Student: _____ Class: _____ Date: _____

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Music is a large part of many people's lives. Because of this, it is often the subject of study. For example,

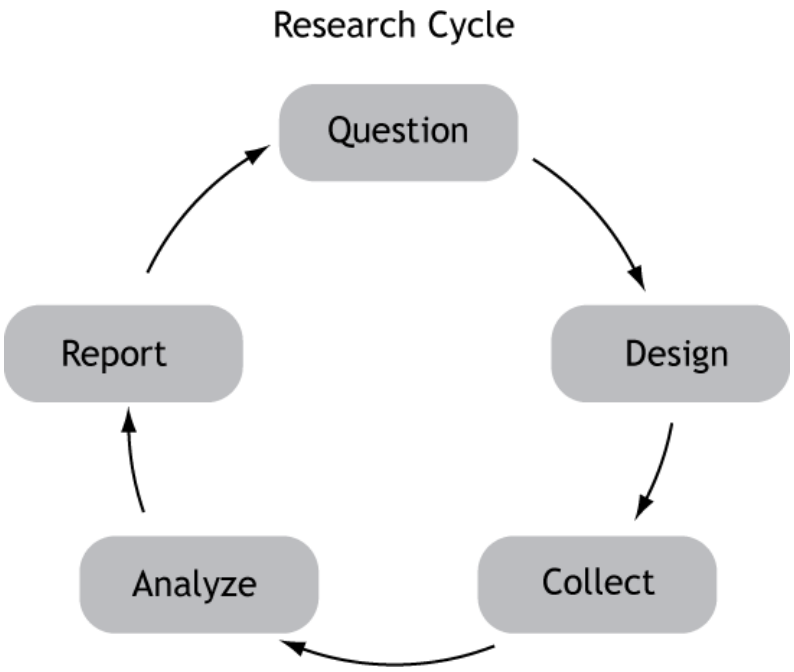
- Music industry executives want to know what will be popular with different age groups.
- Advertisers want to know which radio stations are the most popular.
- Doctors want to know how much hearing damage results from loud music.
- Teachers want to know whether or not listening to classical music helps students perform better on tests.

Statistical investigations are used every day for a variety of reasons.

1. What are the purposes of statistical investigations? Give some examples of statistical investigations with which you are familiar.

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies



This graphic illustrates the process of planning and implementing a statistical investigation. First, a question (or a series of questions) sparks the interest of a researcher. The research team then decides on the best design for investigating the question.

2. The graph shows no obvious ending point (or starting point). What does this mean?

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Consider the following examples of two different types of statistical investigations.

Example 1

Radio rating services sometimes collect data on listenership by asking participants to record the date, time, and station each time they listen to the radio. Other rating services distribute monitoring devices that automatically record this information anytime the participant has the radio turned on. Still others call participants and ask them about their listening habits. The data are then compiled so that advertisers know which stations are the most popular at specific times during the day.

Each of these approaches is an example of an **observational study**, which collects data about some characteristic(s) of the population. The data can be collected by observation, by a survey or interview, or by other means.

3. Describe an observational study in your own words:

An observational study is research in which

Example 2

*A 17-year-old student designed a science fair project with 72 mice randomly assigned to three groups: hard rock music, Mozart, and no music at all (called a **control group**). The mice in the first two groups were exposed to music 10 hours a day. Three times a week, all of the groups were timed as they ran through a maze. An analysis of results showed that the 24 mice in the no-music group averaged about a 5-minute improvement in their maze completion time, while the Mozart mice improved 8.5 minutes. The hard rock mice actually got slower—an average of four times slower! Another interesting fact: The student had to start his experiment over because all the hard-rock mice killed each other. None of the classical mice did that.* (Wertz, M. [1998]. *Why classical music is key to education.* from www.schillerinstitute.org/programs/program_symp_2_7_98_tchor_.html#Music_Mice_Mazes)

This is an example of an **experimental study**. In an experimental study, the researcher separates the participants into one or more groups and applies some sort of treatment. After treatment, the variable of interest is measured and the results are compared.

4. What are the **treatment** and the **variable of interest** in this case?

5. Describe an experimental study in your own words:

An experimental study is research in which

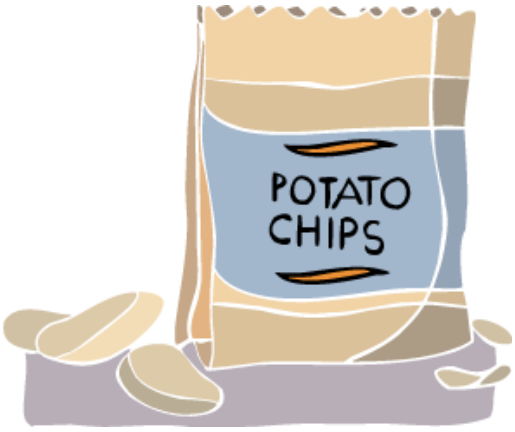
Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Observational and experimental studies have many components that must be planned, such as sampling and data collection procedures. Then the data must be collected, the results analyzed, and the conclusions reported.

6. Referring back to the Research Cycle graphic, why is there an arrow after the Report box?

Question



And what about the Question box? Consider this situation: "This unopened bag of chips is half empty. I wonder if it really contains 28.3 grams as the package says?"

This type of informal question or observation is the beginning of many investigations. Informal questions can turn into more formal problem statements or **research questions**. For example, you may decide to investigate whether there is a scandal in the potato chip industry by checking the following:

"Do Spud Potato Chips contain an average of 28.3 grams of chips per bag?"

7. REFLECTION: Now that you have been introduced to the research cycle process, think of some research questions that you are interested in studying. List at least three ideas of research questions. Consider the following:

- What type of study (experimental or observational) might be best to approach each of your research questions?
- If you only have four weeks to actually implement a research study, is it still possible to study any of your research questions?
- How can you change one of your questions to make it fit into this timeline?

8. Suppose you conduct the investigation into Spud Potato Chips and find that the mean weight of the chips in your sample is 25 grams, rather than 28.3 grams ($\bar{x}$ = 25 grams). Do you think that a difference of 3.3 grams between the actual and advertised weights is large enough that it needs to be reported? If so, how do you report this information and to whom?

In some situations, researchers are even more formal and state **hypotheses.** In a case like this, the **null hypothesis** (H_o) generally states that there is no difference between the true value and the claimed value. The **alternative hypothesis** (H_a) states that something is different or incorrect, or that something has changed.

9. What are the null and alternative hypotheses for the potato chip example?

- H_o: The true mean weight _____
- H_a: The true mean weight _____

Notice that the hypotheses say "The true mean weight." This implies that the statements refer to the **population** of all Spud Potato Chip bags, not just a single bag or even a small sample. When a statistical investigation is conducted, it generally employs a sample that is then used to make a generalization about the population. Notice that in this case (as in many cases), *population* does not refer to people, but to bags of potato chips.

To be concise, researchers often use symbols in place of words. Greek letters are usually used when referring to populations (the entire group being studied, from which a sample or samples will be drawn). English letters are used for samples (the particular items or individuals included in a particular study). For example, when discussing the mean:

- μ = the population mean (Greek letter *mu*—pronounced *mew*)
- $\bar{x}$ = the sample mean (pronounced ***x****-bar*)

So the hypotheses for a study can be stated in words or symbols. When using symbols, you must identify what your symbols represent.

- H_o: $\mu \geq 28.3$ grams, where μ is the true mean weight of a bag of Spud Potato Chips
- H_a: $\mu < 28.3$ grams

Statistical Studies: Statistical Investigations

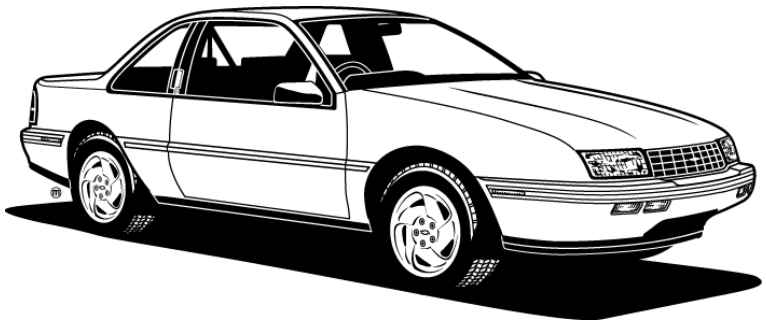
III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Statistical studies are designed with carefully selected measures that ensure (within error margins) that, if the sample is well selected and the study is well designed and conducted, the mean and other measures of the sample are likely to be similar to the corresponding measures of the population being studied. Sometimes, if the population is small (such as high school seniors in a small town), it may be possible that the sample studied is the entire population. However, often a sample is a smaller subset of a population (such as a research question that might target the entire population of high school seniors in a state or in the nation).

For Questions 11 and 12, practice writing hypotheses. Write them in words and then convert them to symbols. Finally, sketch or outline a simple study design that might help study the hypotheses.

10. A local pizza shop advertises "an average delivery time of 20 minutes or less," but it does not offer a guarantee such as a free pizza. The national manager, Su Lin, wonders if her employees are fulfilling the claim.

11. James believes that his mother's houseplants would grow taller if she watered with rainwater instead of tap water.



Not all studies lend themselves to hypothesis writing. For example, if Giancarlo is interested in cars, he might want to collect information about the most popular type of car among teenagers, the most popular color, or prices of various models. He would not design a study that tests a hypothesis. He could write and test hypotheses, however, if he wanted to know whether or not the advertising for the gas mileage of a certain model is true.

