Lecture Notes

# Chapter 3: Measures of Central Tendency

## Learning Objectives

1. Explain the importance of measures of central tendency.
2. Calculate and interpret the mode, the median, and the mean.
3. Identify the relative strengths and weaknesses of the three measures.
4. Determine and explain the shape of the distribution

## Chapter Outline

1. Introduction
	1. Often need to describe a large set of multivariate data for which graphs and tables may not be the most efficient tools.
	2. Numbers that describe what is average or typical of the distribution are called **measures of central tendency**.
	3. The three measures of central tendency:
		1. The mode
		2. The median
		3. The mean
	4. The choice of an appropriate measure of central tendency for representing a distribution depends on three factors:
		1. The way the variables are measured (their level of measurement)
		2. The shape of the distribution
		3. The purpose of the research.
2. The Mode
	1. The **mode** is the category or score with the largest frequency or percentage in the distribution.
	2. Of all the measures of central tendency, the mode is the easiest one to identify.
	3. The mode is not necessarily the category with the majority (i.e., more than 50%) of cases, it is simply the category in which the largest number (or proportion) of cases fall.
	4. The mode is the only measure of central tendency that can be used with nominal-level variables.
		1. With nominal variables, we are only able to classify respondents based on a qualitative and not on a quantitative property.
		2. The mode can also be used to describe the most commonly occurring category in any distribution.
3. The Median
	1. The **median** is a measure of central tendency that can be calculated for variables that are at least at an ordinal level of measurement.
	2. It is the score that divides the distribution into two equal parts so that half the cases are above it and half below it.
	3. The median is a suitable measure for those variables whose categories or scores can be arranged in order of magnitude from the lowest to the highest.
	4. Therefore, the median can be used with ordinal or interval-ratio variables, for which scores can be at least rank-ordered but cannot be calculated for variables measured at the nominal level.
	5. Finding the Median in Sorted Data
4. In most cases, a simple inspection of the sorted data can do it.
5. The location of the median score differs somewhat, depending on whether the number of observations is odd or even.
6. An Odd Number of Cases
7. To locate the median, first arrange the responses in order from the lowest to the highest (or the highest to the lowest).
8. The median is the response associated with the middle case. Find the middle case when *N* (number of observations) is odd by adding 1 to *N* and dividing by 2.
9. The median is the response associated with the middle case.
10. An Even Number of Cases
11. To locate the median, first arrange the number of cases in order from the lowest to the highest.
12. When N is even, we no longer have a single middle case.
13. The median is therefore located halfway between the two middle cases.
14. Find the two middle cases by using the previous formula:

(*N* + 1)/2, where *N* is the number of cases.

