Lecture Five

Dr.Shaymaa Adil
MSc.Rana Adnan

Descriptive Biostatistics

The best way to work with data is to summarize and organize them.

Numbers that have not been summarized and organized are called raw data.

A descriptive measure is a single number that is used to describe a set of data.

Descriptive measures include measures of central tendency and measures of dispersion.

Descriptive Biostatistics

Commonly Used Symbols

For a Sample

$\bar{x}$	sample mean
s^2	sample variance
s	sample standard deviation

For a Population

μ	population mean
σ^2	population variance
σ	population standard deviation

Measures of central tendency

1. **Mean** (generally not part of the data set).
2. **Median** (may be part of the data set).
3. **Mode** (always part of the data set).

Arithmetic mean

The *arithmetic mean* is the "average" which is obtained by adding all the values in a sample or population and dividing them by the number of values.

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

μ = population mean

Σ = summation sign

x_i = value of element i of the sample

N = population size

Formula for Sample mean

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Properties of the mean

1. Uniqueness: For a given set of data there is one and only one mean.
2. Simplicity: The mean is easy to calculate.
3. Affected by extreme values : The mean is influenced by each value. Therefore, extreme values can distort the mean.

For example: A set of data (5, 7, 9, 5, 4).

$$X = \frac{5 + 7 + 9 + 5 + 4}{5} = 6$$

Other set of data with extreme value; (5, 7, 9, 5, 24)

