Module 3: Descriptive Statistics

The Applied Research Center
Module 3 Overview

- Measures of Central Tendency
- Measures of Variability
- Frequency Distributions
- Running Descriptive Statistics
Measures of Central Tendency

- Three measures of central tendency are available
  - The Mean
  - The Median
  - The Mode

- Unfortunately, no single measure of central tendency works best in all circumstances
  - Nor will they necessarily give you the same answer
Example

- SAT scores from a sample of 10 college applicants yielded the following:
  - Mode: 480
  - Median: 505
  - Mean: 526

- Which measure of central tendency is most appropriate?
The Mean

- The mean is simply the arithmetic average
- The mean would be the amount that each individual would get if we took the total and divided it up equally among everyone in the sample
- Alternatively, the mean can be viewed as the balancing point in the distribution of scores (i.e., the distances for the scores above and below the mean cancel out)
The Median

- The median is the score that splits the distribution exactly in half
- 50% of the scores fall above the median and 50% fall below
- The median is also known as the 50th percentile, because it is the score at which 50% of the people fall below
A desirable characteristic of the median is that it is *not* affected by extreme scores

Example:
- Sample 1: 18, 19, 20, 22, 24
- Sample 2: 18, 19, 20, 22, 47

Thus, the median is not distorted by skewed distributions
The Mode

- The mode is simply the most common score
- There is no formula for the mode
- When using a frequency distribution, the mode is simply the score (or interval) that has the highest frequency value
- When using a histogram, the mode is the score (or interval) that corresponds to the tallest bar
Distribution Shape and Central Tendency

- In a normal distribution, the mean, median, and mode will be approximately equal
Skewed Distribution

- In a skewed distribution, the mode will be the peak, the mean will be pulled toward the tail, and the median will fall in the middle.

![Diagram showing skewed distribution with mode (Mo), median (Med), and mean (\(\bar{X}\))]
Choosing the Proper Statistic

- **Continuous data**
  - Always report the mean
  - If data are substantially skewed, it is appropriate to use the median as well

- **Categorical data**
  - For nominal data you can only use the mode
  - For ordinal data the median is appropriate (although people often use the mean)
Example

- SAT scores from a sample of 10 college applicants yielded the following:
  - Mode: 480
  - Median: 505
  - Mean: 526

- Which measure of central tendency is most appropriate?
Measures of Variability

- The fluctuation of scores about a central tendency is called “variability.”
- We can use measures of variability to compare two sets of scores.
- Although the means may be the same, the distribution may be different.

Measure of Variability
- Range
- Standard Deviation
- Variance
Range

- Range is the distance between two extreme scores.
- It informs us about the dispersion of our distribution.
- The larger the range the larger the dispersion from the mean value.
- Although the mean of the scores of two distributions can be identical their ranges may be different.
Drawbacks to the Range

- Good preliminary measure, but one single extreme value can influence the range significantly.
- The calculation of the range is derived from the highest and lowest values and doesn’t tell us anything about the variability of the different values.
Standard Deviation

- Defined as the variability of the scores around the mean
- Each score in a distribution varies from the mean by a greater or lesser amount, except when the score is the same as the mean.
- Deviations from the mean can be noted as either positive or negative deviations from the mean.
- The average of these deviations would equal “zero.”
Standard Deviation (cont’d)

- Large SD

- Small SD
Variance

- The variance and the closely-related standard deviation are measures of how **spread** out a distribution is.
Frequency Distribution Tables
After collecting data, researchers are faced with pages of unorganized numbers, stacks of survey responses, etc.

The goal of descriptive statistics is to aggregate the individual scores (datum) in a way that can be readily summarized.

A frequency distribution table can be used to get “picture” of how scores were distributed.
Frequency Distributions

- A frequency distribution displays the number (or percent) of individuals that obtained a particular score or fell in a particular category.
- As such, these tables provide a picture of where people respond across the range of the measurement scale.
- One goal is to determine where the majority of respondents were located.
When To Use Frequency Tables

- Frequency distributions and tables can be used to answer **all** descriptive research questions.
- It is important to **always** examine frequency distributions on the IV and DV when answering comparative and relationship questions.
Three Components of a Frequency Distribution Table

- **Frequency**
  - the number of individuals that obtained a particular score (or response)

- **Percent**
  - The corresponding percentage of individuals that obtained a particular score

- **Cumulative Percent**
  - The percentage of individuals that fell at or below a particular score (not relevant for nominal variables)
Example

- What are the ages of students in an online course?
- Are students likely to recommend the course to others?
- Step 1: Input the Data into SPSS

<table>
<thead>
<tr>
<th>Age</th>
<th>Recommend</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>31</td>
<td>5</td>
</tr>
<tr>
<td>38</td>
<td>4</td>
</tr>
<tr>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>37</td>
<td>3</td>
</tr>
<tr>
<td>43</td>
<td>4</td>
</tr>
<tr>
<td>41</td>
<td>5</td>
</tr>
<tr>
<td>49</td>
<td>4</td>
</tr>
<tr>
<td>40</td>
<td>2</td>
</tr>
</tbody>
</table>
Example (cont’d)

