AP Statistics Notes – Unit Four: Statistical Studies: Observation and Experimentation

Syllabus Objectives: 2.3 – The student will distinguish between populations and samples.
2.4 – The student will distinguish between parameters and statistics.

A primary goal of statistical studies is to collect data that can then be used to make informed decisions. The data collection step is critical to obtaining reliable information; both the type of analysis that is appropriate and the conclusions that can be drawn depend on how the data are collected. In this unit, we first consider two types of statistical studies and then focus on two widely used methods of data collection: sampling and experimentation.

- Planning and conducting a study
    - Understand the nature of the problem
    - Decide what to measure and how to measure it
    - Collect the data
    - Summarize the data and perform a preliminary analysis
    - Do the formal data analysis
    - Interpret the results

- Populations and Samples
    - The entire group of individuals that we want information about is called the **population.**
        - A population characteristic or a number that describes the population is called a **parameter**. In statistical practice, the value of a parameter is unknown.

    - The part of the population that we actually examine in order to gather information is called the **sample**.
        - A sample characteristic or number that can be computed from the sample data without making use of any unknown parameters is called a **statistic**. In practice, we often use a statistic to estimate an unknown parameter.

    - **Ex:** The ball bearings in a large container have mean diameter **2.5003** centimeters (cm). This is within the specifications for acceptance of the container by the purchaser. By chance, an inspector chooses 100 bearings from the container that have a mean diameter of **2.5009** cm. Because this is outside the specified limits, the container is mistakenly rejected.
        The value 2.5003 is the parameter because It is the mean of the entire population of ball bearings. 2.5009 is the statistic because it is the mean of the sample of 100 ball bearings. Using variables: $\mu = 2.5003$ and $\bar{x} = 2.5009$

    - Sampling versus Census
        - **Sampling** involves studying a part in order to gain information about the whole.
        - **Census** attempts to contact every individual in the entire population.

Syllabus Objective: 2.1 – The student will identify methods of data collection.

- Observational Study vs. Experiment

    o An **observational study** observes individuals and measures variables of interest but does not attempt to influence responses.
      - Difficult to measure or gauge the effect of an action or procedure
      - Lurking variables are uncontrolled so the study may be confounded
      - Can use available data
      - **Sample Surveys** are one kind of observational study
        - Sample surveys are designed to ask questions of a small group of people in the hope of learning something about the population.

    o An **experiment**, on the other hand, deliberately imposes some **treatment** on individuals in order to observe their responses.
      - Experiments are used when attempting to determine the effect of the manipulation of the variables being controlled.
      - Experimentation is the only way we can determine a cause-and-effect relationship between two variables.

    o **Examples:**
      - A survey on youth and smoking contacted by telephone 300 smokers and 300 nonsmokers, all 14 to 22 years of age. *This is an example of a sample survey. It is a stratified sample as both smokers and nonsmokers were contacted. (Stratified is discussed on page five.)*

      - Does air pollution induce DNA mutations in mice? Starting with 40 male and 40 female mice, 20 of each sex were housed in a polluted industrial area downwind from a steel mill. The other 20 of each sex were housed at an unpolluted rural location 30 kilometers away. *This is an example of an experiment. We are imposing a treatment (polluted air or nonpolluted air) on the mice. It is also a block design. (Block design is discussed on page 11).*

      - The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and in people who eat less than one apple a week. They are going to conduct a study with 50 people in each group. Fifty clinic patients who report that they routinely eat an apple a day and 50 clinic patients who report that they eat less than one apple a week will be identified. The dentists will examine the patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups. Why is this an observation study and not an experiment? *This is not an experiment because we are just observing and not assigning the patients to treatment groups. We are only identifying the patients as apple eaters or nonapple eaters by what they tell us. We are not forcing them to eat or not eat apples. This is an example of an observational study.*

Syllabus Objectives: 2.2 – The student will describe the characteristics of a well-designed and well-conducted survey.   2.5 – The student will identify sources of bias in sampling and surveys.

Bias in sampling and surveys

- **Bias –** A sampling method is biased if it systematically favors certain outcomes. Bias in sampling is the tendency for samples to differ from the corresponding population in some systematic way. Bias can result from the way in which the sample is selected or from the way in which information is obtained once the sample has been chosen.

