Sampling and Data Analysis
Week 4: Describing Data

Instructor:
Dr. Ali AJDER

Week-4: Describing Data
Numerical Measures

➢Explain the concept of central tendency.

➢Identify and compute the arithmetic mean.

➢Compute and interpret the weighted mean.

➢Determine the median.

➢Identify the mode.

➢Explain and apply measures of dispersion.

➢Compute and explain the variance and the standard deviation.

➢Explain Chebyshev's Theorem and the Empirical Rule.

Parameter VS Statistics
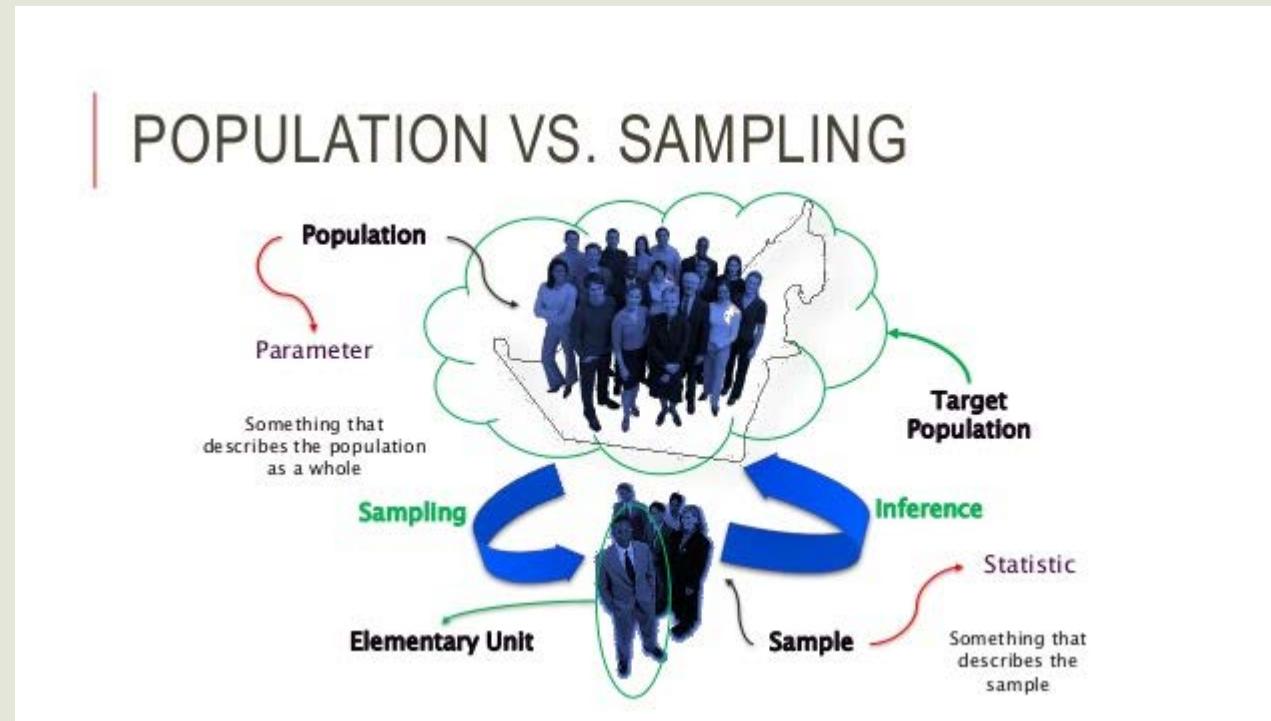
# What is the *measure of location*?

Central Tendency

CENTRAL TENDENCY

- The average U.S. home changes ownership every 11.8 years.
- An American receives an average of 568 pieces of mail per year.
- The average American home has more TV sets than people. There are 2.73 TV sets and 2.55 people in the typical home.
- The average American couple spends $20,398 for their wedding, while their budget is 50% less. This does not include the cost of a honeymoon or engagement ring.
- The average price of a theater ticket in the United States is $7.50, according to the National Association of Theatre Owners.
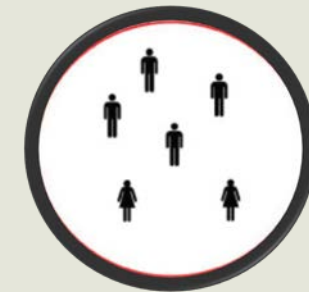
Measures of Location