12. REFLECTION: Recall the potato chip hypotheses:

- H_o: $\mu \geq 28.3$ grams, where μ is the true mean weight of a bag of Spud Potato Chips
- H_a: $\mu < 28.3$ grams

What would you do next to determine which of these hypotheses is true?

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Design

The following cases are examples of observational studies and experimental studies. Consider the type and design of each study.

13. Identify the type of study for each case.

- How do you know?
- What is the variable of interest in each case?
- What are some advantages and disadvantages of each teacher's plan?

a. Mrs. Johnson teaches American History and wanted to help her students do their best on exams. After failing to find any research on different test formats, she decided to conduct her own research. She flipped a coin for each student in her classes. If the coin landed heads up, the student took a multiple-choice test. If the coin landed tails up, the student received a fill-in-the blank exam. Afterward, Mrs. Johnson compared the averages for the two test formats.

b. In World History, Mr. McDonald had a similar concern. He decided, however, to ask his students. He put a question at the bottom of an exam: "Which do you prefer, multiple-choice or fill-in-the-blank questions?" Afterward, Mr. McDonald tallied the results.

c. Mr. Mitchell was interested in the effects of music on student performance. At the bottom of his exam, he asked students to circle their favorite type of music: rock, country, or hip-hop. He then computed the averages for the students who liked each type of music and compared the results.

d. Mrs. Knox's senior English classes were working on their term themes. During 2nd period, she allowed students to listen to their choice of music through earphones while working, but her 4th-period class was required to work quietly without music. Mrs. Knox averaged this major grade for each class and compared 2nd period's average to 4th period's average.

e. Mr. Paul, the guitar teacher, sat at the food court in the mall and made a tally sheet that noted each t-shirt he saw with a musical group illustrated on it. He compiled the results and posted an entry to his blog about the most popular groups.

14. In the experimental studies:

- Describe the treatment(s)
- Who were the participants?
- How was the assignment of treatment(s) accomplished?

15. A group of participants that the treatment group is being compared to is called the ***control group***. Give an example of a treatment, the treatment group, and the control group.

Researchers are often concerned that participants in a study show improvement simply because they are in the study and not because they are receiving an effective treatment. This is called the *placebo effect*.

Example 1: Half of the participants in a study for a new headache remedy receive the new pill, while the remaining participants receive a pill containing only inactive ingredients. Participants receiving the inactive pill (the placebo) report that their headaches have been somewhat relieved. These participants believed they were being treated, and this belief may have affected their perception.

Example 2: Half of the athletes in a study received a new lotion for strained muscles, while the other half received a lotion with only inactive ingredients. Both groups report improvement in their muscle pain.

16. The improvement in Example 2 could be psychological, as in Example 1. Can you think of any other reason for the improvement?

17. REFLECTION: Consider Mrs. Johnson and Mr. McDonald's exam situations. Suppose Mrs. Johnson's results overwhelmingly favor fill-in-the-blank exams and Mr. McDonald's results strongly favor the multiple-choice format. Are these results in conflict with each other? What could be the cause(s) of this difference?

Suppose Mrs. Johnson takes her results to the school board and asks the board to require that all teachers in the district give fill-in-the-blank exams. If a school board member asks you what should be done, what recommendations would you give?

18. REFLECTION: Mrs. Johnson applied two different treatments to the participants in her study—some students received a multiple-choice test, and the rest received a fill-in-the-blank exam. Can you think of two treatments that could be used in a medical experiment? In a cooking experiment? Would a situation with three or more treatments be possible? Explain your thinking.

19. EXTENSION: Read the summary of actual studies on the following pages. For one of the studies, determine the following information. Be prepared to share your findings in a short presentation.

a. Is the study observational or experimental? Explain your answer.

b. Who/what are the participants or experimental units?

c. If experimental, what was the treatment(s)?

d. If there was a treatment, how was it assigned?

e. Was there a control group and/or a placebo?

f. If observational, what was being observed and why?

g. Are there any statements that appear to be opinions?

h. Are there any stated or implied limitations to the study?

Student: _____ Class: _____ Date: _____

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Tobacco and middle school and high school students

The Centers for Disease Control and Prevention conducted the National Youth Tobacco Survey (NYTS) in 2004 to measure current use of tobacco products and selected indicators related to tobacco use, including youth exposure to tobacco-related media and access to cigarettes.

The survey was distributed to 267 U.S. public and private schools; 14,034 middle school students and 13,738 high school students completed the survey. Participation was voluntary and anonymous, and school parental permission procedures were followed. Some results included:

A. 11.7% of middle school students and 28% of high school students reported current use of a tobacco product.
B. 77.9% of middle school students and 86.5% of high school students reported seeing actors using tobacco on television or in movies.
C. 70.6% of current cigarette smokers in middle school and 63.9% in high school said they were not asked to show proof of age when they purchased or attempted to purchase cigarettes from a store.

These results indicate very little change from the results of the 2002 survey. The lack of substantial decreases in the use of tobacco products among students indicates the need to

- increase the retail price of tobacco products,
- implement smoking-prevention media campaigns, and
- decrease minors' access.

The findings in this report are subject to limitations. First, these data apply only to youths who attended middle school or high school. Among 16- and 17-year-olds in the United States, approximately 5% were not enrolled in a high school program and had not completed high school in 2000. Second, the questionnaire was offered only in English. Thus, comprehension might have been limited for students with English as a second language.

(Adapted from Centers for Disease Control. Tobacco Use, Access, and Exposure to Tobacco in Media Among Middle and High School Students—United States, 2004. from www.cdc.gov/mmwr/preview/mmwrhtml/mm5412a1.htm)

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies

Scientists in Turkey undertook a study of people with epilepsy. The research included analyzing scalp hair samples from 22 participants with epilepsy and 23 participants without epilepsy, checking for differences in levels of copper, iron, zinc, magnesium, and calcium. (The researchers speculate that such differences could indicate metabolic differences that may contribute to epilepsy.) Results indicated that the epileptic group had significantly lower levels of copper and iron compared to the nonepileptic group.

(Avci, H., Kizilkan, N., & Yaman, M. [2008]. Comparison of Trace Element Concentrations in Scalp Hair of Epileptic and Normal Subjects. *Trace Elements and Electrolytes*. Vol. 25, No. 3, pp.147–155)

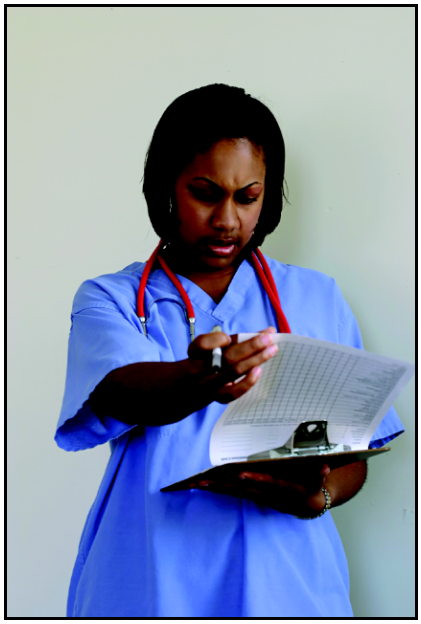
Scientists in Nigeria, in an effort to find an inexpensive method of raising rabbits for food, designed a study to test the effect of replacing some of the rabbits' soybean diet with *Gliricidia sepium* Leaf Meal (GLM). Twenty-five young rabbits were randomly assigned to receive either 0%, 5%, 10%, 15%, or 20% GLM. The groups showed no significant difference in the amount of harvestable meat, while decreasing costs of raising the meat.

(Amata, I., & Bratte, L. [2008]. The Effect of Partial Replacement of Soybean Meal… *Asian Journal of Animal and Veterinary Advances*. Vol. 3, No. 3, pp. 169–173. from www.scialert.net/qredirect.php?doi=ajava.2008.169.173&linkid=pdf)

Student: _____ Class: _____ Date: _____

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 2: Treatment of Subjects



There are two ways to categorize data according to its source—as **primary data** or **secondary data.**

Primary data are data that you collect directly. Sources of this type of data include the following:

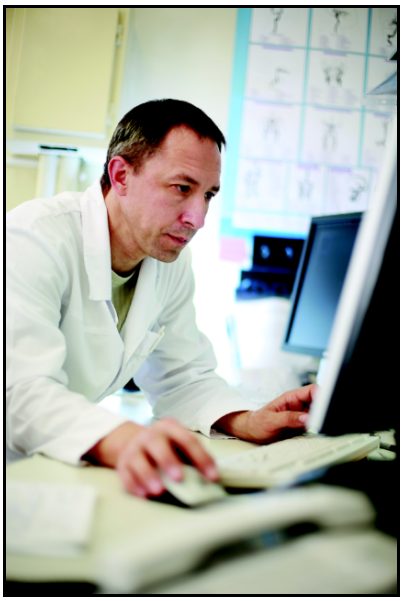
- experiments,
- survey instruments, and
- observation instruments.

When collecting primary data, most researchers run a **pilot study** first. This is a small-scale version of the research plan. What are some reasons to run a pilot study?

Secondary data are data that have been collected by someone else and are available to the researcher. The following are examples of this type of data:

- the Internet, including websites such as the U.S. Census Bureau and other government sites;
- printed materials, such as books, almanacs, newspapers, and magazines; and
- historical documents.

When using secondary data, part of the research plan must include ensuring that the data are reliable. When collecting primary data, you are responsible for doing everything possible to ensure that your participants are well informed and safe.



