \sqrt{177 / 15}$
- So, the standard deviation of the combined data set of size 15 is $\sqrt{177/15}$.
- Additionally, it's also important to note that this calculation assumes that the two data sets are independent and not related.
- If the two data sets are related in some way, the calculation would be different and may require more information to be accurately calculated.

69. Answer: b

Explanation:

The correct answer is **Stratified sampling**.

★ Key Points

- **Stratified sampling** is a sampling technique where the population is divided into subgroups (strata) based on certain attributes
 - and then a sample is taken from each stratum. In the given scenario
 - the population of MBA students is divided into three strata based on their background (commerce, science and engineering)
 - and then five students are selected at random from each stratum.
- This helps to ensure that the sample is representative of the population and reduces sampling bias.

★ Additional Information

- **Stratified sampling** is often used when the population is heterogeneous and you want to ensure that each subgroup is adequately represented in the sample.
- This helps to minimize sampling error and improve the accuracy of the results.
- In general, the sample size for each stratum should be proportional to the size of the stratum in the population.
- In this way, each stratum has a fair chance of being selected, and the results are more representative of the population.

★ Important Points

- It's important to note that stratified sampling is not always the best choice
 - as it can be time-consuming and costly to divide the population into strata and select a sample from each.
- Other factors, such as the nature of the research question and the availability of resources
- It should be considered when choosing the most appropriate sampling technique.

70. Answer: c

Explanation:

The correct answer is **121.62**.

★ Key Points

- The **simple aggregative method** is a price index calculation method that combines the prices of multiple commodities into a single index.
- The formula for the simple aggregative method is as follows:
- **Price index = (Price of basket in current year) / (Price of basket in base year) * 100**

★ Additional Information

- Where the basket consists of the prices of all commodities being considered.
- To calculate the **basket price**, we can use the weighted average price, where the weight is the quantity of each commodity.
- In the given scenario, the basket price in the base year is $20 + 18 + 12 + 24 = 74$
- the basket price in the current year is $25 + 22 + 15 + 28 = 90$. Therefore, the price index is:

$$\text{Price index} = (90)/(74) * 100 = 121.62$$

This means that the average price of the basket has increased by 21.62% between the base year (2020) and the current year (2022).

71. Answer: d

Explanation:

The correct answer is **2**.

Key Points



- The **first moment** about the origin represents the mean of a distribution
 - It can be calculated using the third moment about the mean and the second moment about the origin.
- The **second moment** about the origin represents the variance of a distribution, while the third moment about the mean represents the skewness of a distribution.



Additional Information

- These moments provide important information about the shape and spread of a distribution.
- It is important to note that these moments are typically calculated for a continuous distribution
 - as discrete distributions may have different properties. Additionally, when calculating the moments
 - it is important to choose the appropriate reference point (i.e. the origin or the mean).

72. Answer: b

Explanation:

The correct answer is **8.33**.



Key Points

- The F-statistic is used to test the null hypothesis that the means of multiple groups are equal, versus the alternative hypothesis that at least one mean is different.
- The F-statistic follows an F-distribution with degrees of freedom (df) for the numerator (between treatments) and degrees of freedom for the denominator (error).
- The F-statistic is often used to determine if the variability between groups is significantly greater than the variability within groups.
- The decision to reject or fail to reject the null hypothesis depends on the F-statistic and the level of significance chosen by the researcher, which is often set at 0.05.
- The higher the F-statistic, the stronger the evidence against the null hypothesis, and in favor of the alternative hypothesis.

★ Additional Information

- The **F-statistic** can be calculated as the ratio of the mean squares between treatments to the mean squares of error.
- The formula is:
 - $F = \text{MS between} / \text{MS error}$
- where **MS between** = Sum of Squares between treatments / Degrees of Freedom between treatments
- **MS error** = Sum of Squares error / Degrees of Freedom error
- So, substituting the values from the question, we get:
- $F = 75/3 / (48/16) = 75/3 * 16/48 = 25/16 = 8.33$
- So the F-statistic is 8.33, which is option (a).

73. Answer: d

Explanation:

The correct answer is mean - mode

★ Key Points

- Relative measures of **skewness** aim to quantify the degree of asymmetry in a given dataset with respect to some reference point.
- The first three options, $((Q3 - Q2) - (Q2 - Q1))/(Q3 - Q1)$, $((P90 - 2P50) + P10)/(P90 - P10)$, and $((D9 - 2D5) + D1)/(D9 - D1)$, are all measures of skewness based on quantiles. They quantify the asymmetry of a dataset by comparing the difference between its median ($Q2, P50, D5$) and its extremes ($Q1, Q3, P10, P90, D1, D9$).

