STATISTICS

INFORMED DECISIONS USING DATA

Fifth Edition

Chapter 3

Numerically Summarizing Data

Pearson

Thursday 2/27

3.1 Measures of $\boxed{\text{Central Tendency}}$

↳ "middle"

Learning Objectives

1. Determine the arithmetic **mean** of a variable from raw data
2. Determine the **median** of a variable from raw data
3. Explain what it means for a statistic to be **resistant**
4. Determine the **mode** of a variable from raw data

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (1 of 9)

QN

The **(arithmetic) mean** of a variable is computed by adding all the values of the variable in the data set and dividing by the number of observations.

- Notation:
  – Population mean vs Sample mean

Greek

μ "mu"

$\bar{x}$ "x bar"

Pearson

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (2 of 9)

- The **population (arithmetic) mean**, μ (pronounced "mew"), is computed using all the individuals in a population.
  - The population mean is a parameter.

μ = "sum all data & ÷ by total"

- The **sample (arithmetic) mean**, $\bar{x}$ (pronounced "x bar"), is computed using only the sample data.
  - The sample mean is a statistic.

$\bar{x}$ = sum all sample data & ÷ by # in sample

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (4 of 9)

population size

If $x_1, x_2, \ldots, x_N$ are the N observations of a variable from a population, then the population mean, μ, is

$x_3 + x_4 + \cdots +$

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\Sigma x_i}{N}$$

Data

$\underbrace{x_1, x_2, x_3, \ldots, \ , x_N}_{\text{total is } N}$

sample size

If $x_1, x_2, \ldots, x_n$ are the n observations of a variable from a sample, then the sample mean, $\bar{x}$, is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\Sigma x_i}{n}$$

Data

$x_1, x_2, \ldots, x_n$

Σ = (capital Sigma) = means "sum or add"

Greek

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (6 of 9)

EXAMPLE Computing a Population Mean and a Sample Mean

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

(a) Compute the population mean of this data.

(b) Then take a simple random sample of *n* = 3 employees. Compute the sample mean. Obtain a second simple random sample of *n* = 3 employees. Again compute the sample mean.

(a) $\mu = 24.9$ min per employee

(b) $\bar{x}_1 = 28.0$ min/employee $\bar{x}_2 = 28.0$ min/employee

$\bar{x}_3 = 27.3$ min/employee

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (7 of 9)

EXAMPLE Computing a Population Mean and a Sample Mean

(a)

$$\mu = \frac{\Sigma x_i}{N}$$

$$= \frac{x_1 + x_2 + \ldots + x_7}{7}$$

$$= \frac{23 + 36 + 23 + 18 + 5 + 26 + 43}{7}$$

$$= \frac{174}{7}$$

$$= 24.9 \text{ minutes}$$

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (8 of 9)

EXAMPLE Computing a Population Mean and a Sample Mean

(b) Obtain a simple random sample of size *n* = 3 from the population of seven employees. Use this simple random sample to determine a sample mean. Find a second simple random sample and determine the sample mean.

```
123→rand
                 123
randInt(1,7)
                   5
                   2
                   6
```

1	2	3	4	5	6	7
23,	36,	23,	18,	5,	26,	43

$$\bar{x} = \frac{5+36+26}{3} = 22.3$$

```
789→rand
                 789
randInt(1,7)
                   2
                   3
                   6
```

$$\bar{x} = \frac{36+23+26}{3} = 28.3$$

Pearson

Thursday
Monday

3.1 Measures of Central Tendency

3.1.1 Determine the Arithmetic Mean of a Variable from Raw Data (9 of 9)

$N = 22$ $\quad \bar{x}_1 = 65.5$ bpm $\quad \mu = 74.8$ bpm

$n = 4$ $\quad \bar{x}_2 = 78$ bpm

IN CLASS ACTIVITY

Population Mean versus Sample Mean

Treat the students in the class as a population. All the students in the class should determine their pulse rates (beats per minute).

a) Compute the population mean pulse rate.

b) Obtain a simple random sample of *n* = 4 students and compute the sample mean. Does the sample mean equal the population mean?

c) Obtain a second simple random sample of *n* = 4 students and compute the sample mean. Does the sample mean equal the population mean?

d) Are the sample means the same? Why?

No!

sampling variability

Pearson

"middle"

3.1 Measures of Central Tendency

3.1.2 Determine the Median of a Variable from Raw Data (1 of 5)

The **median** of a variable is the value that lies in the middle of the data when arranged in ascending order.

Notation: We use *M* (*or Med*) to represent the median.

Important Point: the median cuts the data set into two halves!

- Half of the data is lower than the ~~mean~~ median
- Half of the data is higher than the ~~mean~~ median

3.1 Measures of Central Tendency

3.1.2 Determine the Median of a Variable from Raw Data (2 of 5)

Steps in Finding the Median of a Data Set

Step 1 Arrange the data in ascending order.

