Stat 145

Chapter 2

DESCRIBING DATA –NUMERICAL SUMMARIES

Central Tendency:

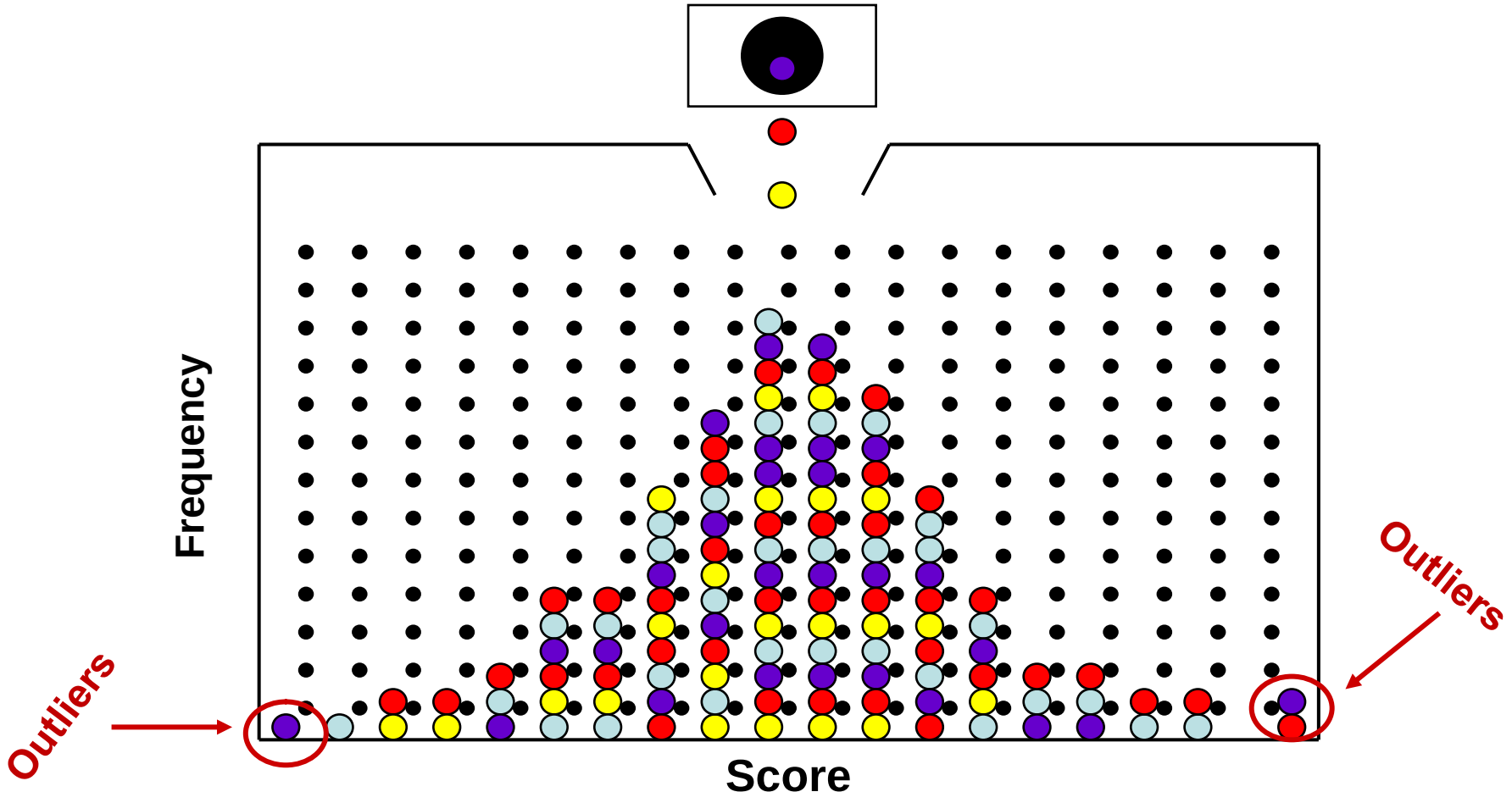
1. There is a natural tendency for data to cluster around a central point

2. Data that falls further away from this central point is less representative of the sample (or population)

3. Data that is located in the tails are typically referred to as extreme scores or outliers
   - Depending on the size and number of outliers, the statistics generated from a sample may be less accurate and reliable

DESCRIBING DATA – NUMERICAL SUMMARIES

Central Tendency:



Measures of central tendency:

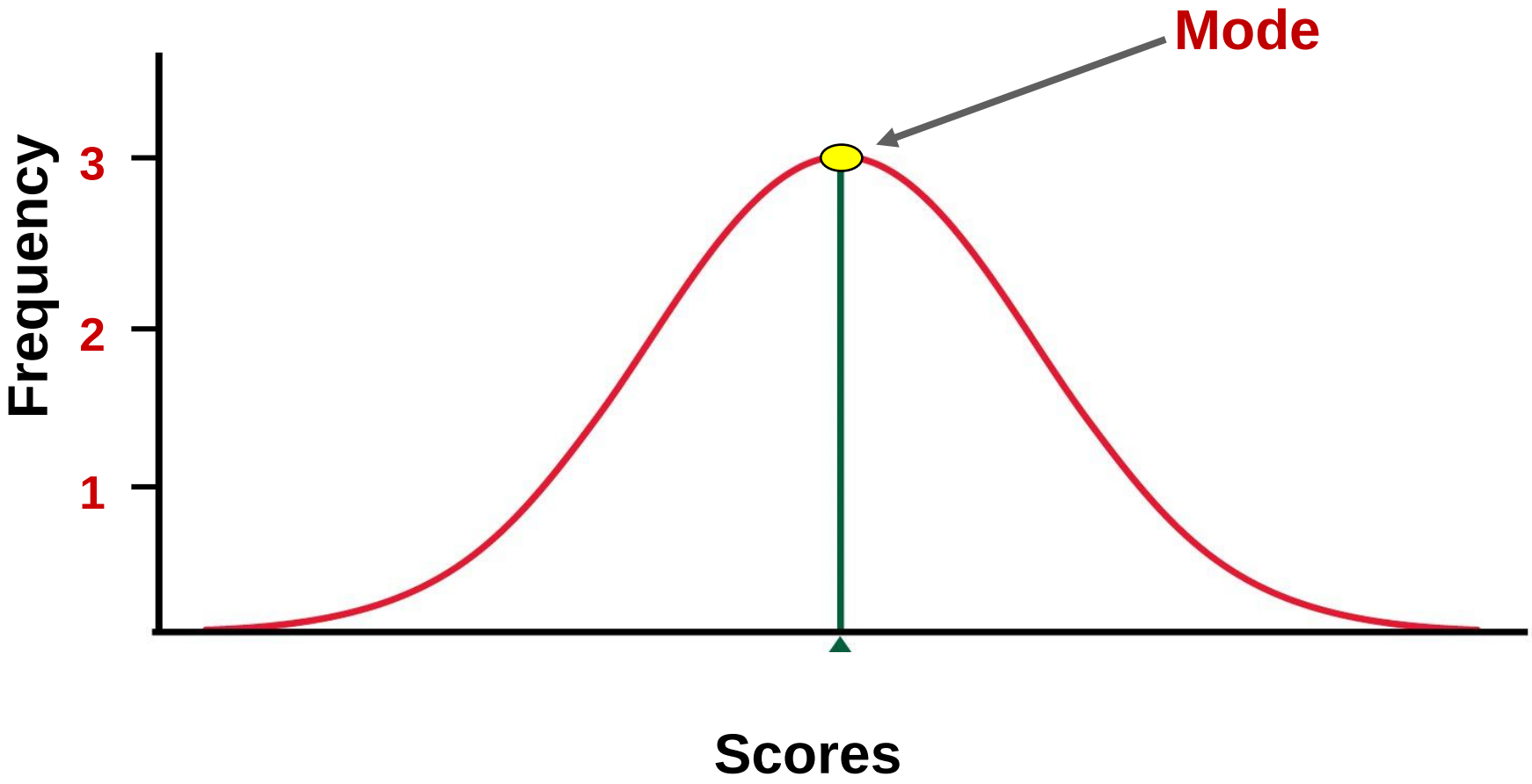
1. The phenomenon of central tendency is often described using three numerical summaries
   - The Mode
   - The Median, and
   - The Mean

Measures of central tendency: The Mode

1. The **Mode:**
   - **Indicates the most frequent or numerous score in a distribution**
   - **Provides no summary of the scores as a group**
   - **Not sensitive to the size (or weight) of each score**
   - **Is the tallest point in any shaped distribution**
   - **Applies to all 4 scales of measurement (N.O.I.R.)**

Measures of central tendency: The Mode

1. The mode is the tallest point(s) in a histogram



Measures of central tendency: The Mode

❶ **EXAMPLE:** Sixteen students are asked how far they drive one-way to campus. The mode indicates the most frequent distance traveled.

