25-question test for Descriptive statistics: (30 mins)



Introduction to Statistics

1. What is the primary difference between a parameter and a statistic?

a) Parameters describe populations, while statistics describe samples.

b) Parameters are used for qualitative data, while statistics are for quantitative data. c) Parameters are always smaller than statistics.

d) Parameters are calculated from samples.

The correct answer is:

**a) Parameters describe populations, while statistics describe samples.**

2. Which of the following is an example of unstructured data? a) A CSV file of student grades.

b) Images collected from social media.

c) A database table of customer transactions.   
d) A sales report spreadsheet.

The correct answer is:

**b) Images collected from social media.**

3. Categorical data can be further classified into: a) Nominal and ordinal data.

b) Interval and ratio data.

c) Discrete and continuous data.   
d) Qualitative and quantitative data.

The correct answer is:

**a) Nominal and ordinal data.**

4. Which of the following is NOT a characteristic of structured data? a) Fixed schema.

b) Organized in rows and columns.

c) Easy to store in relational databases.   
d) Difficult to analyze with statistical tools.

The correct answer is:

**d) Difficult to analyze with statistical tools.**

5. What type of data is represented by the "number of cars owned by a family"?

a) Nominal   
b) Ordinal   
c) Discrete   
d) Continuous

The correct answer is:

**c) Discrete**



Descriptive Statistics: Measures of Central Tendency

6. What measure of central tendency is most affected by extreme values (outliers)? a) Mean   
b) Median   
c) Mode   
d) Range

The correct answer is:

**a) Mean**

7. If the mean and median of a dataset are equal, the data is likely to be: a) Positively skewed.

b) Negatively skewed.

c) Symmetrically distributed.   
d) Bimodal.

The correct answer is:

**c) Symmetrically distributed**

8. Which measure of central tendency is appropriate for categorical data?

a) Mean   
b) Median   
c) Mode   
d) Standard deviation

The correct answer is:

**c) Mode**

9. A dataset has values 2, 2, 3, 4, and 10. What is the median?

a) 2   
b) 3   
c) 4   
d) 10

The correct answer is:

**b) 3**

10. The mode of a dataset is defined as:

a) The middle value.

b) The most frequently occurring value.

c) The average of all values.

d) The difference between the highest and lowest value.

The correct answer is:

**b) The most frequently occurring value.**



Descriptive Statistics: Measures of Dispersion

11. What is the formula for calculating the range of a dataset? a) Maximum value −Minimum value   
b) Mean + Median

c) Sum of values/Number of values   
d) Median × 2

The correct answer is:

**a) Maximum value − Minimum value**

12. Standard deviation is a measure of:

a) Central tendency.

b) Variability or dispersion.   
c) Correlation.

d) Distribution symmetry.

The correct answer is:

**b) Variability or dispersion.**

13. What does a high coefficient of variation indicate?

a) Low relative variability.

b) High relative variability.   
c) Perfect correlation.

d) Symmetrical distribution.

The correct answer is:

**b) High relative variability.**

14. What is the interquartile range (IQR)?

a) Difference between the first quartile (Q1) and third quartile (Q3). b) Difference between the minimum and maximum value.

c) Average of all quartiles.

d) Sum of Q1 and Q3.

The correct answer is:

**a) Difference between the first quartile (Q1) and third quartile (Q3).**

15. A dataset has a standard deviation of 0. What does this mean?

a) All values are identical.

b) The data has no variability.   
c) Both a and b.

d) The mean is zero.

The correct answer is:

**c) Both a and b.**



Descriptive Statistics: Correlation and Skewness

16. A correlation coefficient of -1 indicates:

a) Perfect positive correlation.   
b) Perfect negative correlation.   
c) No correlation.

d) Weak negative correlation.

The correct answer is:

**b) Perfect negative correlation**

17. Positive skewness in a dataset means:

a) The tail is longer on the right side.   
b) The tail is longer on the left side.

c) The data is symmetrically distributed.   
d) The data has no outliers.

The correct answer is:

**a) The tail is longer on the right side.**

18. What does kurtosis measure in a dataset?

a) Central tendency.

b) The spread of data.

c) The sharpness of the peak of a distribution.   
d) Symmetry of the distribution.

The correct answer is:

**c) The sharpness of the peak of a distribution**

19. What is the range of the correlation coefficient (r)?

a) -2 to 2   
b) -1 to 1   
c) 0 to 1   
d) -0.5 to 0.5

The correct answer is:

**b) -1 to 1**

20. Which of the following indicates the strongest correlation? a) ( r = 0.75 )   
b) ( r = -0.85 )   
c) ( r = 0.50 )   
d) ( r = -0.65 )

The correct answer is:

**b) ( r = -0.85 )**



Descriptive Statistics: Five Point Summary and Visualization

21. What are the components of the five-point summary? a) Mean, median, mode, variance, and standard deviation. b) Minimum, Q1, median, Q3, and maximum.

c) Mean, range, skewness, kurtosis, and IQR.   
d) Minimum, Q1, Q3, IQR, and range.

The correct answer is:

**b) Minimum, Q1, median, Q3, and maximum.**

22. What type of chart is best for visualizing the distribution of a numerical dataset? a) Line chart   
b) Histogram   
c) Pie chart   
d) Bar chart

The correct answer is:

**b) Histogram**

23. Which visualization is best for identifying outliers?

a) Box plot   
b) Histogram   
c) Line chart   
d) Scatter plot

The correct answer is:

**a) Box plot**

24. What does a positively skewed box plot look like?   
a) The median is closer to Q3, with a longer tail on the left. b) The median is closer to Q1, with a longer tail on the right. c) The box is symmetrical.

d) There are no outliers.

The correct answer is:

**b) The median is closer to Q1, with a longer tail on the right.**

25. Scatter plots are primarily used to:   
a) Show frequency distribution.

b) Display correlation between two variables.   
c) Visualize the central tendency.

d) Summarize the five-point data.

The correct answer is:

**b) Display correlation between two variables.**