★ Additional Information

- "MEAN- MODE" is simply the difference between the mean and mode of a dataset and does not provide any information about the degree of asymmetry.
- The measure "Mean - Mode" is not considered as a relative measure of skewness because it is not expressed in terms of standard deviations, quartiles or other standard measurements.
- Skewness is the measure of asymmetry of a probability distribution and relative measures of skewness are expressed in terms of the standard units of measurement.

74. Answer: c

Explanation:

The correct answer is (2,24,29).

★ Key Points

- In a **two-way ANOVA**, the degree of freedom (df) is the number of values in a sample that are free to vary.
- The df for source of variation due to interaction is calculated by multiplying the degrees of freedom (df) for each factor and subtracting 1.
- In this case, there are 2 races and 2 genders, so df for interaction = $(2 - 1) * (2 - 1) = 2$.
- The df for error is calculated as the total number of observations minus the number of parameters estimated. In this case, there are 5 observations in each group

- the total number of observations is $3 \times 5 = 30$.
- The number of parameters estimated is 2 (for race and gender) plus 2 (for interaction), so df for error = $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
- The total df is equal to the total number of observations minus 1, which is $30 - 1 = 29$.

★ Additional Information

- In a 2-way **ANOVA**, the degrees of freedom for the source of variation due to interaction is calculated as (number of levels of one factor - 1) * (number of levels of the other factor - 1).
- In this case, the number of levels of one factor is 3 (races) and the number of levels of the other factor is 2 (genders)
- so the degree of freedom for the interaction is $(3 - 1) * (2 - 1) = 2$
- The **degree of freedom for error** is calculated as the total number of observations minus the number of treatments.
- In this case, the total number of observations is 3 races * 2 genders * 5 in each treatment group = 30
- so the degree of freedom for error is $30 - 3 = 27$.
- The total degree of freedom is calculated as the sum of the degrees of freedom for interaction and error. In this case, the total degree of freedom is $2 + 27 = 29$.
- So the answer is (2, 24, 29)

75. Answer: d

Explanation:

The correct answer is **zero**

★ Key Points

- The **skewness of a population** is a measure of its asymmetry and can be calculated as the third standardized moment of the distribution.
- A **skewness of zero** indicates that the distribution is symmetrical

- ie, meaning that the mean, median, and mode are equal and the data is evenly distributed on either side of the center.

★ Additional Information

- In the given data, here have 3 classes with the respective frequencies.
- However, it is not enough information to determine the skewness of the population.
- The skewness can be calculated using the formula which requires the mean, standard deviation and the individual values of the population data.

★ Important Points

- In general, positive skewness indicates that the tail of the distribution on the right side is longer or fatter compared to the left side, and the mode is on the left side of the center.
- On the other hand, negative skewness indicates that the tail on the left side is longer or fatter and the mode is on the right side of the center.
- It is important to note that the skewness is sensitive to outliers, and even a small number of outliers can have a significant impact on the skewness calculation.

76. Answer: b

Explanation:

The correct answer is **6.923**

★ Key Points

- Arithmetic mean is a measure of central tendency that gives the average value of a set of observations.
- To calculate the mean of a frequency distribution, we multiply each value of x by its frequency, and then sum all the products and divide by the total number of observations.
- Arithmetic mean is sensitive to outliers, meaning that if there are extreme values in the data set, it can greatly affect the mean.

- Mean is only appropriate for interval/ratio scale data, where the difference between the values has meaning.
- There are other **measures of central tendency** such as mode (most frequent value), median (middle value), and harmonic mean (used in certain fields like physics).
- Each of these measures can give a different result depending on the distribution of the data.

★ Additional Information

- The arithmetic mean of a **frequency distribution** can be calculated by summing the product of each value of x and its frequency
 - then dividing by the total number of observations.
- The mean for the given frequency distribution is:
 - $(3 * 1 + 5 * 4 + 6 * 2 + 8 * 1 + 9 * 3 + 10 * 2) / (1 + 4 + 2 + 1 + 3 + 2) = 6.923$
 - So, the arithmetic mean of the number of members in the families in the society is 6.923.

77. Answer: c

Explanation:

The correct answer is **-0.80**

★ Key Points

- **Pearson's correlation coefficient** measures the linear relationship between two variables.
- The value of the Pearson's correlation coefficient ranges from **-1 to 1**
 - -1 indicates a perfect negative correlation
 - 0 indicates no correlation
 - 1 indicates a perfect positive correlation.
- To **calculate Pearson's correlation coefficient**, we first standardize the observations by subtracting the mean and dividing by the standard deviation, then we calculate the product of the standardized variables.

- Pearson's correlation coefficient only measures **linear relationships**, and cannot detect non-linear relationships or complex relationships between variables.
- Other measures of association, such as Spearman's rank correlation or Kendall's tau, can be used to capture non-linear relationships.
- When interpreting Pearson's correlation coefficient, it is important to consider the strength and direction of the relationship, as well as the sample size and the presence of outliers.