UNSORTED

Student 1	6
Student 2	3
Student 3	1
Student 4	3
Student 5	0.5
Student 6	2
Student 7	5
Student 8	6
Student 9	0.5
Student 10	2
Student 11	1
Student 12	3
Student 13	3
Student 14	5
Student 15	2
Student 16	4

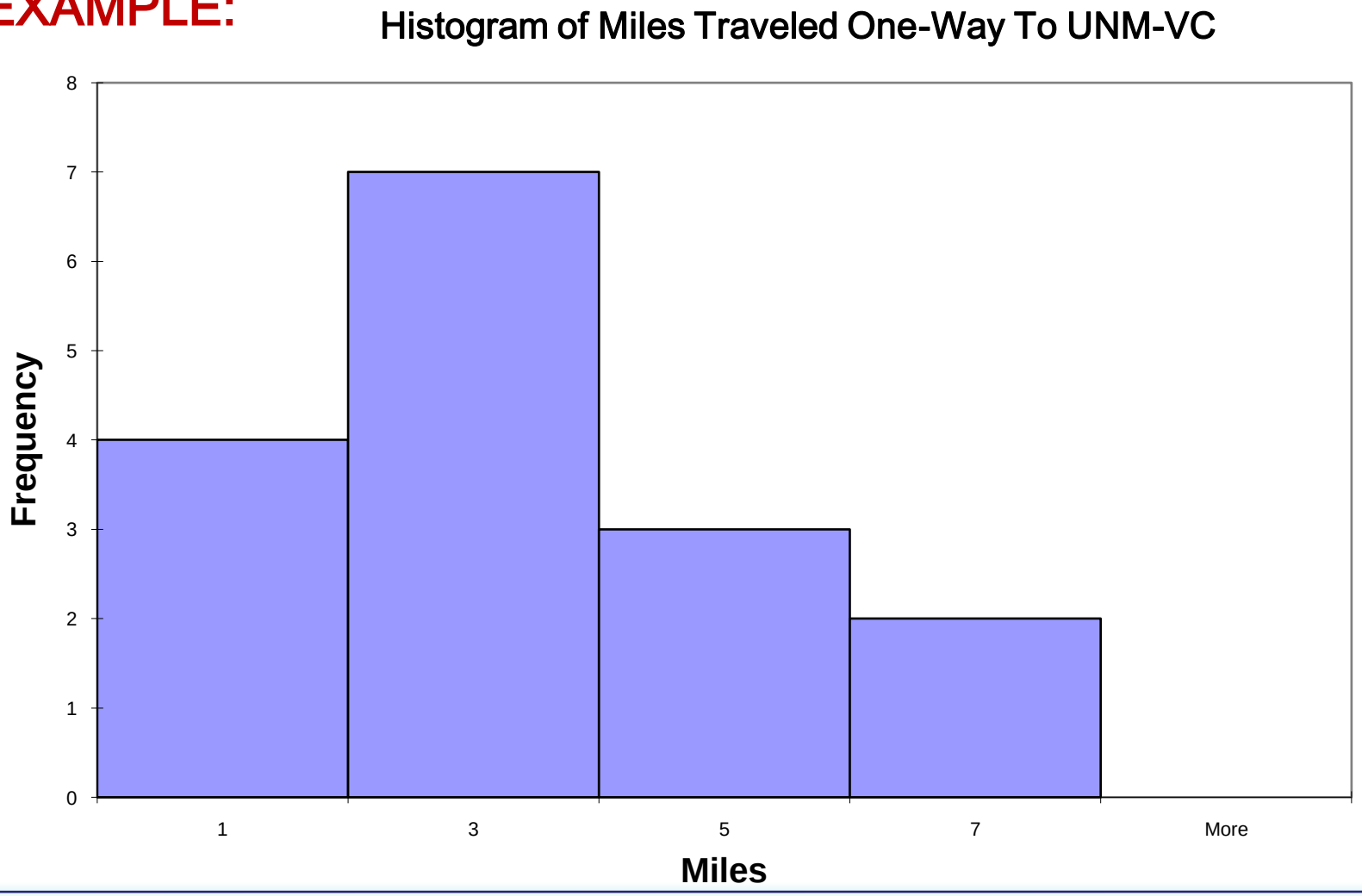
SORTED

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

3 miles is the most frequent distance traveled. Thus the mode is 3 miles.

Measures of central tendency: The Mode

❶ EXAMPLE:



Measures of central tendency: The Median

1. The **Median:**
   - Indicates the middle point in a distribution, where half of the scores fall at or above and half of the scores fall at or below
   - Not affected by the size of the scores, but only the number of scores
   - Represents the 50th percentile in a distribution
   - Applies to all four scales of measurement (N.O.I.R.)
   - More affected by outliers compared to the mode, but less affected compared to the mean

Measures of central tendency: The Median

❶ **EXAMPLE:** Sixteen students are asked how far they drive one-way to campus. The median indicates the middle score or 50th percentile.

