

Measures of Central Tendency and Measures of Location (Part 2)

Munmun Akter

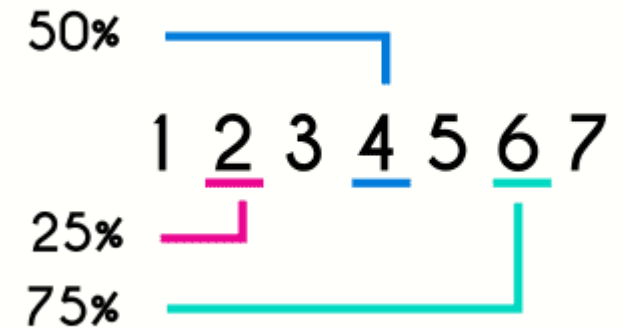
Lecturer

NFE, DIU

Quartiles

- are a set of three values that divide an ordered dataset into four equal parts, each containing approximately 25% of the data.
- They are used to analyze the spread and distribution of data, especially when dealing with large datasets or data that may have outliers.
- To find the quartiles of a dataset, the data must be arranged in ascending order first.

Quartiles



3 quartiles are denoted as Q1, Q2 (the median), and Q3.

1. Q1 (First Quartile): This is the value that separates the lowest 25% of the data from the rest (the remaining 75%). It is located at the 25th percentile of the dataset.

2. Q2 (Second Quartile or Median): This is the middle value of the dataset when arranged in ascending order. It separates the lower 50% of the data from the upper 50%. Q2 is also the 50th percentile.

3. Q3 (Third Quartile): This is the value that separates the lowest 75% of the data from the highest 25%. It is located at the 75th percentile of the dataset.

Quartile Formula

**The Quartile Formula
For Q1** = $\frac{1}{4} (n + 1)^{\text{th}}$ term

**The Quartile Formula
For Q3** = $\frac{3}{4} (n + 1)^{\text{th}}$ term

**The Quartile Formula
For Q2** = $Q3 - Q1$ (Equivalent to Median)

Consider the ordered dataset: {10, 20, 25, 30, 40, 50, 60, 70, 80, 90}

Percentile

- is a statistical measure used to describe the relative position of a particular data point within a dataset.
- It represents the value below which a given percentage of the data falls.
- Useful in understanding the distribution of data and identifying how a particular value compares to the rest of the dataset.
- For example, the 25th percentile (also known as the first quartile, Q1) indicates the value below which 25% of the data points fall.
- Percentiles are widely used in various fields, such as
 - education (e.g., standardized test scores),
 - finance (e.g., income distribution), and
 - health (e.g., growth percentiles for children)

- To calculate a specific percentile for a dataset, you first need to sort the data in ascending order. Then, you can use the following formula to find the value at a given percentile:

$$\text{Position} = (\text{Percentile} / 100) * (N + 1)$$

- If the position calculated is a whole number, the corresponding data point is the value at the desired percentile.
- If the position is not a whole number, you can use linear interpolation to estimate the value at the percentile.
- $\text{Interpolation} = \text{Value at } n\text{th position} + (\text{Position} - n) * \{\text{Value at } (n+1)\text{th position} - \text{Value at } n\text{th position}\}$

Example 2

0 1 5 6 7 8 9 10 12 12 13 14 16 19 19

- Find the P_{50} of the following set of data.

19 12 16 0 14

9 6 1 12 13

10 19 7 5 8

$$P_{50} = \left(\frac{i \cdot (n + 1)}{100} \right)^{th}$$

$$P_{50} = \left(\frac{50 \cdot (15 + 1)}{100} \right)^{th}$$

$$P_{50} = \left(\frac{800}{100} \right)^{th}$$

$$P_{50} = 8^{th}$$



Consider the sorted dataset: {10, 15, 20, 25, 30, 35, 40, 45, 50, 55}

Desired Percentile: 75th percentile (Q3)

Deciles

- used to divide an ordered dataset into ten equal parts, each containing approximately 10% of the data.
- used to analyze the spread and distribution of data, similar to percentiles, but with a higher level of granularity.
- To find the deciles of a dataset, the data must be arranged in ascending order first.
- The ten deciles are denoted as D1, D2, D3, ..., D10.
- D1 (First Decile): This is the value that separates the lowest 10% of the data from the rest (the remaining 90%).
- D2 (Second Decile): This is the value that separates the lowest 20% of the data from the rest (the upper 80%) and so on.

