AP STATISTICS **Name: _____ Block: _____**
Review: Collecting Data and Designing Studies

This is a quick review of key 1st semester ideas, before we get too deep into the 2nd ☺

- In "the real world", whenever we want to know something about a population, we take a **sample** (hopefully a **RANDOM** sample) from our population of interest.
- WHY take a sample instead of a census? Because we have a limited amount of time in our lives, and taking a census of a population can take forever.

The PARAMETER OF INTEREST *("what" we want to know about the population)* can be one of two things:

- a **PROPORTION** (usually for **categorical** data. For instance: "What is the **proportion** of people in the survey that have blonde hair?")
- a **MEAN** (or an average, for **numerical** data. For instance: "What is the **mean** height, in inches, of the people in our survey?")

I. MEANS VS PROPORTIONS

For each of the following scenarios, determine if we are dealing with **PROPORTIONS** or **MEANS**. Then answer the questions that follow *(you may want to dig up your **Unit 1/Chapter 12 notes** on statistics vs parameters)*.

1. It is thought that about 21% of the U.S. population plays video games on a daily basis. In a recent random sample of 450 U.S. teenagers, 257 responded that they play video games every day, for a percentage of 57.1%.

    a. **MEANS** or **PROPORTIONS** ?
    b. "21%" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____
    c. "57.1%" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____

2. Among all U.S. adults who own a "Playstation 3" video game system, it is thought that the mean number of "Playstation 3" video games owned is about 9.2. In a convenience sample of 34 responses collected from an internet discussion forum, the average number of "Playstation 3" games owned was 18.4.

    a. **MEANS** or **PROPORTIONS** ?
    b. "9.2 games" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____
    c. "18.4 games" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____

3. The mean carbon footprint for a random sample of 120 people in Austin is 25 tons of carbon dioxide equivalent per year (CO_2 eq/year). According to The Nature Conservancy, the mean carbon footprint for all persons in the U.S. is 27 tons CO_2 eq/year. Does this provide evidence that people in Austin are more environmentally-conscious than the rest of the U.S.?

    a. **MEANS** or **PROPORTIONS** ?
    b. "25 CO_2 eq/year" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____
    c. "27 CO_2 eq/year" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____

4. A teacher believes that no more than 10% of high school students ever cheat on an exam, but a confidential survey found that 14 of 88 (or 15.9%) randomly selected students admitted having cheated at least once. Is this strong evidence that the teacher was wrong?

    a. **MEANS** or **PROPORTIONS** ?
    b. "15.9%" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____
    c. "10%" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____

5. In a simple random sample of 48 East PHS students conducted in 2013, the mean GPA for the sample was 3.27. According to the registrar at East Podunk High School, the mean GPA for all students at East PHS in 2010 was 3.09. Does this provide evidence that the mean GPA at East PHS has increased?

    a. **MEANS** or **PROPORTIONS** ?
    b. "3.27" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____
    c. "3.09" is a **PARAMETER** or **STATISTIC** ? What symbol should be used to represent it? _____

II. DESIGNING STUDIES: RANDOM SAMPLE VS. RANDOM ASSIGNMENT

When dealing with observational/experimental studies, remember these two things:

- In order to show **CAUSATION** (aka, "cause-and-effect") between the explanatory and response variables, the subjects in a study must be **RANDOMLY ASSIGNED** to the different treatment groups. Without random assignment of subjects to treatments, the study is an OBSERVATIONAL STUDY, and you CANNOT show causation.
- Additionally, if the subjects in an experimental study are **RANDOMLY SELECTED** from the population of interest, then you may **GENERALIZE THE RESULTS** of the study to the larger population of interest. HOWEVER: Obtaining a RANDOM SAMPLE of subjects for an experimental study is often difficult, sometimes even unethical, so this is NOT a required element of a good experiment.

Consider each of the following scenarios and answer the questions that follow (use separate paper if necessary).

1. A researcher wonders if meat in the diet may be a factor in high blood pressure. She selected a random sample of 80 adults in San Francisco, and compares the blood pressures of the 40 adults who choose to be vegetarians, to those of the 40 adults who choose to eat meat. The mean blood pressure of the 40 vegetarians was found to be statistically significantly lower than the mean blood pressure of the 40 meat-eaters.

   a) Based on this study, is it reasonable to conclude that being a vegetarian will cause an adult to have lower blood pressure? Explain why or why not.

   b) If your response to part (a) was "yes": Can the results of this study be generalized to all adults in San Francisco? Explain.

2. High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, the mean cholesterol reduction for the males in group B was significantly higher than for the males in group A.

   a) Would it be reasonable to conclude that taking this cholesterol drug (in addition to exercise and diet) will lead to a greater reduction in cholesterol level than exercise and diet alone? Explain.

   b) If your response to part (a) was "yes": Can the results of this study be generalized to all such male patients at this large local hospital? Explain.

3. A group of scientists conduct a study in which 50 young adult volunteers in Colorado had their cholesterol levels measured over a 6-month period of time. About half of the subjects were randomly assigned to play video games for about an hour a day, while the other half of the subjects were not allowed to play any video games at all. At the end of the 6-month study, the group that played video games every day had a significantly lower mean cholesterol level than the group that did not play video games.

   a) Does this study indicate that playing video games will directly lead to a reduction in cholesterol levels? Explain.

   b) If your response to part (a) was "yes": Can the results of this study be generalized to all young adults in Colorado?

4. Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation. The proportion of men with high blood pressure in the group who meditate is found to be statistically significantly lower than the proportion of men with high blood pressure in the group who do not meditate.

   a) Would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

   b) If your response to part (a) was "yes": Can the results of this study be generalized to all men in this large retirement community?