- Three Major Types of Bias:

   o **Undercoverage (or Selection) Bias –** This occurs when some groups in the population are left out of the process of choosing the sample.
      - **Ex:** Taking a sample of opinion in a community by selecting participants from phone numbers in the local phone book would systematically exclude people who choose to have unlisted numbers, people who do not have phones, and people who have moved into the community since the telephone directory was published.

   o **Nonresponse Bias –** This occurs when an individual chosen for the sample can't be contacted or does not cooperate. As with undercoverage bias, nonresponse bias can distort results if those who respond differ in important ways from those who do not respond.
      - **Ex:** In a study that asks questions of a personal nature, many individuals that are selected might refuse to answer the survey questions. People that do not answer their phones for a phone survey or do not return mail surveys are also examples of nonresponse bias.

   o **Response (or Measurement) Bias –** This is the tendency for samples to differ from the corresponding population because the method of observation tends to produce values that differ from the true value.
      - **Ex:** Taking a sample of weights of a type of apple when the scale consistently gives a weight that is 0.2 ounces high.
      - It does not have to involve a measurement – the behavior of the respondent or of the interviewer can also cause response bias.
         - Respondents may lie, especially if asked about illegal or unpopular behavior
         - Respondents tendency to tailor their responses to please the interviewer
         - Race or sex of the interviewer can influence responses
         - Poor wording of survey questions

- It is important to note that bias is introduced by the way in which a sample is selected or by the way in which the data are collected from the sample. Increasing the size of the sample, although possibly desirable for other reasons, does nothing to reduce bias if the method of selecting the sample is flawed or if the nonresponse rate remains high. The remedy for bias in choosing a sample is to allow impersonal chance to do the choosing. Randomization reduces bias.

- **Ex:** A television news editor would like to know how local registered voters would respond to the question, "Are you in favor of the school bond measure that will be voted on in an upcoming special election?" A television survey is conducted during a break in the evening news by listing two telephone numbers side by side on the screen, one for viewers to call if they approve of the bond measure, and the other to call if they disapprove. This survey method could produce biased results for a number of reasons. Which one of the following is the most obvious reason?

    (A)   It uses a stratified sample rather than a simple random sample.
    (B)   People who feel strongly about the issue are more likely to respond.
    (C)   Viewers should be told about the issues before the survey is conducted.
    (D)   Some registered voters who call might not vote in the election.
    (E)   The wording of the question is biased.

The correct answer is B. This study is biased as people will be selecting to be part of the sample, instead of being chosen. This is an example of a voluntary response sample (discussed on page 6) and since people are selecting themselves to be part of the sample, the people that most often respond are people with strong and often negative opinions on the issue.

- **Ex:** Which of the following sample designs does NOT contain a source of bias?

    (A)   A politician would like to know how her constituents feel about a particular issue. As a result, her office mails questionnaires about the issue to a random sample of adults in her political district.
    (B)   A company uses the telephone directory to randomly select adults for a telephone survey to gauge their feelings toward items manufactured by the company.
    (C)   An interviewer selects a random sample of individuals to question about a particular issue. Since some of the individuals are not informed about the issue, the interviewer gives background and his personal view on the issue before recording their responses.
    (D)   A news show asks viewers to call a toll-free number to express their opinions about a recent high-publicity trial.
    (E)   One thousand numbered tickets are sold as a fund-raiser. Five numbers are chosen randomly, and the individuals with the winning ticket numbers each win $10.

The correct answer is E. This is an example of a simple random sample. Each individual has the same probability of having his/her number selected, and every sample of size 5 has an equal probability of being selected. The other choices contain the biases previously discussed.

Syllabus Objective: 2.6 – The student will sample from a population using various sampling methods, including simple random sampling, stratified random sampling, and cluster sampling.

- Types of Sampling Designs
    1. **Simple Random Sample (SRS) –** is a sample that is selected in a way that ensures that every different possible sample of the desired size has the same chance of being selected.
        - A common method of selecting an SRS is to first create a list, called a **sampling frame**, of the individuals in the population. Each item on the list can then be identified by a number and a table of random digits or a random number generator can be used to select the sample.

    2. **Stratified Random Sample –** An entire population is divided into subpopulations called strata. Stratified sampling entails selecting a separate simple random sample from each of the strata.
        - **Ex:** Suppose we want to learn how students feel about funding for the football team at a large university. The campus is 60% men and 40% women, and we suspect that men and women have different views on the funding. If we use an SRS and select 100 people, we could end up with 70 men and 30 women or 65 men and 35 women and our estimates could vary widely. To help balance this, we could separate the population into men and women and then take a random sample from each group. To represent the university, we would choose 60 men and 40 women.

