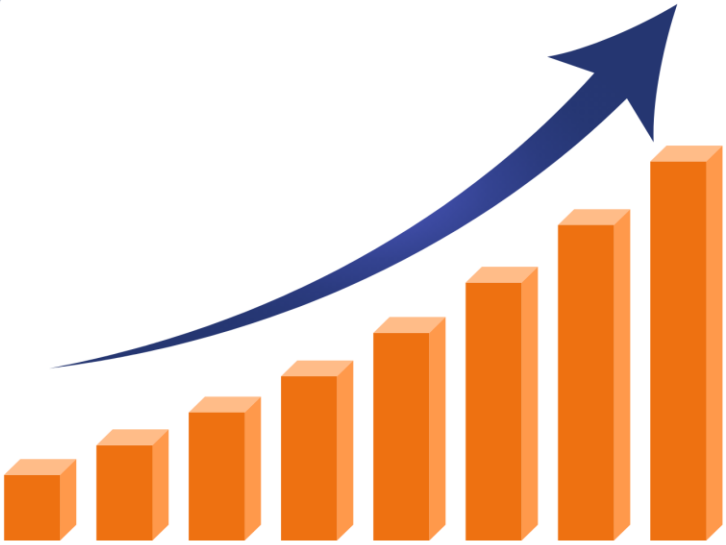
iNeuron



Statistics

31. Provide the formula for Bayes' Theorem.
Ans.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

32. What is a random variable?

Ans. A random variable is a variable that takes on different values based on chance or randomness.

33. Define the expectation (mean) of a random variable.

Ans. The expectation of a random variable is the average or mean value it takes over many repetitions.

34. What is simple random sampling?

Ans. Simple random sampling is a method where each member of the population has an equal chance of being selected.

35. Explain stratified sampling.

Ans. Stratified sampling involves dividing the population into subgroups (strata) and then randomly sampling from each stratum.

36. What is a null hypothesis?

Ans. The null hypothesis is a statement that there is no significant difference or effect.

37. Define Type I error in hypothesis testing.

Ans. Type I error occurs when a true null hypothesis is incorrectly rejected.

38. Define a confidence interval.

Ans. A confidence interval is a range of values used to estimate an unknown population parameter, with a specified level of confidence.

39. How do you interpret a 95% confidence interval?

Ans. It means that if the same population is sampled on numerous occasions, 95% of the calculated confidence intervals will contain the true population parameter.

40. Explain the Central Limit Theorem.

Ans. The Central Limit Theorem states that the distribution of the sum (or average) of many independent, identically distributed random variables approaches a normal distribution, regardless of the original distribution.

41. What is the Law of Large Numbers?

Ans. The Law of Large Numbers states that as the size of a sample increases, the estimate of the population parameter approaches the true value.

42. What does the standard deviation measure in a dataset?

a. Central tendency
b. Variability or dispersion
c. Skewness
d. Median
Ans. b) Variability or dispersion

43. Which measure of dispersion is least affected by outliers?

a. Range
b. Variance
c. Standard Deviation
d. Interquartile Range (IQR)
Ans. d) Interquartile Range (IQR)

44. The coefficient of variation is useful for comparing variability between datasets with different:

a. Skewness
b. Means
c. Medians
d. Ranges
Ans. b) Means.

45. Which of the following measures is affected by extreme values (outliers)?
a. Median
b. Mode
c. Range
d. Variance
Ans. d) Variance

46. What does the median represent in a dataset?

a. The most frequently occurring value
b. The average value
c. The middle value when the data is ordered
d. The sum of all values divided by the count
Ans. c) The middle value when the data is ordered

47. The mode is the value that occurs:

a. Least frequently
b. Most frequently
c. Exactly once
d. Only in the center of the dataset
Ans. b) Most frequently

48. What is the primary advantage of using the median over the mean?
a. Simplicity in calculation
b. Sensitivity to outliers
c. Robustness to outliers
d. Compatibility with all types of data
Ans. c) Robustness to outliers

49. The harmonic mean is most suitable for calculating averages of:
a. Datasets with extreme values
b. Datasets with symmetrical distributions
c. Rates or ratios
d. Integer values only
Ans. c) Rates or ratios

50. Which of the following measures is a resistant statistic, meaning it is not sensitive to outliers?
a. Median
b. Mean
c. Mode
d. Range
Ans. a) Median

51. What is the primary purpose of calculating the mid-range in a dataset?

a. Identifying the most common value
b. Estimating the center of the distribution
c. Measuring the spread of data
d. Providing a simple average
Ans. b) Estimating the center of the distribution