Why study Dispersion ?

- In the previous chapter we studied Mean, Median and Mode. These are called “Measures of Central Tendency”.
- It is necessary to describe the variability or dispersion.

Find out the mean of each of these series

Series A	Series B	Series C
100	100	1
100	105	489
100	102	2
100	103	3
100	90	5

Mean of each of these series is 100. But a close examination will reveal that the distributions vary widely.

	Series A	Series B	Series C
	100	100	1
	100	105	489
	100	102	2
	100	103	3
	100	90	5
Total	500	500	500
$\overline{X}$	100	100	100

Measures of variation

- Measures of variation point out as to how far an average is representative of mass.
- When dispersion is small, the average is the typical value.
- When the dispersion is large, average is not so typical.

Methods of studying variation

1. The Range
2. The Interquartile Range, and the Quartile Deviation.
3. The Mean Deviation or Average Deviation.
4. The Standard Deviation.

Methods of studying variation

1. The Range
2. The Interquartile Range, and the Quartile Deviation.
3. The Mean Deviation or Average Deviation.
4. The Standard Deviation.

1. Range

- Range is the simplest method of studying dispersion.
- It is defined as the difference between the value of the smallest item and the value of the largest item in the distribution.

Range formula

$$\text{Range} = L - S$$

L = Largest item, and

S = Smallest item.

Range Q – Individual Observation

Illustration 1. The following are the prices of shares of *XYZ* Co. Ltd. from Monday to Saturday :

Day	Price (Rs.)	Day	Price (Rs.)
Monday	200	Thursday	160
Tuesday	210	Friday	220
Wednesday	208	Saturday	250

Calculate range and its coefficient.

Solution.

$$\text{Range} = L - S$$

Here $L = 250$ and $S = 160$

$$\text{Range} = 250 - 160 = \text{Rs. } 90$$

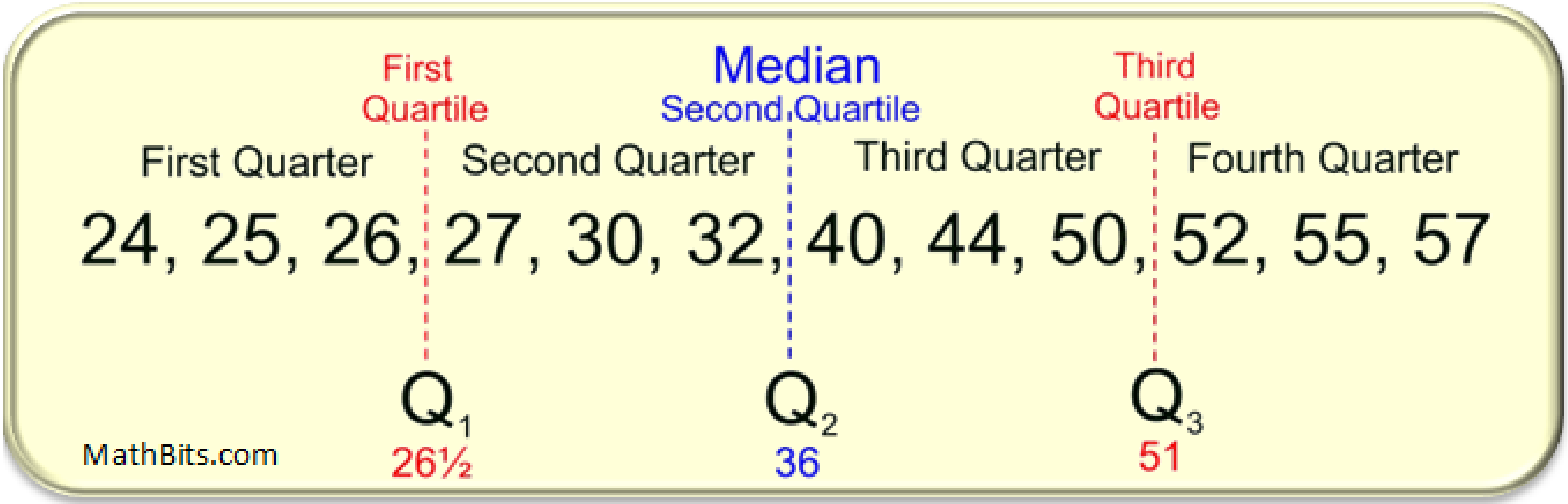
$$\text{Coefficient of Range} = \frac{L - S}{L + S} = \frac{250 - 160}{250 + 160} = \frac{90}{410} = 0.22.$$

Series	Values
A	46, 6, 46, 46, 46, 46, 46, 46
B	6, 10, 6, 6, 46, 46, 46, 46
C	6, 6, 15, 25, 30, 32, 40, 46

Find the Range of each of the three series. You will find that the Range is 40 in each of the three cases. Although the series are very different when it comes to individual elements and their variation.