The purpose of a measure of location is to **pinpoint the center** of a distribution of data.
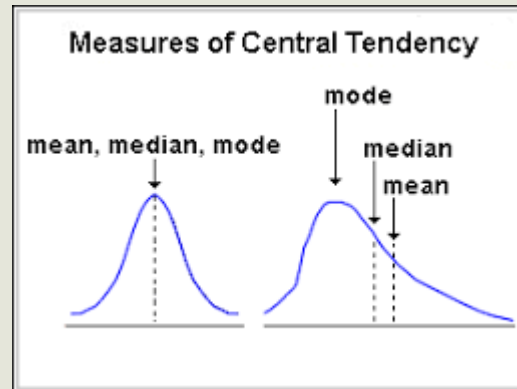
Arithmetic mean:
Population mean vs Sample mean

Population Mean	Sample Mean
$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$	$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$
N = number of items in the population	n = number of items in the sample

Example

There are 42 exits on I-75 through the state of Kentucky.
Listed below are the distances between exits (in miles).

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

Does this information belong to a population or a sample?
What is the mean number of miles between exits?

Example

SunCom is studying the number of minutes used monthly by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.
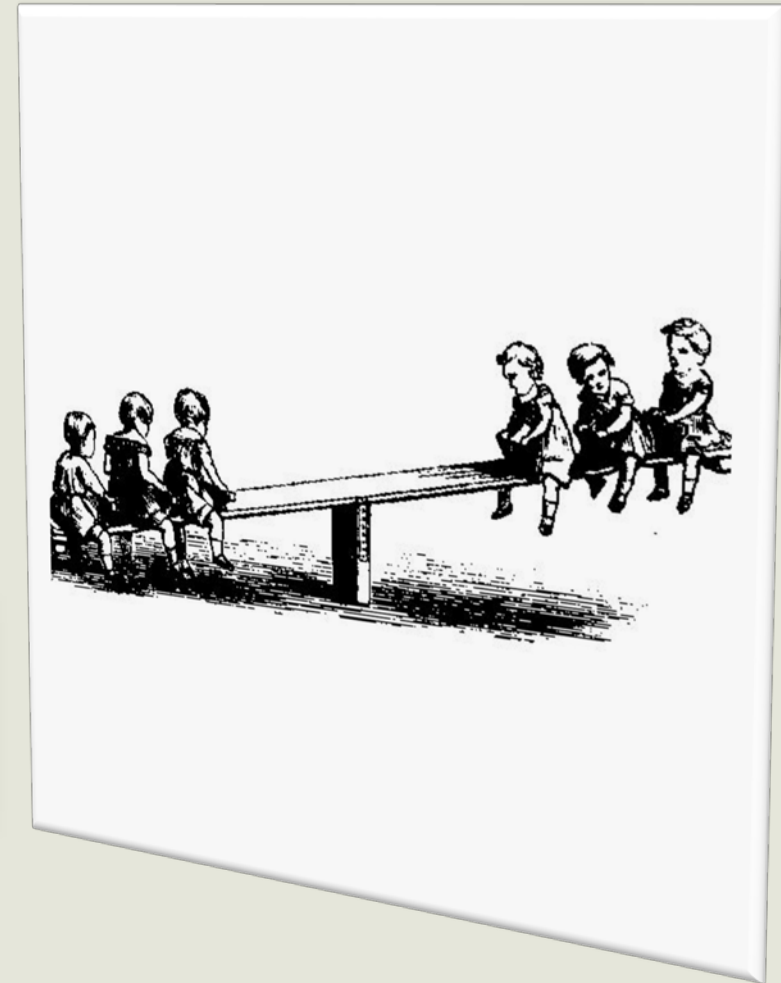
90	77	94	89	119	112
91	110	92	100	113	83

What is the arithmetic mean number of minutes used?