- Step 2: Run the Frequencies
- Analyze → Descriptive Statistics → Frequencies
- Move variables to the Variables box (select the variables and click on the arrow).
- Click OK.
Example

- Frequency distribution showing the ages of students who took the online course

<table>
<thead>
<tr>
<th>Age</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>18.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
</tr>
<tr>
<td>26.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>14.3</td>
</tr>
<tr>
<td>31.00</td>
<td>2</td>
<td>14.3</td>
<td>14.3</td>
<td>28.6</td>
</tr>
<tr>
<td>32.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>35.7</td>
</tr>
<tr>
<td>35.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>42.9</td>
</tr>
<tr>
<td>37.00</td>
<td>2</td>
<td>14.3</td>
<td>14.3</td>
<td>57.1</td>
</tr>
<tr>
<td>38.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>64.3</td>
</tr>
<tr>
<td>40.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>71.4</td>
</tr>
<tr>
<td>41.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>78.6</td>
</tr>
<tr>
<td>43.00</td>
<td>1</td>
<td>7.1</td>
<td>7.1</td>
<td>85.7</td>
</tr>
<tr>
<td>49.00</td>
<td>2</td>
<td>14.3</td>
<td>14.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>
Example (cont’d)

- Student responses when asked whether or not they would recommend the online course to others

<table>
<thead>
<tr>
<th>Recommend</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>3</td>
<td>21.4</td>
<td>21.4</td>
<td>21.4</td>
</tr>
<tr>
<td>Probably Would Not</td>
<td>2</td>
<td>14.3</td>
<td>14.3</td>
<td>35.7</td>
</tr>
<tr>
<td>May or May Not</td>
<td>6</td>
<td>42.9</td>
<td>42.9</td>
<td>78.6</td>
</tr>
<tr>
<td>Probably Would</td>
<td>3</td>
<td>21.4</td>
<td>21.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Definitely Would</td>
<td>3</td>
<td>21.4</td>
<td>21.4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

- Most would recommend the course
Running Descriptive Statistics
Example

- Are there differences in the anxiety levels of students who have had statistics before versus students who have never had statistics?
Example (cont’d)

- Step 1: Input the data into SPSS

<table>
<thead>
<tr>
<th>Stats History</th>
<th>Anxiety Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
</tr>
</tbody>
</table>
Example (cont’d)

- **Step 2: Run the descriptive statistics**
  - Analyze → Compare Means → Means
  - Anxiety = Dependent List
  - Stats History = Independent List
  - Click Options
    - Move Median over
    - Move Minimum over
    - Move Maximum over
  - Click Continue
  - Click OK
Example (cont’d)

<table>
<thead>
<tr>
<th>statshist</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>84.0000</td>
<td>5</td>
<td>11.40175</td>
<td>85.0000</td>
<td>65.00</td>
<td>95.00</td>
<td>30.00</td>
<td>130.000</td>
</tr>
<tr>
<td>yes</td>
<td>57.0000</td>
<td>5</td>
<td>16.43168</td>
<td>65.0000</td>
<td>35.00</td>
<td>75.00</td>
<td>40.00</td>
<td>270.000</td>
</tr>
<tr>
<td>Total</td>
<td>70.5000</td>
<td>10</td>
<td>19.50071</td>
<td>70.0000</td>
<td>35.00</td>
<td>95.00</td>
<td>60.00</td>
<td>380.278</td>
</tr>
</tbody>
</table>
Example (cont’d)

- Step 3: Create a Histogram for Anxiety with a normal curve option
  - Graphs → Legacy Dialogs → Histogram
  - Variable = anxiety
  - Check the “Display normal curve” check box
  - Click Ok
Histogram for Anxiety

- Mean = 70.50
- Std. Dev. = 19.501
- N = 10
Step 4: Write up the results

Descriptive statistics revealed that students who had previous experience with statistics ($M = 57.00$, $SD = 16.43$) had lower anxiety at the beginning of the semester than students who did not have any previous experience with statistics ($M = 84.00$, $SD = 11.40$).
Module 3 Summary

- Measures of Central Tendency
- Measures of Variability
- Frequency Distributions
- Running Descriptive Statistics
Review Activity and Quiz

- Please complete the Module 3 Review Activity: Descriptive Statistics Terminology located in Module 3.
- Upon completion of the Review Activity, please complete the Module 3 Quiz.
- Please note that all modules in this course build on one another; as a result, completion of the Module 3 Review Activity and Module 3 Quiz are required before moving on to Module 4.
- You can complete the review activities and quizzes as many times as you like.
Upcoming Modules

- Module 1: Introduction to Statistics
- Module 2: Introduction to SPSS
- Module 3: Descriptive Statistics
- Module 4: Inferential Statistics
- Module 5: Correlation
- Module 6: t-Tests
- Module 7: ANOVAs
- Module 8: Linear Regression
- Module 9: Nonparametric Procedures