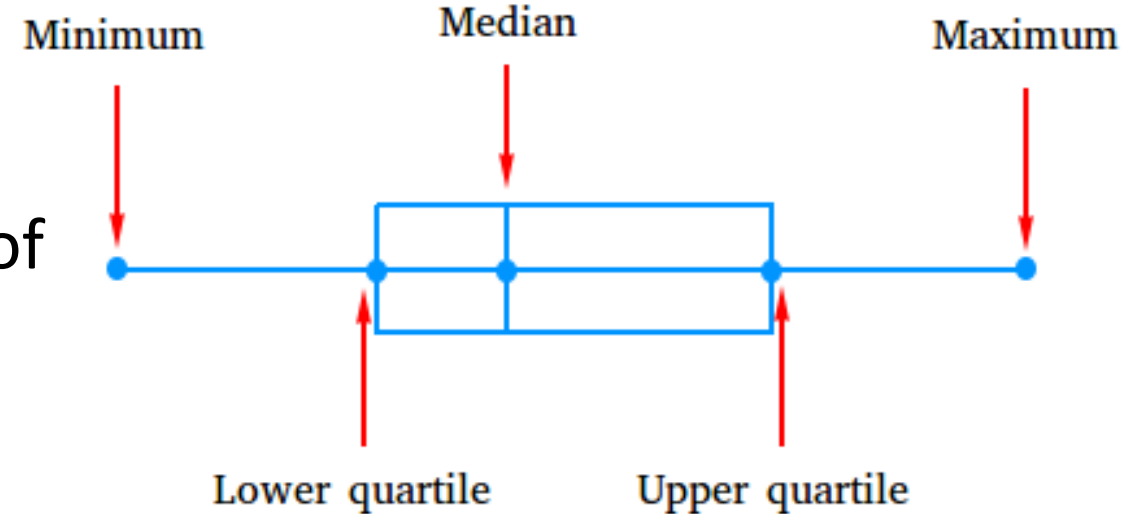
$$k^{\text{th}} \text{ decile} = \left[\frac{k \cdot (n + 1)}{10} \right]^{\text{th}} \text{ data}$$

Consider the ordered dataset: {80, 27, 25, 13, 64, 50, 96, 70, 38, 90, 55, 78, 15} and calculate 8th decile.

Box-Whisker plot

- A box plot, also known as a box-and-whisker plot
- is a graphical representation of the distribution of a dataset.
- It displays the five-number summary of the data, such as:
 1. Minimum values ,
 2. First quartile (Q1),
 3. Median (second quartile, Q2),
 4. Third quartile (Q3), and
 5. Maximum values
- The plot is particularly useful for identifying the central tendency, spread, and potential outliers in the data.

Box and Whiskers Plot



A box plot consists of the following components:

1.Box: The box represents the interquartile range (IQR), which is the range between the first quartile (Q1) and the third quartile (Q3). It spans the middle 50% of the data, with the median (Q2) marked as a line inside the box.

2.Whiskers: The whiskers extend from the edges of the box to the minimum and maximum values within a certain range. The range is determined by considering outliers.

3.Outliers: Data points that fall outside the whiskers are considered outliers and are represented as individual points on the plot. They are typically displayed separately from the main box plot to draw attention to their unusual nature.

How to create a box plot?

1. Order the dataset in ascending order.
2. Calculate Q1, Q2 (median), and Q3.
3. Calculate the interquartile range (IQR) as $Q3 - Q1$.
4. Determine the minimum and maximum whisker values:
 1. Minimum whisker: $Q1 - 1.5 * IQR$
 2. Maximum whisker: $Q3 + 1.5 * IQR$
5. Identify outliers, which are values outside the minimum and maximum whisker values.
6. Draw the box between Q1 and Q3, with the line representing the median (Q2).
7. Draw the whiskers extending from the box to the minimum and maximum non-outlier values.
8. Display the outliers as individual points outside the whiskers.

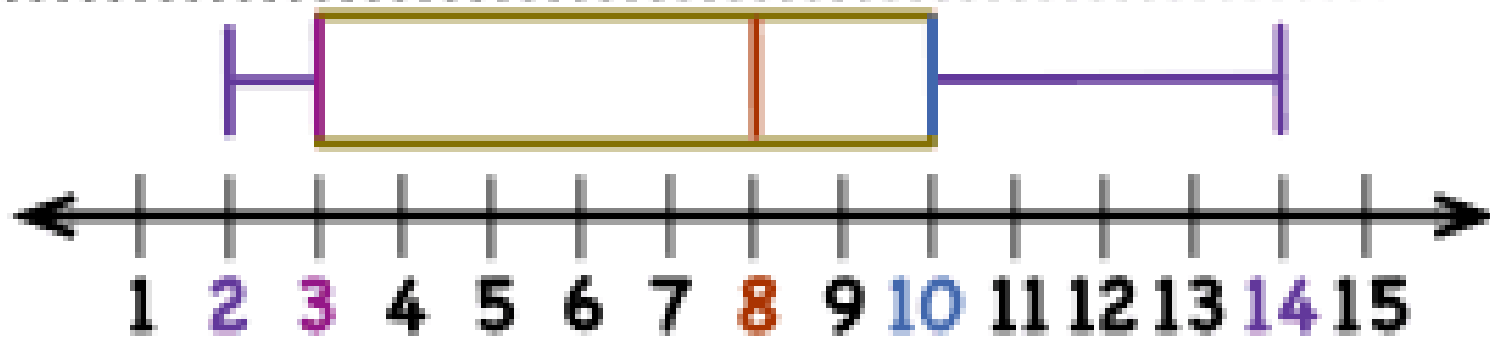
Make a Box-and-Whisker Plot

Hours Spent on Homework

9, 3, 10, 2, 6, 8, 3, 11, 14, 3, 4, 8, 9, 12, 5

2, 3, 3, 3, 4, 5, 6, 8, 8, 9, 9, 10, 11, 12, 14

↑ min ↑ Q_1 ↑ median ↑ Q_3 ↑ max



Thank You

Any Question?