Weighted mean

weighted mean center   mean center
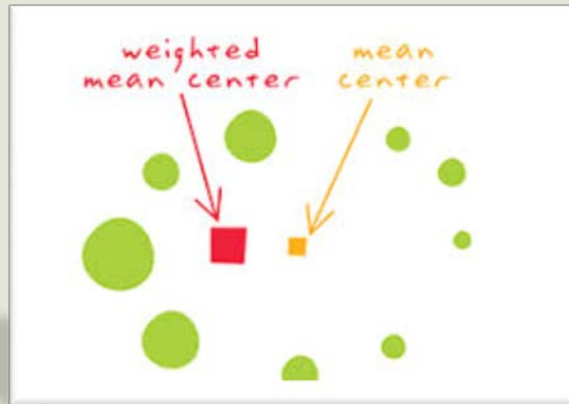
WEIGHTED MEAN

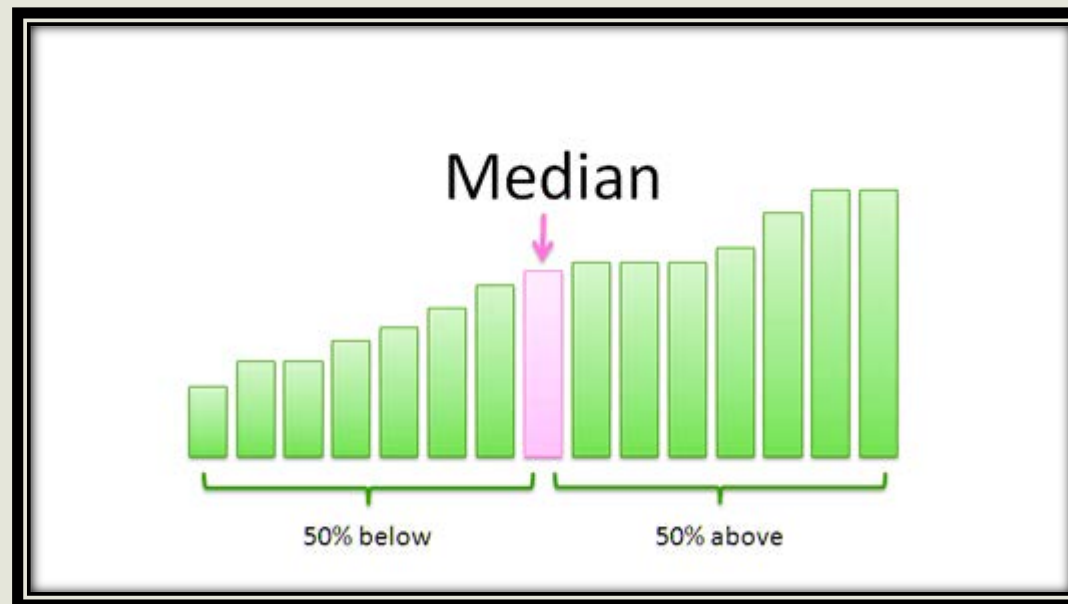$$\overline{X}_w = \frac{w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n}{w_1 + w_2 + w_3 + \cdots + w_n}$$

Example

The Carter Construction Company pays its hourly employees $16.50, $19.00, or $25.00 per hour. There are 26 hourly employees, 14 of whom are paid at the $16.50 rate, 10 at the $19.00 rate, and 2 at the $25.00 rate.

What is the mean hourly rate paid the 26 employees?

Median

Examples

The ages for a sample of five college students are:

21, 25, 19, 20, 22

Arranging the data in ascending order gives:

19, 20, 21, 22, 25.
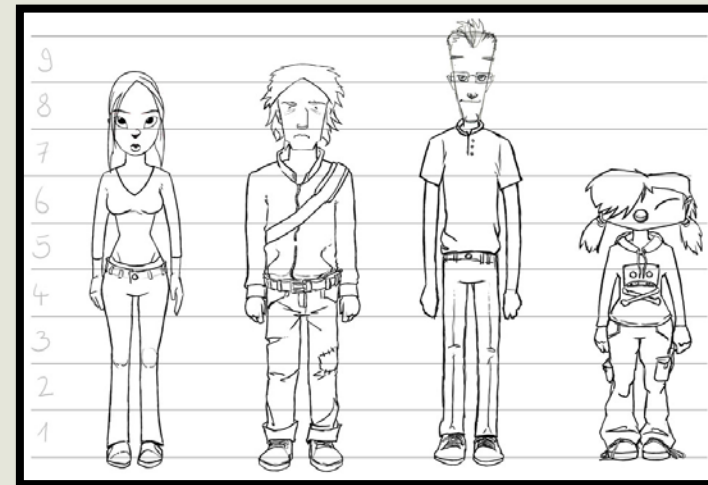
Thus the median is **21.**

The heights of four basketball players, in inches, are:

76, 73, 80, 75

Arranging the data in ascending order gives:

73, 75, 76, 80.