★ Additional Information

- Pearson's correlation coefficient measures the linear relationship between two variables. It can range from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation.
- Let's consider the original values of X and Y:
 - $X = [1, 2, 3, 4], Y = [3, 4, 2, 1]$
- After halving X and doubling Y:
 - $X = [0.5, 1, 1.5, 2], Y = [6, 8, 4, 2]$
- To calculate Pearson's correlation coefficient, we first find the mean of X and Y, then standardize the observations by subtracting the mean and dividing by the standard deviation. Finally, we calculate the product of the standardized X and Y.
- Using the formula for Pearson's correlation coefficient
 - $r = \frac{\sum((X - \text{mean}(X)) * (Y - \text{mean}(Y)))}{(n - 1) * (\text{std}(X) * \text{std}(Y))}$
- We can find that the new Pearson's correlation coefficient equals to -0.80.
- So, the Pearson's correlation coefficient between the halved X and doubled Y observations is -0.80

78. Answer: a

Explanation:

The correct answer is 6:5

★ Key Points

- **Odds** are a ratio that express the **likelihood of an event happening** or not happening.
- The odds in favor of an event are the ratio of the number of ways the event can happen to the number of ways it cannot happen.
- The odds against an event are the opposite of the odds in favor, obtained by flipping the ratio.
- Odds can be converted to probabilities by dividing the number of favorable outcomes by the total number of outcomes.
- Odds are commonly used in betting, gambling, and statistical probability.
- Understanding odds can help in making informed decisions, evaluating risks, and estimating the likelihood of events.

★ Additional Information

- Odds are a ratio that express the likelihood of an event happening or not happening.
- The odds in favour of an event are the ratio of the number of ways the event can happen to the number of ways it cannot happen.
- If the odds in favor of an event A are 5:6, this means that for every 5 ways the event can happen
 - there are 6 ways it cannot happen.
- To find the odds against the event, we simply flip the ratio: Odds against the event = 6:5
- So, the odds against event A are 6:5.

79. Answer: b

Explanation:

The correct answer is $1/(\text{Priceindexnumber})^2 1/\text{Priceindexnumber}$

★ Key Points

- The **purchasing power of money** refers to the amount of goods or services that can be purchased with a given amount of money.
 - It is a measure of the value of money, and how much it can buy.

- The purchasing power of money is inversely related to the price level in an economy
 - as measured by a price index such as the **Consumer Price Index (CPI)** or the **Producer Price Index (PPI)**.
- When prices are rising, the purchasing power of money decreases, and when prices are falling, the purchasing power of money increases.
- The relationship between the purchasing power of money and the price level can be expressed as:
 - $Purchasingpower = 1/Priceindexnumber$
 - So, the purchasing power of money is equal to $1/Price\ index\ number$.
- This means that the purchasing power of money is directly proportional to the reciprocal of the price index number, which means that as the price index number goes up, the purchasing power of money goes down, and vice versa.

★ Additional Information

- **The purchasing power of money** is the amount of goods or services that can be purchased with a given amount of money.
 - It is a measure of the value of money and its ability to buy goods and services.
 - It is inversely related to the price level in an economy, as measured by a price index such as the Consumer Price Index (CPI) or the Producer Price Index (PPI).
- As prices rise, the purchasing power of money decreases, and as prices fall, the purchasing power of money increases.
- The purchasing power of money can be expressed as $1/Price\ index\ number$, which means that it is directly proportional to the reciprocal of the price index number.
- Changes in the purchasing power of money can have significant impacts on the economy, such as inflation or deflation, which can affect the cost of living, the demand for goods and services, and the overall level of economic activity.

80. Answer: b

Explanation:

The correct answer is **6.625**

★ Key Points

- **Arithmetic mean** is a commonly used measure of central tendency that gives an average of a set of numbers.
- The **formula** for finding the arithmetic mean is:
 $(\text{sum of data points} \times \text{frequency}) / \text{total frequency}$.
- The arithmetic mean can be sensitive to extreme values in the data set.
- There are other measures of central tendency, such as the median and mode, that can provide a better representation of the data in some cases.

★ Additional Information

- To find the **arithmetic mean** (also known as the **average**), we first need to find the sum of the products of each data point and its frequency, and then divide by the total frequency.
- This can be done as follows:
 - $(2 \times 3) + (4 \times 4) + (6 \times 2) + (8 \times 1) + (10 \times 4) + (12 \times 2) = 6 + 16 + 12 + 8 + 40 + 24 = 106$
- Total frequency = $3 + 4 + 2 + 1 + 4 + 2 = 16$
- Arithmetic mean = $106 / 16 = 6.625$
- Understanding the concept of **central tendency** and the **various measures** used to describe it is important for data analysis and interpretation.
- It is important to choose the appropriate measure of central tendency depending on the type and distribution of the data being analyzed.

