

Central Tendency

Ray Block, Jr.
PS #585 (Research Methods)
Fall 2003

Today's Blueprint

Last Class(es)

- Hypothesis Testing
- Confidence Intervals
 - The General Idea
 - The Notion of Error

Today's Class

- Univariate Data Analysis (Part 1)
 - Statistical Models
 - Measures of Central Tendency

Statistical Models

Statistical Models

- What are Models?
 - Abstractions from reality that order and simplify our view of reality
- What Purpose Do They Serve?
 - Discuss significant relationships among *concepts*
 - Enable researchers to form testable propositions between *variables*
 - Summarize *data*
- Remember:
 - You *cannot* test theories
 - You *can* test models based on theories

Statistical Models

- Model Building:
 - Models are symbolic representations of real-world phenomena
 - The goal is to build a model that best represents the real-world phenomena of interest

Statistical Models

- How to Build Models:
 - Observe some facts about the world
 - Speculate about the process(es) that produced those facts (i.e. the data generating process)
 - Collect data that represent the process(es)
 - Reduce the process(es) to a statistical model using the data you collected

Statistical Models

- We use the models we build to make *inferences* and *predictions* about real-world processes
- We want our models to be accurate so that our inferences and predictions will also be accurate
- If we want our inferences and predictions to be accurate, then the models we build must accurately represent the data we collect
- The degree to which a statistical model represents the data collected is known as the *fit* of the model to the data

Statistical Models

- We will discuss several simple statistical models
- These models fall into one of the following general categories:
 - Measures of Central Tendency
 - Measures of Spread

Measures of Central Tendency

Think: The 4 “Ms”

Measures of Central Tendency

- Measures of Central Tendency (The 4 “Ms”):
 - The Midpoint
 - The Mode
 - The Median
 - The (Arithmetic) Mean

Measures of Central Tendency

- The *midpoint* (?) is the value that falls equidistant from the lowest and highest points in a scale
 - The Midpoint is not used very often
 - It is a very rough estimate of the average

Measures of Central Tendency

- The *mode* (“Maximum Frequency” or “Mo”) is the most frequently occurring number in a list of numbers
 - It is the closest thing to what people mean when they say something is “average” or “typical”
- Calculating Mo:
 - The mode can easily be found by inspection, rather than through computation

Measures of Central Tendency

- The *median* (“middlemost value” of “Mdn(x)”) is the number that falls in the middle of a range of numbers
- Calculating Mdn(x):
 - The median position can be found by inspection or by the following formula:

$$\text{Position of Midpoint} = \frac{N+1}{2}$$

- Where:
 - N = Total number of scores (observations)
- Interpreting Mdn(x):
 - It’s not the average; it’s the halfway point
 - There are always just as many numbers above the median as below it

Measures of Central Tendency

- The most commonly used measure of central tendency is the (arithmetic) *mean* (“Center of Gravity” or “x-bar”)
- Calculating X-Bar:
 - The mean = sum of scores divided by the total number of scores:

Measures of Central Tendency

- Calculating X-Bar:
 - The mean = sum of scores divided by the total number of scores:

$$\bar{X} = \frac{\sum X_i}{N}$$

- Where:
 - X-Bar = [Arithmetic] Mean

- Σ = Sum
 - X_i = Each individual value of X
 - N = Total number of scores (observations)
- Interpreting X-Bar:
 - It's not the average nor a halfway point
 - It is a kind of center that balances high numbers with low numbers

Measures of Central Tendency

- Uses:
 - The mean is the most important measure of central tendency in statistics
 - Most measures of spread are based on the mean
 - Why? Because the mean is the number which has the smallest squared distance from all other numbers in the distribution

Measures of Central Tendency

- Step-by-Step Illustration:
 - Suppose that a volunteer canvasses houses in her neighborhood collecting money for a local charity. She receives the following donations (in dollars):

5
10
25
15
19
2
5

- Here are the steps you would use to calculate the mode, median and the mean:
- **Step 1:** Arrange the scores from highest to lowest:

25
18
15
10
5
5
2

- **Step 2:** Find the most frequently occurring score:
 - By inspection, mode = \$5

25
18
15
10
5
5
2

- **Step 3:** Find the middlemost score:
 - By inspection: Because there are 7 scores (an odd number) the fourth score from either end is the median
 - By the formula: $(N-1)/2 = (7-1)/2 = 3$, so the median is the fourth score from either end
 - In both cases, median score = \$10