SORTED

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

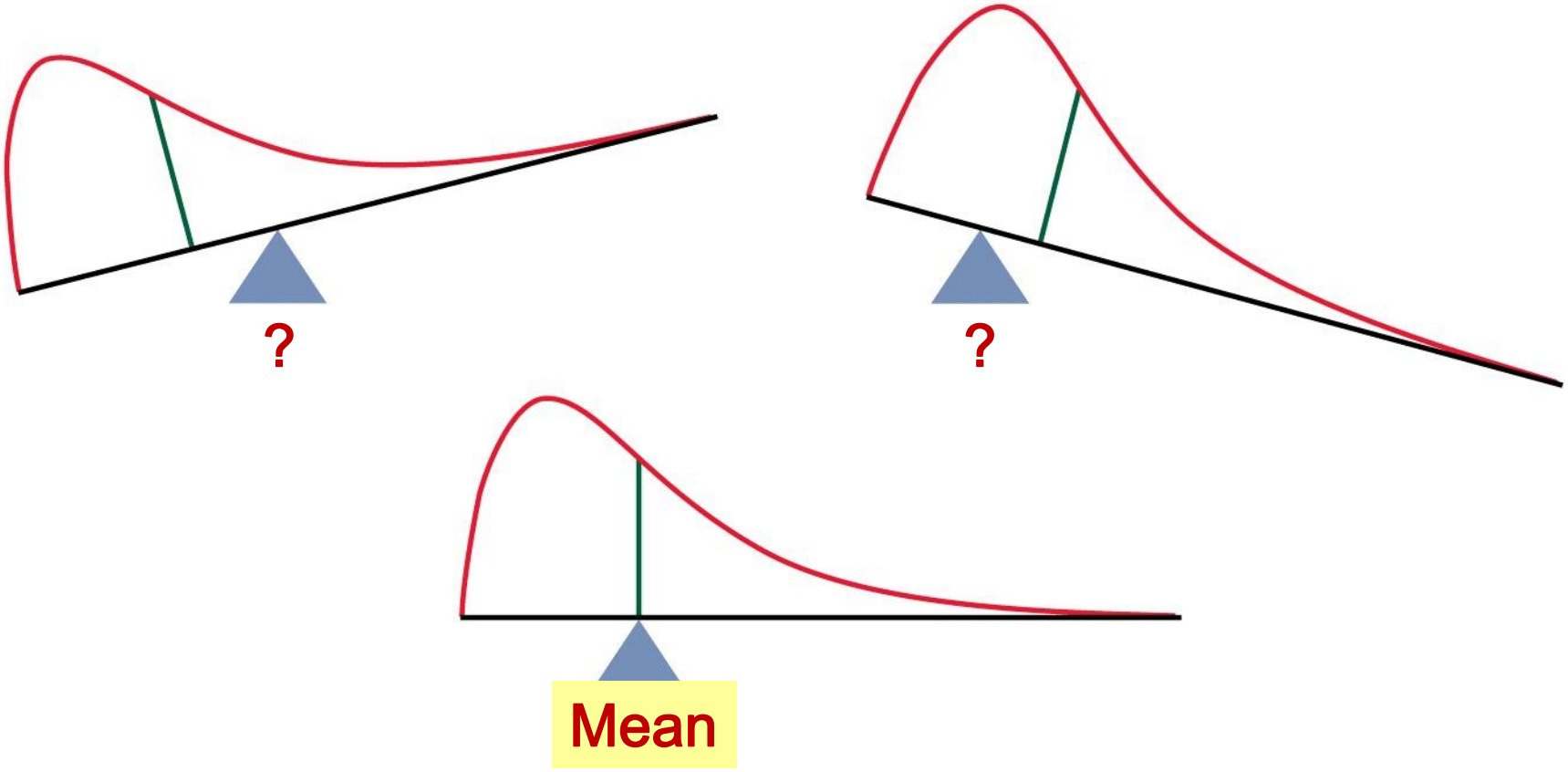
The median is the middle score. There are an even number of scores, so we add the 8th and 9th scores and divide by two. The median distance traveled is 3 miles.