Thus the median is **75.5**

Examples

Facebook is a popular social networking website. Users can add friends and send them messages, and update their personal profiles to notify friends about themselves and their activities. A sample of 10 adults revealed they spent the following number of hours last month using Facebook.
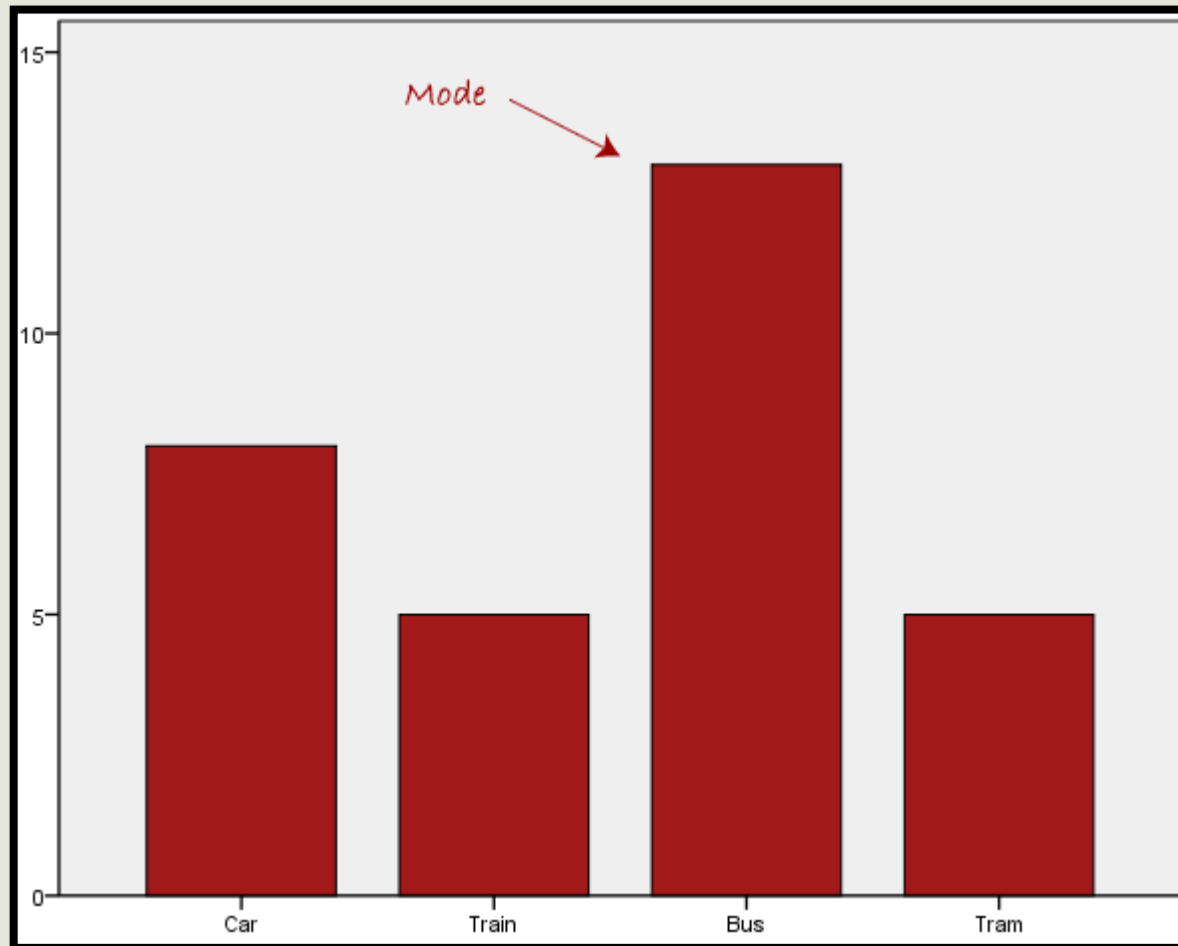
Find the median number of hours.

