15 problems based on Descriptive Stats: 30 mins



Introduction to Statistics

1. Problem:

A researcher collects the following data on the heights (in cm) of a sample of five plants:

120, 125, 130, 135, 140 .  **[C]**

Classify the type of data as:

a) Structured b) unstructured

c) Numerical d) categorical

2. Problem:

A survey records the following data for 10 individuals: their age, favorite color, and hours

spent on social media per day.

Identify the types of data for:

a) Age

b) Favorite color

c) Hours spent on social media

**Ans: a) Numerical**

**b) Categorical**

**c) Numerical**



Measures of Central Tendency

3. Problem:

Calculate the mean, median, and mode for the dataset:

3, 7, 7, 10, 15, 20 .

**Ans:**

**Mean = 10.3**

**Median = 8.5**

**Mode = 7**

4. Problem:

The weights (in kg) of five parcels are: 12, 15, 18, 21, 25 .

Add an outlier weight of 50 . How does this affect the mean and median?

**Ans:**

**Mean = 18.2 Median = 18**

**Mean(after adding an outlier) = 23.5 Median(After adding an outlier) = 19.5**

* **Mean increases more than median**

**Mean is more sensitive to outliers**



Measures of Dispersion

5. Problem:

Find the range and interquartile range (IQR) for the dataset:

5, 10, 15, 20, 25, 30, 35 .

**Ans:**

**Range = Max value – Min value**

**= 35- 5 = 30**

**IQR = Q3-Q1**

**Q1 = Median of first half of the data = 10**

**Q3 = Median of second half of the data = 30**

**IQR = 30-10 =20**

6. Problem:

A dataset has a standard deviation of . If all values in the dataset are doubled, what is the 5

new standard deviation?

**Ans:**

**If all the values in the data set are doubled then standard deviation also doubled.**

**Given standard deviation is 5 then new standard deviation is 10**

7. Problem:

Calculate the coefficient of variation for a dataset with a mean of 50 and a standard

deviation of .

Ans : Standard deviation is not mentioned



Correlation and Skewness

8. Problem:

|  |  |  |  |
| --- | --- | --- | --- |
| Two variables, X | and Y | , have a correlation coefficient of 0.85 | . Interpret this value. |

**Ans:**

**The correlation coefficient value is positive so it is strong positive relation**

9. Problem:

A dataset has a positive skew. Which measure of central tendency (mean, median, or mode) is likely the largest?

**Ans :**

**Mean will be likely to be the largest**

**For positive skew : Mean > Median>Mode**

10. Problem:   
Calculate the Pearson correlation coefficient for the following paired data: X : 1, 2, 3, 4   
Y : 2, 4, 6, 8

**Ans:**

**Here, Y = 2X**

**The relation is positively skewed and Pearson coefficient 1.**



Five Point Summary and Visualization

11. Problem:   
Determine the five-point summary for the dataset: 5, 8, 12, 14, 18, 20, 24 .

**Ans: Min value = 5**

**Max value= 24**

**Median = 14**

**Q1 = 8**

**Q3 = 20**

12. Problem:   
A box plot shows the median closer to Q1, with a long tail extending to the right. What does this indicate about the dataset's skewness?

**Ans: This indicates that the data is Positively skewed**

13. Problem:   
Construct a histogram for the following dataset: 2, 2, 3, 3, 3, 4, 5, 6, 6, 7 .

Suggest appropriate bin sizes.

**Ans :**

**Bin size : 2-3, 4-5, 6-7**



Application Problems

14. Problem:   
A factory measures daily production output (units): 200, 210, 190, 220, 230, 240, 205 .

Find the standard deviation.

**Ans:**

**Mean = 213.57**

**Variance = 262.17**

**Standard deviation = 16.19**

15. Problem:   
 You are analyzing sales data for two products.

Product A: Mean sales = 100 , Standard deviation = 20 , Standard deviation = 30 Product B: Mean sales = 150   
Which product has higher relative variability?

**Ans:**

**Variability of product A = (20/100) \* 100%=20%**

**Variability of Product B= (30/150) \* 100% = 20%**

**Both are giving 20% both products have same relative**

