Measures of Central Tendency

Descriptive Statistics Part 1

Cal State Northridge
Ψ320
Andrew Ainsworth PhD

What is a typical score like?

- There are three indices of this central tendency:
  - Mode
  - Median
  - Mean

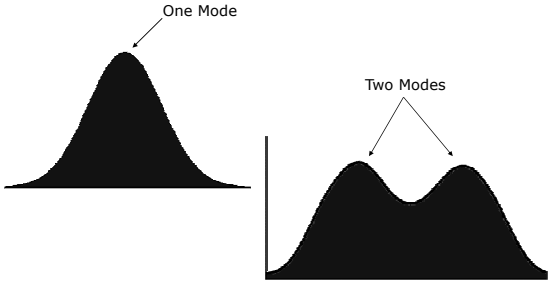
Measures of Central Tendency

Measure	Definition	Level of Measurement	Disadvantage
Mode	Most frequent value	nom., ord., int./rat.	Crude
Median	Middle value	ord., int./rat.	Only two points contribute
Mean	Arithmetic average	int./rat.	Affected by skew

The Mode

- Mode: most common (frequent) number
- Peak of distribution
- Always an observed score
- With grouped data is merely the *midpoint* of the most frequent measurement class.
- Thus, if a case were drawn at random from the distribution, that case is more likely to fall in the modal class than any other.
  - Important interpretation

How to find the mode?



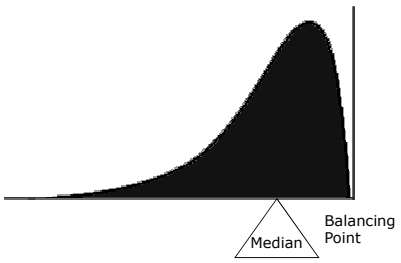
Problems with the Mode

- However, there are three disadvantages with the mode:
  1. Multiple modes (especially with grouped data)
  2. The mode is very sensitive to the size and number of class intervals (different intervals = different modes)
  3. The mode of a sample undependable when estimated population

The Median

- Median: middle of distribution (50th percentile)
- Not applicable to nominal data; ordinal, interval & ratio data only
- Is considerably less sensitive when grouping into class intervals.
- More useful for making inferences (although not the best)
- Not affected by *outliers.*

Median



Median

- The median requires 1 or 2 pieces of information
- If there are an odd number of scores the median is the center score
- If there is an even number the median is midpoint between the two middle numbers

Median

- Examples
  - Data Set 1: {**7, 2, 9, 3, 4, 5, 8**}
    - Ordered: _____
    - **Median =** _____
  - Data Set 2: {**11, 15, 10, 9, 5, 13, 12**}
    - Ordered: _____
    - **Median =** _____

Median

Data set 1: {12, 9, 8, 5, 3, 2}
Data set 2: {77, 50, 8, 5, 5, 5}

- Example: In a six-score set of data the only scores that matter when calculating the median are the two middle scores
- The rest of the data is ignored
- Because of this the median is not sensitive to outlying scores

Median

- The average **absolute deviation** is smallest when taken from the median.

$$A.A.D = \frac{\sum_i \left| X_i - Mdn \right|}{n}$$

The Mean

- Only used for interval & ratio data.

$$Mean = M_X = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Major advantages:
  - The sample value is a very good estimate of the population value.
  - Mean of the sample = $\overline{X}$
  - Mean of the population = μ

The Mean

- Example
  - Data Set: {3, 4, 6, 10}
  - Formula: $\overline{X} = \frac{\sum X}{n}$

$$\overline{X} = \frac{\_\_ + \_\_ + \_\_ + \_\_}{\_\_} = \_\_\_\_$$

The Mean and Deviations

- For the mean the following is always true.

$$\sum \left(X_i - \overline{X} \right) = 0$$

$$\sum \left(X_i - \overline{X} \right)^2 = \text{minumum}^*$$

*When compared to the other 2 measures of central tendency

Mean: Demonstration

- Since $\overline{X} = 5.75$

Score	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$
3	(3 - 5.75) = **-2.75**	**7.563**
4	(4 - 5.75) = **-1.75**	**3.063**
6	(6 - 5.75) = **0.25**	**0.063**
10	(10 - 5.75) = **4.25**	**18.06**
Σ	**0**	**28.75**

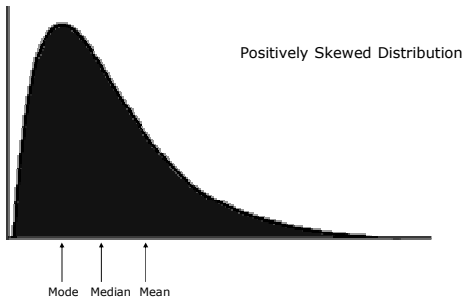
The Mean

- **Mental Note:** If the mean is guessed as the value of any case drawn at random from a distribution, on average the amount of signed error will be 0.
  - This is a most important interpretation of the mean and why it's so most often!
- Disadvantage
  - Highly affected by outliers

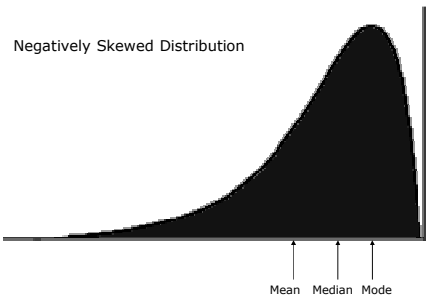
The Influence of Outliers

Data	Ordered Data	Data	Ordered Data
8	6	8	6
10	6	10	6
9	8	9	8
9	9	9	9
6	9	6	9
6	9	6	9
9	9	9	9
13	10	33	10
10	10	10	10
9	13	9	33
Mode =		Mode =	
Median =		Median =	
Mean =		Mean =	

Mode, Mean and Median

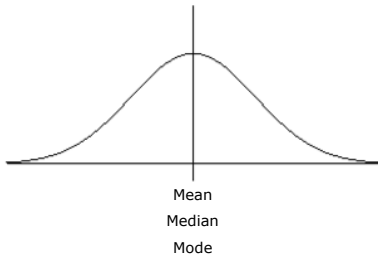


Mode, Mean and Median



Mode = median = mean?

- When the distribution is symmetric



The mean vs. the median

- The degree of discrepancy between them indicates the skewness of the data
- The closer the two values are the more symmetric the data