$$X = \frac{5 + 7 + 9 + 5 + 24}{5} = 10$$

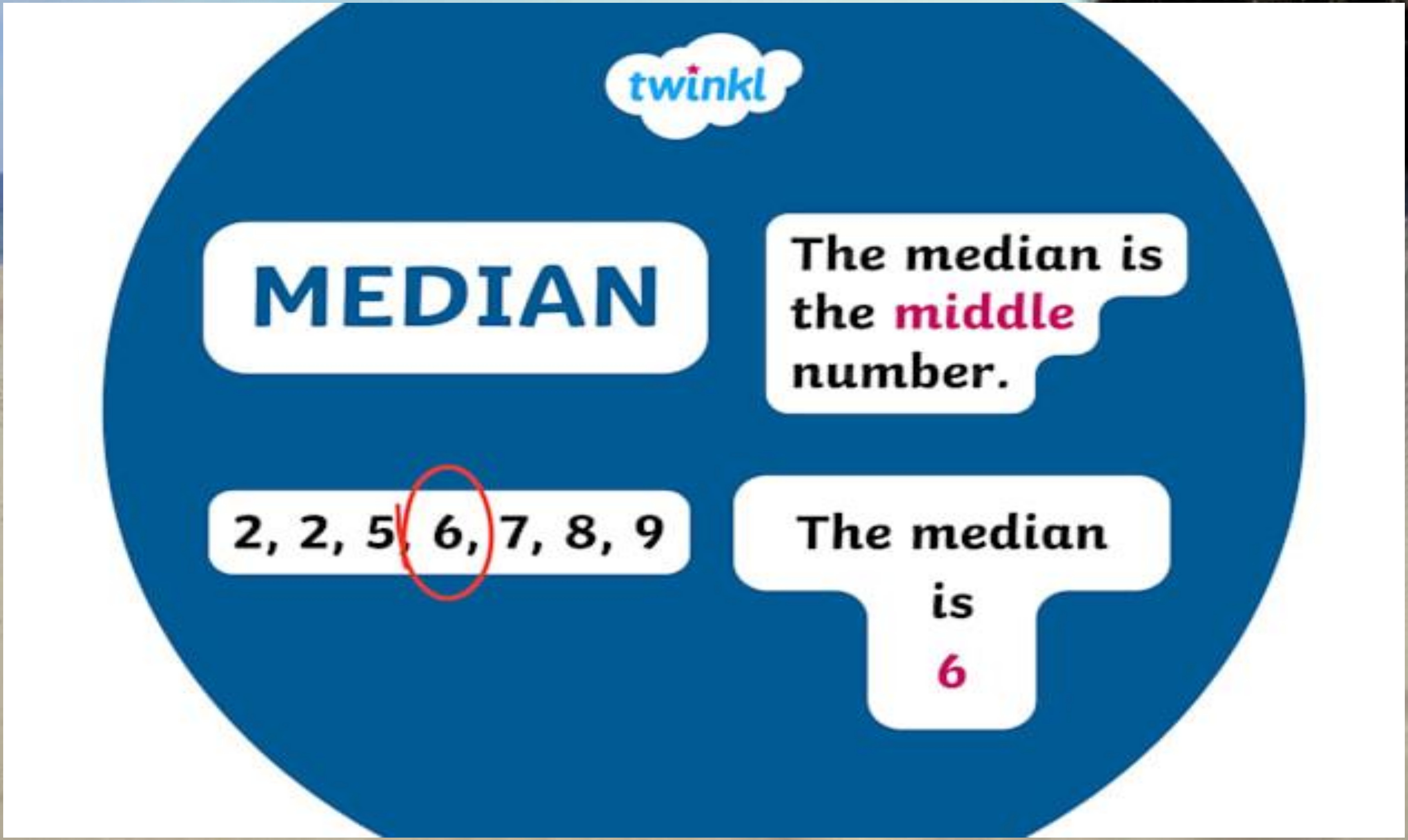
Median

The median is the value that divides the set of data into two equal parts. The number of values equal to or greater than median equals the number of values less than or equal the median.

Properties of the median:

1. Uniqueness: There is only one median for each set of data.
2. Simplicity: It is easy to calculate.
3. The median is not affected by extreme values as the mean.

Median



MEDIAN
The median is the middle number.
2, 2, 5, 6, 7, 8, 9
The median is 6

Finding the median

1. Arrange the data in order of increasing value in a sorted list.

2. Find the median.

a. Odd number of values (n is odd).

Median = X (n+1/2).

b. Even number of values (n is even).

Median = average of the two values in the middle.

Example: A set of data (10, 54, 10, 33, 21, 53).

1.Step one:

Ordered by size. (10, 10, 21, 33, 53, 54), (Don’t forget to order the values!).

2.Step two:

$$Median = \frac{X_{\frac{6}{2}} + X_{\frac{6}{2}+1}}{2} = \frac{X_3 + X_4}{2} = \frac{21 + 33}{2} = 27$$