Inter-Quartile Range

- Quartiles are values that divide a dataset into four equal parts, each containing 25% of the data points. They help describe the distribution of data by showing where the data is centered and how it spreads out. The key quartiles are:


First Quartile
Median
Second Quartile
Third Quartile
First Quarter
Second Quarter
Third Quarter
Fourth Quarter
24, 25, 26, 27, 30, 32, 40, 44, 50, 52, 55, 57
Q_1
Q_2
Q_3
MathBits.com
26½
36
51

Inter-Quartile Range

- The difference between Q3 and Q1 is called the **Interquartile Range (IQR)**, which measures the spread of the middle 50% of the data.

$$\text{Interquartile range} = Q_3 - Q_1$$

Coefficient of Quartile Deviation

- Interquartile Quartile deviation is an absolute measure of dispersion.
- The relative measure corresponding to this measure, called the coefficient of quartile deviation, is calculated as follows: Range is divided by 2 to obtain Quartile Deviation also known as Semi-inter-quartile range.

Merits and Demerits of Quartile Deviation

- Merit - It is superior to Range in some aspects, because we are not relying on only 2 values (Largest and Smallest, as in Range). But we are selecting 50% of data between Q1 and Q3.

- Demerit – It ignores 50% of data (Data before Q1 and Data after Q3. Also, it cannot be used in Mathematical calculations, because it is a position.

Variance and Standard Deviation

- Recall that while calculating mean deviation about mean or median, the absolute values of the deviations were taken.
- Another way to overcome this difficulty which arose due to the signs of deviations, is to take squares of all the deviations.

Variance σ^2

Variance (σ^2):

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

where:

- N is the number of data points in the population,
- X_i represents each data point,
- $\bar{X}$ is the mean of the population.

The Standard Deviation

- It is the square root of Variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

Variance and Standard Deviation

They are always calculated from Mean.

Therefore, first step always is to find the Mean.

Calculating Variance and Standard Deviation (Individual Observations)

Data: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24

Step 1: Calculate the Mean ($\bar{x}$)

We already know:

$$\bar{x} = 15$$

Step 2: Organize the Data and Calculate Deviations and Squared Deviations

We'll use a table with the following columns:

1. Data Point (x_i)
2. Deviation from Mean ($x_i - \bar{x}$)
3. Squared Deviation ($(x_i - \bar{x})^2$)

x_i	Deviations from mean $(x_i - \bar{x})$	$(x_i - \bar{x})^2$
6	–9	81
8	–7	49
10	–5	25
12	–3	9
14	–1	1
16	1	1
18	3	9
20	5	25
22	7	49
24	9	81
		330

$$\text{Variance } (\sigma^2) = \frac{1}{n}\sum_{i=1}^{10}(x_i - \overline{x})^2 = \frac{1}{10}\times 330 = 33$$

$$\text{Standard deviation } (\sigma) = \sqrt{33} = 5.74$$

Next Topic - Moments

In statistics, moments are quantitative measures used to understand the shape of a distribution. Moments are calculated from the data and are used to describe various characteristics of the distribution.

- **First Moment (Mean)**: This is simply the average of the data. It provides a measure of the central location of the distribution.
- **Second Moment (Variance)**: This measures the spread or dispersion of the data around the mean. It’s the average of the squared differences from the mean.
- **Third Moment (Skewness)**: This measures the asymmetry of the distribution around the mean. It can indicate whether the data leans to the left or right.

Fourth Moment (Kurtosis): This measures the "tailedness" of the distribution. It describes the shape of the distribution in terms of how heavy or light the tails are compared to a normal distribution.