    3. **Cluster Sample –** An entire population is divided into non-overlapping subgroups called clusters. Cluster sampling entails selecting clusters at random and ALL individuals in the selected clusters are included in the sample.
        - **Ex:** In a large university, a professor wants to find out about student attitudes and randomly selects a number of classes to survey and he includes ALL the students in those classes. Note: The ideal situation occurs when it is reasonable to assume that each cluster reflects the general population. If that is not the case or when clusters are small, a large number of clusters must be selected to get a sample that reflects the population.

    4. **Systematic Sample –** This sampling is a procedure that can be employed when it is possible to view the population of interest as consisting of a list or a sequential arrangement. A value, k, is specified (like 25, 50, 100). One of the first k individuals in the sequence is randomly selected to be included in the sample and then each kth individual is chosen after that.
        - **Ex:** In a large university, a professor wants to select a sample of students to determine the average age. The professor takes the student directory (an alphabetical list) and randomly chooses one of the first 100 students on the list. Then, the professor takes every 100th student from that point on in the list.

    5. **Convenience Sample –** This is a type of nonrandom sampling. It uses the easiest, most available or convenient group to form a sample.
        - **Ex:** Trying to gauge freshman opinion about the food served on campus, a student stands outside the school cafeteria outside at lunchtime and stops people to ask them questions.

6. **Voluntary Response Sample –** This sample consists of people who choose themselves by responding to a general appeal.
   - **Ex:** These samples are often taken by television news programs and internet services. Viewers are encouraged to go to a website and "vote" yes or no on some issue. The commentator then would announce the results of the survey. It is highly unlikely that the responses would be accurately representative of the opinion of the public at large.

- Things to avoid: Voluntary response samples are almost always biased and should be avoided and distrusted. Convenience samples are likely to be flawed for similar reasons. Even with a reasonable design, sample frames may not be representative.

- You try:

  Ex: Determine the type of sampling used below.

  1. A soccer coach selects 6 players from a group of boys aged 8 to 10, 7 players from a group of boys aged 11 to 12, and 3 players from a group of boys aged 13 to 14 to form a soccer team.   (Stratified Random Sample)
  2. A pollster interviews all human resource personnel in five different high tech companies.   (Cluster Sample)
  3. A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers.   (SRS)
  4. A medical researcher interviews every third cancer patient from a list of cancer patients at a local hospital.   (Systematic Random Sample)
  5. A student interviews classmates in his Algebra class to determine how many pairs of jeans a student owns.   (Convenience Sample)
  6. An engineering researcher interviews 50 women engineers and 50 men engineers.   (Stratified Random Sample).
  7. A question posted on the Lycos Web site on June 18, 2000, asked visitors to the site to say whether they thought that marijuana should be legally available for medicinal purposes.   (Voluntary Response Sample)

More on Simple Random Samples

- **Simple Random Samples** offer the best method of gathering unbiased data. An SRS insures that every possible *sample* of the size we plan to draw has an equal chance to be selected, not only that each *person* has an equal chance of being selected. Every individual has an equal chance and every possible group of *n* individuals has an equal chance of being our sample.
  - The idea of an SRS is to choose our sample by drawing names from a hat. Put all individuals in the population in the "hat" and choose your sample. This method, of course is not practical.
  - To select our SRS, we can use random digits from a calculator, table or computer.
    - **Table of Random Digits** - a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these two properties.
      1. Each entry is equally likely to be any one of the 10 digits 0 through 9.
      2. The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

3. This table can be found in the text.

4. Choosing an SRS: First, LABEL – assign a numerical label to every individual. Must be the same number of numbers for each person. Next, use the TABLE – use the table to select people at random.

- **Ex:** An accounting firm serves 90 business clients. They want to interview a sample of 8 clients in detail to find ways to improve client satisfaction. To avoid bias, they choose an SRS of size 8. Let's choose an SRS using the table of random digits above.

First, number the 90 clients. We must use two digit numbers so all individuals have the same number of numbers, either 00 – 89 or 01 – 90 will work. Let's use 01 – 90. Now start on ANY line of the table – we'll choose line 105 and read across the line, two digits at a time, ignoring repeats and numbers that are not in our interval of numbers. First is 95 (ignore), next is 59, then 29, 40, 07, 69, 97 (ignore), 19, 14 and 81. That means that the business clients with the labels of 59, 29, 40, 07, 69, 19, 14 and 81 will be interviewed.