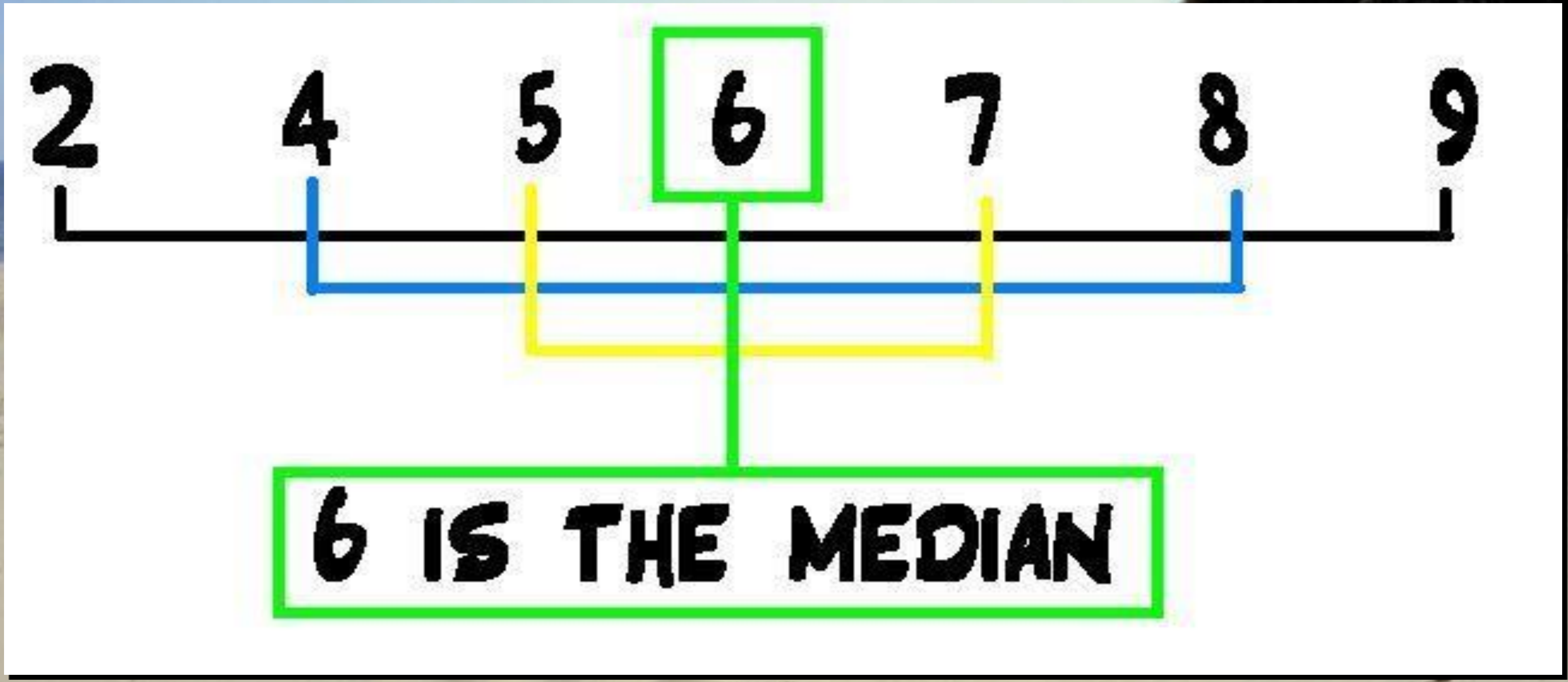
Example

Set of data (10,10,33,21,53)

Ordered by size (10,10, 21,33,53),(don’t forget to order the values).

$$Median = X_{\frac{5+1}{2}} = X_3 = 21$$

Median



Example:

The following data give the weight lost (in pounds) by a sample of five members of a health club at the end of two months of membership:

(10,5,19,8,3) . Find the median.

First, we rank the given data in increasing order as follows:

(3,5,8,10,19).

There are five observations in the data set. Consequently, n = 5 and

Median = X(n+1/2) = X(5+1/2)=X3

Therefore, the median is the value of the third term in the ranked data.

(3,5,8,10,19).

The median weight loss for this sample of five members of this health club is 8 pounds.

Mode

The mode is the value that occurs with the highest frequency in a data set.

It is possible to have more than one mode or no mode.

<u>Example:</u>

(5, 6, 2, 7, 0, 11, 4) → (no Mode)

(5, 5, 10, 10, 6, 7, 6, 6) → (one Mode,6)

(0, 2, 5, 4, 2, 0, 2, 4, 4) → (two Mode, 2, 4)

The data set with only one mode is called uni modal.

