

Descriptive statistics



Measures of central tendency

Measure	Formula	When to use
Mean (\bar{x})	$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$	No outliers; symmetric data
Median	Middle value when data is ordered	Outliers present; skewed data
Mode	Most frequent value	Categorical data; finding peaks

Example: Data: 3, 5, 7, 7, 9, 12, 28

$$\text{Mean} = \frac{3 + 5 + 7 + 7 + 9 + 12 + 28}{7} = \frac{71}{7} \approx 10.1 \quad \text{Median} = 7 \text{ (middle value)} \quad \text{Mode} = 7 \text{ (appears twice)}$$

The outlier (28) pulls the mean up, but the median is resistant.

Measures of spread

Measure	Formula
Range	Range = Max – Min
Variance (sample)	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$
Standard Deviation (sample)	$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$
Interquartile Range (IQR)	IQR = $Q_3 - Q_1$

Note: For a **population**, use N instead of $n - 1$ in the variance formula: $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$

Example: Data: 2, 4, 6, 8, 10 Mean: $\bar{x} = 6$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
2	-4	16
4	-2	4
6	0	0
8	2	4
10	4	16
		$\Sigma = 40$

$$s^2 = \frac{40}{5 - 1} = 10 \quad s = \sqrt{10} \approx 3.16$$

Student Learning Support, Teaching and Learning Centre

studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning



This document is licensed under Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Quartiles and the five-number summary

Order the data, then find:

- Q_1 (25th percentile): median of the lower half
- Q_2 (50th percentile): the median
- Q_3 (75th percentile): median of the upper half

Five-number summary: Min, Q_1 , Median, Q_3 , Max

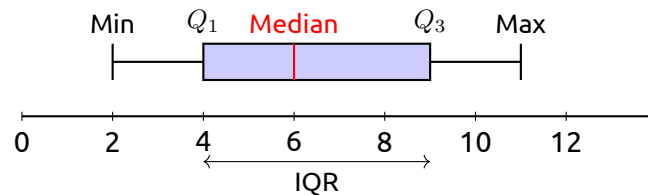
Example: Data: 1, 3, 5, 7, 9, 11, 13, 15, 17

Min = 1, $Q_1 = 4$, Median = 9, $Q_3 = 14$, Max = 17

IQR = $Q_3 - Q_1 = 14 - 4 = 10$

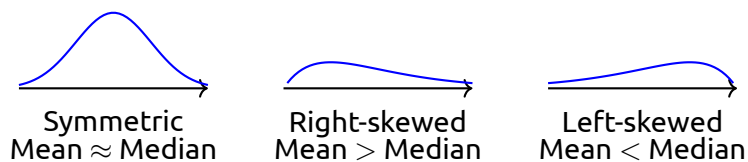
Boxplots

A boxplot displays the five-number summary graphically.



Outliers: Values below $Q_1 - 1.5 \cdot \text{IQR}$ or above $Q_3 + 1.5 \cdot \text{IQR}$ are outliers (plotted as individual points).

Shape of distributions



Tip: The mean is pulled toward the tail (toward the outliers/skew).