| 3 | 5 | 7 | 5 | 9 | 1 | 3 | 9 | 17 | 10 |

Mode

Examples

Using the data regarding the distance in miles between exits on I-75 through Kentucky. The information is repeated below. What is the modal distance?
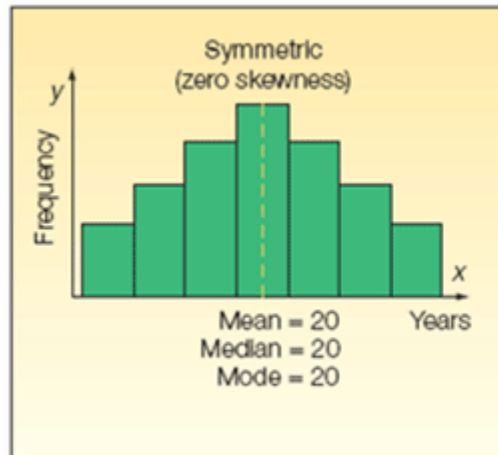
Organize the distances into a frequency table.

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

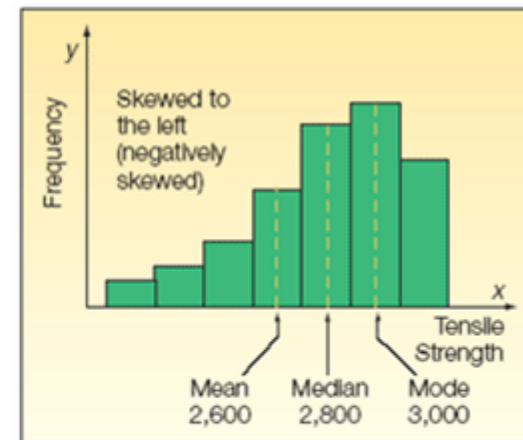| Distance in Miles between Exits	Frequency
1	8
2	7
3	7
4	3
5	4
6	1
7	3
8	2
9	1
10	4
11	1
14	1
Total	42

The Relative Positions of the Mean, Median, and the Mode

zero skewness
mode = median = mean
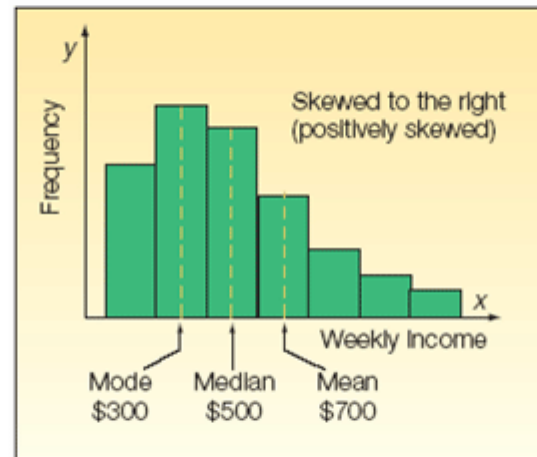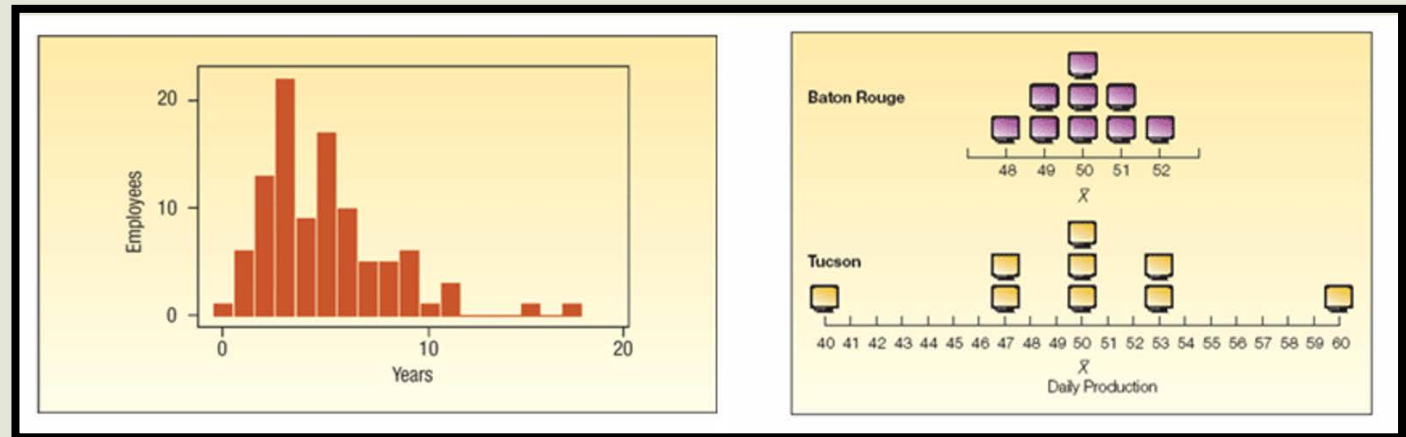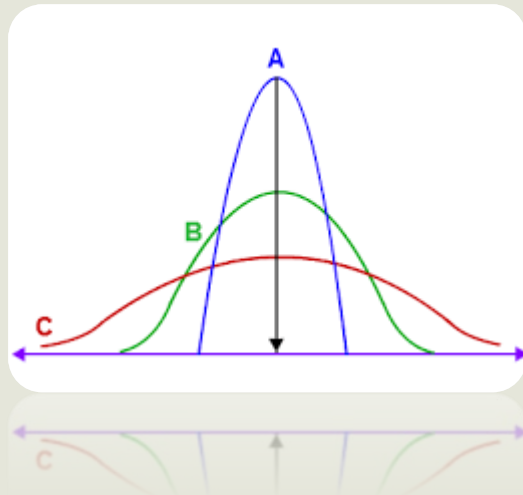
positive skewness
mode < median < mean

