

Statistical Analysis

- Statistics
 - Are tools that help us to build meaningful research studies
 - Statistics help us make sense of raw data

Descriptive Statistics

- Tools used to describe a set of numbers that describe key characteristics of the larger data set
- Allows researchers to learn about entire data set without looking at each piece of data

■ Two main types of Descriptive Stats

- 1. Measures of Central Tendency
 - Describes the center point of a set of raw data
 - Is there a number that best represents the general response subjects gave?
 - Means, median, and mode

■ Mean

- The arithmetic average – add all scores together and divide by the total number of scores
- Most commonly reported statistic
- Example: 2, 4, 5, 6, 7, 9, 9 mean = 6
- Advantages
 - Can be manipulated algebraically
 - Neither the mode or median can be
- Disadvantages
 - Can be influenced by an outlier
 - 6, 5, 6, 4, 8, 67, 4, 4, 5, 7 mean = $116/10 = 11.6$

- **Median**

- The scores where half of the observations are above it, and half of the observations are below it
- The scores that divides the distribution in half
- Example: 3 4 5 6 6 6 7 8 9
 - Mean is 6
 - Median is also 6-because half of the number fall above and below
- Advantages
 - Insensitive to outliers, especially with smaller numbers
 - Example: 3 4 5 6 6 6 7 8 18
 - Mean is 7 but the median is 6
- Disadvantages
 - Does not really help with sophisticated data analysis
 - Example: 100, 50, 0 median = 50

- **Mode**

- The number or score that occurs most often
- 2 4 5 3 6 6 7 6 6 5 4 6 6 8 7 6 6 mode = 6
- Advantages
 - It is a score that actually occurred
- Disadvantages
 - May not be representative of all our data

- Measures of central tendency fail to tell us the whole story
 - Tells us the average score, the score that occurs the most often, and the half-way point
 - We need additional measures so we can better understand the degree to which observations are clustered about or deviate from the central value-how scores vary

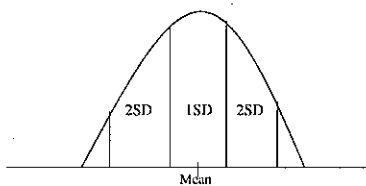
■ 2. Measures of Variability (Dispersion)

- Helps us understand how the scores vary from the center point or average
- Also, were scores pretty close together or did they go all over the place?
- Range, variance, and standard deviation

■ Standard Deviation

- Tells us exactly how much each raw score deviates from the mean, on average
- Example: If the mean is 5 and the SD is 2, it tells us that each individual score varies, on average, 2 points from the mean.
- If you have a normal distribution of numbers (bell curve), 68% of individual scores will be within 1 SD on either side of the mean
- 95% of the numbers fall within 2 SD's of the mean
- 99% of the numbers fall within 3 SD's of the mean

■ Standard Deviation-normal distribution



■ Example:

- The mean starting salary for police officers this year is about \$25,000 with a SD of about \$4,000. What can we conclude about 68% of the police officers' starting salaries?
- 68% of the officers' starting salaries range from \$21,000 to \$29,000

■ Variance

- How scores are spread out, or dispersed around the mean of the data
 - Are they spread out or are they tightly clustered around the mean
 - Example: We know the average shoe size for women in the US is an 8 1/2
 - Do women typically wear sizes close or far from 8 1/2?
 - Variance tells shoe store owners what to keep in stock
- A: 4 5 6
 ■ B: 3 5 7
 ■ C: 2 5 8
 ■ D: 1 5 9
- Mean for each one is 5
 - But, each next set has greater variability from the mean

■ Variance