Student: _____ Class: _____ Date: _____

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 2: Treatment of Subjects

One nice thing about a potato chip study is that researchers do not have to worry about hurting the chips, and the chips' permission is not needed! With human participants, you have to be more careful.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was created by the 1974 enactment of the National Research Act. Its purpose is "to identify the basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects and to develop guidelines which should be followed to assure that such research is conducted in accordance with these principles."
(from Office of Human Subjects Research on the National Institutes of Health website: ohsr.od.nih.gov/guidelines/belmont.html)

The commission developed the following guidelines:

A. Ethical Principles and Guidelines for Research Involving Human Subjects
B. Boundaries Between Practice and Research
C. Basic Ethical Principles
    1. Respect for Persons
    2. Beneficence
    3. Justice
D. Applications
    1. Informed Consent
    2. Assessment of Risk and Benefits
    3. Selection of Subjects

Issues related to these guidelines include the protection of "vulnerable subjects," recruiting volunteers, payment for volunteers, and so on. To ensure that researchers follow these guidelines, research facilities have **Institutional Review Boards** (IRBs) that must approve all study design plans. If a facility does not have its own IRB, it must contract with an outside source.

The U.S. Department of Health and Human Services' Food and Drug Administration performs inspections of IRBs to ensure that they are working effectively.
(from Food and Drug Administration document at www.fda.gov/oc/ohrt/irbs/reviewboard.pdf)

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 2: Treatment of Subjects

You will explore the guidelines further through a class share activity. Discuss the following questions and record your responses.

1. Describe why special classes of human subjects need special protection. Select your choices from the following list: fetuses, women, children/minors, cognitively impaired persons, prisoners, traumatized or comatose patients, terminally ill patients, elderly/aged persons, minorities, students, or employees.

2. Could excluding classes of people from research studies be dangerous?

3. How could a survey or questionnaire be dangerous to participants?

4. Why is *informed consent* of human subjects important to a researcher?

5. What factors should a researcher consider in deciding whether to pay volunteers to participate in a study?

6. **REFLECTION:** How well do you think the 1974 National Research Act protects the rights and privacy of individuals? In what ways does it support or interfere with the goals of researchers in answering important questions?

7. **EXTENSION:** Find out about the Institutional Review Board at a research university in your area or online. Do the IRB's guidelines reflect the provisions of law? Do they go further? Interview a researcher to get his or her perspective on the IRB's provisions.

Student: _____ Class: _____ Date: _____

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 3: Margin of Error

Recall the study from Student Activity Sheet 1 (Question 19) that analyzed scalp hair samples from 22 participants with epilepsy and 23 participants without epilepsy, checking for differences in levels of copper, iron, zinc, magnesium, and calcium. The results are summarized in the table below.

	Copper	Iron	Zinc	Magnesium	Calcium
Males, epileptic	14	6	211	329	947
Males, nonepileptic	19	9	200	259	960
Females, epileptic	10	7	218	444	1,143
Females, nonepileptic	16	15	218	505	1,162

Average trace element concentrations (μg/g) in scalp hair

1. If it were possible to measure the presence of copper in the hair of *all* males with epilepsy in the world, do you think the average would be exactly 14 μg/g? Explain your thinking.

2. The journal article that contains the results of the study actually reports that males with epilepsy have an average of 14 ± 9 μg/g of copper in their scalp hair. What do you think *14 ± 9* means in this situation?

3. The $\pm$ *9* is called the *margin of error*. This wording, however, does not mean that someone messed up the research. It simply means that no sampling method can guarantee that the sample exactly matches the population, but that the sampling techniques (when used correctly) can be trusted to give results that are accurate within a certain range.

Since the males with epilepsy in the sample showed an average of 14 μg/g of copper in their scalp hair, the researchers are fairly confident that the true average copper concentration for all males with epilepsy in the study is between 5 and 23 μg/g.

Have you ever seen a news report that mentions *margin of error*? What was the report about?

Politician Paul and Candidate Carl are running for governor, and the election is next week. The latest poll shows that Politician Paul has 46% of the vote, while Candidate Carl has 43% of the vote. The news report, however, states that this poll contains a 3% margin of error.

4. What does this mean for Politician Paul?

5. What does this mean for Candidate Carl?

6. What do these poll results tell you about the upcoming election?

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 3: Margin of Error

Recall the study from Student Activity Sheet 1 (Question 19) that tested the effect of replacing rabbits' soybean diet with *Gliricidia sepium* Leaf Meal (GLM). The rabbits were randomly assigned to receive either 0%, 5%, 10%, 15%, or 20% GLM. The resulting effect on weight gain is summarized in the table below.

	0% GLM	5% GLM	10% GLM	15% GLM	20% GLM
Weight gain	958 ± 40	887 ± 59	992 ± 24	972 ± 33	919 ± 44

Increase in weight during an eight-week period, measured in grams

7. Fill in the interval of the true mean weight gain for each treatment. Based on these results alone, what do you recommend to the farmers in the area? Why?

GLM	True Mean Weight Gain
0%	
5%	
10%	
15%	
20%	

8. **REFLECTION:** Remember that increasing the concentration of GLM in the rabbit food decreases the cost of the food. Does this change your recommendation? Why or why not?

9. **EXTENSION:** Review an article about not trusting pollsters' interpretations and the need to see the actual questions. Write a short summary of the article. The following is one such article:

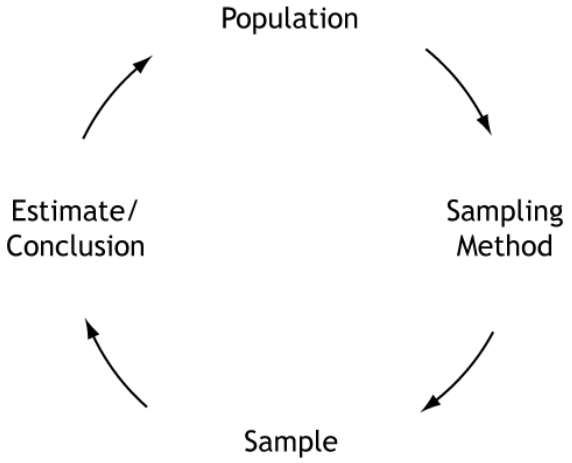
   Leo, J. (1999, October 10). Polls and poll-emics. *U.S. News and World Report*. from www.usnews.com/usnews/opinion/articles/991018/archive_002116.htm.

Student: _____ Class: _____ Date: _____

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 4: Sampling Design and Methods

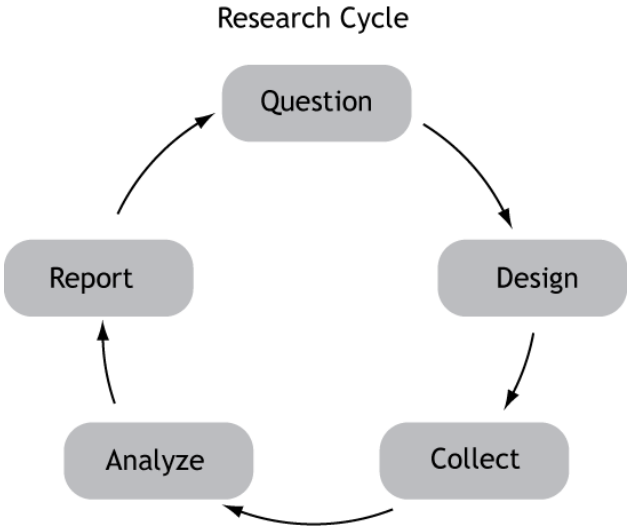
Recall the Spud Potato Chips scenario from Student Activity Sheet 1. You hypothesized that the true mean weight of bags of Spud's might be less than the 28.3 grams advertised on the bags. Discuss and make some notes on how you might collect a sample of bags to test your hypothesis. Remember that the sample should be representative of the population. What do you mean by "the population of Spud Potato Chips" that you are interested in testing?



This leads to another circular relationship. You want to know something about a population. You choose a sampling method and obtain a sample. Collecting data from the sample provides some sort of estimate or conclusion about the population of interest.

It may seem that the hard work in statistical investigations comes during data collection. Notice, however, that the other four boxes in the research cycle are the same size as the Collect box—they require the same time and attention as the actual data collection.

In particular, researchers cannot skimp on the time and care given to planning sample selection. If the sample is not representative of your population, the results may be worthless!



Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 4: Sampling Design and Methods

Recall two of the teachers from Student Activity Sheet 1 (Question 13) who were interested in the best test format for student achievement. Mrs. Johnson flipped a coin to decide whether a student would take a multiple-choice or fill-in-the blank exam. Mr. McDonald asked students which format they preferred.

1. These teachers chose to study the population of *all* students in their classes. This approach is called a *census*. The U.S. government conducts a census every 10 years. List some things you know about the U.S. Census.

2. Rather than go through the time and expense of a census, researchers usually choose to *sample* the population. There are a variety of sampling techniques. Define the following common techniques with help from your teacher or by researching other resources.

a. Simple random sampling

b. Stratified random sampling

c. Systematic sampling

d. Cluster sampling

e. Convenience sampling

3. Suppose the school board wished to see whether the age of the student affects test achievement. The testing coordinator separated the roster of high school students into freshmen, sophomores, juniors, and seniors and randomly selected 20 students from each classification. She then flipped a coin to determine which test format each student would receive and then compared the results as follows:

- freshmen who took the multiple-choice test compared to freshmen who took the fill-in-the blank exam,
- multiple-choice sophomores compared to fill-in-the-blank sophomores,
- multiple-choice juniors compared to fill-in-the-blank juniors, and
- multiple-choice seniors compared to fill-in-the-blank seniors.

Randomization still occurred because the testing coordinator flipped the coin to assign the test format to each student. This approach is an example of what type of sampling technique? Explain your thinking.

