Measures of Central Tendency

EDP 613

Week 3

Basic Idea

The mean, median and mode are measures of central tendency and attempt to summarize the typical value of a variable.

Why?

These may help us draw conclusions about a specific group or compare different groups using a single numerical value.

Recall Distributions

Skewed Left

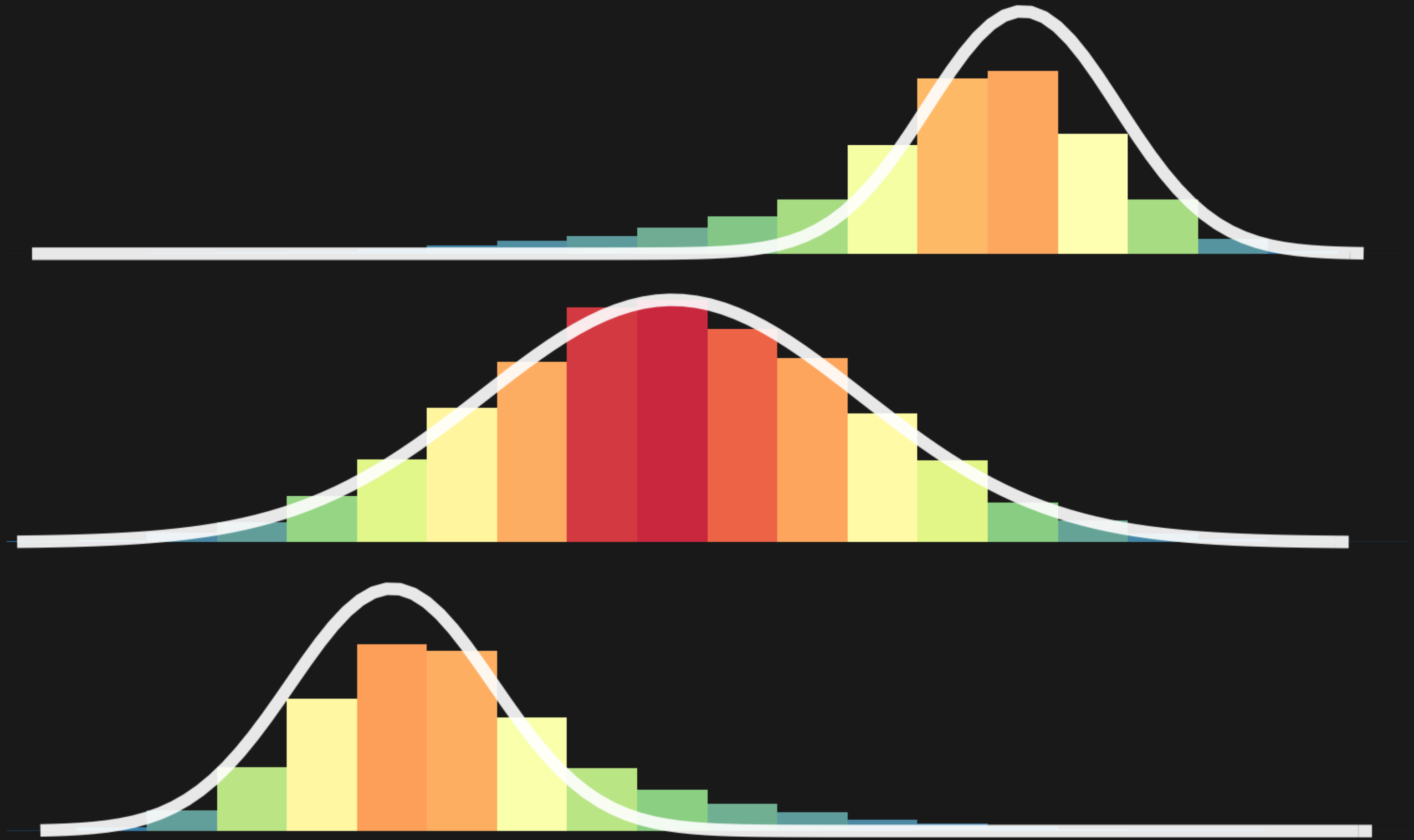
mean < median < mode

Normal

mean = median = mode

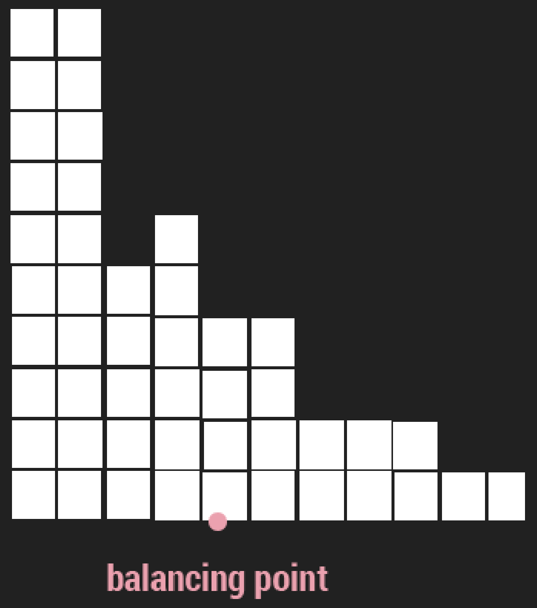
Skewed Right

mean > median > mode



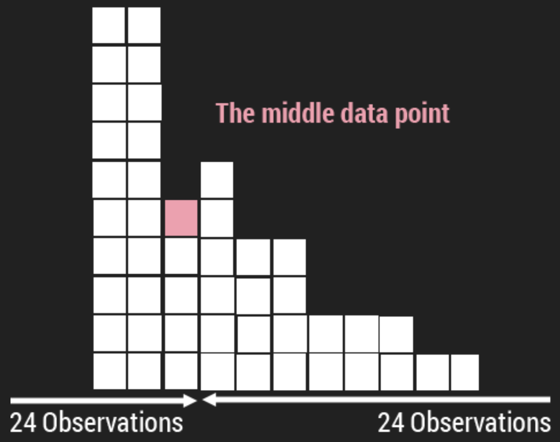
Measures of Central Tendency: The Mean

- The *average* number.
- There are other types of means (e.g. geometric, harmonic, etc.) but we are only using the *arithmetic* mean.
- Essentially the **balancing point** or center of mass of a distribution
- Found by adding all data points and dividing by the number of data points



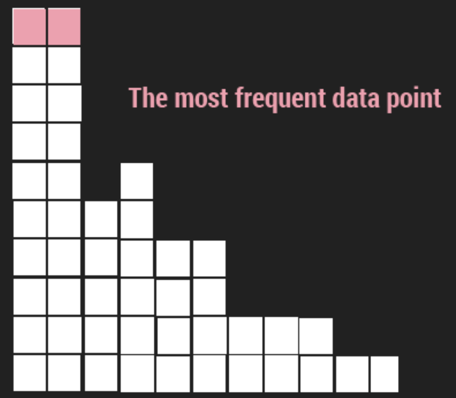
Measures of Central Tendency: The Median

- The *middle* number
- Essentially the **point that cuts a data set in half**
- Found by ordering data points from least to greatest or greatest to least and locating the middle number - if there are two middle data points, they are averaged



Measures of Central Tendency: The Mode

- The *most frequent* number
- Essentially the point that occurs the most
- Found by determining the data point(s) that appear the most - if none exists, then there is no mode



Basic Procedure: The Mean

> - **Mean**
>   - Add the numbers up, divide by the total number of values in the set.
>   - Denoted by $\overline{Y}$

Example

Compute the mean for the following sample: $\{21.3, 31.4, 12.7, 41.6\}$

Solution

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

Give it a Try

Compute the mean for the following sample: $\{2, 5, 5, 7, 7, 8, 9\}$

Solution

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

Basic Procedure: The Median

> - **Median**
>   - Put the numbers in order from least to greatest or greatest to least and find the middle number.