Measures of central tendency: The Mean

1. The **Mean**:
   - Indicates the arithmetic average of a group of scores
   - **Summarizes all values** into a single score
   - Sensitive to the size (or weight) of each score
   - Is the **balancing point** in any shaped distribution
   - Only applicable to Interval and Ratio scales of data
   - Most affected by outliers compared to the median and mode

Measures of central tendency: The Mean

1. The value of the mean is the balancing point in a histogram



Measures of central tendency: The Mean

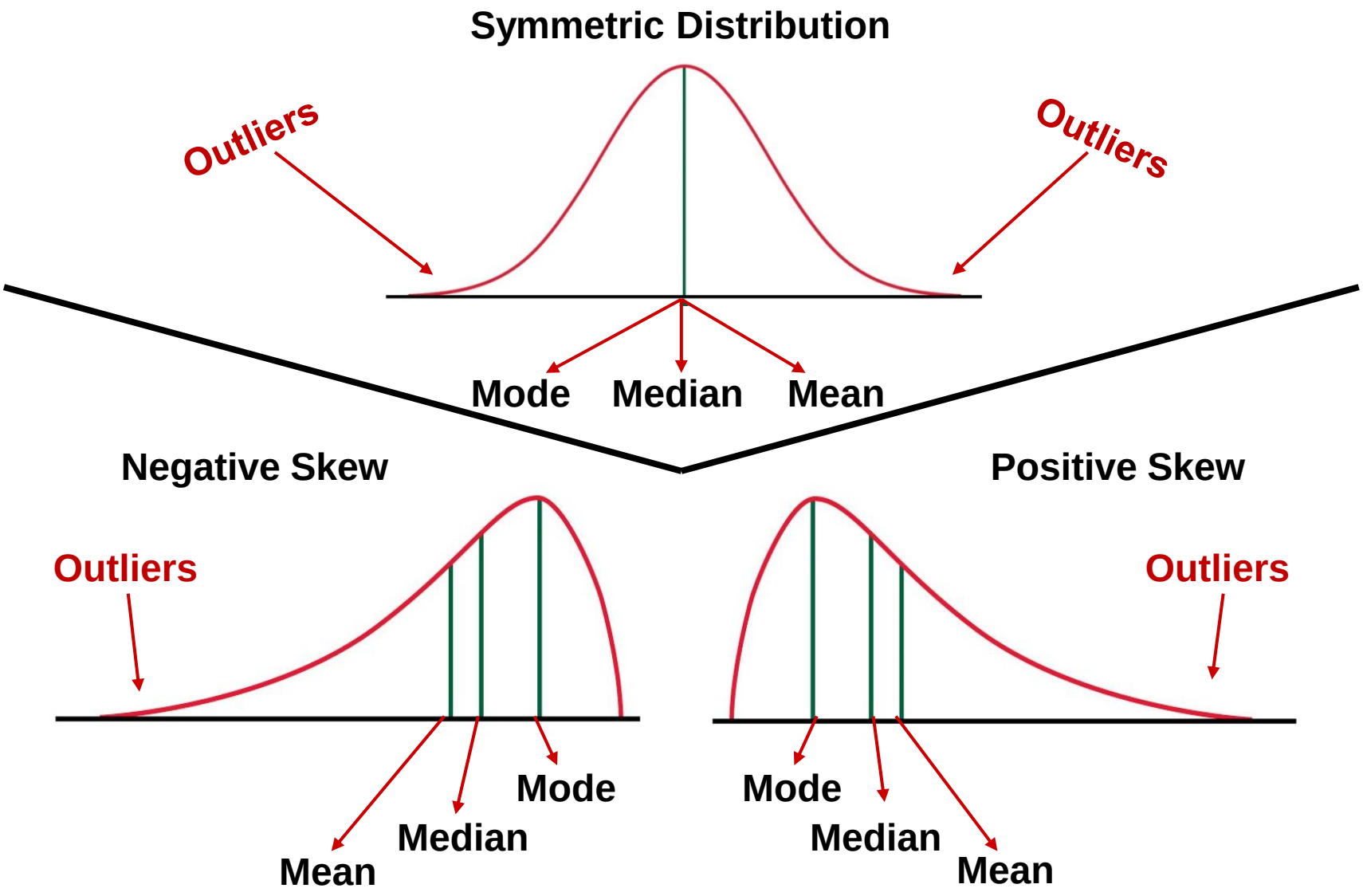
❶ EXAMPLE: Sixteen students are asked how far they drive one-way to campus. The mean tells us as a whole, the 16 students drive 2.94 miles one-way to campus

The mean summarizes all individual scores into a single value. In other words, if the group was treated as one subject, the mean would represent that subject's score.