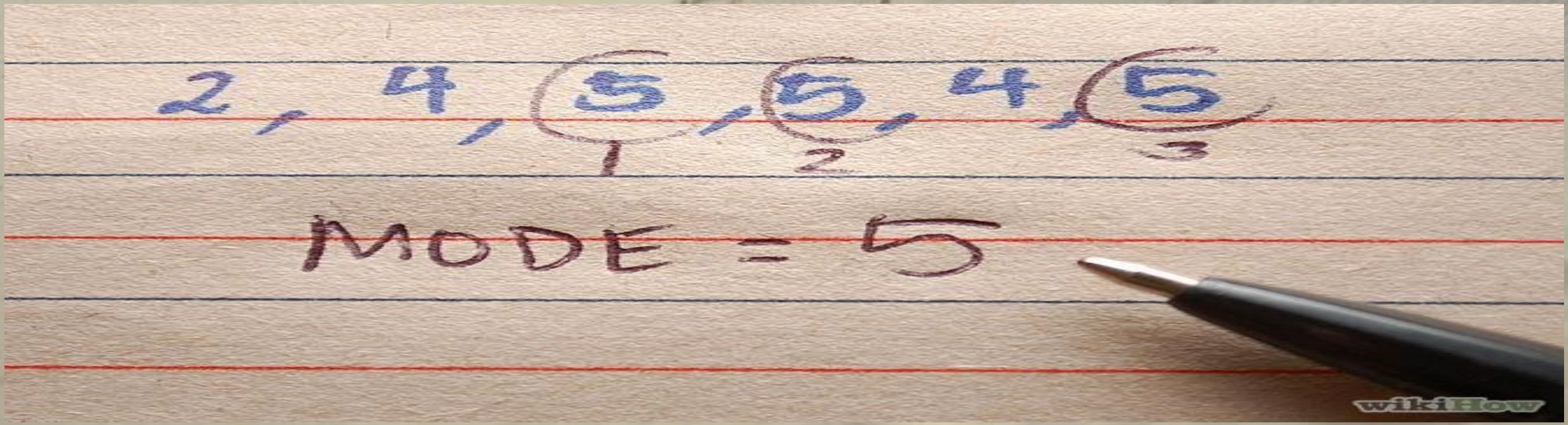
The data set with two modes is called bimodal.

The data set with more than two modes is called multimodal.

Example:

The ages of 10 randomly selected students from a class are 21, 19, 27, 22, 29, 19, 25, 21, 22 and 30. Find the mode.

This data set has three modes: 19, 21 and 22. Each of these three values occurs with a (highest) frequency of 2.



One advantage of the mode is that it can be calculated for both kinds of data, quantitative and qualitative, whereas the mean and median can be calculated for only quantitative data.

Example:

The status of five students who at a college are good , bad, very good, very good, very good. Find the mode.

Answer:

Because very good occurs more frequently than the other categories, it is the mode for this data set.

We cannot calculate the mean and median for this data set.

Comparison of the mean, median, and mode

Data set (2, 2, 4, 5, 7, and 9).

Mean = 4.833; Median = 4.5; Mode = 2

If we change the last data point from 9 to 28, Mean = 8, but the median and the mode remain the same.

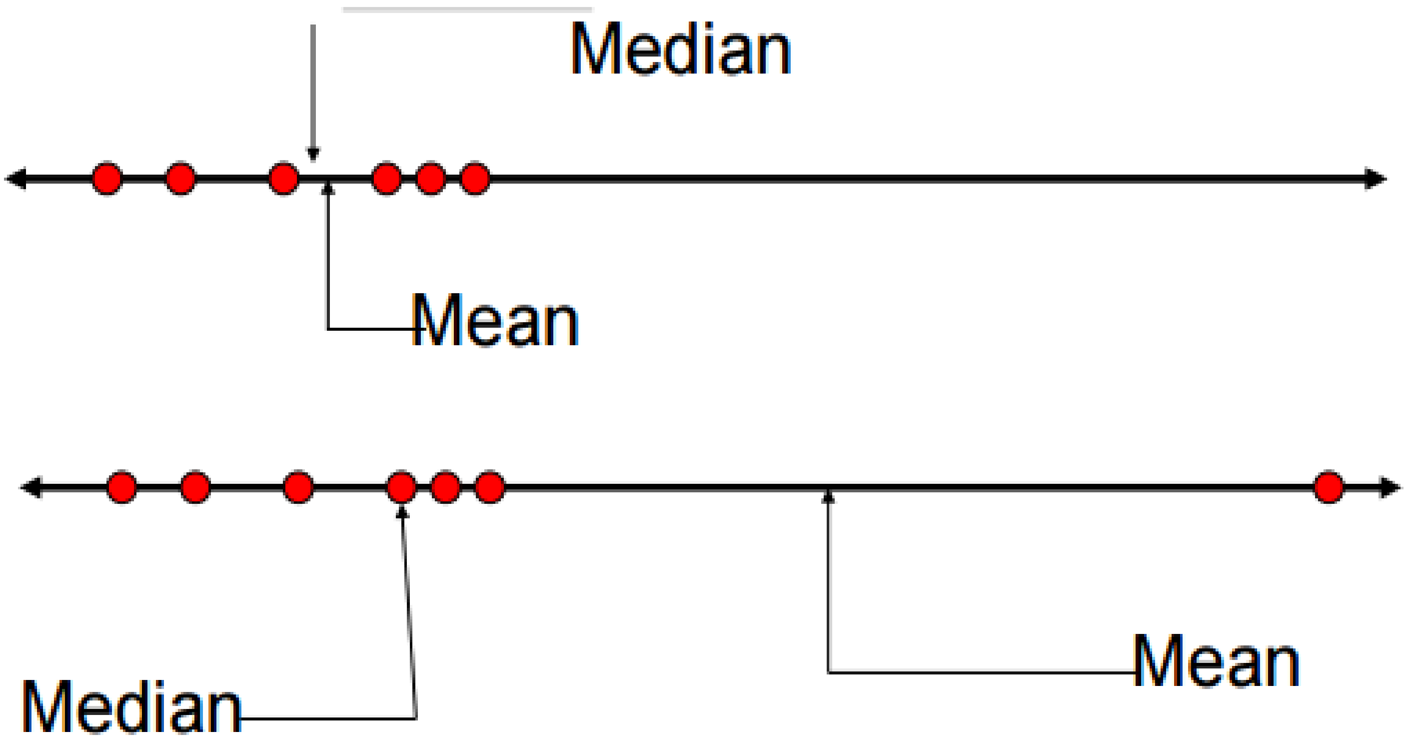
1.The mean is sensitive to extremes value but the median is not.