Step 2 Determine the number of observations, *n*. Is n even or odd?

Step 3 Determine the observation in the middle of the data set.

- When n is odd: there is only one "middle" number so pick this as the Median. Note: It will be the $\frac{n+1}{2}$ position.
- When n is even: there is no one "middle" number so pick the two in the "middle" and their their mean. Note: They will be the $\frac{n}{2}$ and $\left(\frac{n}{2}\right)+1$ positions.

3.1 Measures of Central Tendency

3.1.2 Determine the Median of a Variable from Raw Data (4 of 5)

EXAMPLE Computing a Median of a Data Set with an Odd Number of Observations

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Determine the median of this data.

Ascending: (5 18 23) $\boxed{23}$ (26 36 43) (n=7)

even or odd : n=7 odd

Median

3.1 Measures of Central Tendency

3.1.2 Determine the Median of a Variable from Raw Data (4 of 5)

EXAMPLE Computing a Median of a Data Set with an Odd Number of Observations

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Determine the median of this data.

Step 1: 5, 18, 23, 23, 26, 36, 43

Step 2: There are $n = 7$ observations.

Step 3: $\frac{n+1}{2} = \frac{7+1}{2} = 4$ $M = 23$

5, 18, 23, **23**, 26, 36, 43

3.1 Measures of Central Tendency

3.1.2 Determine the Median of a Variable from Raw Data (5 of 5)

EXAMPLE Computing a Median of a Data Set with an Even Number of Observations

Suppose the start-up company hires a new employee. The travel time of the new employee is 70 minutes. Determine the median of the "new" data set.

23, 36, 23, 18, 5, 26, 43, 70

Med 24.5

new

ascending order: (5 18 23 23 26 36 43 70)

even or odd?

n = 8

Median

mean of theses

$$\frac{23+26}{2} = \frac{49}{2} = 24.5$$

3.1 Measures of Central Tendency

3.1.2 Determine the Median of a Variable from Raw Data (5 of 5)

EXAMPLE Computing a Median of a Data Set with an Even Number of Observations

Suppose the start-up company hires a new employee. The travel time of the new employee is 70 minutes. Determine the median of the "new" data set.

23, 36, 23, 18, 5, 26, 43, 70

Step 1: 5, 18, 23, 23, 26, 36, 43, 70

Step 2: There are n = 8 observations.

Step 3: $\frac{n+1}{2} = \frac{8+1}{2} = 4.5 \qquad M = \frac{23+26}{2} = 24.5$ minutes

$$\begin{array}{c} 5,\ 18,\ 23,\ 23,\ 26,\ 36,\ 43,\ 70 \\ \uparrow \\ M = 24.5 \end{array}$$

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (1 of 6)

EXAMPLE Computing a Median of a Data Set with an Even Number of Observations

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43

Suppose a new employee is hired who has a 130 minute commute. How does this impact the value of the mean and median? Use calc.

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (1 of 6)

EXAMPLE Computing a Median of a Data Set with an Even Number of Observations

The following data represent the travel times (in minutes) to work for all seven employees of a start-up web development company.

23, 36, 23, 18, 5, 26, 43, 130

Suppose a new employee is hired who has a 130 minute commute. How does this impact the value of the mean and median?

Mean before new hire: 24.9 minutes
Median before new hire: 23 minutes

Mean after new hire: 38 minutes
Median after new hire: 24.5 minutes

Resistance!
↳ Median

Mean is NOT resistant

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (2 of 6)

Importance of previous example:

- Median is resistant
- Mean is not resistant

A numerical summary of data is said to be **resistant** if extreme values (very large or small) relative to the data do not affect its value substantially.

Extreme values relative to the data are called **outliers**.

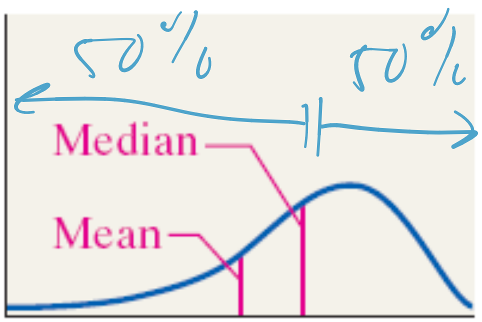
most

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (3 of 6)

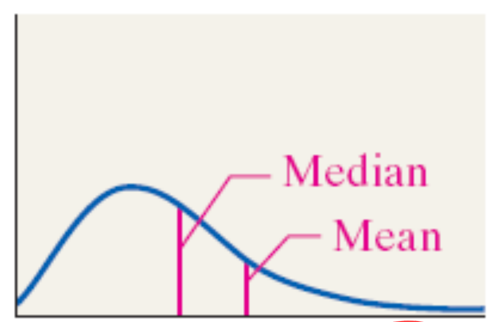
tail pulls the mean (either up or down)

