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 .

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 - Numerical

b) Favorite color - Categorical

c) Hours spent on social media - Continuous



Measures of Central Tendency

3. Problem:

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

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

Mean = (3+7+7+10+15+20) /6 = 62/6 = 10.3

Median = (7+10) / 2 = 8.5 as dataset is having even numbers

Mode = 7 as it is having more freq compared to other data in the data

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?

Mean without outlier = (12+15+18+21+25)/5 = 18.2

Median without outlier = 18 as dataset contain odd no. of values.

Mean with outlier = (12+15+18+21+25+50) / 6 = 23.5

Median with outlier = (18+21) / 2 = 39/2 = 19.5

Mean and Median values with outlier have shown increase in the value.



Measures of Dispersion

5. Problem:

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

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

Range = 35 – 5 = 30

Median = 20

Q3 = 30

Q1 = 10

IQR = Q3 – Q1 = 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?

7. Problem:

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

deviation of 15 .

CV = (SD/Mean) \* 100 = (15/50) \*100 = 30 %



Correlation and Skewness

8. Problem:

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

It means values of X and Y are highly dependent on each other.

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

Mean > Median

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

X = np.array([1,2,3,4])

Y = np.array([2,4,6,8])

np.corrcoef(X,Y)

array([[1., 1.],

[1., 1.]])

Indicates perfect positive relationship.



Five Point Summary and Visualization

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

Min = 5

Q1 = 8

Q2 = 14

Q3 = 20

Max = 24

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?

Data is left skewed

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

Suggest appropriate bin sizes.

sns.histplot(data=df1, bins=5)



Application Problems

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

Find the standard deviation.

data = [200, 210, 190, 220, 230, 240, 205]

Sd = np.std(data)

Sd = 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?

Both have same coefficient of variability.