2.For some data sets, the mean can give a misleading picture of the observations.

Example: (2, 2, 2, 2, 17) Mean (5) is not very representative of the data set.

3.The median sometimes ignores potentially useful information because only the middle value (or two middle values) affects the median.

Potential Problem with Means



Relationships among the Mean, Median, and Mode

1. For symmetrical histogram and frequency curve with one peak (Figure 1), the values of the mean, median, and mode are identical, and they lie at the center of the distribution.
2. For a histogram and a frequency curve skewed to the right (Figure 2), the value of the mean is the largest, that of the mode is the smallest, and the value of the median lies between these two. Notice that the mode always occurs at the peak point. The value of the mean is the largest in this case because it is sensitive to outliers that occur in the right tail. These outliers pull the mean to the right.
3. If a histogram and a distribution curve are skewed to the left (Figure 3), the value of the mean is the smallest and that of the mode is the largest, with the value of the median lying between these two. In this case, the outliers in the left tail pull the mean to the left.

Figure1: Mean, median, and mode for a symmetric histogram and frequency curve

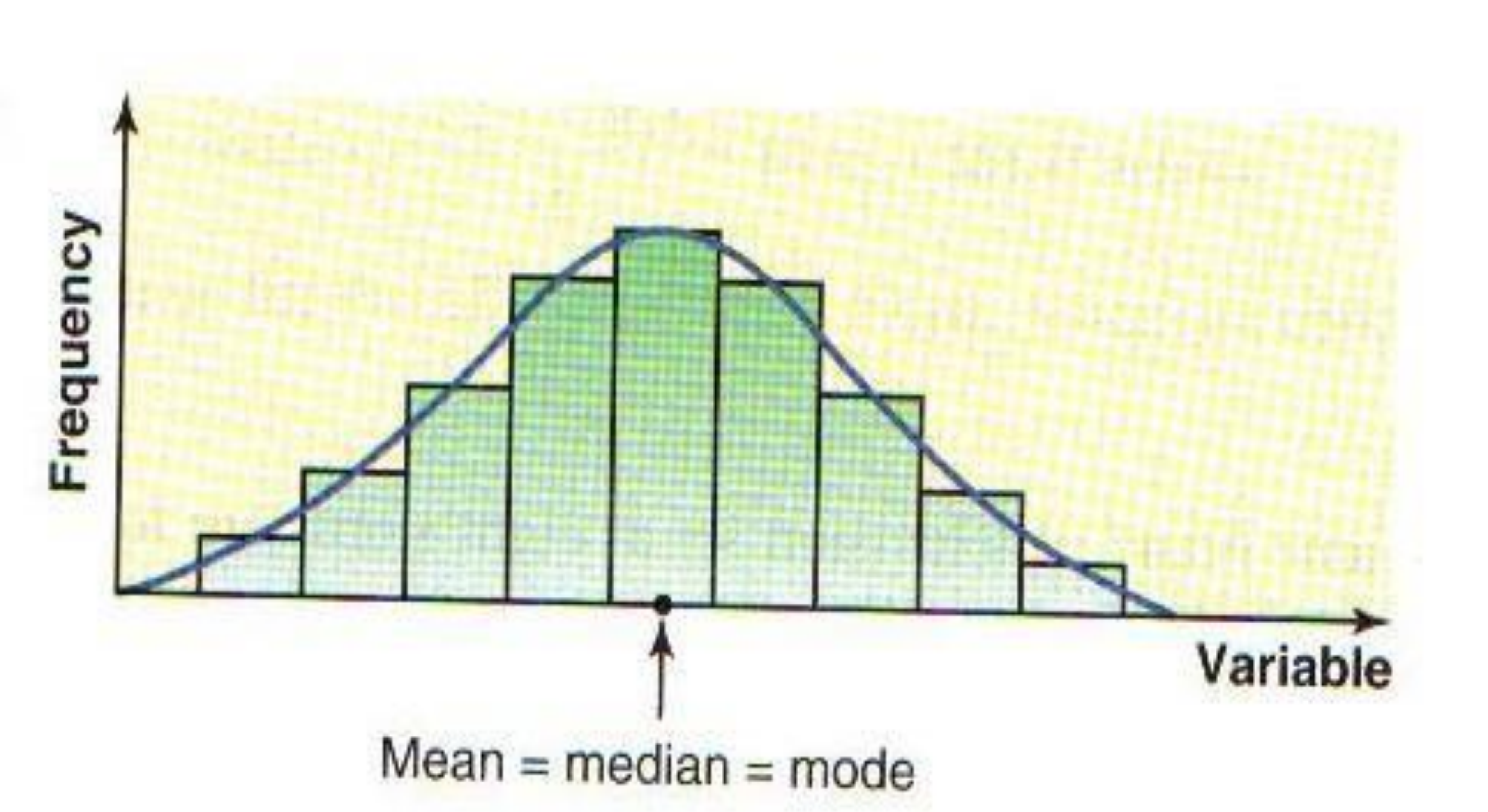


Figure 2: Mean, median, and mode for a histogram and frequency curve skewed to the right.

