

Summary of Lecture 2: Measure of Location of Data

Unit 2 : Introduction to Measures of Location

Objective: Understand and apply measures of location such as median, quartiles, and percentiles to interpret data, with examples from sport business statistics.

Unit 2.1: Median

2.1.1 Definition and Formula

- **Median:** The median is the middle value of a data set when the values are arranged in ascending order. If the number of observations (n) is odd, the median is the middle number. If n is even, it is the average of the two middle numbers.

Formula: • For an odd number of observations:

$$\text{Median} = X_{(\frac{n+1}{2})}$$

- For an even number of observations:

$$\text{Median} = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$$

2.1.2 Worked Example

Example: Consider the following data set representing the scores of a football team in 10 matches: [1, 2, 3, 5, 8, 9, 12, 14, 16, 18]

- Arrange in ascending order: [1, 2, 3, 5, 8, 9, 12, 14, 16, 18]
- Number of observations (n) = 10 (even)
- Median : $\frac{8+9}{2} = 8.5$

Interpretation: we have the data set of the scores of a football team in 10 matches. The median score is 8.5, indicating that half of the team's scores are below or equal to 8.5 and the other half are above or equal to 8.5.

2.1.3 Try-It Problem

Consider a data set representing the scores of a basketball team in 10 games:

[3, 7, 8, 10, 12, 15, 16, 18, 21, 25] Calculate the median of the above data set. Interpret your findings.

Unit 2.2: Quartiles

2.2.1 Definition and Formula

- **Quartiles:** Quartiles divide the data into four equal parts. They include:
 - **Q1** (First Quartile): The 25th percentile.
 - **Q2** (Second Quartile): The median or 50th percentile.
 - **Q3** (Third Quartile): The 75th percentile.

Formula for Quartiles:

- To find Q_k (where $k = 1, 2, 3$):

$$Q_k = X_{\left(\frac{k(n+1)}{4}\right)}$$

2.2.2 Worked Example

Example: Using the previous data set of football scores: [1, 2, 3, 5, 8, 9, 12, 14, 16, 18]

- Q1:** Position = $\frac{10+1}{4} = 2.75$ (interpolate between 2nd and 3rd values)

$$Q1 = 2 + 0.75 \times (3 - 2) = 2.75$$

- Q3:** Position = $\frac{3 \times (10+1)}{4} = 8.25$ (interpolate between 8th and 9th values)

$$Q3 = 14 + 0.25 \times (16 - 14) = 14.5$$

Interpretation:

- Q1** = The data of football scores is given, we found the first quartile **2.75**, indicating that 25% of the scores are below or equal to 2.75 and 75% of the scores are above or equal to 2.75
- Q3** = The data of football scores is given, we found the first quartile **14.5**, indicating that 75% of the scores are below or equal to 14.5 and 25% of the scores are above or equal to 14.5

2.2.3 Try-It Problem

- A hospital is analyzing the recovery times (in days) for 12 patients after a routine surgical procedure. The medical staff wants to understand the distribution of recovery times by finding the first and third quartiles.

Recovery Times for 12 Patients (in days):

7,8,9,10,11,12,13,15,16,18,20,22

- Calculate Q1 and Q3. Interpret your results

Unit 2.3 : Percentiles

2.3.1 Definition and Formula

- Percentiles:** Percentiles divide the data into 100 equal parts. The p th percentile is the value below which $p\%$ of the data falls.

Formula for the p th Percentile: $P_p = X_{\left(\frac{p(n+1)}{100}\right)}$

2.3.2 Worked Example

Example: Using the football scores data set: [1, 2, 3, 5, 8, 9, 12, 14, 16, 18]

$$\text{20th Percentile (P}_{20}\text{)} = \frac{20 \times (10+1)}{100} = 2.2 \text{ (interpolate between 2nd and 3rd values)}$$

$$P_{20} = 2 + 0.2 \times (3 - 2) = 2.2$$

Interpretation: The 20th percentile for the football score is **2.2**, meaning 20% of the scores are below or equal to 2.2 and 80% of the scores are above or equal to 2.2

4.3 Try-It Problem

- A sports merchandise store is reviewing the daily sales of jerseys (number of jerseys sold) over 12 days. The store manager wants to find the 30th percentile (P_{30}) to see the lower end of their sales, representing the days with the least demand.

Calculate the 30th percentile to identify the lower bound of daily jersey sales. Interpret your result

Homework Questions

1. A fitness center is evaluating the performance of its clients based on the time taken (in minutes) to complete a standard endurance test over 14 sessions. The trainer wants to identify the 40th percentile (P_{40}) to see what time marks the bottom 40% of performance.

Endurance Test Completion Times (in minutes):

40,42,44,46,48,50,51, 51, 52, 54, 56,58,60,62

Objective: Find the 40th percentile to understand the performance threshold for the bottom 40% of clients. Provide interpretation

2. A sports Analytics Company is evaluating the performance of a basketball player based on the number of points scored in 12 consecutive games. The analyst wants to determine the third quartile (Q_3) to see how the player's top 25% of performances compare to the rest.

Points Scored in 12 Games:

12,15,18,20,22,24,26,28,30,32,35

Objective: Find the third quartile (Q_3) to identify the score below which 75% of the games fall. Provide interpretation