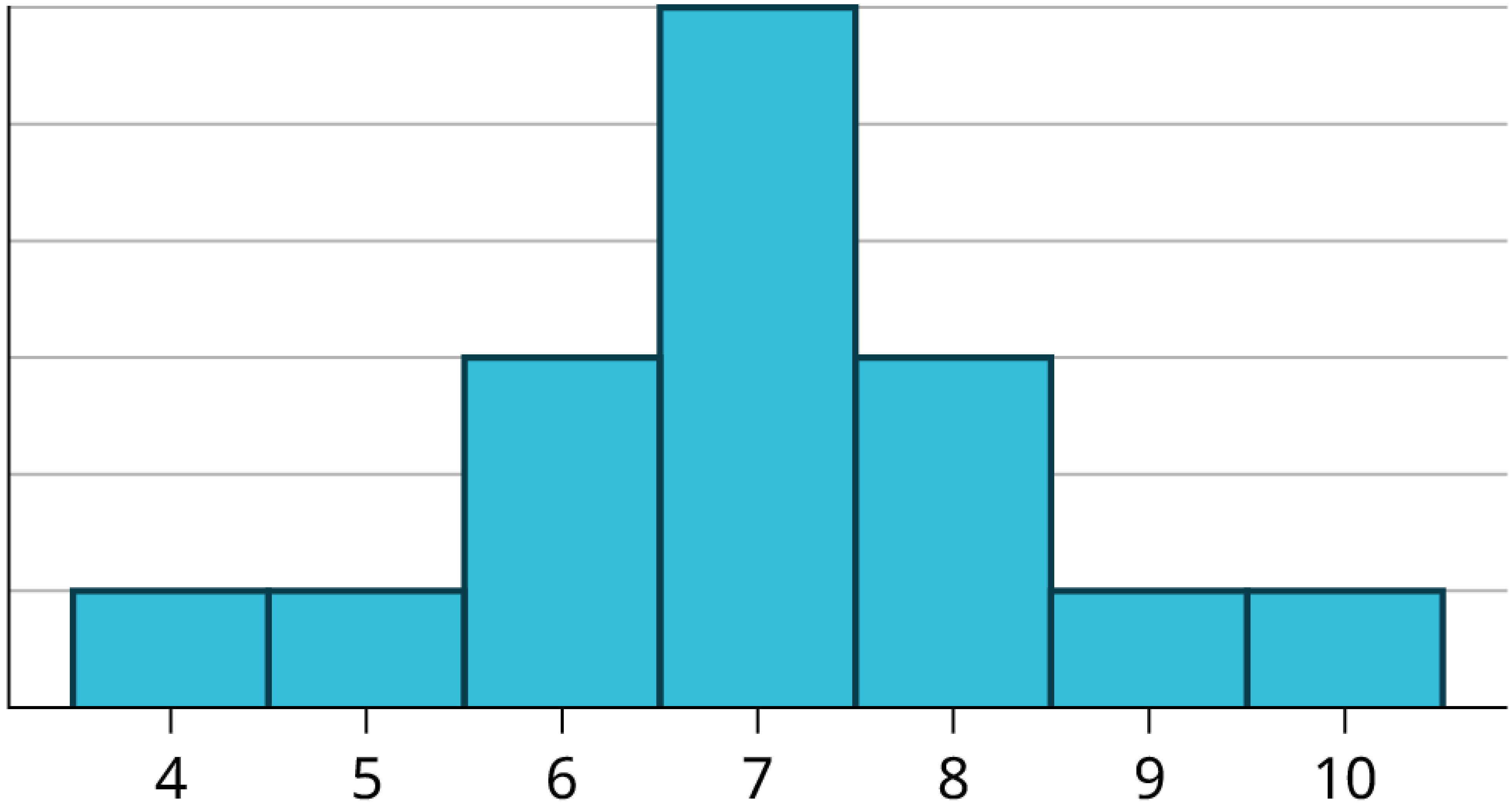
Skewness

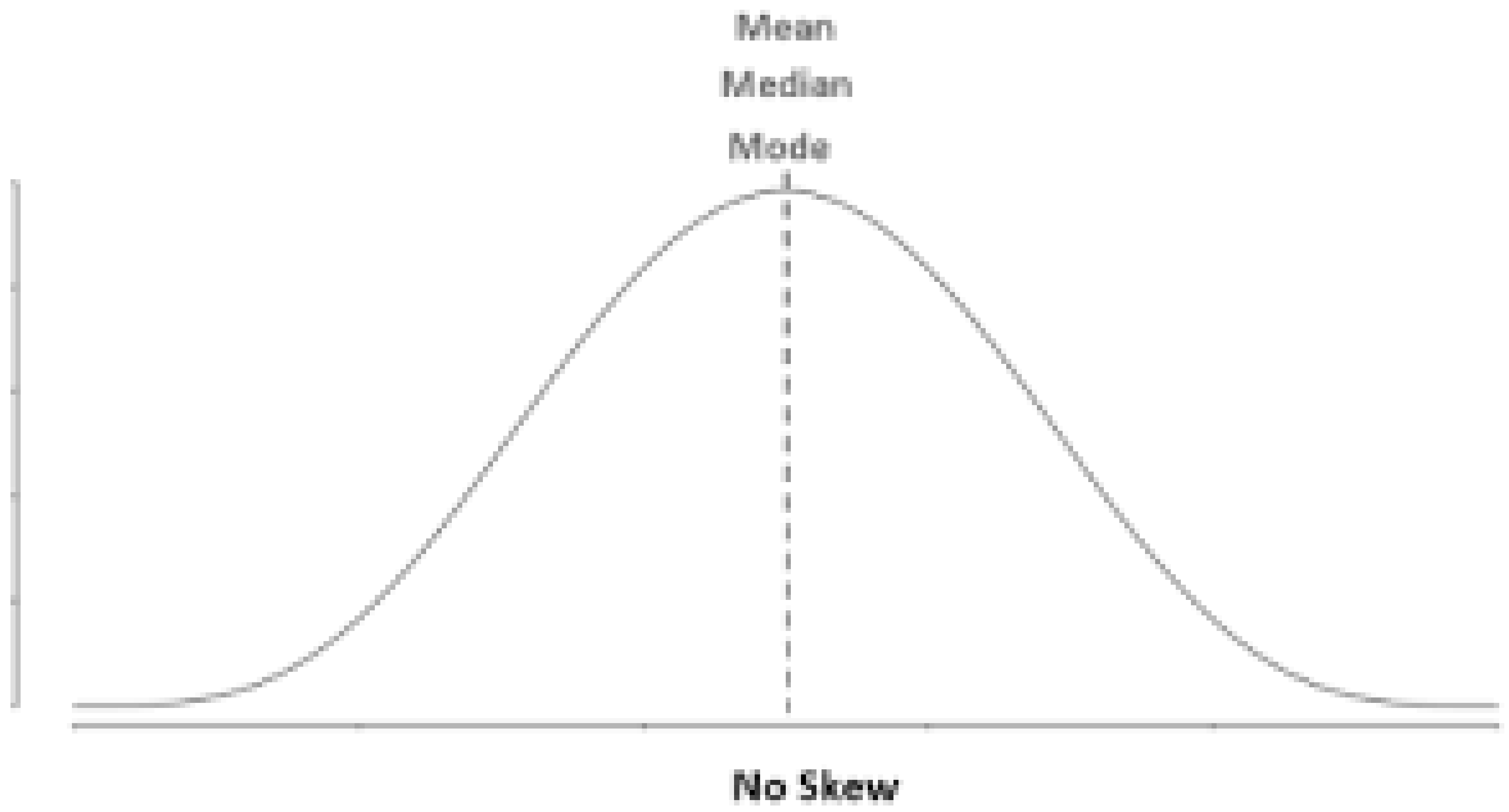
- Lack of symmetry is called skewness.

Perfectly symmetric distribution

4; 5; 6; 6; 6; 7; 7; 7; 7; 7; 7; 8; 8; 8; 9; 10

Perfectly symmetric distribution (No Skew)