- **Ex:** The student government at a high school wants to conduct a survey of student opinion. It wants to begin with a simple random sample of 60 students. Which of the following survey methods will produce a simple random sample?
  (A)   Survey the first 60 students to arrive at school in the morning.
  (B)   Survey every 10th student entering the school library until 60 students are surveyed.
  (C)   Use random numbers to choose 15 each of first-year, second-year, third-year, and fourth-year students.
  (D)   Number the cafeteria seats. Use a table of random numbers to choose seats and interview the students until 60 have been interviewed.
  (E)   Number the students in the official school roster. Use a table of random numbers to choose 60 students from this roster for the survey.
  The only method that will produce an SRS is E. A is a convenience sample, B is a systematic random sample, C is a stratified random sample and D is biased as not all students in the population will be in the cafeteria and students are left out of the sampling frame.

- Random Digit Table Example:

Every Monday a local radio station gives coupons away to 50 people who correctly answer a question about a news fact from the previous day's newspaper. The coupons given away are numbered from 1 to 50, with the first person receiving coupon 1, the second person receiving coupon 2, and so on, until all 50 coupons are given away. On the following Saturday, the radio station randomly draws numbers from 1 to 50 and awards cash prizes to the holders of the coupons with these numbers. Numbers continue to be drawn without replacement until the total amount awarded first equals or exceeds $300. If selected, coupons 1 through 5 each have a cash value of $200, coupons 6 through 20 each have a cash value of $100, and coupons 21 through 50 each have a cash value of $50.

(a) Explain how you would conduct a simulation using the random number table provided below to estimate the distribution of the number of prize winners each week.

I will use 2-digit numbers to represent each ticket (for example, the numbers 01 – 05 represent the $200 tickets, 06 – 20 represent the $100 tickets and the numbers 21 – 50 represent the $50 tickets). I will read 2 digits at a time from the random number table, ignoring 00 and numbers above 50 and also ignoring repeats. First, I will pick a number on the table and determine the amount of the prize associated with the chosen ticket. I will then repeat, adding this amount to the total amount awarded so far. If the total amount awarded is still less than $300, I will repeat this process until $300 or just over is reached. Once $300 is obtained, I will then count the total number of winners.

b) Perform your simulation 3 times. (That is, run 3 trials of your simulation.) Start at the leftmost digit in the first row of the table and move across. <u>Report the number of winners in each of your 3 trials</u>.

72749 13347 65030 26128 49067 02094 49953 74674 94617 13317

81638 36566 42709 33717 59943 12027 46547 61303 46699 76423

38449 46438 91579 01907 72146 05764 22400 94490 49833 09258

Reading across the table, beginning with the first line, we have 72, 74, 91, 33…
Trial 1: Winners: 33, 47, 03 - $50 + $50 + $200 = $300.   3 winners.
Trial 2: Winners: 02, 28, 48 - $200 + $50+ $50 = $300.   3 winners.
Trial 3: Winners: 06, 29, 04 - $100 + $50 + $200 = $350   3 winners

Syllabus Objectives: 2.7 – The student will describe the characteristics of a well-designed and well-conducted experiment. 2. 8 – The student will discuss the practice of assigning treatments within an experiment.

Experiments – Deliberately imposes some treatment on individuals in order to observe their responses.
- Allows the measurement of effect of a treatment
- Can help to control lurking variables
- Can give good evidence of causation
- May not measure realistic effects. Not necessarily workable in real life.

Experimentation Vocabulary
1. The **design** of an experiment is the overall plan for conducting an experiment.
2. The individuals on which the experiment is done on are the **experimental units**.
3. When the experimental units are human beings, they are called **subjects**.
4. A specific experimental condition applied to the units is called a **treatment**.
5. A **response variable** measures an outcome of a study.
6. An **explanatory variable** helps explain or influences changes in a response variable.
7. The explanatory variables in an experiment are often called **factors**.
8. More on **factors:**
   a. An **extraneous factor** is one that is not of interest in the current study but is thought to affect the response variable.
   b. Two factors are **confounded** if their effects on the response variable cannot be distinguished from one another.
9. The specific values that the experimenter chooses for a factor are called the **levels** of the factor.
10. The combination of specific levels from all the factors that an experiment unit receives is known as its **treatment.**
11. More on **treatments:**
    a. A **placebo** is a treatment that resembles the other treatments in an experiment, but which has no active ingredients. In an experiment to test the effectiveness of a new vaccine (by injection), a placebo treatment would consist of injection of a neutral substance such as saline solution.
    b. A **control group** is a group that receives no treatment or a placebo treatment.