81. Answer: a

Explanation:

The correct answer is [32.29, 36.21].

★ Key Points

- **Confidence interval** is a range of values that is likely to contain the true population mean with a certain level of confidence.

- The formula for calculating a confidence interval for the mean is:
- $\text{mean} \pm (\text{critical value} \times \text{standard deviation} / \text{square root of sample size})$.
- The **level of confidence** is often expressed as a percentage (e.g. 95% confidence) and determines the width of the interval.
 - A higher level of confidence results in a wider interval, while a lower level of confidence results in a narrower interval.

★ Additional Information

- A 95% confidence interval provides a range of values that is likely to contain the true population mean with a probability of 95%.
- For a 95% confidence interval, the critical value for a t-distribution with 99 degrees of freedom (100-1) is approximately 1.98.
- So, the 95% confidence interval can be calculated as:
 - $\text{mean} \pm (1.98 \times \text{standard deviation} / \text{square root of sample size}) = 34.25 \pm (1.98 \times 10 / \text{sqrt}(100)) = 34.25 \pm 1.98 = [32.29, 36.21]$.
- So, the 95% confidence interval for the mean age of accidents in a city during last year is [32.29, 36.21].

★ Important Points

- The **critical value** used in the calculation of the confidence interval depends on the level of confidence and the degrees of freedom (sample size minus 1).
- The **sample mean and standard deviation** are estimated values and may not exactly reflect the true population mean and standard deviation.
- Understanding the concept of confidence intervals and how to calculate them is important for data analysis and interpretation, as it provides a way to estimate and quantify the uncertainty in a measurement or estimate.

82. Answer: b

Explanation:

The correct answer is $r_{12}=0.69$, $r_{13}=0.22$, $r_{23}=0.23$, $R_{1.23}=0.69$

★ Key Points

- **Correlation coefficient (r)** measures the linear association between two variables.
- **Multiple correlation coefficient (R)** measures the strength of the association between multiple predictor variables and a single criterion variable.
- The values of r and R range between -1 and 1 .
- Values close to 1 indicate a strong positive association (or relationship), values close to -1 indicate a strong negative association (or relationship), and values close to 0 indicate no or weak association (or relationship).
- Correlation coefficients (r) and multiple correlation coefficients (R) are useful for understanding the relationships between variables in a data set.

★ Additional Information

- If r_{12} represents the correlation between X_1 and X_2 , r_{13} represents the correlation between X_1 and X_3 , and r_{23} represents the correlation between X_2 and X_3 , then the values of r_{12} , r_{13} , and r_{23} should range between -1 and 1 , with values close to 1 indicating a strong positive association, values close to -1 indicating a strong negative association, and values close to 0 indicating no or weak association.
- The multiple correlation coefficient (R) measures the strength of the association between multiple predictor variables (X_1, X_2, X_3) and a single criterion variable.
- In this case, $R_{1.23}$ represents the multiple correlation between X_1, X_2 , and X_3 with respect to the criterion variable.
- Like r , R values range between -1 and 1 , with values close to 1 indicating a strong relationship and values close to 0 indicating a weak relationship.
- So, based on the given values, $r_{12}=0.69$, $r_{13}=0.22$, $r_{23}=0.23$, $R_{1.23}=0.69$

★ Important Points

- It is important to note that **correlation** does not imply causation, meaning that just because two variables are correlated does not mean that one causes the other.
- Correlation and multiple correlation coefficients are based on linear relationships and may not capture non-linear relationships between variables.

- Interpreting the strength and direction of relationships between variables is an important aspect of data analysis and interpretation.

83. Answer: b

Explanation:

The correct answer is $k = 3$.

★ Key Points

- **The joint density function**
 - It is a mathematical representation of the distribution of two random variables.
 - It is defined over a region of support, which is the region in which the random variables can take on values.
- The normalization of the joint density function ensures that it integrates to 1 over the region of support, meaning that the total probability of the joint distribution is equal to 1.
 - It is used to describe the relationships between two random variables, including their dependence and independence.
- The **integral of the joint density function** over the region of support gives the probability of observing the two random variables within that region.
- The joint density function is an important tool in probability and statistics, and is used in many applications, including Bayesian analysis, multivariate statistics, and risk assessment.

★ Additional Information

- The **joint density function $f(x,y)$** describes the distribution of two random variables X and Y , which represent the waiting time and service time of customers at a shopping mall.
- The function $f(x,y) = kx$ is defined for $0 \leq y \leq x \leq 1$, meaning that the possible values of X and Y are limited to this region.
- To find the value of k , we need to normalize the density function so that it integrates to 1 over the region of support.