4. Mr. McDonald expanded his study to the entire school. He collected a student roster from the office and used the random number generator on his calculator to select one of the first 50 students on the list. Mr. McDonald then selected every 50th student on the list after this initial student for his sample. The calculator generated the number 32. Which students on the roster are the first five in his sample? What type of sampling technique is this? Explain your thinking.

5. Coach Smith wants to know whether students would pay for the privilege of parking their cars in the lot closest to the school. He surveyed students getting on the buses while he monitored bus loading each afternoon. What type of sampling technique is this? What do you think of his plan?

6. A large university wants to find out whether it is adequately serving the needs of its students who live off campus. The campus is surrounded by a large number of apartment complexes. The researchers randomly selected three of the complexes that seemed to contain a diverse group of residents who adequately reflect the student body as a whole, and they surveyed these residents about campus services. What type of sampling technique is this? Why do you think the university chose this method? Explain your thinking.

7. Recall the research from Student Activity Sheet 1 (Question 19) in which scientists analyzed the scalp hair samples from 22 participants with epilepsy and 23 participants without epilepsy, checking for differences in levels of copper, iron, zinc, magnesium, and calcium. The scientists were concerned about previous research that showed conflicting results. They speculated that other differences in the study group, besides the presence or absence of epilepsy, could have caused these mixed results. The scientists attempted to control some of these other differences by gathering all participants from the same region of Turkey (indicating similar dietary habits) and separating the participants into the following groups:

- males with epilepsy,
- males without epilepsy,
- females with epilepsy, and
- females without epilepsy.

This is an example of what kind of sampling? Explain your thinking.

8. Recall the study from Student Activity Sheet 1 (Question 19) that tested the effect of replacing rabbits' soybean diet with *Gliricidia sepium* Leaf Meal (GLM). Twenty-five young rabbits were randomly assigned to receive either 0%, 5%, 10%, 15%, or 20% GLM. Suppose, rather than random assignment, the scientists chose the following method:

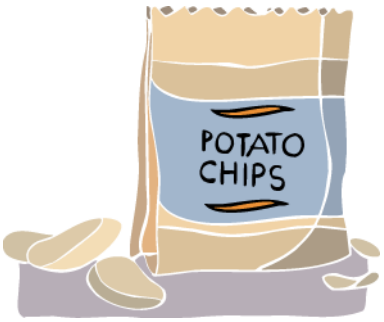
> The research assistant who was in charge of gathering rabbits went in the barn and assigned the first five rabbits he could catch to the 0% group. He assigned the next five that he caught to the 5% group, and so on.

What type of sampling technique is this? What do you think of his plan?

9. **REFLECTION:** During busy political seasons, many opinion polls are conducted. In a presidential race, how do you think the participants in polls are generally selected? Discuss any issues regarding simple random, stratified, systematic, cluster, and convenience sampling in these polls. What about other types of polls, besides political?

Statistical Studies: Statistical Investigations

III.A Student Activity Sheet 4: Sampling Design and Methods



10. Describe briefly how each technique could be used in the potato chip investigation. Which techniques are the most appropriate?

- Simple random sampling

- Stratified random sampling

- Systematic sampling

- Cluster sampling

- Convenience sampling

11. The most feasible method is to ask other U.S. schools to help with cluster sampling. On a smaller scale, you could mimic this by cluster sampling from grocery stores in the area. The easiest choice (but least reliable) is convenience sampling.

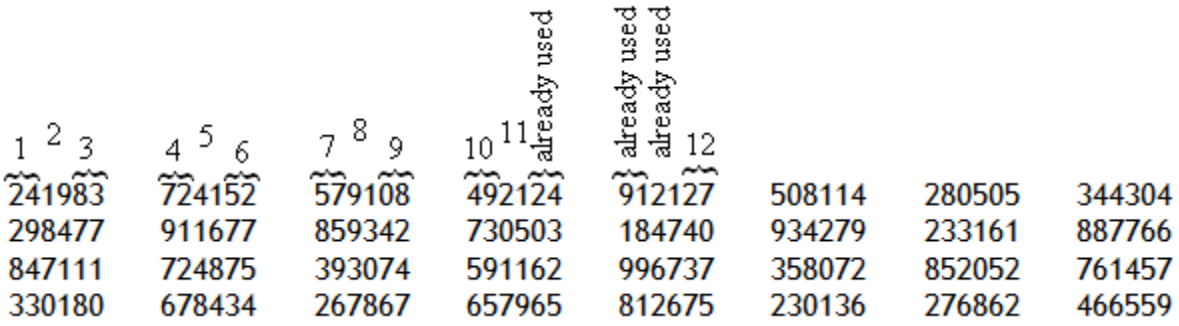
Pursue the idea of cluster sampling with other schools. To do this, you simply modify random selection techniques to choose your schools. Following are two ways to use random numbers for selection.

Method 1: Random Number Tables

Such tables can be found on the Internet by typing "random number tables" in the search box. Here is an example of what you might get:

241983	724152	579108	492124	912127	508114	280505	344304
298477	911677	859342	730503	184740	934279	233161	887766
847111	724875	393074	591162	996737	358072	852052	761457
330180	678434	267867	657965	812675	230136	276862	466559
711671	103110	259433	112082	050556	058988	273557	154354
533800	188724	706462	730447	481964	913634	627872	459415
564428	022187	445787	920698	352175	172443	463765	576652
981420	452204	121938	693647	353129	156861	696173	707475

If you have 100 schools and wish to select 12 of them to help with the study, you assign each school a number between 00 and 99. Then read the table from left to right and select the schools as their numbers appear. The 12 schools are Nos. 24, 19, 83, 72, 41, 52, 57, 91, 8, 49, 21, and 27.



a. If you have 1,000 schools on the list, number them from 000 to 999 and read the table three digits at a time. What are the first five schools in this sample?

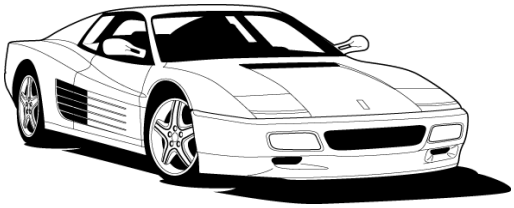
b. To avoid using the same numbers for every study, you can start from a different row. If you read three digits at a time starting with the *third row*, what five schools do you put in your sample?

Method 2: Random Number Generators

The same task can be accomplished with your graphing calculator or a random number generator found on the Internet.

c. If you use the calculator to choose a sample from the list of 1,000 schools, what are the first five schools in your sample? (***Hint:*** Remember that you need three-digit numbers.)

12. EXTENSION



Dream Car

Students' tastes vary by grade level

Do you remember what car you liked best when you were in 7th grade? What car do you like best now? Are the cars different because your tastes have matured or because new car styles have been introduced?

To investigate whether students' tastes vary by grade level, students from every grade level (7–12) must be represented fairly. You decide to use a sample size of 60. Since there are six grade levels, you could randomly select 10 students from each grade. Notice, however, that the 8th-grade class is much larger than the 12th-grade class. Should both of these grade levels have the same number of students in the sample? One way to fairly represent each grade is to use a stratified sample. Following are sample enrollment numbers.

Grade Level	No. of Students	Percentage of Total	No. of Students for Sample	No. of Students for Sample (Rounded)
7th	476	17.539	10.52	
8th	511			
9th	492			
10th	473			
11th	425			
12th	337			
Total	2,714			

a. Seventh-graders represent $\frac{476}{2{,}714}$, or 17.539% of the total. Fill in the table column showing the percentage of the total represented by each grade.

b. Since you plan to collect data from a sample size of 60, fill in the fourth column of the table indicating how many of the sample should be allocated to each grade. For example, the 7th graders get 0.17539 • 60 = 10.52 representatives in the sample.

c. In the fifth column of the table, round each number in the normal way. What sample size does this give?

d. How should you award the extra person(s)?

13. EXTENSION

Researchers must disclose their methods and any potential problems to readers of their study. Researchers can also make recommendations for future studies that improve on their methods.

Look at the article, "Are Women Really More Talkative Than Men?" and answer the following questions. There are two different studies discussed in the article, so some questions may have two answers. You may have to infer some answers from the reading.

a. What was the population of interest?

b. How were the samples obtained?

c. What were some problems that researchers noted about their study designs?

d. What were some recommendations for future study?

14. EXTENSION

Cavities and Kids

Does eating candy increase cavities?

Have you ever seen a news report on the effect that eating candy has on children's teeth? To study this subject further, what are some issues you would have to consider?

a. Describe an observational design to study this issue. Be sure to include a description of the population of interest, your sample selection technique, and what variables will be measured.

b. Describe an experimental design to study this issue. Be sure to include a description of the population of interest, your sample selection technique, and what variables will be measured.

c. What is a key difference in what can be reported for each design type?

d. Are there any other issues to consider in this study? For example, do you want to put all children in a group together? Or can you think of a way in which stratified sampling is more appropriate?

15. EXTENSION

"When I Grow Up…"

What careers are of most interest to teens?

a. What do you think a news report on teens' career plans would say?

b. If you wanted to study this subject further, what are some issues you must consider?

