NOTES--Types of Surveys

Surveys can gather information through different methods of observation. However, most surveys employ a questionnaire to measure specific characteristics of the population. There are two main ways to gather this information.

A **census survey** collects complete information from all participants in the population. General criteria of a census survey include:

A **sample survey** uses a representative group of a given population to determine characteristics of the entire population. Sample surveys are used when it is not possible or practical to conduct a census to count each individual of an entire population. Everyday examples of sample surveys include political polling, health statistics, and television ratings. Sample surveys are a proven, effective method for gathering accurate information if they are properly designed and the sample design is accounted for in the estimation methods.

General requirements of a sample survey include:

- The sample must be representative of the entire population.
- The data collected from the representative sample is expanded to produce an estimate of the total population.
- Assumptions should be tested to identify and measure potential bias if possible.

Example Let's say that the school is deciding to get some new tables and chairs, and they want to make sure that they are made so that the "average height" student will find them comfortable. They've asked you to figure out the average height of students in the school. Doing a census would mean getting the height of every single person in your school, and then averaging it. Doing a sample would mean getting the height of a smaller number of people (e.g. 50 people), and using their average height as an estimate of the average height of the people in the school. If you had to choose, which would you prefer to do, a census or a sample? Why?

Smaller # of people
easier calculations

Problems and Bias with Surveys

--you happen to choose a bad group of people to use as your "average"

- *Undercoverage bias* occurs when some members of the population are inadequately represented in the sample
- *Nonresponse bias* occurs when individuals chosen for the sample are unwilling or unable to participate in the survey
- *Voluntary response bias* occurs when sample members are self-selected volunteers

--Measurement biases, which include issues related to how the outcome of interest was measured

--Intervention (exposure) biases, which involve differences in how the treatment or intervention was carried out, or how subjects were exposed to the factor of interest

Example See if you can figure out what is wrong with using these groups of people to find the average height of all students in your school:

1. Measure the heights of all the people in 9th grade
   still growing
2. Measure the heights of all the boys
   generally taller than girls
3. Measure the heights of the first 50 people who walk out of the front door once school finishes
   carpool - probably 9th & 10th grade nonresponsebias
4. Measure the heights of the members of the school basketball team
   tallest students - outliers

All of these samples would be easier to do than a census, but would probably not get the right answer for the average height of students in the whole school. These are called *biased samples*, and need to be avoided at all costs when trying to figure out a statistic. There is no point in doing something quickly if you don't get the right answer! Biased samples are a serious problem, and can appear in a surprisingly large number of scenarios.

The History of the U.S. Census

The word *census* is Latin in origin and means to tax. The first U.S. census took place over 200 years ago, but the United States is certainly not the first country to implement a census. Based on archaeological records, it appears that the ancient Egyptians conducted a census as early as 3000 B.C.E.

The U.S. census is mandated by the U.S. Constitution in Article I, Section 2, which states, in part, "Representatives and direct Taxes shall be apportioned among the several States ... according to their respective Numbers .... The Number of Representatives shall not exceed one for every thirty thousand, but each State shall have at Least one Representative ...." The Constitution then specifies how to calculate the number of people in each state and how often the census should take place.

The U.S. census has been conducted every ten years since 1790, but as time has passed, our census has evolved. Not o have the types of questions changed but also the manner in which the data are collected and tabulated. Originally, the census had only a few questions, the purpose of which was to discern the number of people in each household and the ages. Presumably, these data were used to determine the number of men in each state who were available to go to wa Federal marshals were charged with the task of conducting this first census. After collecting data from their respective jurisdictions, the marshals sent the data to President Washington.

As time has passed, more questions have been added to the U.S. census. Today, the census includes questions designed to collect data in various fields such as manufacturing, commerce, and transportation, to name a few. Data that were once manually tabulated are now processed by computers. Home visits by census officials were once the norm, but now the census is conducted primarily through the U.S. Postal Service. Each household in the United States receives in the mail a copy of the census questionnaire to be completed by its head of household who then mails it back to the Census Bureau. Home visits are paid only to those individuals who do not return the questionnaire by the specified deadline.

The census is an important part of our Constitution. Today, the census not only tells us the population of each state, thereby determining the number of representatives that each state will have in the House of Representatives, but it also provides the U.S. government with very useful data that paint a picture of the current state of our population and how it has changed over the decades.