- That is, we need to find the constant k such that the integral of $f(x,y)$ over the region of support is equal to 1.
- The integral can be calculated as follows:
 - $\iint f(x,y) dx dy = \int x * k x dx dy = k * \iint x^2 dy dx = k * (1/3) * \int x^3 dx = k * (1/3) * (x^4/4)$ evaluated over the region of support.
- Since the integral of $f(x,y)$ over the region of support should equal 1, we can set the above expression equal to 1 and solve for k :
 - $1 = k * (1/3) * (1^4/4) = k * (1/3) * (1/4)$
 - $k = 3 / ((1/3) * (1/4)) = 3.$
- So, the value of k is 3.

★ Important Points

- It is important to ensure that the joint density function is correctly normalized to ensure that the probabilities are correctly calculated.
- The choice of the region of support should reflect the range of possible values of the random variables.
- The joint density function should be chosen based on the relationships between the random variables, and the problem context.

84. Answer: b

Explanation:

The correct answer is (0.1,0.1)

★ Key Points

- The **probability mass function (PMF)** of a discrete random variable gives the probabilities associated with each possible value of the random variable.
- The **mean of a random variable** can be calculated by taking the weighted sum of the possible values of the random variable, where the weights are given by the PMF.

★ Additional Information

- The **PMF** must satisfy the properties of a probability function, i.e. it must be non-negative for all values of the random variable, and the sum of the probabilities over all possible values must equal 1.
- The **mean of a random variable** provides a measure of the center of the distribution, and is commonly used in many statistical applications.
- To determine the PMF given a mean, it is necessary to know the possible values of the random variable, and their probabilities.
- The mean of a random variable is a parameter of the distribution, and can be used to describe the center of the distribution.
- The calculation of the mean requires knowledge of the PMF.

85. Answer: c

Explanation:

The correct answer is: $y = -0.8x + 5.5$

★ Key Points

- Regression line represents the best fit line for the given data points, which means that it describes the relationship between X and Y as accurately as possible.
- The slope of the regression line (b) represents the change in Y for a unit change in X, and the y-intercept (a) represents the value of Y when X is equal to 0.
- The regression line is used in various fields such as finance, economics, and engineering to model and make predictions based on the relationship between two variables.
- It is important to note that the regression line is not always the exact representation of the relationship between X and Y, but it is the closest possible approximation.
- There are different types of regression lines, including simple linear regression (one independent variable), multiple linear regression (more than one independent variable), and non-linear regression (a non-linear relationship between X and Y).

★ Additional Information

- The regression line is used to describe the relationship between two variables X and Y.
- It can be found using the formula:
 - $b = (\sum xy - (\sum x * \sum y)/n) / (\sum x^2 - (\sum x)^2/n)$ $a = (\sum y - b * \sum x)/n$
- where b is the slope and a is the y-intercept. Substituting the given values in the formula, we can find the equation of the regression line.

86. Answer: d

Explanation:

The correct answer is 4/25.

★ Key Points

- The **variance of a random variable** X with probability density function f(x) is defined as the expected value of the squared deviation of X from its mean, and can be calculated as $E[(X - E[X])^2]$.
- To find the variance of X in this case, you would need to know the full form of f(x) and calculate $E[X]$ and $E[(X - E[X])^2]$.
- **Variance measures** the spread or dispersion of a set of data or a random variable around its mean.
- A smaller variance indicates that the data is clustered closely around the mean, while a larger variance indicates that the data is more spread out.
- Variance is an important quantity in statistics and probability, and is used in many statistical models and hypothesis tests.
- Variance is always non-negative and its square root, the standard deviation, is a commonly used measure of spread.

★ Additional Information

- **Variance is a measure** of the second moment or central moment about the mean of a set of data or a random variable.

- The variance of a random variable is also related to the expected value (mean) of the squared deviation of the random variable from its mean.
- The variance is an important quantity in the study of random variables, as it provides information about the distribution and spread of the data.
- In some cases, the variance may not provide a complete picture of the spread of the data. In such cases, other measures such as the coefficient of variation or interquartile range may be used.
- Variance is a key parameter in many statistical models, including regression analysis, analysis of variance (ANOVA), and others.
- Understanding and being able to calculate the variance is a fundamental skill for anyone working in the field of statistics and data analysis.

87. Answer: a

Explanation:

The correct answer is $P(A \cup B) = P(A) + P(B)$.

★ Key Points

- **Mutually exclusive events** have no elements in common and cannot occur simultaneously.
- The union of two mutually exclusive events is the sum of their individual probabilities.
- The union of two events, A and B, is the event that either A or B occurs, or both occur. It is denoted as $A \cup B$.