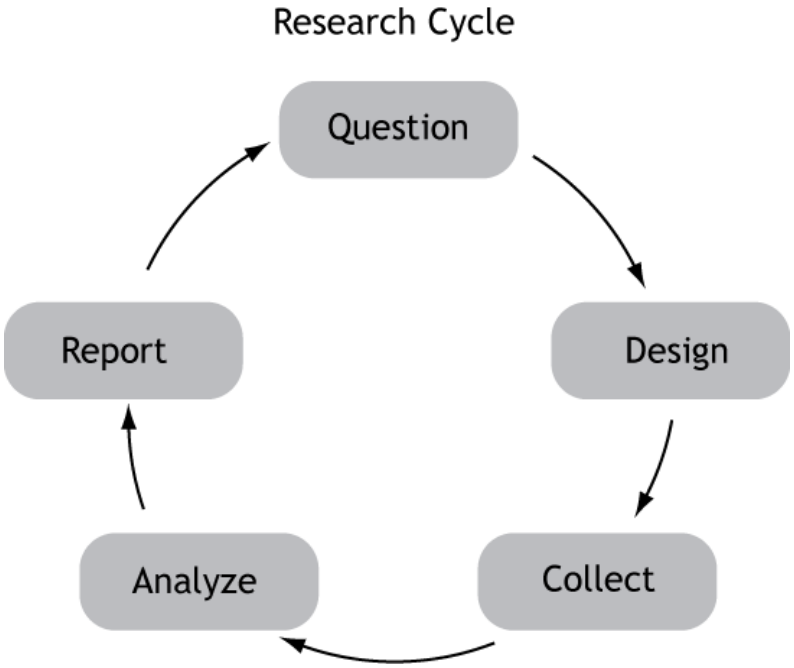
c. Write three questions that could address those issues.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 5: Histograms

1. What does the histogram in the opening of the lesson represent? Describe the distribution as completely and accurately as possible with regard to center, shape, and spread.

2. Refer to the research cycle. Where in the cycle is examining histograms located?



So, let's collect some data!

3. Without discussing with your classmates, answer the following questions—write one answer on each of your 12 slips of paper. Place each slip in the appropriate paper bag.

a. What year were you born?

b. What is your gender?

c. How many text messages did you send yesterday?

d. What does your cell plan charge for texting?

e. How many people under the age of 18 live in your house?

f. Which is your favorite food from the following choices: pizza, hamburgers, sushi, salad, chicken, other?

g. Do you have a job that pays by the hour?

h. If yes, how many hours do you work in an average week? (If no, put *0.*)

i. List how many hours you work a week (again) and then the average number of hours that you study a week.

j. List your gender (again) and your shoe size.

k. List your gender (again) and number of text messages (again).

l. What is your favorite kind of music?

4. While your classmates finish Question 3, jot down your thoughts about the following questions:

- What do you think ***categorical data*** means?

- What do you think ***quantitative data*** means?

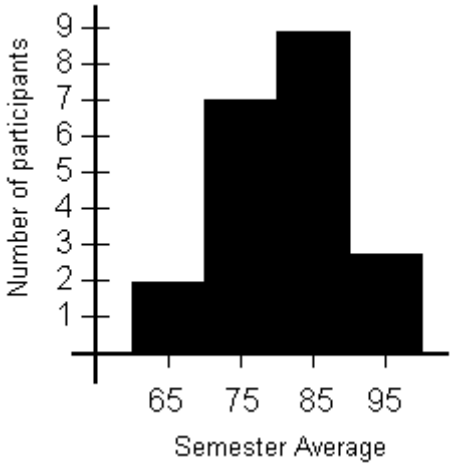
- What do you think ***univariate*** means?

- What do you think ***bivariate*** means?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 5: Histograms

Use the following histogram to answer Questions 5-7.



5. Does the histogram illustrate univariate data or bivariate data? Categorical or quantitative?

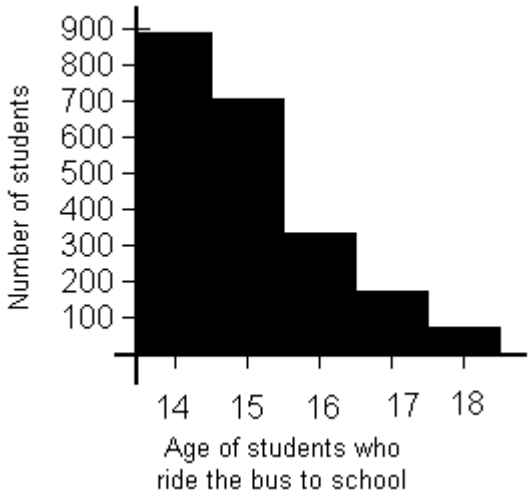
6. Would you describe the histogram as symmetric? Why or why not?

7. Describe the distribution of grades for the class. Justify any estimates you make.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 5: Histograms

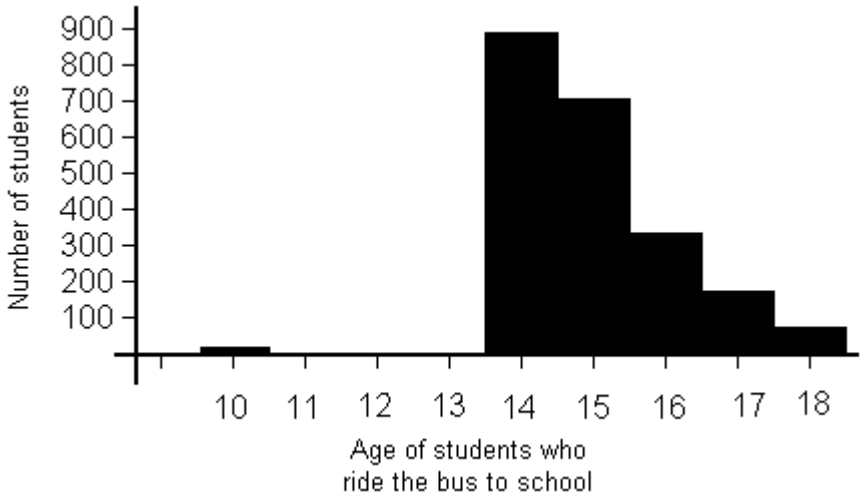
Use the following histogram to answer Questions 8 and 9.



8. Consider the new histogram showing bus ridership at a local high school. This histogram is not symmetric; it is skewed to the right. You know this because the distribution has a tail out toward the right side. What is happening with this population that causes the distribution to be skewed to the right?

9. Estimate the average age of students who ride the bus. This average is the center of the distribution. Justify your estimate.

10. Suppose the histogram of bus ridership looked like this instead. Data values that are distant from most of the other values are generally thought of as ***outliers***. What may have happened? Does this data distribution affect your answers to Questions 8 and 9?



11. **REFLECTION:** Make a histogram that represents the number of students of each age at the same school who drive to school each day (either in their own car or riding with friends). Describe the center, shape, and spread of your histogram.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 5: Histograms

Several of the previous histograms, including the one in Question 10, had a set interval width (for example, the ages of bus riders in intervals of one year). Now you will investigate the effect that interval width (also known as ***bin size***) can have on a histogram as well as on a reader of the histogram.

Go to the following website and select the "Colleges' SAT Math Scores" data set:

illuminations.nctm.org/ActivityDetail.aspx?id=78

(National Council of Teachers of Mathematics. [NCTM] Histogram Tool. In Illuminations: Resources for Teaching Math.)

12. The default interval width for this data set is 19.1. Use the slider under the histogram to change the interval width to approximately 30. What does this do to the appearance of the graph? Explain.

13. Change the interval width to approximately 50. Comment on the new graph.

14. Try an interval of approximately 75. Interpret your result.

15. Try an interval of 250. What do you think of this graph and the information it communicates?

16. Reset the interval size back to approximately 75. Change the value on the Maximum Frequency on Histogram box to 300 (click the Update Histogram button to enact the change). What do you think of this change to the graph and the information it communicates?

17. EXTENSION: Above the histogram is a drop-down menu from which you can choose another data set. Select one of the other data sets and explore a variety of interval widths for the graph. Sketch three histograms that show how changing the interval width results in widely differing perceptions for the reader of the graph. Explain each choice.

18. EXTENSION: Use the drop-down menu to choose a third data set. Once again, sketch three histograms that show how changing the interval width results in widely differing perceptions for the reader of the graph. Explain each choice.

19. EXTENSION: Use the drop-down menu to choose My Data. Go to the Describe Your Data box below the histogram and give your histogram a name. Below that is a data input box (*Enter your data below, one per line*). Enter your data and try a variety of interval widths. Choose two graphs to sketch—one that accurately communicates the data to the reader and another that distorts the data. Explain each choice.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 6: Analyzing Graphical Displays

The Phoenix Mercury of the Women's National Basketball League had 14 players on the roster for the 2008 season. The players and their average points per game (PPG) are shown below.

Player	Diana Taurasi	Cappie Poindexter	Tangela Smith	Le'coe Willingham	Kelly Miller	Kelly Mazzante	LaToya Pringle
PPG	24.1	21.2	11.1	10.1	8.3	5.8	4.4

Player	Brooke Smith	Barbara Farris	Olympia Scott	Yuko Oga	Allie Quigley	Willnett Crockett	Jennifer Derevjanik
PPG	4.1	3.5	2.7	2.4	2.1	1.8	0.8

(Source: www.wnba.com/mercury/stats)

1. The smallest value listed in a data set is called the ***minimum***. The minimum of this data set is _____. Which player has the minimum value?

2. The largest value listed in a data set is called the ***maximum***. The maximum of this data set is _____. Which player has the maximum value?

3. The middle value in a data set is called the **median.** The median of this data set is _____. Which player has the median value? (***Note:*** This problem requires some work because a middle value does not exist when there is an even number of data points. In this case, you must average together the <u>two</u> middle values.)

4. Now list the data set horizontally from smallest to largest, and write the median in the list in the appropriate location. Circle the minimum, median, and maximum**.**

5. Cover up the right side of the list in Question 4 so that you can only see the seven values below your median. Find the median of these seven numbers and circle it; under that number write *Q1*. Repeat this process with the other half of the data by covering up the left side so that you can only see the seven highest values. Find the median of these seven numbers and circle it; under that number write *Q3*.

