“In God we trust. All others must bring data.”

W. Edwards Deming
Measure of Central Tendency

a single or few values that summarize the information of a variable

examples:
- GPA
- unemployment rate
- passing yard per game

three commonly used measures of central tendency
- mean, mode, and median

Mean

mean: the arithmetic average of a continuous variable's scores

$$\bar{Y} = \frac{Y_1 + Y_2 + Y_3 + \ldots + Y_N}{N} = \frac{\sum Y}{N}$$

where the variable Y has N observations;
- \(Y_i\) the value of the variable Y for the \(i\)th observation;
- \(\sum\) add all elements from \(i=1\)st to \(N\)th.

Mean

Example 1: the mean number of children per family

$$\bar{Y} = \frac{7 + 2 + \ldots + 1}{24} = \frac{70}{24}$$

Example 2: the mean age of US presidents when they took office

$$\bar{Y} = \frac{57 + 61 + \ldots + 54 + 57}{44} = 54.7$$

Mode

mode: the value(s) that appears most often

example 1: the modal number of children per family...?

example 2: the most popular baby girl names in 2014? Baby boy names?

Sophia
Jackson
(source: www.babyscenter.com)
Median

Median: the value(s) that appears exactly half of all observations fall above that point and half fall below that point.

The value at which the cumulative percentage reaches 50%.

To compute median:

1. Arrange the observations in order of size from lowest to highest;
2. If N (the number of observations) is odd, then the median is the middle observation;
3. If N is even, then the median is the mean of the middle two observations.

Median example 1: when N is odd...

5, 8, 10

Median example 2: when N is even...

5, 8, 10, 15

The median is (8+10)/2=9

Outliers

The implications of outliers, or extreme values, for mean, median, and mode...

<table>
<thead>
<tr>
<th>Values</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2, 2, 5</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2, 2, 5, 7</td>
<td>4</td>
<td>3.5</td>
<td>2</td>
</tr>
<tr>
<td>2, 2, 5, 100</td>
<td>27.25</td>
<td>3.5</td>
<td>2</td>
</tr>
</tbody>
</table>

Outliers strongly influence the mean, but not much the median or the mode.
Why Median Household Income?

<table>
<thead>
<tr>
<th>Rank</th>
<th>State</th>
<th>Median Income</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mississippi</td>
<td>35,521</td>
<td>1,294</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>51</td>
<td>Maryland</td>
<td>76,165</td>
<td>4,272</td>
</tr>
</tbody>
</table>

Measure of Central Tendency for Continuous Variable

data from a survey about study hours:

12.5 10 10 3 10 12 3.5 5 2.5 25 20 2 15 4 10 20 22.5 11 7.5 15 15 7.5

mean? mode? median?

Measure of Central Tendency for Continuous Variable

data from a survey about work hours:

27.5 20 30 40 0 27.5 0 30 35 15 27.5 0 30 10 0 40 15 30 35

mean? mode? median?

Measure of Central Tendency for Continuous Variable

data from a survey about work hours:

27.5 20 30 40 0 27.5 0 30 35 15 27.5 0 30 10 0 40 15 30 35... depends...

mean? mode? median?
For Discrete Variable

nominal variables: mean? median? mode (yes)

ordinal: mean? median (yes-ish), mode (yes)

Measure of Central Tendency for Discrete Variable

data from a survey about home town:

DE DE DE DE CHN DE DE PA DE NY NJ DE NJ DE DE MD NJ DE PA

mode?

Measure of Central Tendency for Discrete Variable

data from a survey about astrological sign:

Taurus, Cancer, Scorpio, Aquarius, Scorpio, Gemini, Pisces,
Aquarius, Taurus, Gemini, Aquarius, Aries, Gemini, Libra, Sagittarius,
Gemini, Leo, Pisces, Gemini, Leo, Libra

mode?