★ Additional Information

- The **probability of a union** of events is equal to the sum of the probabilities of the individual events, as long as the events are mutually exclusive.
- It is important to determine if events are mutually exclusive before applying the addition rule.
- The addition rule for mutually exclusive events is a fundamental rule in probability theory and is widely used in various applications.

88. Answer: d

Explanation:

The correct answer is $46/3$.

★ Key Points

- The mode of a **binomial distribution** is the value of x that has the highest probability of occurrence.
- In a binomial distribution with parameters n and p , the mode can be found using the formula: $\text{mode} = \text{floor}(n * p)$, where "floor" represents the greatest integer less than or equal to the argument.
- The **mean and variance** of a binomial distribution can be used to find the values of n and p , and subsequently the mode.
- The mode, mean and variance are important characteristics of a distribution that describe the shape, central tendency and spread of the data.
- Understanding the properties of binomial distribution and how to find the mode, mean and variance is important in many areas of statistics and data analysis.

★ Additional Information

- The **mode of a binomial distribution** is the value of x that has the highest probability of occurrence.
- In a binomial distribution with parameters n and p , the mode can be found using the formula:
- **$\text{mode} = \text{floor}(n * p)$**
- Where "floor" represents the greatest integer less than or equal to the argument.
- Given that the mean of the distribution is 15 and the variance is 10, we can find n and p by solving two equations.
- The mean of a binomial distribution is equal to $n * p$, and the variance is equal to $n * p * (1 - p)$.
- Solving for n and p and substituting into the formula for the mode, we get $46/3$ as the answer.

89. Answer: a

Explanation:

The correct answer is 25

★ Key Points

Given:

$$\text{Mean} = np = 5$$

$$\text{Variance} = np(1-p) = 4$$

Calculation:

$$\text{Variance} = 4$$

$$\Rightarrow np(1-p) = 4$$

$$\Rightarrow 1-p = 4/5$$

$$\Rightarrow 1-p = 4/5$$

$$\Rightarrow p = 1/5$$

$$\text{Since, } np = 5$$

$$\Rightarrow n \times 1/5 = 5$$

$$\Rightarrow n = 25$$

∴ The value of n is 25.

Given the mean and variance of a binomial distribution, we can find the values of n and p. The mean of a binomial distribution is equal to $n * p$, and the variance is equal to $n * p * (1 - p)$. We can use these two equations to solve for n and p.

★ Additional Information

- The **binomial distribution** is a discrete probability distribution that describes the number of successes in a fixed number of independent **Bernoulli trials**.
- The **parameters of a binomial distribution** are n (the number of trials) and p (the probability of success in each trial).
- The **mean and variance** of a binomial distribution are $n * p$ and $n * p * (1 - p)$, respectively.
- Given the mean and variance of a binomial distribution, we can use these two equations to find the values of n and p .
- Understanding the **properties of binomial distributions** and how to find the mean and variance is important in many areas of statistics and data analysis.

90. Answer: a

Explanation:

The correct answer is -2.25

★ Key Points

$$\mu_1' = 2; \mu_2' = 8; \mu_3' = 14 \text{ (Given)}$$

If the first, second, and third moment about the origin is 2, 8, and 14, respectively, then Karl Pearson gamma coefficient of skewness Y_1 is:

$$Y_1 = \pm \sqrt{B_1} \text{ --- (1)}$$

$$\text{where } B_1 = \frac{(\mu_3')^2}{(\mu_2')^3}$$

$$\text{Now, } \mu_2 = \mu_2' - (\mu_1')^2$$

$$\Rightarrow 8 - (2)^2$$

$$\Rightarrow 4$$

$$\mu_3 = \mu_3' - 3(\mu_2')(\mu_1') + 2(\mu_1')^3$$

$$\Rightarrow 14 - 3(8)(2) + 2(2)^3$$

$$\Rightarrow 14 - 48 + 16$$

$$\Rightarrow -18$$

By putting the above two values in (1)

$$\Rightarrow Y_1 = -\sqrt{\frac{(-18)^2}{(4)^3}}$$

$$\Rightarrow Y_1 = -\sqrt{5.0625}$$

$$\Rightarrow Y_1 = -2.25$$

★ Additional Information

Important points about Karl Pearson coefficient of skewness:

- It measures the **asymmetry of a distribution** about its mean.
- A **positive value of skewness** indicates that the distribution is positively skewed, i.e., it has a long tail on the right side of the mean. A negative value of skewness indicates that the distribution is negatively skewed, i.e., it has a long tail on the left side of the mean.
- A value of 0 for skewness indicates that the distribution is symmetric about its mean.
- The **skewness** of a normal distribution is always 0.
- The skewness can be sensitive to outliers, so it should be used with caution in case of a distribution with extreme values.
- The skewness can be used in combination with other measures of central tendency and variability, such as mean, median, mode, variance, and standard deviation, to describe the shape of a distribution.