NOTES--Measures of Central Tendency

measures of central tendency-- an average or typical value - generally located near the center of the distribution

the 4 types of central tendency we will study:

1. mean
2. median
3. mode
4. midrange

THE MEAN

The **mean** is the sum of the data items divided by the number of items.

$$\text{Mean} = \frac{\Sigma x}{n}$$

where Σx represents the sum of all the data items and n represents the number of items.

CALCULATING THE MEAN FOR A FREQUENCY DISTRIBUTION

$$\text{Mean} = \frac{\Sigma xf}{n}$$

where

x represents each data value.

f represents the frequency of that data value.

Σxf represents the sum of all the products obtained by multiplying each data value by its frequency.

n represents the *total frequency* of the distribution.

THE MEDIAN

To find the **median** of a group of data items,

1. Arrange the data items in order, from smallest to largest.
2. If the number of data items is odd, the median is the data item in the middle of the list.
3. If the number of data items is even, the median is the mean of the two middle data items.

POSITION OF THE MEDIAN

If n data items are arranged in order, from smallest to largest, the median is the value in the

$$\frac{n+1}{2}$$

position.

10

THE MODE

The **mode** is the data value that occurs most often in a data set. If no data items are repeated, then the data set has no mode. If more than one data value has the highest frequency, then each of these data values is a mode.

THE MIDRANGE

The **midrange** is found by adding the lowest and highest data values and dividing the sum by 2.

$$\text{Midrange} = \frac{\text{lowest data value} + \text{highest data value}}{2}$$

Example 1 Suppose your six exam grades in a course are

52, 69, 75, 86, 86, and 92.

Compute your final course grade (90–100 = A, 80–89 = B, 70–79 = C, 60–69 = D, below 60 = F) using the

a. mean. b. median. c. mode. d. midrange.

a. $\frac{460}{6} = 76.7$ C

c. 86 B

b. $\frac{75+86}{2} = 80.5$ B

d. $\frac{52+92}{2} = 72$ C

Example 2 *Consumer Reports* magazine gave the following data for the number of calories in a meat hot dog for each of 17 brands:

173, 191, 182, 190, 172, 147, 146, 138, 175, 136, 179, 153, 107, 195, 135, 140, 138.

107 135 136 138 138 140 146 147 153

Find the mean, median, mode, and midrange for the number of calories in a meat hot dog for the 17 brands. If necessary, round answers to the nearest tenth of a calorie.

a. mean $= \frac{2697}{17} = 158.6$

c. mode = 138

b. median = 153

d. midrange $= \frac{107+195}{2} = 151$

Example 3 Find each measure of central tendency for the frequency distribution.

Score x	Frequency f
1	2
2	4
3	3
4	1

a. Find the mean. $\frac{23}{10} = 2.3$

b. Find the median. $\frac{2+2}{2} = 2$

c. Find the mode. 2

d. Find the midrange. $\frac{1+4}{2} = 2.5$

1, 1, 2, 2, 2, 2, 3, 3, 3, 4

Example 4 Find each measure of central tendency for the data below.

Ten Highest-Paid Baseball Players

Player	Team	Length of Contract	Average Annual Salary in Millions of Dollars
Alex Rodriguez	Texas Rangers	2001–10	$25.2
Manny Ramirez	Boston Red Sox	2001–08	$20.0
Derek Jeter	New York Yankees	2001–10	$18.9
Barry Bonds	San Francisco Giants	2002–06	$18.0
Sammy Sosa	Chicago Cubs	2002–05	$18.0
Jason Giambi	New York Yankees	2002–08	$17.1
Jeff Bagwell	Houston Astros	2002–06	$17.0
Carlos Delgado	Toronto Blue Jays	2001–04	$17.0
Todd Helton	Colorado Rockies	2003–11	$15.7
Roger Clemens	New York Yankees	2001–02	$15.5

Source: Sports Illustrated

a. *mean.*
b. *median.*
c. *mode (or state that there is no mode)*
d. *midrange.*

a. mean $= \frac{182.4}{10} = \$18.24$ (millions)

b. $\frac{17.1+18.0}{2} = \$17.55$ (millions)

c. \$17 mill & \$18 mill

d. $\frac{25.2+15.5}{2} = \$20.35$ mill