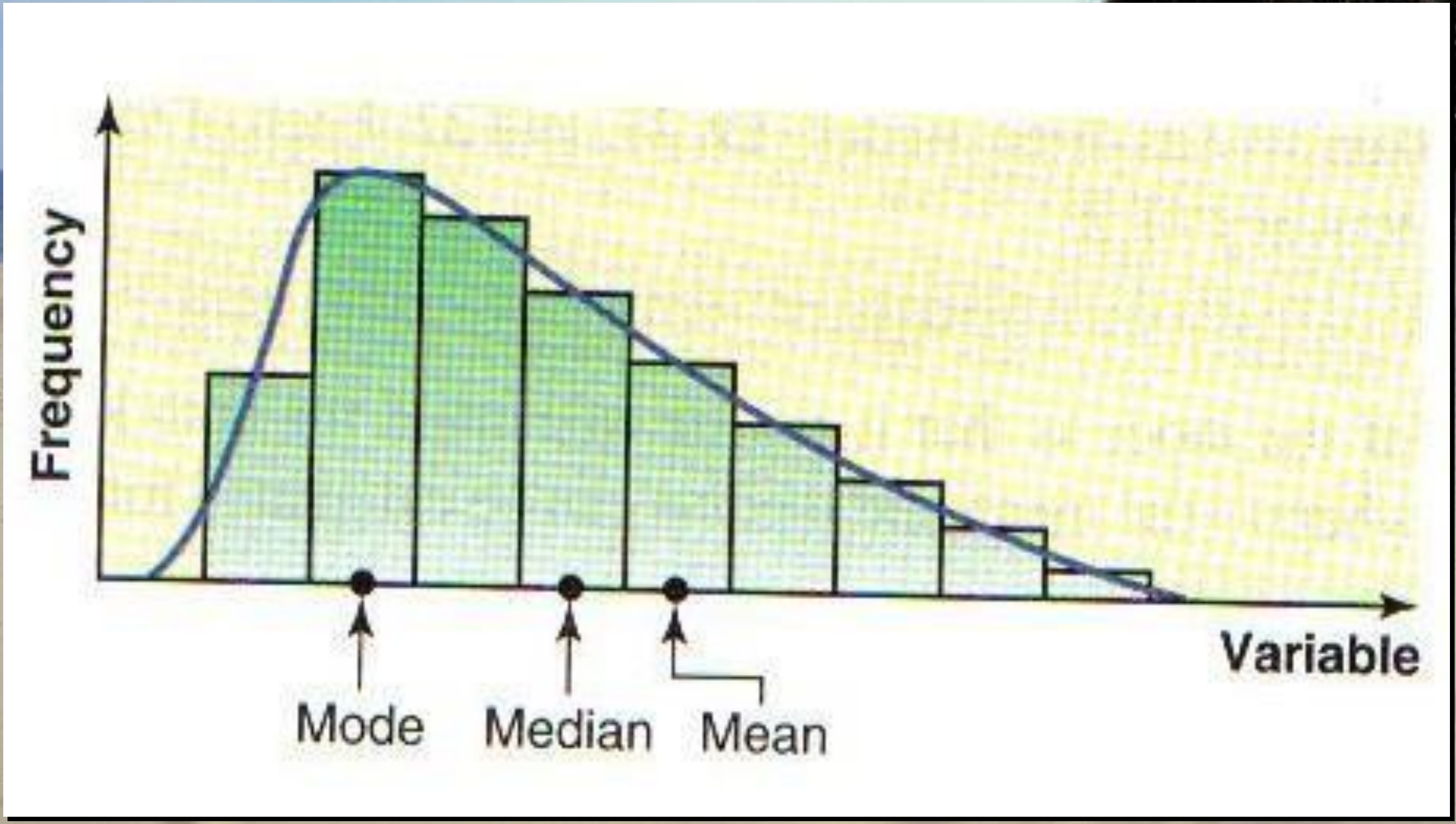