>   - If there are two middle numbers, average them.

Example

Compute the median for the following sample: $\{2, 5, 5, 7, 7\}$

Solution

> - Since these data point are already in numerical order, we can use them as is without reordering.
> - $n = 5$ which is an odd number so we can locate the median by
>
> $$\frac{n+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$$
>
> telling us to look in the *third position* from either side of the list of numbers.
>
> - In
>
> $$\{2, 5, 5, 7, 7\}$$
>
> the middle number is 5 so that must be the median!

Give it a Try

Compute the mean for the following sample: $\{21.3, 31.4, 12.7, 41.6\}$

Solution

> - Since these data point are NOT already in numerical order, we must reorder them.
>
> $$\{12.7, 21.3, 31.4, 41.6\}$$

> - $n = 4$ which is an even number so we can locate the median by taking the mean of the the numbers in

- $\frac{n}{2} = \frac{4}{2} = 2$, or the *second position*
- $\frac{n}{2} + 1 = \frac{4}{2} + 1 = 3$, or the *third position*
- So the median is

$$\frac{21.3 + 31.4}{2} = 26.35$$

Basic Procedure: The Mode

> - **Mode**
>   - Find the number(s) that appear the most.
>   - If none exists, then there is no mode.

Example

Compute the mode for the following sample: $\{2, 5, 5, 7, 7\}$

Solution

> -
>
> | Data point | Frequency |
> |---|---|
> | 2 | 1 |
> | 5 | 2 |
> | 7 | 2 |

> - The data points 5 and 7 repeat twice while 2 only appears once.
> - The modes are 5 and 7.
> - Known as *bimodal*. Three modes would be *trimodal* and so on.

Give it a Try

Compute the mode for the following sample: $\{21.3, 31.4, 12.7, 41.6\}$

Solution

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

Something to Think About

- A statistic is **resistant** if its value is not affected by extreme values (large or small) in the data set.
- Which of the measures of central tendency are resistant?

That's it. We will work more with R next week!