25
18
15
10
5
5
2

- Step 4: Determine the Sum of the Scores

25
18
15
10
5
5
2
+
80

- Therefore, $\sum X_i = 80$
- Step 5: Determine the Mean by dividing the Sum by the Number of the Scores

$$\bar{X} = \frac{\sum X_i}{N} = \frac{\$80}{7} = 11.43$$

Measures of Central Tendency

- The mode, median, and mean provide different pictures of “charitable” giving in the neighborhood
 - The mode suggests that donations are typically small
 - The median suggests that the average donation is more generous
 - The mean paints the most generous picture of the average donation

Measures of Central Tendency

- Which Measure You Use Depends On:
 - The Data’s Level of Measurement
 - The Distribution of Data
 - The Research Objective

Measures of Central Tendency

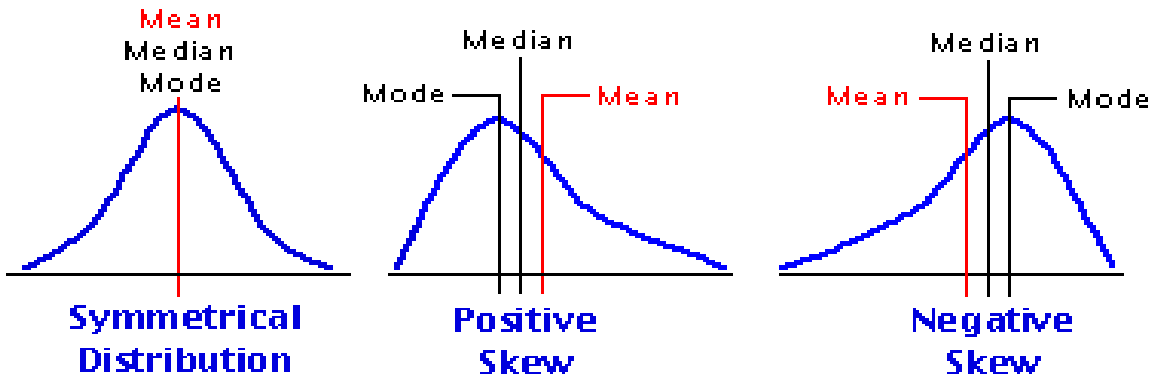
- The Data’s Level of Measurement:

	Mode	Median	Mean
Nominal	✓		
Ordinal	✓	✓	
Interval/Ratio	✓	✓	✓

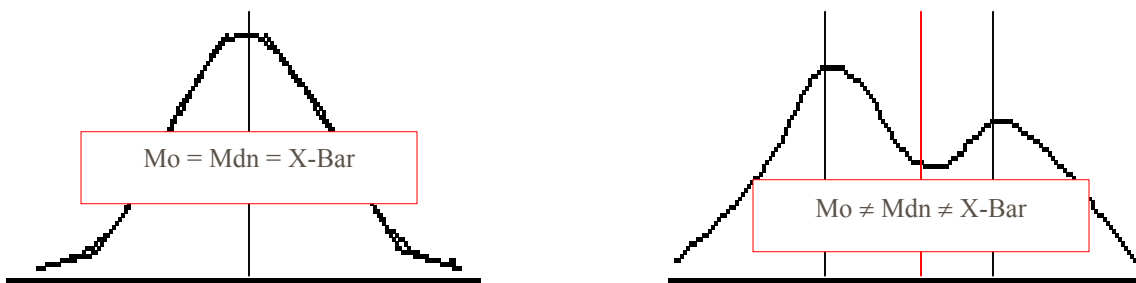
- For nominal level data, you can only use the mode
- For ordinal level data, you can use the mode or the median
- For interval level data, you can use the mode, median, or mean

Measures of Central Tendency

- Distribution of Data = Shape of the Distribution
 - Symmetric
 - Skewed
 - Positive (right)
 - Negative (left)
 - Unimodal, multimodal



- Unimodal = One mode
- Multi modal = More than one mode



- In a unimodal symmetric distribution (see image above to the left) the mean, median, and mode are identical
- In a bimodal distribution (see image to the right) the mode, median, and mean differ

Measures of Central Tendency

- Research Objective
- For the Mode, the goal is to obtain fast, simple, but rough measure of central tendency
- For the Median, The goal is to obtain precise measure of central tendency
 - Sometimes can be used for more advanced statistical operations or for splitting distributions into categories (for example, low versus high)
- For the Mean, The goal is to obtain precise measure of central tendency
 - Often can be used for more advanced statistical operations, including hypothesis tests

References

FYI:

- Levin, Jack and James Alan Fox. 2003. Elementary Statistics in Social Research, 9th Edition. Boston, MA: Pearson Education Group, Inc.