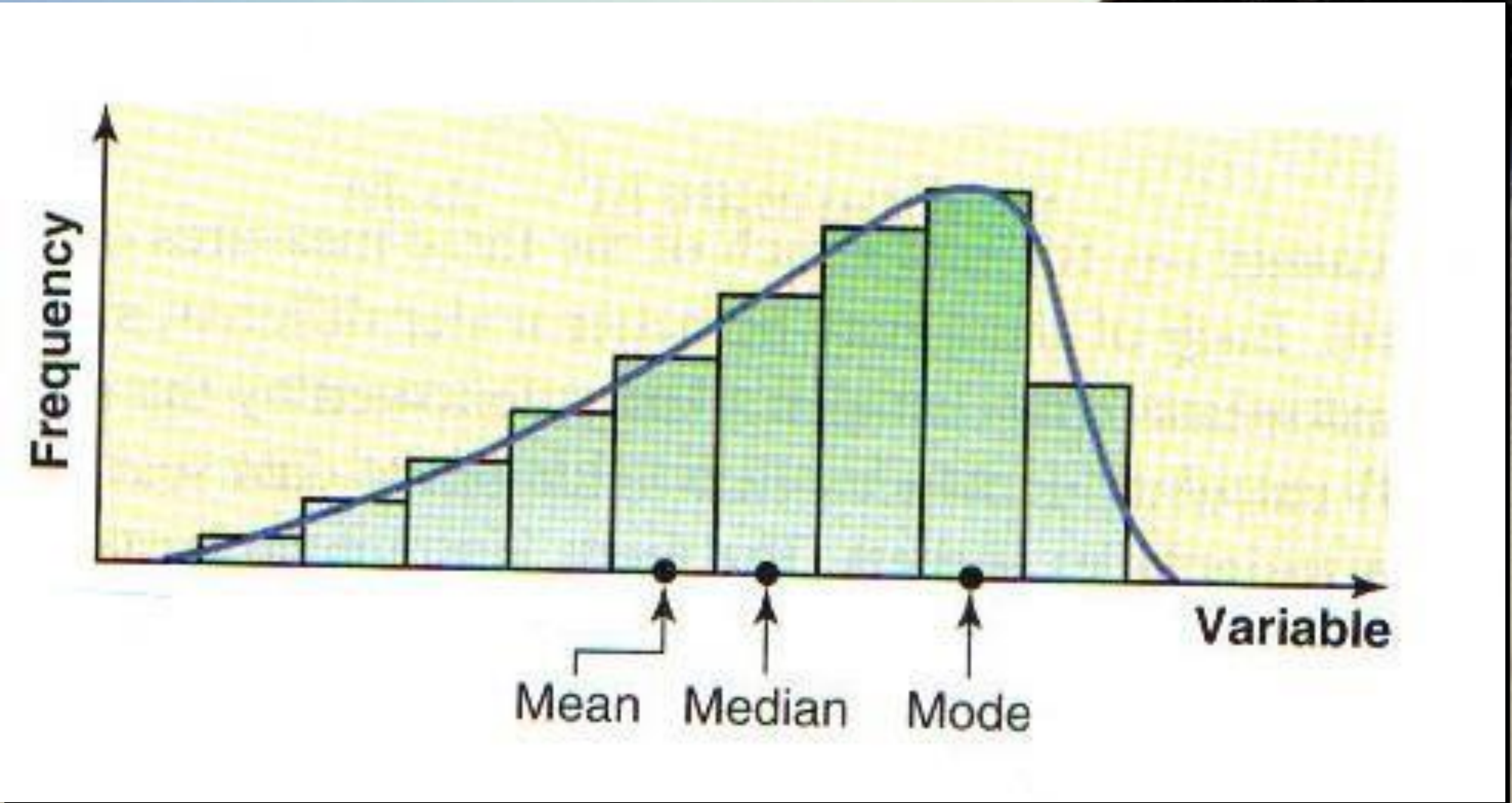