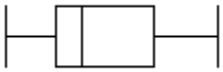
The numbers you have circled are called the ***five-number summary***. These numbers separate your data into four quartiles, or 25% sections.

- The data between the minimum and Q1 are the first quartile.
- The data between Q1 and the median are the second quartile.
- The data between the median and Q3 are the third quartile.
- The data between Q3 and the maximum are the fourth quartile.

6. The five-number summary allows you to make a graphical display called a ***boxplot*****,** or a ***box-and-whisker plot***. The reason for this interesting name becomes obvious as you construct the graph. First you need to decide on a scale. What would be a good scale for these data—to count by 1s, 10s, 100s, or something in between?

7. Construct a box-and-whisker plot. The following steps are provided for your reference.

- Plot your scale on the line below.
- Place an appropriate label below the line.
- Place dots for your five-number summary values about an inch above the line.
- Put a small vertical line, about the size of this *l*, on each dot.
- Use these lines to construct a box-and-whiskers like this one:



8. Interpret the "box" part of your box-and-whisker plot:

50% of the Phoenix Mercury players _____.

9. You can also create boxplots on your graphing calculator. Sketch your new graph—compare and contrast it to your hand-drawn boxplot.

There may also be an option on your calculator for creating a modified box plot, which reveals any outliers. If so, sketch this graph. What do you think this graph is showing that is different from your previous one? (If you have access to this information using technology, research information about this plot that distinguishes it from a box-and-whisker plot.)

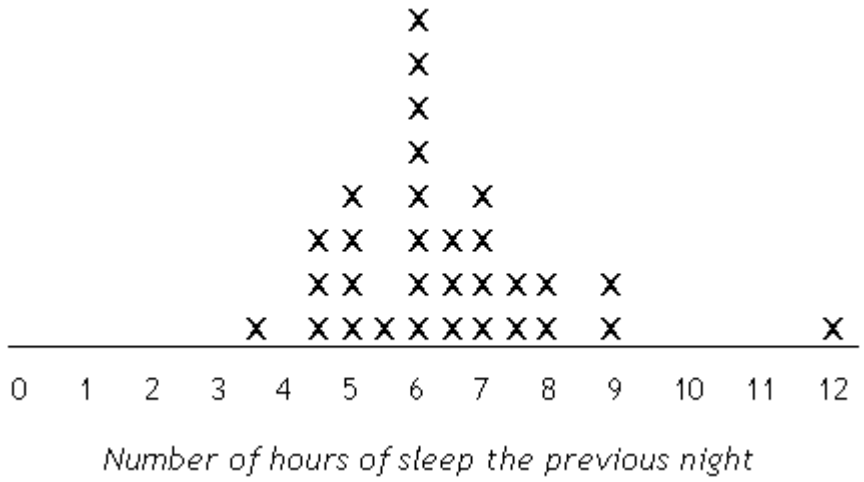
10. **EXTENSION:** After you finish the hand-constructed and calculator graphs, look up statistics for another WNBA team (or NBA team), compute the five-number summary, and add it to the same graph from earlier teams. Thus, you have created a side-by-side boxplot. Compare and contrast the two boxplots.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 6: Analyzing Graphical Displays

Take a look at other graphical displays. Consider the characteristics of the center, shape, spread, and any unusual features.

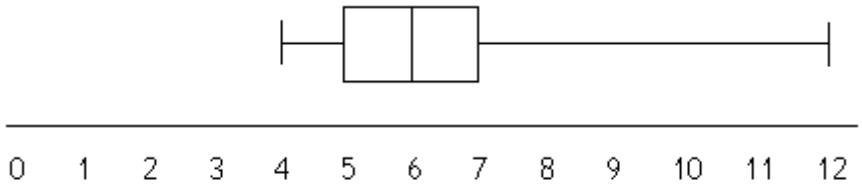
11. The school newspaper conducted a survey in which 31 randomly selected students were asked a variety of questions. The responses to one question are shown in the following dotplot (lineplot). Discuss what you now know about these students.



12. Some members of the newspaper staff wanted to report the sleep data in a frequency table as shown below. Discuss the advantages and disadvantages of this option.

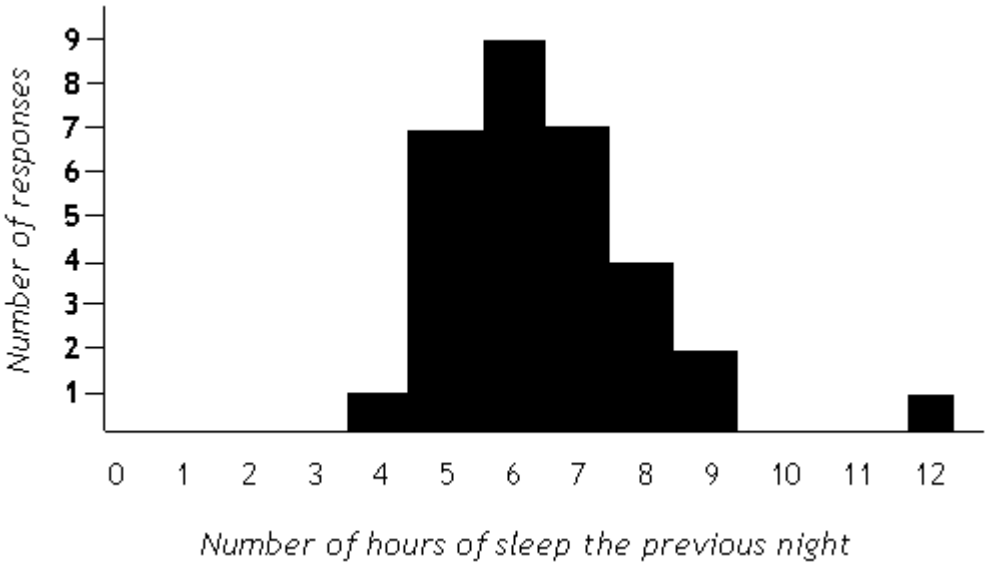
Number of hours of sleep	Frequency
4	1
5	7
6	9
7	7
8	4
9	2
12	1

13. Other staff members voted for a boxplot. Compare and contrast the usefulness of this boxplot with that of the previous two graphical displays.

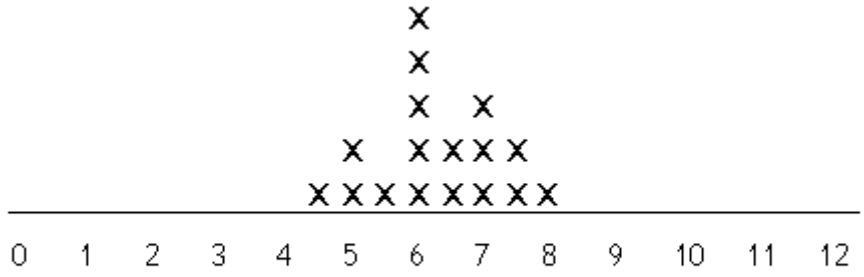


Number of hours of sleep the previous night

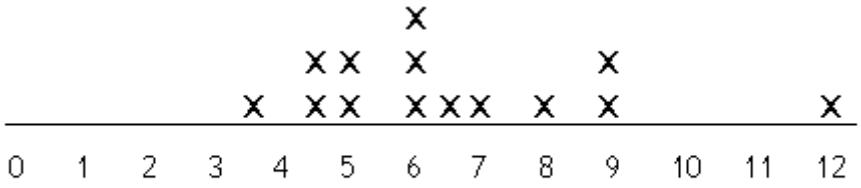
14. Another group of staffers argued for the following graph. What were their reasons for preferring a histogram? What are the arguments against using one?



15. The following dotplots show the effect of separating the data on male students' hours of sleep from the data on female students' hours of sleep. Compare and contrast the two plots.



Number of hours of sleep the previous night - females



Number of hours of sleep the previous night - males

16. Choose the frequency table, boxplot, or histogram format to report the male and female sleep data.

17. **REFLECTION:** Is it easier to compare and contrast the male and female sleep data from the dotplot or from your new display in Question 16? Explain. Refer to the information regarding limitations and differences of various graphical representations to support your thinking as needed.

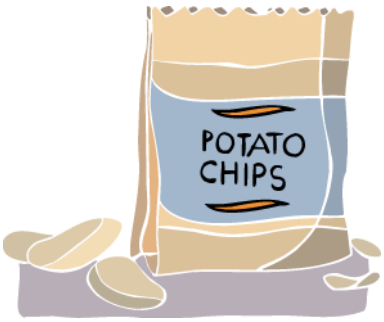
18. EXTENSION: For one of the sets of data in Question 3 in Student Activity Sheet 5, represent the data in as many different graphical displays as possible.

19. EXTENSION: Write two reports—an informative paper on student sleep patterns (or other sets of data of their choice) and a persuasive paper that utilizes the data and graph.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Suppose data were collected on 25 bags of Spud Potato Chips. The weight (to the nearest gram) of the chips in each bag is listed below.



25	28	23	26	23
25	25	24	24	27
23	24	28	27	24
26	24	25	27	26
25	26	24	25	25

1. Create a dotplot of the potato chip data and describe the distribution.

2. Does this distribution appear to support or contradict the manufacturer's claim of an average weight of 28.3 grams of chips per bag? Explain your reasoning.

3. Use your calculator or spreadsheet software to create a more formal display of these data. Make a sketch of the result here and save an electronic copy for your formal report. (If you create a histogram, remember to carefully consider your bin size.)

4. You can also use your calculator or spreadsheet software to create more precise numerical descriptions of the data. This approach is quicker and usually more accurate than computing by hand.