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

SORTED

DESCRIBING DATA – NUMERICAL SUMMARIES



Measures of Variability or Dispersion

1. In addition to central tendency, we can also describe how scores are spread out from the low end of the scale to the high end
2. Histograms that are flat and wide indicate more variability between scores whereas tall and narrow histograms indicate less variability between scores
    - Outliers can also increase variability, especially when they are extreme and numerous
3. Data that is highly variable is more difficult to predict and is more likely to lead to experiments with non-significant results

Measures of Variability or Dispersion

1. Three common measures of variability:
   - The Range
   - The Standard Deviation, and
   - The Five Number Summary

Measures of Variability or Dispersion: The Range

1. The Range:
   - Indicates the difference between the highest score and lowest score
   - Quick and easy to calculate
   - Tells nothing about the average distribution of scores around a central data point (e.g., the mean)
   - Can be VERY misleading when extreme outliers are present

Measures of Variability or Dispersion: The Range

❶ **EXAMPLE:** The range for the 16 students asked how far they drive one-way to campus is 5.5 miles.

SORTED

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

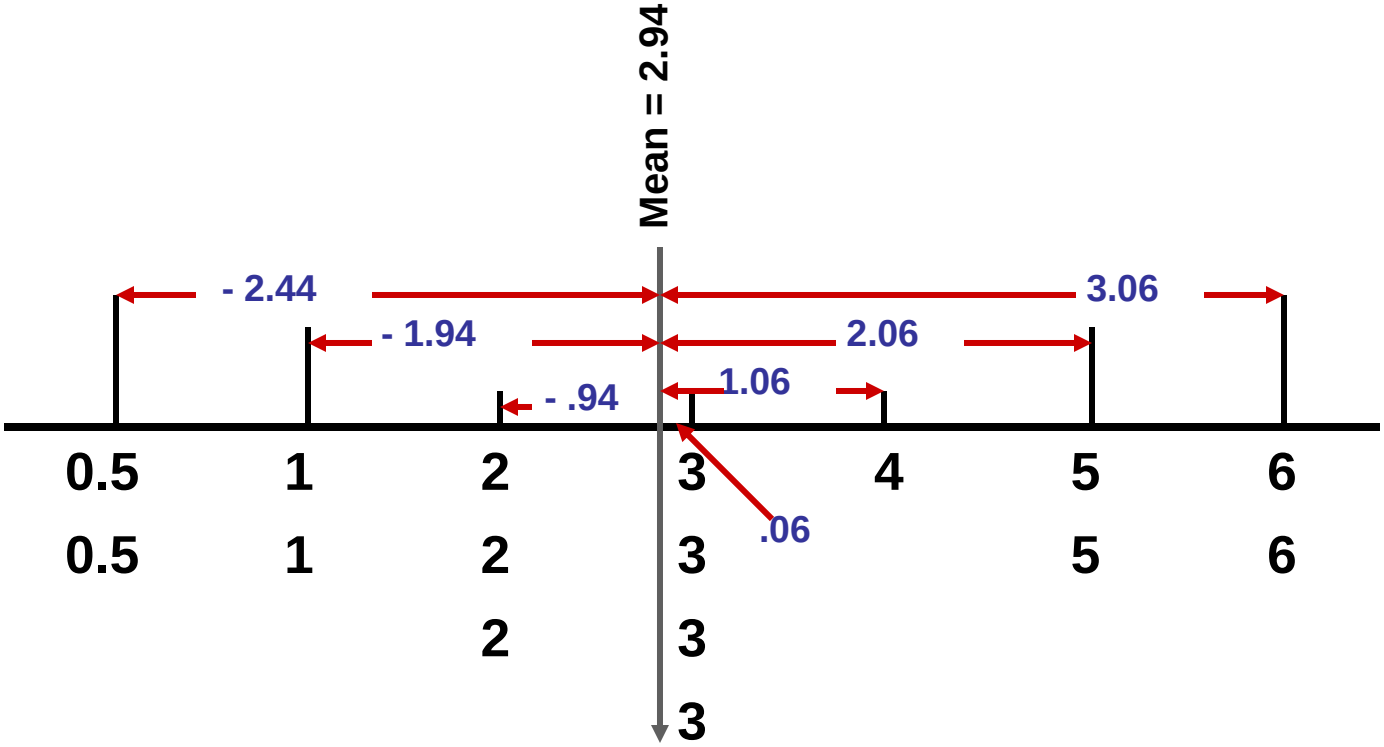
Range → 6 - .5 = 5.5 miles

Measures of Variability or Dispersion: The Standard Deviation

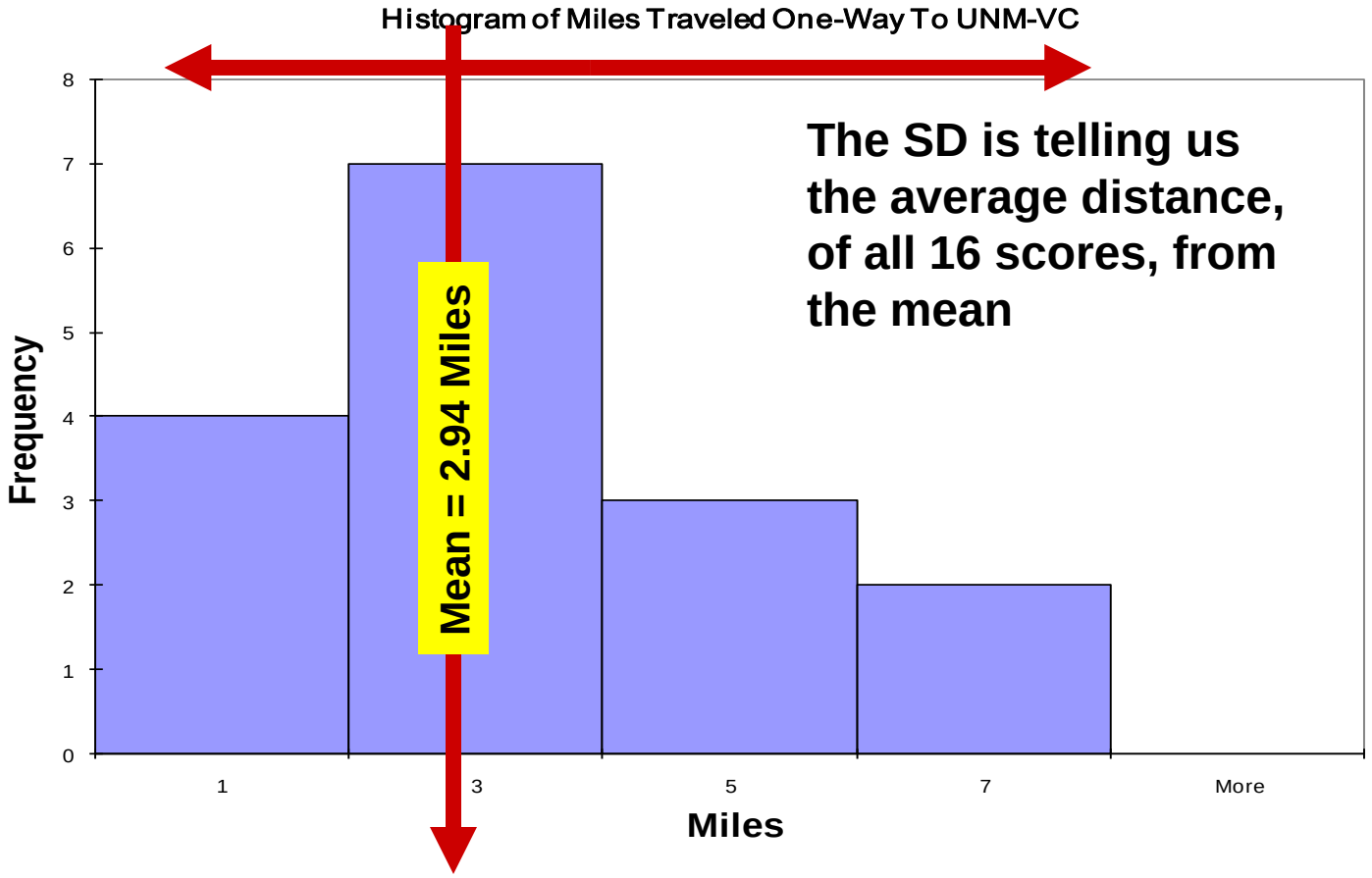
1. The Standard Deviation:
   - Indicates the average distance scores are from the mean in the original units of the dependent variable
   - The larger the standard deviation, the flatter the distribution, and the more variable and dissimilar the scores are in the distribution
   - Is sensitive to outliers, especially when outliers are extreme and numerous

Measures of Variability or Dispersion: The Standard Deviation

❶ **EXAMPLE:** On the average, how different are the distances traveled by the 16 students from the mean? The standard deviation will tell us that answer.



❶ **EXAMPLE:** For this example, the standard deviation (SD) is 1.83 miles. Thus, we can say "On the average, the 16 students surveyed travel 1.83 miles one-way to campus different than the mean of 2.94 miles."