91. **Answer: b**

Explanation:

The correct answer is: "A and B are not independent."

★ Key Points

- **mutually exclusive events:**
 - Mutually exclusive events are also known as disjoint or incompatible events.
 - Mutually exclusive events are events that cannot occur simultaneously.
 - In other words, if one event occurs, the other event cannot occur.
- If two events A and B are mutually exclusive, then the probability of their union is equal to the sum of their individual probabilities:
 - $P(A \cup B) = P(A) + P(B)$.
- If two events are mutually exclusive, they are not independent. Independence means that the occurrence of one event does not affect the probability of the other event.
- A proper subset is a subset in which the set is not equal to the original set. So, $A \subset B$ is not a correct answer in this case.

★ Additional Information

- Mutually exclusive events are also known as disjoint or incompatible events
- Independence means that the occurrence of one event does not affect the probability of the other event
- Understanding the concept of mutually exclusive events is important in probability theory as it helps to calculate the probability of the union of two or more events.
- Understanding the relationship between mutually exclusive events and independence is also important as it helps to determine the relationship between different events in a probability model.

92. Answer: c

Explanation:

The correct answer is F-statistics is 7.5.

★ Key Points

- **F-statistics** is a ratio of the between-treatment mean square to the error mean square

- It's used in **ANOVA** to test the null hypothesis of equal population means between different groups or treatments.
- A larger F-statistic indicates larger difference between the groups, and a smaller value indicates more similarity between the groups.
- In ANOVA, the degrees of freedom for the between-treatment mean square is equal to the number of groups minus 1.
- The **degrees of freedom** for the error mean square is equal to the total number of observations minus the number of groups.

★ Additional Information

- 45/3 The **F-statistic** is calculated as the ratio of the between-treatment mean square (MSB) to the error mean square (MSE).
- $MSB = 45/3 = 15$ and $MSE = 32/16 = 2$, therefore $F = MSB/MSE = (15/2) = 7.5$.
- The F-statistic is used in ANOVA to test the null hypothesis that the population means are equal between different groups or treatments.
- A larger F-statistic value indicates a larger difference between the groups, and a smaller value indicates that the groups are more similar.

93. Answer: d

Explanation:

The correct answer is **(100, 20)**.

★ Key Points

- The **sample mean** is an estimator of the population mean, but the true population mean is usually not known.
- The **standard deviation** of the sample mean decreases as the sample size increases.
- In general, **larger sample sizes** lead to more accurate estimates of population parameters.

★ Additional Information

- The sample mean follows a normal distribution with mean 100 and standard deviation $4/3$.
- The sample size is 225.
- The standard deviation of the population, σ , can be calculated using the formula $\sigma = \text{standard deviation of the sample mean} * (\text{square root of the sample size})$, where the standard deviation of the sample mean is $4/3$.
- $\sigma = 4/3 * (\text{square root of } 225) = 20$
- So, the population mean is $\mu = 100$ and the standard deviation is $\sigma = 20$.

94. Answer: d

Explanation:

The correct answer is "Pre-election poll by media."

★ Key Points

- In a **census**, the entire population is studied and counted, so there is no need for inference or estimation.
- A **pre-election poll by media** is an example of using a sample to make inferences about a population.
 - The pollsters take a sample of people and ask them who they plan to vote for in an election.
 - Based on the responses from the sample, the pollsters estimate the opinions and voting preferences of the larger population.

★ Additional Information

- **Sampling is a statistical technique** used to make inferences about a population based on a sample.
- In **statistical inference**, the goal is to estimate population parameters, such as the mean or proportion, based on the sample data.
- The accuracy of the inference depends on several factors, including the sample size, the sampling method, and the representativeness of the sample.
- Sampling is often used in surveys, polls, and experiments to gather information about populations that are too large or difficult to study in their entirety.

95. Answer: b

Explanation:

The correct answer is **8.8**.

★ Key Points

- A **decile** is a **statistical measure** that divides the data into 10 equal parts.
- The **seventh decile (D7)** is the value that separates the first six parts (60% of the data) from the last four parts (40% of the data).
- To calculate the decile, the data must be sorted in ascending or descending order, and then the decile value can be found by taking the average of the two closest values that correspond to the desired decile.

★ Additional Information

- Deciles are commonly used to describe the distribution of data and to identify specific points or values in the data set.
- There are nine deciles in total: D1, D2, D3, D4, D5, D6, D7, D8, D9, and D10.
- D1 is the first decile, and it separates the lowest 10% of the data from the rest.
- D10 is the last decile, and it separates the highest 10% of the data from the rest.
- Deciles can also be used to calculate other summary statistics, such as percentiles and quartiles.

