EDP 613 Fall 2020

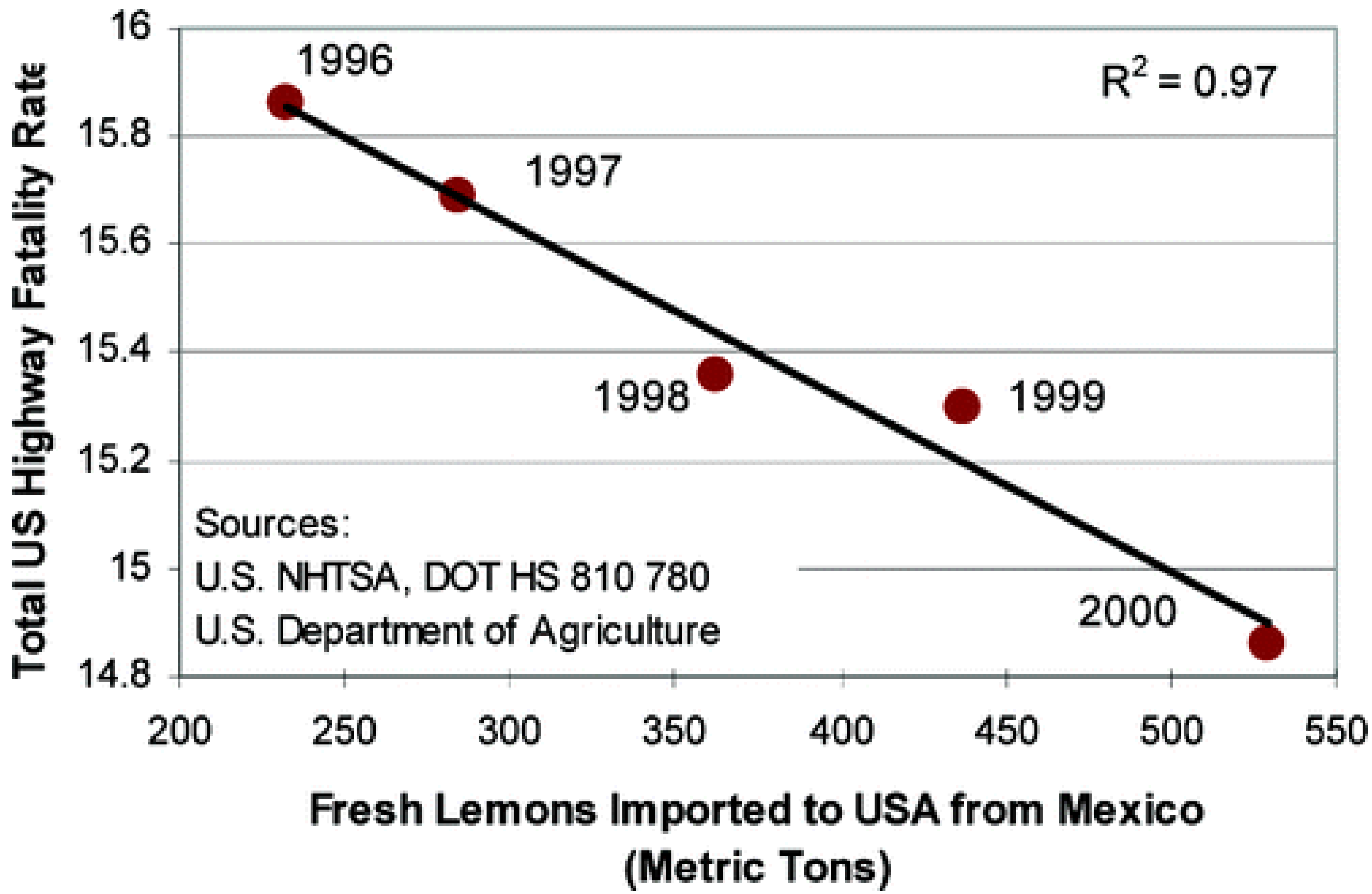
Chapter 3: Measures of Central Tendency

Abhik Roy

Abhik.Roy@mail.wvu.edu

West Virginia University

Always Remember!



CORRELATION DOES NOT PROVE CAUSATION!

Example

Idea: The **average**

- is a measure of central tendency.
- can refer to a Mode, Median or the the Mean.

Notation:

1. If x is a bunch of numbers, then $\sum x$ is the sum of those numbers.

Example: If $x = \{1, 3, -1, 5\}$, find $\sum x$.

Example: If $x = \{1, 3, -1, 5\}$, find $\sum x$.

Solution:

$$\begin{aligned} \sum x &= 1 + 3 - 1 + 5 \\ &= 8 \end{aligned}$$

Definition:

1. **Raw data** is data that has is untouched since its been collected.

The Mode

- of a data set is the value or values that appear most frequently.
- can have more than one value if two or more data points are tied for most frequent.
- can have no value if no value appears more than once.
- gives you at least one number if it exists.