negative skewness
mode > median > mean

# Why do we need *measures of dispersion*?

Dispersion

A **measure of location**, such as the mean or the median, only describes the **center** of the data. It is valuable from that standpoint, but it does not tell us anything about the *spread* of the data.

For example, if your nature guide told you that the river ahead averaged 3 feet in depth, would you want to wade across on foot without additional information? Probably not. You would want to know something about the variation in the depth.

A second reason for studying the dispersion in a set of data is to compare the spread in two or more distributions.

Measures of Dispersion
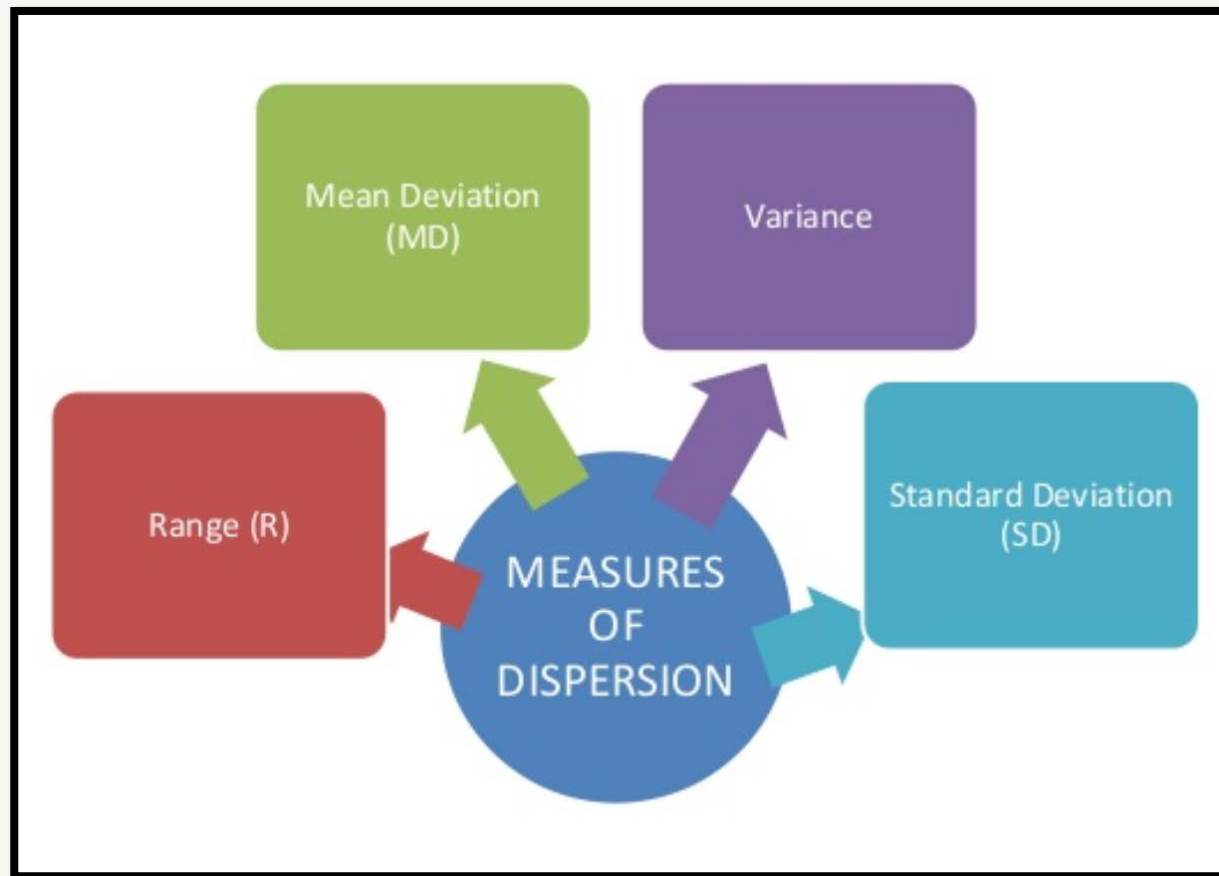
Examples

The number of cappuccinos sold at the Starbucks location in the Orange Country Airport between 4 and 7 p.m. for a sample of 5 days last year were 20, 40, 50, 60, and 80.

Determine the range for the number of cappuccinos sold.

Examples

The number of cappuccinos sold at the Starbucks location in the Orange Country Airport between 4 and 7 p.m. for a sample of 5 days last year were 20, 40, 50, 60, and 80.

