

Exploring Data with Graphs and Numerical Summaries | 2

2.3 Describe the Center of Quantitative Data

Aug 8-9:15 AM

Learning Objectives

- Calculating the **mean**
- Calculating the **median**
- Comparing the mean & median
- Definition of **resistant**
- Know how to identify the **mode** of a distribution

flowjo.typepad.com

Aug 8-9:15 AM

Mean

- The mean is the sum of the observations divided by the number of observations
- It is the center of mass

$$\bar{x} = \frac{\sum x}{n}$$

The following table shows the mean of the cereal sodium data:

Cereal	Sodium
Frosted Flakes	0
Raisin Bran	210
All Bran	260
Apple Jacks	152
Capt Crunch	220
Cherries	290
Cinnamon Toast	210
Crackling Oat Bran	140
Crisps	220
Frosted Flakes	300
Fruit Loops	125
Grape-Nuts	170
Honey Nut Cheerios	250
Life	150
Quaker Raisin Crisp	170
Sugar Smacks	70
Special K	230
Wheaties	200
Corn Flakes	290
Honeycomb	180

Aug 8-9:15 AM

Median

Midpoint of the observations when ordered from least to greatest

- Order observations
- If the number of observations is:
 - Odd, the median is the middle observation
 - Even, the median is the average of the two middle observations

Order	Data
1	78
2	91
3	94
4	98
5	99
6	101
7	103
8	105
9	114

Order	Data
1	78
2	91
3	94
4	98
5	99
6	101
7	103
8	105
9	114
10	121

Aug 8-9:15 AM

Comparing the Mean and Median

- Mean and median of a **symmetric** distribution are close
- Mean is often preferred because it uses all
- In a **skewed** distribution, the mean is farther out in the skewed tail than is the median
- Median is preferred because it is better representative of a typical observation

Symmetric Distribution: Mean = Median
 Right-Skewed Distribution: Median < Mean
 Left-Skewed Distribution: Mean < Median

Aug 8-9:15 AM

Resistant Measures

- A measure is **resistant** if extreme observations (outliers) have little, if any, influence on its value
- Median is resistant to outliers
- Mean is not resistant to outliers


www.stat.psu.edu

Aug 8-9:15 AM

1,2,3,4,5	1,2,3,4,500
Med=	Med=
\bar{X} =	\bar{X} =

Aug 28-10:44 AM

Mode



- Value that occurs **most often**
- Highest bar in the histogram
- Mode is most often used with **categorical** data

Aug 8-9:15 AM

2.30 Median versus mean: The mean and median describe the center.

- Why is the median sometimes preferred? Give an example.
- Why is the mean sometimes preferred? Give an example.

Aug 8-9:15 AM

2.37 Public transportation—center: The owner of a company in downtown Atlanta is concerned about the large use of gasoline by her employees due to urban sprawl, traffic congestion, and the use of energy inefficient vehicles such as SUVs. She'd like to promote the use of public transportation. She decides to investigate how many miles her employees travel on public transportation during a typical day. The values for her ten employees (recorded to the closest mile) are

0 0 4 0 0 0 10 0 6 0

- Find and interpret the mean, median, and mode.
- She has just hired an additional employee. He lives in a different city and travels 90 miles a day on public transport. Re-compute the mean and median. Describe the effect of this outlier.

Aug 8-9:15 AM

[Empty box for student work]

Aug 14-3:15 PM