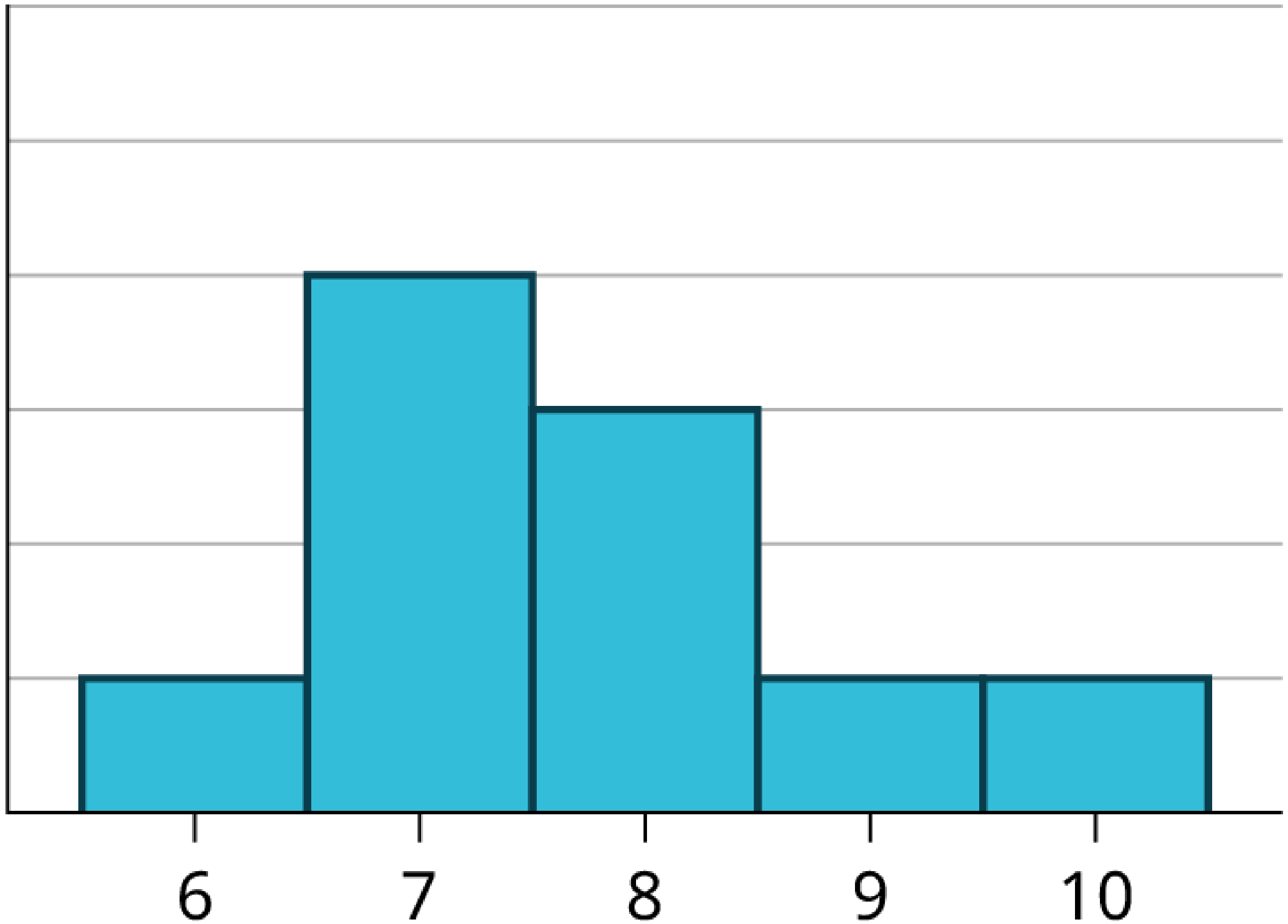

Mean
Median
Mode
No Skew

Perfectly symmetric distribution (No Skew)

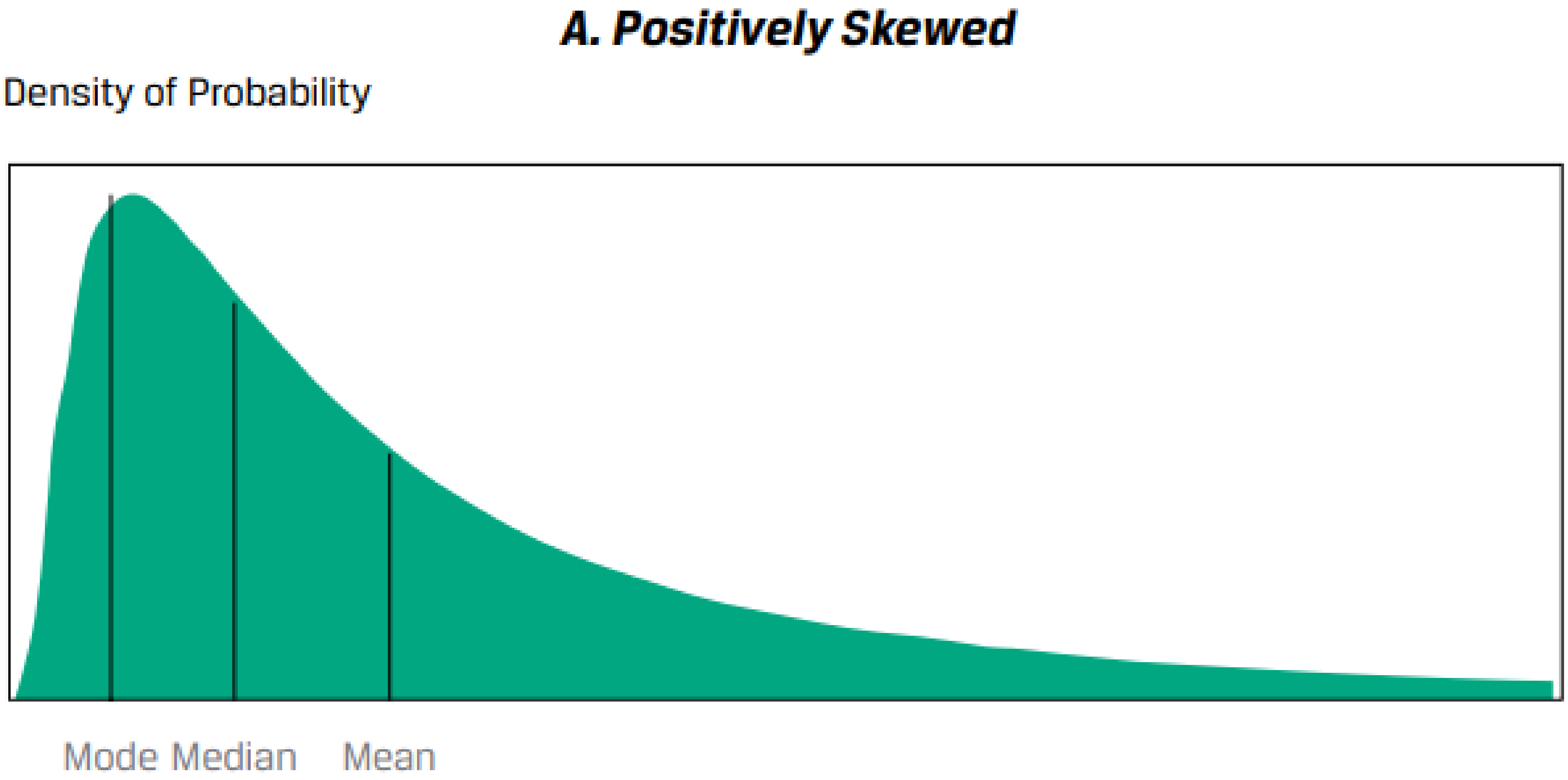
- In general, for perfectly symmetric and unimodal distributions (i.e., having a single mode), the mean, median, and mode are equal or very close to each other.

