Chapter 6, Unit A: Characterizing Data

The **distribution** of a variable (data set) describes the values taken on by the variable and the frequency (or relative frequency) of these values.

Example: The distribution of the grades shows how many students received grades A, B, C, D, and F.

**Mean** is the average value.

**Median** is the middle value in the sorted data set.

**Mode** is the most common value in a distribution.

Example 1: Seven stores sell the energy drink for the following prices: $1.29 $1.29 $1.35 $1.39 $1.49 $1.59 $1.79

What is the mean price?

What is the median?

What is the mode?

Example 2: Eight stores sell the energy drink for the following prices: $1.09 $1.29 $1.29 $1.35 $1.39 $1.49 $1.59 $1.79

What is the mean price?

What is the median?

What is the mode?

Example 3: Imagine that 5 graduating seniors on a college football team receive the follow first-year contract offers to play in the National Basketball Association (zero indicates that the player did not receive a contract offer).

$0 $0 $0 $0 $3,500,000

What is the mean contract offer?

What is the median?

What is the mode?

**Outlier** is a data value that is much higher or much lower than almost all other values.

Confusion about “Average” – Mean or Median

Example 4: A track coach wants to determine an appropriate heart rate for her athletes during their workouts. She chooses 5 best runners and asks them to wear heart rate monitors during the workout. In the middle of the workout, she reads the following heart rates for the five athletes: 130, 135, 140, 145, and 325. Which is a better measure of the average in this case: the mean or the median? Why?

Example 5: A newspaper surveys wages for assembly workers in regional high-tech companies and reports and average of $22 per hour. The workers at one large firm immediately request a pay raise, claiming that they work as hard as employees at other companies but their wage is only $19. The management rejects their request, telling them that they are overpaid because their average, in fact, is $23. Can both sides be right? Explain

$19 $19 $19 and $39

**Shapes of Distributions**

Our goal is to describe the general shape of the graph in words. We will focus on three characteristics of a distribution: its numbers of peaks, its symmetry or skewness, and its variation.

**Number of peaks:** single-peaked or unimodal, double-peaked or bimodal, other distribution may have no peaks (uniform distributions) or more than two peaks.

Example 6: Height of all women at a college.

Example 7: Height of all students at a college.

Example 8: The numbers of people with particular last digits (0 through 9) in their SS numbers.

A distribution is **symmetric** if its left half is a mirror o its right half.

A distribution is **left-skewed** if its values on are more spread out on the left side.

A distribution is **right-skewed** if its values are more spread out on the right side.

Example 9: Heights of a sample of 100 women.

Example 10: Family income in the US.

Example 11: Speeds of cars on a road where a visible patrol car is using radar to detect speeders.

**Variation** (low, medium and high) describes how widely data values are spread out about the center of a distribution.

Example: How would you expect the variation to differ teen times in the Olympic marathon and times in the New York marathon? Explain.

**Homework: Quick Quiz: 1-10**

**Review Questions: 1-6**

**Basic Skills & Concepts: 13, 15, 17, 19, 20**