Relation Between the Mean, Median, and Distribution Shape	
Distribution Shape	**Mean versus Median**
Skewed left	Mean substantially smaller than median
Symmetric	Mean roughly equal to median
Skewed right	Mean substantially larger than median



(a) Skewed Left
Mean < Median

(b) Symmetric
Mean = Median



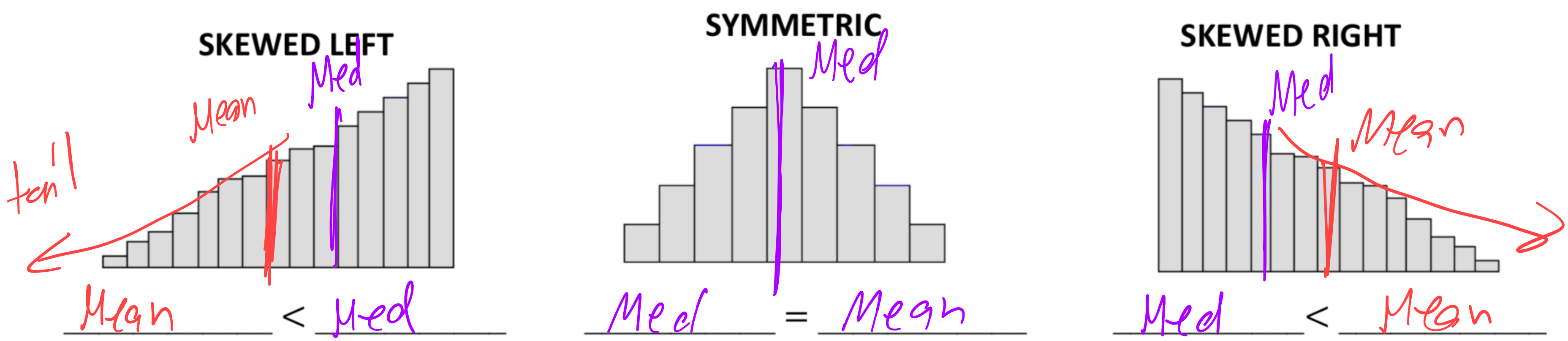
(c) Skewed Right
Mean > Median

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (3 of 6)

Relation Between the Mean, Median, and Distribution Shape

Distribution Shape	**Mean versus Median**
Skewed left	Mean substantially smaller than median
Symmetric	Mean roughly equal to median
Skewed right	Mean substantially larger than median



Pearson

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (4 of 6)

EXAMPLE Describing the Shape of the Distribution

The following data represent the asking price of homes for sale in Lincoln, NE.

79,995	128,950	149,900	189,900
99,899	130,950	151,350	203,950
105,200	131,800	154,900	217,500
111,000	132,300	159,900	260,000
120,000	134,950	163,300	284,900
121,700	135,500	165,000	299,900
125,950	138,500	174,850	309,900
126,900	147,500	180,000	349,900

Source: http://www.homeseekers.com

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (5 of 6)

Find the mean and median.