Positive Skewed or Skewed to Right

6; 7; 7; 7; 7; 8; 8; 8; 9; 10



Positive Skewed or Skewed to Right



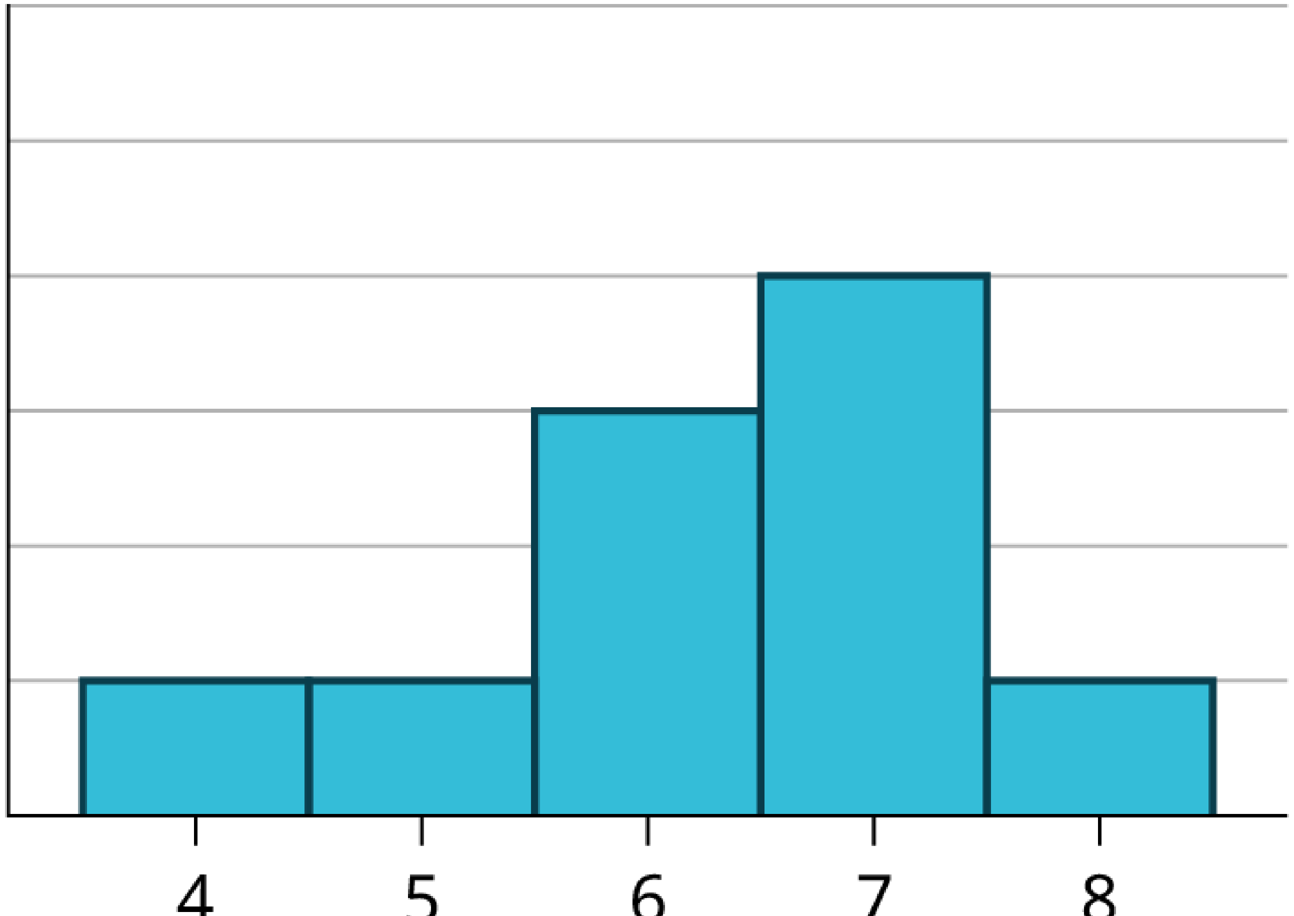
Positive Skewed or Skewed to Right - Example