Example: Compute the mode for the following sample:

$$0 \quad 1 \quad 1 \quad 2 \quad 5 \quad 7$$

Example: Compute the mode for the following sample:

$$0 \quad 1 \quad 1 \quad 2 \quad 5 \quad 7$$

Solution:
The data point 1 appears twice whereas all other points appear once. Therefore the mode is 1.

On Your Own: Compute the mode for the following sample:

0 1 3 2 5 7

On Your Own: Compute the mode for the following sample:

$$0 \quad 1 \quad 3 \quad 2 \quad 5 \quad 7$$

Solution:

No data point appears more than once points appear once. Therefore there is no mode.

The Median

- the numerical value separating the higher half of a data set or distribution from the lower half.
- gives you the position of a number.

Steps to compute the median:

1. Sort your data points from least to greatest in numerical value.
2. Count the number data points. Call it n.
3. If n is
   - *odd*: The median is the middle number. This is in the position $\dfrac{n+1}{2}$.
   - *even*: The median is the average of the middle two numbers. This is the average of the two positions $\dfrac{n}{2}$ and $\dfrac{n}{2}+1$.

Example: Compute the median for the following sample:

$$2 \quad 5 \quad 5 \quad 7 \quad 7$$

Example: Compute the median for the following sample:

$$2 \quad 5 \quad 5 \quad 7 \quad 7$$

Solution:

Since these data point are already in numerical order, we can use them as is without reordering.

So $n = 5$ (*odd*) implying we must take the number in the

$$\frac{n+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3^{\text{rd}}$$

position as our median.

$$\text{Median} \quad = \quad 5$$

On Your Own: Compute the median for the following sample:

21.3   31.4   12.7   41.6

On Your Own: Compute the median for the following sample:

21.3   31.4   12.7   41.6

Reordering from least to greatest, we have $12.7, 21.3, 31.4, 41.6$.

So $n = 4$ (*even*) implying we must take the mean of the middle two numbers in the

$$\frac{n}{2} = \frac{4}{2} = 2^{\text{nd}}$$

and

$$\frac{n}{2} + 1 = \frac{4}{2} + 1 = 2 + 1 = 3^{\text{rd}}$$

positions.

The Mean

- is known as the arithmetic mean.
- typically used to describe central tendency in interval-ratio variables.

Notation:

Y	is a set of raw data points.
$\sum Y$	is the sum of all raw data points.
N	is the number of raw data points.
$\overline{Y}$	is the mean of the raw data points.

Steps to compute the mean:

1. Create a fraction.
2. In the top (*numerator*) add up all of the raw data points and put that number here: aka $\sum Y$
3. In the bottom (*denominator*) count up the number of raw data points and put that number here: aka N.
4. Do arithmetic.
5. Resulting number is the arithmetic mean of all raw data points: aka $\overline{Y}$.

Example: Compute the mean for the following sample:

$$Y = \quad 21.3 \quad 31.4 \quad 12.7 \quad 41.6$$

Example: Compute the arithmetic mean for the following sample:

$$Y = \quad 21.3 \quad 31.4 \quad 12.7 \quad 41.6$$

Solution:

$$\begin{aligned}
\overline{Y} &= \frac{\sum Y}{N} \\
&= \frac{21.3 + 31.4 + 12.7 + 41.6}{4} \\
&= \frac{107}{4} \\
&= 26.75
\end{aligned}$$

On Your Own: Compute the arithmetic mean for the following sample:

$$Y = \quad 2 \quad 5 \quad 5 \quad 7 \quad 7 \quad 8 \quad 9$$

On Your Own: Compute the arithmetic mean for the following sample:

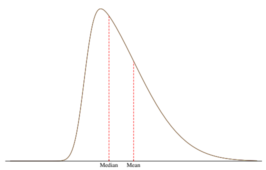
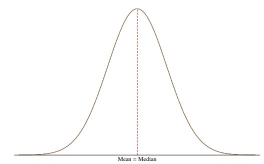
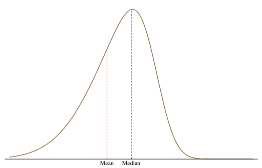
$$Y = \quad 2 \quad 5 \quad 5 \quad 7 \quad 7 \quad 8 \quad 9$$

Solution:

$$\begin{aligned}
\overline{Y} &= \frac{\sum Y}{N} \\
&= \frac{2+5+5+7+7+8+9}{6} \\
&= \frac{43}{7} \\
&\approx 6.14
\end{aligned}$$

Idea

A statistic is **resistant** if its value is not affected by extreme values (large or small) in the data set.

		
Skewed right	Normal	Skewed left
Mean $>$ Median	Mean $=$ Median	Mean $<$ Median

Advantages and Disadvantages of Mean and Median

	Advantages	Disadvantages
Mean	Takes every value into account.	Not resistant.
Median	Resistant.	Dependent*

* on middle value or mean of middle two values.

R Session

Tabled for the in-class portion. A video will post to the website shortly!