The mean of the data is $\bar{x}$ =

The standard deviation of the data (a measure of "spread-outness") is S_x =

The number of data values is n =

The five-number summary is

The numbers in the five-number summary represent

5. Sometimes it makes sense to analyze the proportion of a population that meets some criterion. This is another method for investigating whether the manufacturer is correct in claiming that the average weight of a bag of Spud's is 28.3 grams. Write a statement or hypotheses involving the proportion of bags of chips that meet a weight criterion.

6. What proportion of the bags in your sample were 28.3 grams or more? Does the answer change your opinion about Question 2? Why or why not?

7. Suppose instead that the weight of the chips in each Spud's bag is the following:

29	34	22	27	26
25	28	24	26	33
28	29	31	30	27
28	31	28	32	25
31	28	30	29	27

Produce a histogram and recompute the descriptive statistics for this set of data. What do you notice?

The mean of the data is $\bar{x}$ =

The standard deviation of the data is S_x =

The number of data values is n =

The five-number summary is

8. EXTENSION: Prepare a professional report for the president of Spud Potato Chips based on one of the data sets.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 1

Go to www.billboard.com and select the Hot 100. This list gives the top 100 songs for the week. You will compute statistics on the number of weeks that a random selection of the songs has been on the charts. (If you wish, choose one of the specialized charts such as R&B/Hip-Hop or Country.)

Use your calculator or a random number table to select 10 of the 100 songs. Write down the name of each song, its rank on the list, and the number of weeks it has been on the chart. For example, if your random number generator gives you *3*, write "No. 3, Mary Had a Little Lamb¸ 14 weeks."

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 2

Go to www.imdb.com and select IMDb Top 250. (To find this link, scroll down and look for User's Favorites in the sidebar on the left.) This list ranks the top 250 movies as voted by regular users of the Internet Movie Database (IMDb). (***Note:*** These are listed by ranking, not by number of voters. The voters did not vote for a favorite movie; rather, they ranked the movies on a scale of 0–10). You will compute statistics on the number of voters who provided input on the movies, without actually using 250 pieces of data in your calculation. (If you wish, choose the IMDb Bottom 100 instead.)

Use your calculator or a random number table to select 10 of the 250 movies. Write down the name of each movie, its rank on the list, and the number of voters who ranked it.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 3

For information about state sales taxes, go to the Federation of Tax Administrators' website at www.taxadmin.org/FTA/rate/sales.html. The sales tax for each state and the District of Columbia is listed. You will compute statistics for state sales taxes, without actually using 51 pieces of data in your calculation.

Use your calculator or a random number table to select 10 of the 51 locations. Write down the name of each location and its sales tax.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 4

To view gasoline sales taxes by state, go to the American Petroleum Institute website at www.api.org/statistics/fueltaxes. You will compute statistics for gasoline sales taxes by state, without actually using 51 pieces of data in your calculation.

Use your calculator or a random number table to select 10 of the 51 amounts. Write down the name of each state and its sales tax on a gallon of gas.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 5

To view diesel sales taxes by state, go to the American Petroleum Institute website at www.api.org/statistics/fueltaxes. (Scroll down to find the diesel data, which appear after the gasoline data). You will compute statistics for diesel sales taxes by state without, actually using 51 pieces of data in your calculation.

Use your calculator or a random number table to select 10 of the 51 amounts. Write down the name of each state and its sales tax on a gallon of diesel fuel.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 6

To view the tournament records for men's college basketball teams, go to webpages.charter.net/dbwoerner/coaches/schl109.htm. Because there are hundreds of schools listed, you will take a sample to compute the average winning percentage.

Use your calculator or a random number table to select 25 teams. Write down the name of each school and compute the winning percentage $\left(\frac{\text{wins}}{\text{total games played}}\right)$.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 7

Visit the website for the National Assessment of Educational Progress (NAEP) at nces.ed.gov/NATIONSREPORTCARD/states. Click on State Comparisons. Circle the subject that your teacher assigns to you:

Mathematics Reading Science Writing

Choose Grade 8, your assigned subject, Gender, and the most recent testing year. Then click on the Next Steps button; this generates a data table from the 50 states, the District of Columbia, and Department of Defense schools. Circle the data display that your teacher assigns to you:

Histogram Boxplot

Compute the appropriate statistics for each gender's data set.

Create a comparative graphical display.

Write a technical report comparing and contrasting the data that might be submitted to the NAEP website for publication.

Prepare a nontechnical version of your report to be presented to students and parents.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 8

Visit the website for the U.S. Census Bureau at www.census.gov. In the People & Households category, click Estimates, then Estimates Data, and then Totals. Under Vintage 2009, click Annual Population Estimates. You can view the Excel document with population estimates for each state. Make a list of the 10 least populous states and a list of the 10 most populous states.

Next, circle the website that your teacher assigns to you:

the American Petroleum Institute at www.api.org/statistics/fueltaxes to view gasoline sales taxes by state

the Census Bureau at www.census.gov to view household incomes; in the People & Households category, choose State Median Income

Swivel at www.swivel.com/data_sets/spreadsheet/1006019 to view the number of cars per state

Swivel at www.swivel.com/data_sets/spreadsheet/1000483 to view the crime rate per state

Go to your assigned website and collect data for the 20 states on your list. You will display these data in one of two ways. Circle the data display that your teacher assigns to you:

Histogram Boxplot

Compute the appropriate statistics for your assigned data set.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Create a comparative graphical display.

Write a technical report comparing and contrasting the data that might be submitted to the Census Bureau for publication.

Prepare a nontechnical version of your report to be presented to the local media.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Data Set 9

Visit the website for the National Collegiate Athletic Association (NCAA) at www.ncaa.org/wps/portal. Go to Statistics & Records and choose a sport. Which sport did you choose?

Click on Archived Team-by-Team Final Statistics. (***Note:*** This may be worded slightly differently, depending on the sport you choose.)

Choose two different schools. (***Note:*** Remember that schools do not play every sport, so if you do not get results, pick another school). Which two schools did you choose?

Choose a category such as Points Per Game or Assists and record the data for the players at each school.

Circle the data display that your teacher assigns to you:

Histogram Boxplot

Compute the appropriate statistics for each data set.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 7: Using Technology

Create a comparative graphical display.

Write a technical report comparing and contrasting the data that might be submitted to the NCAA for publication.

Prepare a nontechnical version of your report to be presented to local sportswriters.

Student: _____ Class: _____ Date: _____

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 8: Survey Design

"School uniforms should not be required."

Agree or disagree?

"Standardized dress can promote a productive school environment."

Agree or disagree?

These questions show that the design of a survey can influence the results. Questions should be designed to be neutral and to allow the accurate recording of the opinions or facts given by the participants. For example, the uniform question could be worded as follows:

I am in favor of school uniforms for high school students.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

This is an example of a ***closed question*** because the participants are limited to the response choices that are given. Closed questions are easier (and usually cheaper) to analyze, and sometimes a computer does much of the work. ***Open questions*** allow the participants to give more detailed responses. This approach, however, requires a follow-up analysis that is more labor intensive (and more expensive).

1. Reword the school uniform question so that it is an open question.

Because the design of a survey is so important for capturing the information needed accurately, ***pilot surveys*** are often conducted. A pilot survey is used with a small number of people and then analyzed to look for the following:

- question ambiguity,
- leading questions,
- too many people choosing *Other* or *No opinion*, and
- other data collection problems.

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 8: Survey Design

2. The following questions can be worded more effectively. Describe the question's problem and what effect the problem could have on the results. Then come up with more effective wording for the question. Provide a closed question and an open question.

a. I can always talk to my parents about my problems. True/False

b. Teachers and students like the new school schedule. True/False

c. I like the school cafeteria. True/False

d. I support school activities. True/False

Statistical Studies: Analyzing Data

III.B Student Activity Sheet 8: Survey Design

When designing a survey (or an observation instrument), the researcher must also consider how the data will be collected. The following are some options:

- **a.** A face-to-face interviewer asks questions and records the answers of the participants (on paper or an electronic recording).
- **b.** A telephone interviewer asks questions and records the answers.
- **c.** A computer calls homes and records answers through key presses.
- **d.** A researcher observes behaviors or characteristics and marks the survey.
- **e.** A researcher hand-delivers surveys to participants, who then fill out the surveys and turn them back in.
- **f.** The survey is mailed to participants.
- **g.** The survey is e-mailed to participants.

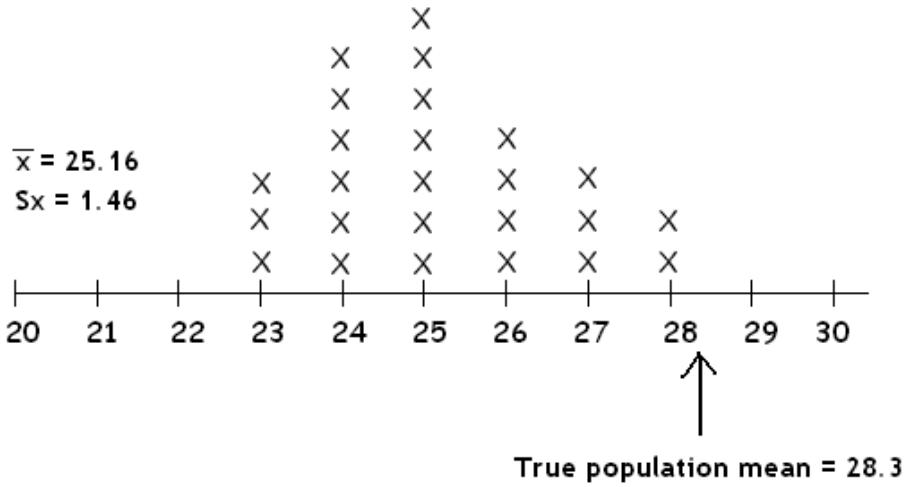
3. Discuss the pros and cons of each method with your partner(s). Record your observations.