Positive Skew Example:

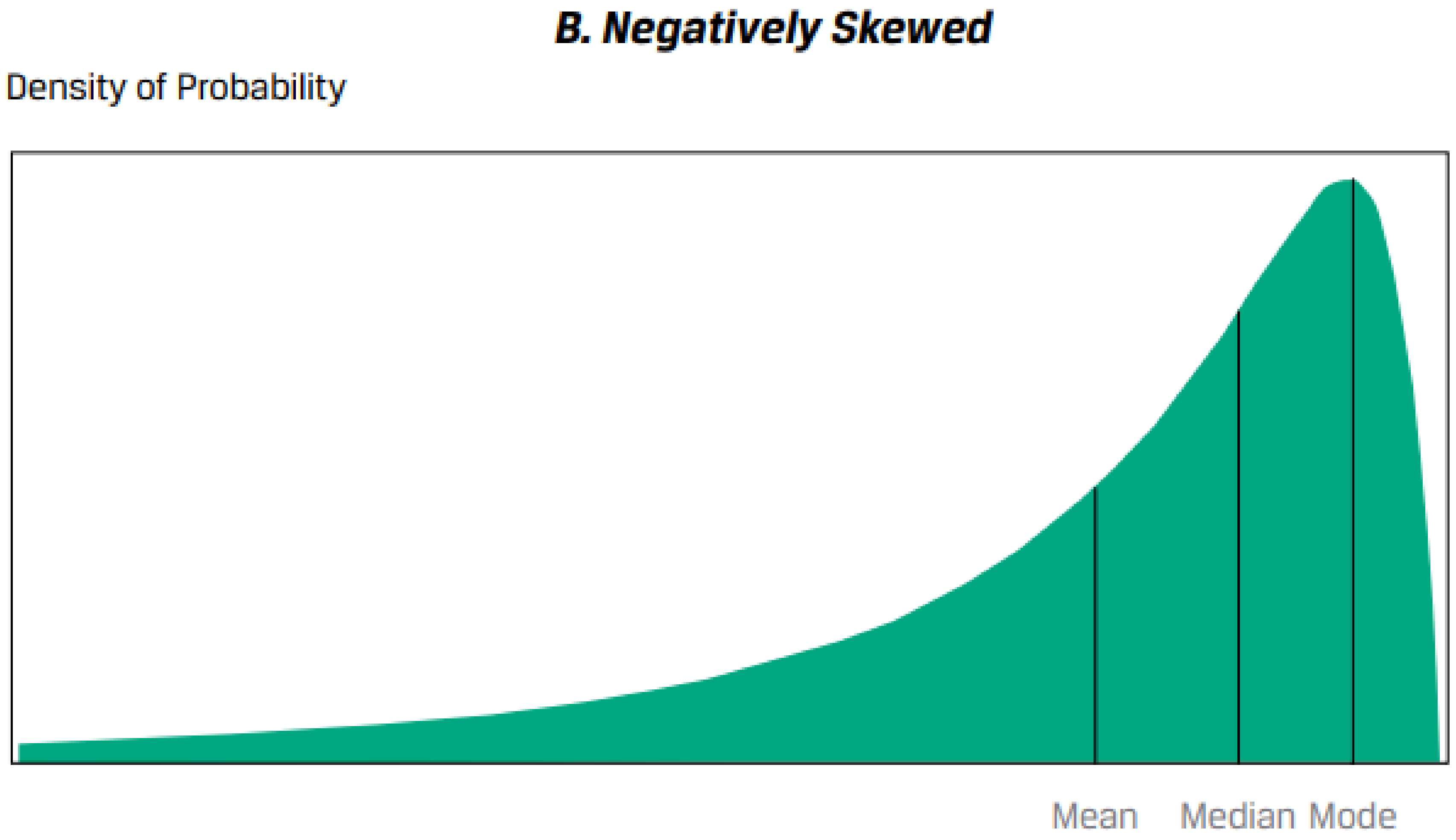
Gambling Winnings: Imagine you're playing a game where you usually lose a little bit each time you play, but occasionally you hit a jackpot and win a lot. Most of the time, your losses are small, but those few huge wins create a long tail to the right on your distribution graph. This is an example of positive skew because the distribution has a lot of small losses and a few extreme gains.