Figure 3: Mean, median, and mode for a histogram and frequency curve skewed to the left



Example

The data sets below give the waiting times (in minutes) of several people at two offices. Find the mean, median, and mode of each data set.

Office A	Office B
14, 17, 18, 19, 20, 24, 24, 30, 32	8, 11, 12, 16, 18, 18, 18, 20, 23

Office A

Mean= 14+17+……..+32 /9 =198/9=22 , Median =20 ,Mode =24.

Office B

Mean=8+11+….+23/9 =144/9=16 , Median=18 Mode=18.

Example:

For a set of data (4,8,12,15,3,2,6,9,8,7)

Mean= 4+8+12+…..+7/10 =74/10 =7.4

First, we rank the given data in increasing order as follows: (2,3,4,6,7,8,8,9,12,15)

Median=7.5

Mode =8

Thank you

Mean, Median, Mode and Range   www.cazoommaths.com/us

Mean

Add all the numbers then divide by the amount of numbers

9, 3, 1, 8, 3, 6

9 + 3 + 1 + 8 + 3 + 6 = 30

30 ÷ 6 = 5

The mean is 5

Median

Order the set of numbers, the median is the middle number

9, 3, 1, 8, 3, 6

1, 3, 3, 6, 8, 9

The median is 4.5

Mode

The most common number

9, 3, 1, 8, 3, 6

The mode is 3

Range

The difference between the highest number and lowest number

9, 3, 1, 8, 3, 6

9 - 1 = 8

The range is 8