Determine the mean deviation for the number of cappuccinos sold.

Examples

The number of traffic citations issued during the last five months in Beaufort County, South Carolina, is reported below:

| Month	January	February	March	April	May	June	July	August	September	October	November	December
Citations	19	17	22	18	28	34	45	39	38	44	34	10

What is the population variance?

Examples

The hourly wages for a sample of part-time employees at Home Depot are: $12, $20, $16, $18, and $19.

What is the sample variance?

Chebyshev's Theorem

CHEBYSHEV'S THEOREM For any set of observations (sample or population), the proportion of the values that lie within k standard deviations of the mean is at least $1 - 1/k^2$, where k is any constant greater than 1.
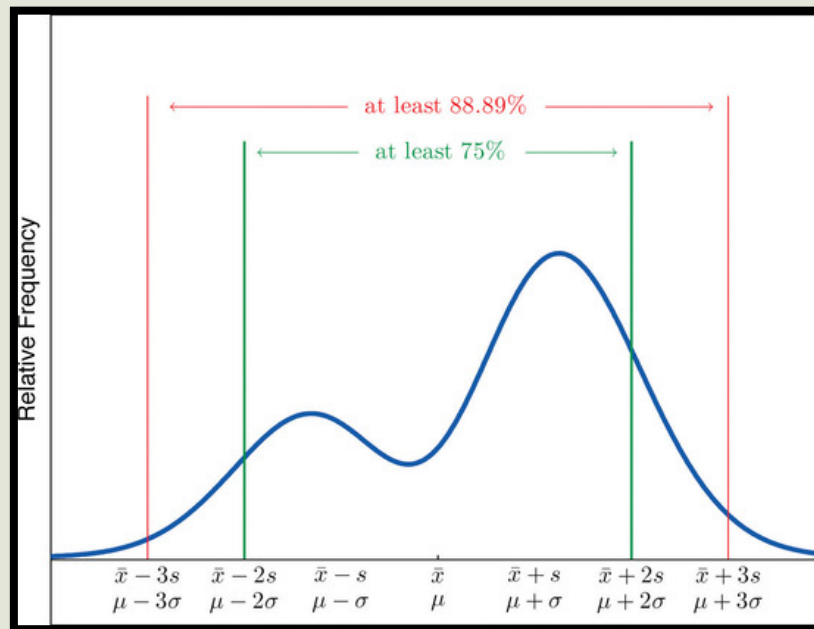
The Empirical Rule

EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean; about 95 percent of the observations will lie within plus and minus two standard deviations of the mean; and practically all (99.7 percent) will lie within plus and minus three standard deviations of the mean.