Negative Skewed or Skewed to Left

4; 5; 6; 6; 6; 7; 7; 7; 7; 8



Negative Skewed or Skewed to Left



Negative Skewed or Skewed to Left - Example

Retail Stock Performance: Suppose you invest in a retail stock that generally performs well but occasionally has severe downturns, like during a major economic crisis. Most of the time, you see small gains, but there are rare but significant losses. In this case, the distribution of returns would be negatively skewed, with frequent small gains and a few extreme losses.

Kurtosis

It is a measure of the combined weight of the tails of a distribution relative to the rest of the distribution

Leptokurtic

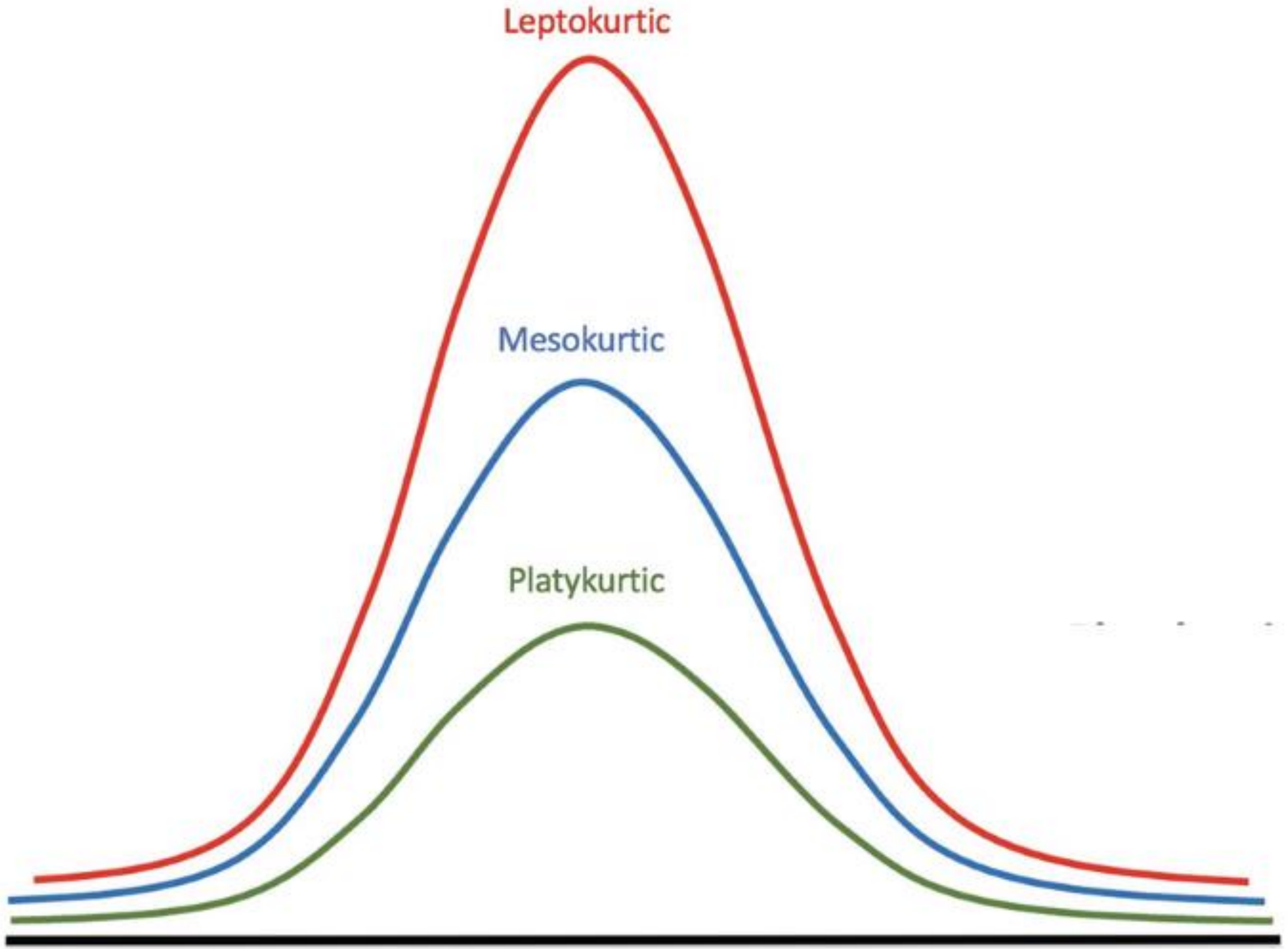
A distribution that has fatter tails than the normal distribution is referred to as **leptokurtic** or **fat-tailed**

Platykurtic

A distribution that has thinner tails than the normal distribution is referred to as being **platykurtic** or **thin-tailed**

Mesokurtic

A distribution similar to the normal distribution as it concerns relative weight in the tails is called **mesokurtic**.


Leptokurtic
Mesokurtic
Platykurtic