Always bear in mind that whatever method of data collection is used, it is vitally important that the results be accurately interpreted and reported. This includes thoroughly disclosing the methods to the reader.

Statistical Studies: Sources of Variability

III.C Student Activity Sheet 9: Introduction to Statistical Bias and Variability

Recall Spud Potato Chips one last time. One sample that was "collected" had the results shown below. If the true mean of ***all*** Spud's bags is really 28.3 grams, as claimed by the company, this sample appears to have problems. The sample mean is far below the true mean and, in fact, all the bags are below the true mean.



This discrepancy between a population parameter and a sample statistic is known as ***statistical bias***, which can result from many different sources. Two broad categories of statistical bias are ***biased sampling method*** and ***biased statistic***.

1. Give an example of a biased sampling method.

2. Sometimes researchers spend vast resources (time, money, effort) to get a great sample and then still end up with a biased statistic. For example, something could be wrong with the data collection method. What could happen after selecting the sample (for example, potato chip bags) and before calculating the statistic that could result in a biased statistic?

If in fact the method of sampling has introduced bias, one reason for this might be that the sample was actually nonrepresentative of the population (***nonrepresentative sampling***).

Nonrepresentative sampling is when the sample does not represent the differences in a population. For example, the students in this class are mainly seniors in high school; therefore, they do not represent all students in the school. If only the football team or only the volleyball team were surveyed at most schools, the sample would have all males or all females.

3. What other situations could be responsible for bias? Provide examples where possible.

4. REFLECTION: Describe a scenario in which different types of sampling methods lead to different kinds of bias in sampling, including yielding nonrepresentative samples or samples with undercoverage of a population.

Statistical Studies: Sources of Variability

III.C Student Activity Sheet 9: Introduction to Statistical Bias and Variability

5. Biased statistics can result from a variety of problems during the data collection process. Record your thoughts or information you have located through research regarding each of the following sources of statistical bias.

a. Response bias

b. Nonresponse bias

c. Observer effect

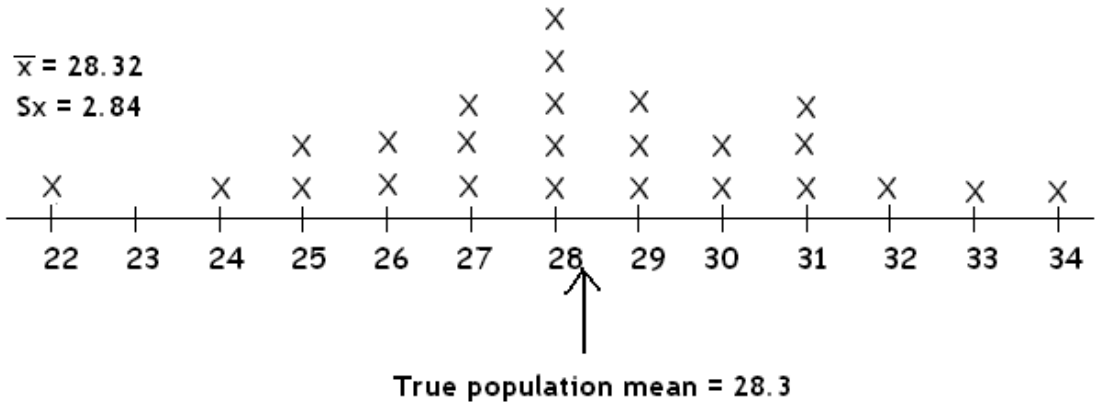
d. Wording of questions

e. Placebo effect

Statistical Studies: Sources of Variability

III.C Student Activity Sheet 9: Introduction to Statistical Bias and Variability

Recall that a second sample of Spud Potato Chips was "collected," and the following results were obtained:



6. Comment on this distribution compared to that of the original Spud's sample.

7. REFLECTION: This sample of chips is an example of ***high variability*** and no statistical bias. This situation can be a big problem for the manufacturer. People who get bags with only 22 grams of chips probably do not really care what the mean is—they just care that they were shorted! Compare the standard deviations that are given for the original distribution on Page 1 and this distribution.

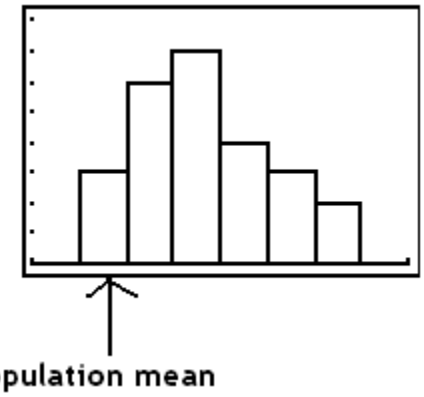
What are your recommendations for the manufacturer?

Statistical Studies: Sources of Variability

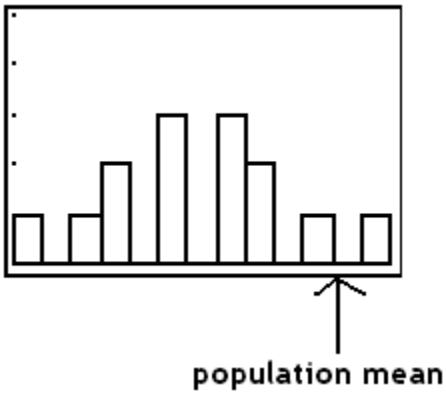
III.C Student Activity Sheet 9: Introduction to Statistical Bias and Variability

8. Identify each of the following as *high* or *low statistical bias* and *high* or *low variability*.

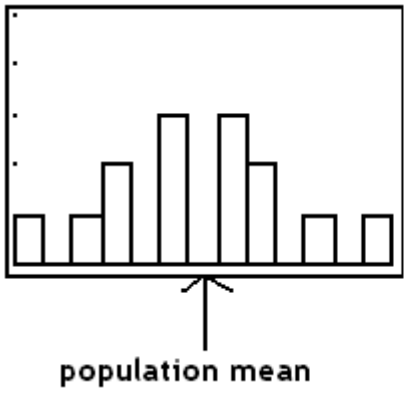
a.



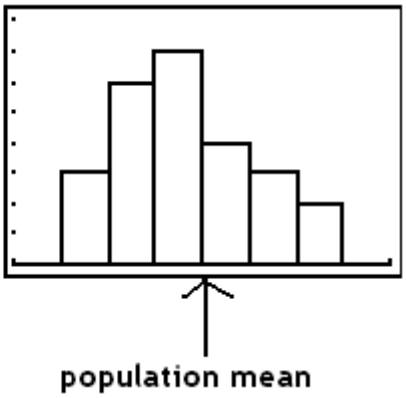
b.



c.



d.



9. REFLECTION: In your opinion, which graph in Question 8 represents the worst situation for researchers? Explain your reasoning.

10. EXTENSION: Variability can be caused by ***natural variability*** or ***induced variability***. Research each, record your findings, and provide some examples.

a. Natural variability

b. Induced variability

Student: _____ Class: _____ Date: _____

Statistical Studies: Sources of Variability

III.C Student Activity Sheet 10: Statistical Bias in Research Studies and Polls

Statistical bias in experimental studies can result from several factors. Look back at your notes about statistical bias and see which sources you think could apply to experimental studies. List your ideas below.

Experimental studies can suffer from nonrepresentative samples and undercoverage, as discussed in Student Activity Sheet 9. Often these problems cannot be fixed.

1. What are some reasons for deliberately using a nonrepresentative sample in an experimental study? What problems could result?

2. Rather than spending a great deal of time and money to ensure a representative sample, researchers often use the techniques listed below to try to eliminate any statistical bias introduced by the sampling method. Discuss how each method can reduce sampling method bias and thus increase the accuracy of a study's results.

a. Random assignment of treatments

b. Blind/double-blind studies

c. Use of control groups

d. Replication

3. Some of the same issues (nonrepresentative samples, undercoverage, replication) affect observational studies. Other issues discussed (response bias, wording of questions) can also specifically affect these studies. What problem could be occurring in each study described?

a. Jade is so tired of sales calls that she refuses to answer the phone if she does not recognize the name or number on the Caller ID.

b. Serena knows the caller will probably keep calling, so she answers the phone. If the call's subject is an opinion poll, she politely tells the surveyor that she does not want to participate.

c. DeShawn does not worry about sales calls since his family uses cell phones and no longer has a land line.

d. At the mall, an older man is interviewing people about their consumer spending habits. Maria does not want to tell him she just bought a swimsuit, so she says her shopping bag contains a new shirt.

e. Mrs. Gibbs normally gets a hamburger for lunch. When she notices that someone who looks like a nurse is sitting in the corner writing down what each person orders, Mrs. Gibbs orders a salad with fat-free dressing.

f. The local television station runs a viewer poll on the nightly news: "Call us or log on to our website to give your opinion about the new skate park."

4. Several television shows have a survey component in which viewers call in and vote for their favorite singer/dancer/entertainer/competitor. Discuss the issues you perceive with this method.

5. Some shows have viewers vote for the contestant whom they want to leave the show. Discuss the issues you perceive with this method.

6. REFLECTION: What suggestions do you have for improving call-in or online polls or contests?

7. **EXTENSION:** Discuss with your partner(s) where in the research cycle is statistical bias most likely to occur. Suggest at least two strategies for avoiding bias and minimizing variability in your project.