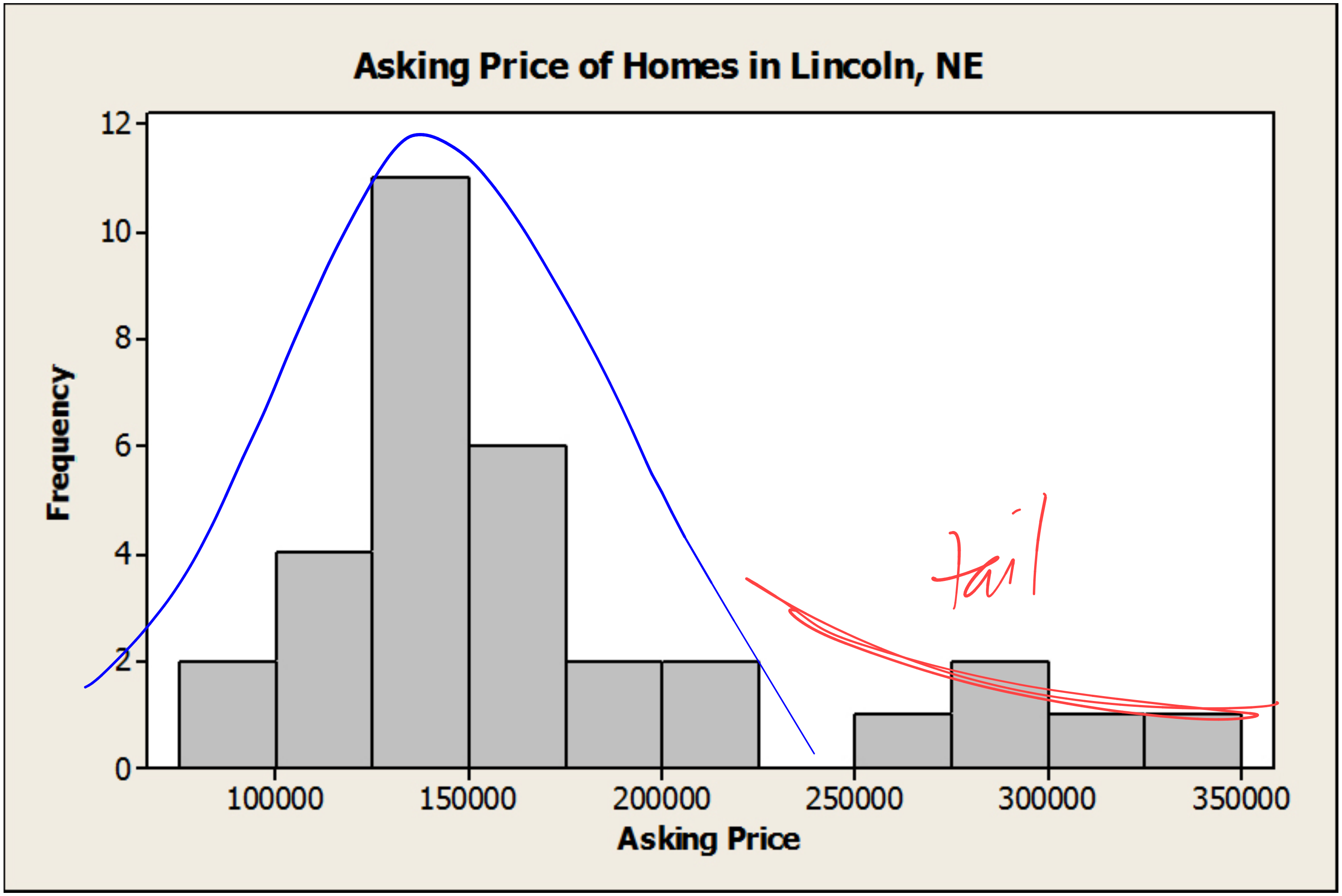
Use the mean and median to identify the shape of the distribution.

Verify your result by drawing a histogram of the data.

The mean asking price is $168,320 and the median asking price is $148,700. Therefore, we would conjecture that the distribution is skewed right.

3.1 Measures of Central Tendency

3.1.3 Explain What It Means for a Statistic to Be Resistant (6 of 6)



3.1 Measures of Central Tendency

3.1.4 Determine the Mode of a Variable from Raw Data (1 of 7)

The **mode** of a variable is the most frequent observation of the variable that occurs in the data set.

A set of data can have no mode, one mode, or more than one mode. "bimodal" "trimodal"

If no observation occurs more than once, we say the data have **no mode**.

3.1 Measures of Central Tendency

3.1.4 Determine the Mode of a Variable from Raw Data (2 of 7)

EXAMPLE Finding the Mode of a Data Set

Brown 5

Ex: Find the mode for each example below.

a) The following data represent the number of O-ring failures on the shuttle *Columbia* for its 17 flights prior to its fatal flight: 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 2, 3	b) The data of the test scores from above: 82, 77, 90, 71, 62, 68, 74, 84, 94, 88 no mode	c) Hair color of ten people in line: Brown, Blonde, Red, Brown, Brown, Blonde, Brown, Blonde, Blonde, Red Brown & Blonde

mode = 0 O-ring failures

units

(each unique)

Hair color

3.1 Measures of Central Tendency

3.1.4 Determine the Mode of a Variable from Raw Data (7 of 7)

Measure of Central Tendency	Computation	Interpretation	When to use
Mean	Population Mean: $\mu = \frac{\Sigma x_i}{N}$ Sample Mean: $\bar{x} = \frac{\Sigma x_i}{N}$	Center of Gravity	When data are quantitative and the frequency distribution is roughly symmetric
Median	Arrange data in ascending order and divide the data set in half	Divides the bottom 50% of the data from top 50% of data	When the data are quantitative and the frequency distribution is skewed left or skewed right
Mode	Tally data to determine most frequent observation	Most frequent observation	When most frequent observation is desired measure of central tendency or the data are qualitative

3.1 Measures of Central Tendency

3.1.4 Determine the Mode of a Variable from Raw Data (7 of 7)

UNITS of:

- **Mean:** $\bar{x} = \dfrac{\sum x}{n}$ ← units of x per units n
- **Median:** same units as x
- **Mode:** same units as x