Measures of Variability or Dispersion: The Standard Deviation

1. For any data set, if we add a score equal to the mean, the value of the mean will not change, but the standard deviation will become smaller
   - If the value added is equal to the mean, it has no distance from the mean and therefore adds nothing to the sum of the differences from the mean
   - However, by adding a score equal to the mean, we are still increasing the total number of scores by one, which means we're dividing the sum of the differences by a total number of scores that has increased by one

Measures of Variability or Dispersion: Five Number Summary

1. The Five Number Summary is a measure of variability represented by:
   - The Lowest Score
   - The 1st Quartile (or 25th percentile)
   - The Median
   - The 3rd Quartile (or 75th percentile)
   - The Highest Score

Measures of Variability or Dispersion: Five Number Summary

1. The Five Number Summary is often used for data sets that are badly skewed
   - Skewed data sets often contain outliers which can lead to misleading interpretations of the mean and standard deviation
     - Remember – the mean and SD are very sensitive to outliers and can be exaggerated by these scores
2. The Five Number Summary allows the researcher to describe central tendency and variability across the scale of the dependent variable

Measures of Variability or Dispersion: Five Number Summary

1. Steps for obtaining the Five Number Summary:

   - Sort data from lowest score to highest score

   - Identify the low score, high score, and median score
     - When the number of scores is even, the median is the middle score
     - When the number of scores is odd, the median is the two middle scores divided by two

Measures of Variability or Dispersion: Five Number Summary

1. Steps for obtaining the Five Number Summary:
   - Obtain the 1st quartile
     - When the total number of scores is even, the 1st quartile will be the median of the first half of the scores
     - When the total number of scores is odd, the 1st quartile will be the median for the first half of the scores, including the original median for all scores

Measures of Variability or Dispersion: Five Number Summary

1. Steps for obtaining the Five Number Summary:
   - Obtain the 3rd quartile
     - When the total number of scores is even, the 3rd quartile will be the median of the second half of the scores
     - When the total number of scores is odd, the 3rd quartile will be the median for the second half of the scores, including the original median for all scores

Measures of Variability or Dispersion: Five Number Summary

❶ **EXAMPLE:** Five number summary for the sixteen students asked how far they drive one-way to campus.

SORTED

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

Lowest Score = .5

1st Quartile = 1.5
(25th percentile)

Median Score = 3
(50th percentile)

3rd Quartile = 4.5
(75th percentile)

Highest Score = 6

❶ **EXAMPLE:** Suppose we add one more person to our survey who drives 2 miles to campus. What is the five number summary for *seventeen* students asked how far they drive one-way to campus?

SORTED

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 17	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

Lowest Score = .5

1st Quartile = 2
(25th percentile)

Median Score = 3
(50th percentile)

3rd Quartile = 4
(75th percentile)

Highest Score = 6

Measures of Variability or Dispersion: Five Number Summary

❶ EXAMPLE: Five number summary for the sixteen students asked how far they drive one-way to campus.

SORTED

Student 5	0.5
Student 9	0.5
Student 3	1
Student 11	1
Student 6	2
Student 10	2
Student 15	2
Student 2	3
Student 4	3
Student 12	3
Student 13	3
Student 16	4
Student 7	5
Student 14	5
Student 1	6
Student 8	6

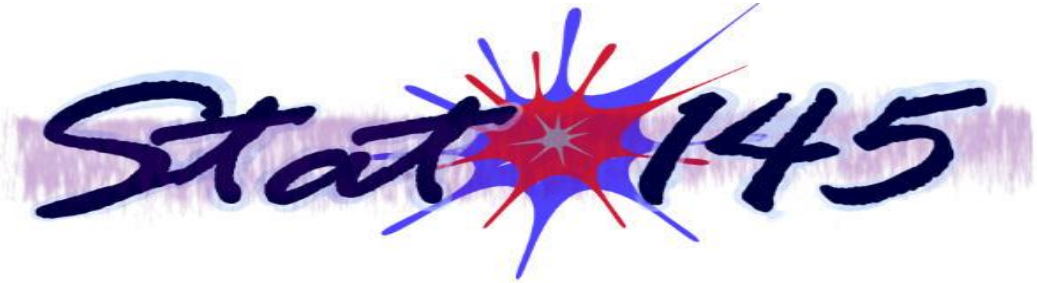
Lowest Score = .5

1st Quartile = 1.5
(25th percentile)

Median Score = 3
(50th percentile)

3rd Quartile = 4.5
(75th percentile)

Highest Score = 6



End of Chapter 2 – Part 2