1. As a note of caution, when data are ordinal, averaging the middle two scores is no longer appropriate. The median simply falls between two middle values.
	1. Finding the Median in Frequency Distributions
		1. To find the median, identify the category associated with the observation located at the middle of the distribution.
		2. Use formula (*N* + 1)/2.
		3. The median is the value of the response category not the frequency.
	2. Locating Percentiles in a Frequency Distribution
2. The median is a special case of a more general set of measures of location called percentiles.
3. A **percentile** is a score at or below which a specific percentage of the distribution falls.
4. The *n*th percentile is a score below which *n*% of the distribution falls.
5. Like the median, percentiles require that data be ordinal or higher in level of measurement.
6. Percentiles are easy to identify when the data are arranged in frequency distributions.
7. Another widely used measure of location is the quartile.
8. The lower quartile is equal to the 25th percentile and the upper quartile is equal to the 75th percentile.
9. The Mean
	1. The arithmetic **mean** is by far the best known and most widely used measure of central tendency.
	2. The mean is typically used to describe central tendency in interval-ratio variables.
	3. To calculate the mean, simply add up all the scores and divide by the total number of scores.
	4. We can define the mean by the following formula:  .
	5. Consider these new symbols:
		1. *Y* represents the raw scores in the distribution of the variable of interest.
		2. is pronounced “Y-bar” and is the mean of the variable of interest.
		3. Greek letter Σ is pronounced “sigma,” is a summation sign (just like the Σ sign) and directs us to sum whatever comes after it. Therefore, Σ*Y* means “add up all the raw *Y* scores.”
		4. The letter *N* represents the number of cases (or observations) in the distribution.
	6. Understanding Some important Properties of the Arithmetic Mean
		1. Interval Ratio Measurement
			1. Because it requires the mathematical operations of addition and division, the mean can be calculated only for variables measured at the interval-ratio level.
			2. This is the only level of measurement that provides numbers that can be added and divided.
		2. Center of Gravity
10. Because the mean incorporates all the scores in the distribution, we can think of it as the center of gravity of the distribution.
11. The mean is the point that perfectly balances all the scores in the distribution.
12. If we subtract the mean from each score and add up all the differences, the sum will always be zero.
	* 1. Sensitivity to Extremes
13. Unlike with the mode or the median, every score enters into the calculation of the mean.
14. This property makes the mean sensitive to extreme scores in the distribution.
15. The mean is pulled in the direction of either very high or very low values.
16. Because of the sensitivity of the mean, it is not suitable as a measure of central tendency in distributions that have a few very extreme values on one side of the distribution.
17. Reading the Research Literature: The Case of Reporting Income
	1. Gender income inequality has not been achieved and can be reviewed in published income data.
	2. Median earnings are routinely reported in scholarly research and government publications, such as those by the U.S. Department of Labor and Census Bureau.
	3. In 2017, women’s median earnings were 82% of men’s.
18. Statistics in Practice: The Shape of the Distribution
	1. The distribution of interval-ratio variables can also be described by their general shape.
	2. Using histograms, produced by SPSS, we will demonstrate how a distribution can be either symmetrical or skewed, depending on whether there are a few extreme values at one end of the distribution.
	3. **The Symmetrical Distribution**
		1. A distribution is symmetrical if the frequencies at the right and left tails of the distribution are identical, so that if it is divided into two halves, each will be the mirror image of the other.
		2. In a unimodal, symmetrical distribution, the mean, median, and mode are identical.
	4. The Positively Skewed Distribution
		1. As a general rule, for **skewed distributions,** the mean, median, and mode do not coincide.
		2. The mean, which is always pulled in the direction of extreme scores, falls closest to the tail of the distribution where a small number of extreme scores are located. This is represented as **positively skewed distribution**.
	5. The Negatively Skewed Distribution
		1. A distribution with a few extremely low values is a **negatively skewed distribution.**
		2. Here the mode has the highest value, the median has the second highest value, and the mean has the lowest value.
	6. Guidelines for Identifying the Shape of a Distribution
		1. In unimodal distributions, when the mode, median, and mean coincide or are almost identical, the distribution is symmetrical.
		2. When the mean is higher than the median (or is positioned to the right of the median), the distribution is positively skewed.
		3. When the mean is lower than the median (or is positioned to the left of the median), the distribution is negatively skewed.
19. Considerations for Choosing a Measure of Central Tendency
	1. In general, one tends to use only one of the three measures of central tendency, and the choice of the appropriate one involves a number of considerations.
	2. These considerations and how they affect our choice of the appropriate measure are presented in the form of a decision tree.
	3. Level of Measurement
		1. The variable’s level of measurement is the primary consideration in choosing a measure of central tendency.
		2. However, with ordinal data, we have two choices: (1) the mode or (2) the median (or sometimes both).
		3. When the data are measured on an interval-ratio level, the choice between the appropriate measures is a bit more complex and is restricted by the shape of the distribution.
	4. Skewed Distribution
		1. When the distribution is skewed, the mean may give misleading information on the central tendency because its value is affected by extreme scores in the distribution.
		2. The median or the mode can be chosen as the preferred measure of central tendency because neither is influenced by extreme scores.
	5. Symmetrical Distribution
		1. When the distribution we want to analyze is symmetrical, we can use any of the three averages.
		2. Our choice depends on the research objective and what we want to know about the distribution.
		3. The mean is our best choice because it contains the greatest amount of information and is easier to use in more advanced statistical analyses.