96. Answer: b

Explanation:

The correct answer is **F-ratio belongs to $[-\infty, \infty]$.**

★ Key Points

- The **F-ratio in ANOVA** (analysis of variance) is a statistical test used to determine if there are significant differences between the means of two or

more groups.

- The F-ratio is calculated as the ratio of two mean square values, which are estimated variances based on the sample data.
- The F-ratio follows the **F-distribution**, which is a continuous probability distribution that is always positive and has its support on the interval $(0, \infty)$.

★ Additional Information

- **ANOVA** is a powerful tool for comparing means and testing hypotheses about population means.
- The **null hypothesis (H_0)** in ANOVA is usually that all population means are equal, which is symbolically expressed as $H_0: \mu_1 = \mu_2 = \dots = \mu_n$.
- If the F-ratio calculated from the sample data is large enough, we can reject the null hypothesis and conclude that there are significant differences between some of the population means.
- The F-ratio is commonly used in a variety of fields, including psychology, economics, biology, and engineering, to compare group means and make inferences about population parameters.

97. Answer: d

Explanation:

The square of the multiple correlation coefficient is **0.64**.

★ Key Points

- The **multiple correlation coefficient (R)** is a statistic that measures the strength of the relationship between multiple independent variables and a dependent variable.
- The **square of the multiple correlation coefficient (R^2)** is a measure of the proportion of variation in the dependent variable that is explained by the independent variables.

★ Additional Information

- In this case, the square of the multiple correlation coefficient of 0.64 means that 64% of the variation in the dependent variable can be explained by the combination of the independent variables X, Y, and Z.

98. Answer: b

Explanation:

The correct answer is Sample size.

★ Key Points

- **Sample size** is the most relevant for deriving a point estimate.
- **Point estimates** are estimates of population parameters derived from sample data.
- To derive a point estimate, one must first take a sample from a population.
- The sample size is crucial because larger sample sizes generally lead to more accurate point estimates, while smaller sample sizes may result in less accurate estimates.
- The population size is not as relevant in deriving a point estimate because it is not directly used in the estimation process. Instead, the sample data is used to calculate the point estimate, which is then used to make inferences about the population.

★ Additional Information

- It's important to note that the accuracy of point estimates also depends on the sampling method used to collect the sample data.
- **Simple random sampling** is considered to be the most appropriate method for deriving point estimates because it provides a representative sample of the population, reducing the chances of bias in the estimation process.
- Additionally, point estimates can also be accompanied by measures of precision, such as **confidence intervals**, which provide information about the range of values that the **population parameter** is likely to fall within with a certain level of confidence. The width of the confidence interval is influenced by both the sample size and the variability of the sample data.

- In summary, sample size is a crucial factor in deriving accurate point estimates, while the population size is a less relevant factor. The accuracy of point estimates can be improved by using appropriate sampling methods and accompanying them with measures of precision.

99. Answer: b

Explanation:

The correct answer is **155.92**

★ Key Points

- The **Marshall-Edgeworth price index equals 155.92**.
- The Marshall-Edgeworth index is a price index used to measure changes in the price level between two periods.
 - It is based on the Laspeyres price index formula and is used to adjust for changes in both the quantity and quality of goods over time.
- The formula for the Marshall-Edgeworth index is:

$$ME = (\sum p_1 q_0) / (\sum p_0 q_0) = 395 / 260 = 155.92$$

★ Additional Information

- In this case, the value of 155.92 means that the price level has increased by 55.92% between the base year (0) and the current year (1).
- This increase in the price level could be due to inflation, changes in production costs, changes in the supply and demand for goods, or other factors.
- It's important to note that the Marshall-Edgeworth index is a simplified version of the price index and may not accurately reflect changes in the price level for certain goods or services.
- In practice, more sophisticated price index formulas, such as the Paasche index or the Fisher index, are used to measure changes in the price level.

100. Answer: d

Explanation:

The probability of type-II error is **0.20**.

★ Key Points

- In a **hypothesis test**, the probability of type-II error, also known as β
- It is the **probability** of failing to reject the null hypothesis when it is false.
- In other words, it is the probability of making a false negative.
- The **type-II error rate** depends on the choice of the critical region and the value of the alternative hypothesis.

★ Additional Information

- The **distribution function** has an unknown **parameter, θ** .
- The hypothesis test compares two values of $\theta : H_0 : \theta = 1 \text{ and } H_1 : \theta = 2$.
- The critical region is defined as $X \geq 0.4$.
- In hypothesis testing, a trade-off exists between the probabilities of type-I and type-II errors.
- As one decreases, the other increases.
- The critical region can be chosen to minimize the probability of type-II error for a given level of type-I error or vice versa.
- The choice of critical region and hypothesis test should be made based on the specific problem and desired outcomes.