Ex: What are the effects of repeated exposure to an advertising message? The answer may depend on both the length of the ad and how often it is repeated. An experiment investigated this question using undergraduate students. They all viewed a 40-minute television program that included ads for a digital camera. Some saw a 30-second commercial; others a 90-second version. The same commercial was shown either 1, 3 or 5 times during the program.
The undergraduate students are the subjects. The experiment has two factors: length of the commercial, with 2 levels (30 sec and 90 sec), and repetitions, with 3 levels (1, 3 and 5). The 6 combinations of one level of each factor form 6 treatments (see the table below). After viewing, all of the subjects answer questions about their recall of the ad, their attitude toward the camera, and their intention to purchase it. These are the response variables.

<div align="center">

Factor B:
Repetitions

</div>

		1 time	3 times	5 times
Factor A: **Length**	30 seconds	1	2	3
	90 seconds	4	5	6

Syllabus Objective: 2. 9 – The student will describe the need for control groups, randomization, and replication within experiments.

- **The Four Principles of Experimental Design**
  1. **Randomization –** Random assignment of subjects to treatments or of treatments to trials is needed to ensure that the experiment does not systematically favor one experimental condition over another. Randomization allows us to equalize the effects of unknown or uncontrollable sources of variation. If experimental units were not assigned to treatments at random, we would not be able to use the powerful methods of Statistics to draw conclusions from an experiment.
  2. **Blocking –** Using extraneous factors to create groups (blocks) that are similar. All experimental conditions are then tried in each block. Sometimes, attributes of the experimental units that we are not studying and that we can't control may nevertheless affect the outcomes of an experiment.
  3. **Control –** Holding extraneous factors constant so that their effects are not confounded with those of the experimental conditions. We control sources of variation other than the factors we are testing by making conditions as similar as possible for all treatment groups. If the purpose of an experiment is to determine whether some treatment has an effect, it is important to include an experimental group that does not receive the treatment. Such a group is called a **control group**. The use of a control group allows the experimenter to assess how the response variable behaves when the treatment is not used. This provides a baseline against which the treatment groups can then be compared.
  4. **Replication –** Ensuring there is an adequate number of observations in each experimental condition.

Syllabus Objectives: 2.12 – The student will conduct a completely randomized experiment for two treatments. 2.13 – The student will conduct an experiment using a randomized block design, including matched pairs.

- **Experimental Design**
  - **Completely Randomized Design (CRD)** – All experimental units are allocated at random among all treatments. This can be shown in a diagram form.
    - **Ex:** Many utility companies have programs to encourage their customers to conserve energy. The company has two approaches. One is to give customers a chart and information about monitoring their electricity use and the other is to place electronic meters in households to show the current cost of electricity at that moment. The experiment below compares these two approaches with each other and also with a control group which receives no treatment. The experiment begins with random assignment and ends with the response variable – comparing electricity use. Note: all 60 homes are randomly assigned to the 3 treatments.

- **Ex:** A new medication has been developed to treat sleep-onset insomnia (difficulty in falling asleep). Researchers want to compare this drug to a drug that has been used in the past by comparing the length of time it takes subjects to fall asleep. Of the following, which is the best method for obtaining this information?
  - (A) Have subjects choose which drug they are willing to use, then compare the results.
  - (B) Assign the two drugs to the subjects on the basis of their past sleep history without randomization, then compare the results.
  - (C) Give the new drug to all subjects on the first night. Give the old drug to all subjects on the second night. Compare the results.
  - (D) Randomly assign the subjects to two groups, giving the new drug to one group and no drug to the other group, then compare the results.
  - (E) Randomly assign the subjects to two groups, giving the new drug to one group and the old drug to the other group, then compare the results.

  The answer is E. This is an example of a completely randomized design.

- **Block Design –** A **block** is a group of experimental units or subjects that are known before the experiment to be similar in some way that is expected to affect the response to the treatments. In a **block design**, the random assignment of units to treatments is carried out separately within each block. That means, we RANDOMIZE AFTER we block.
  - **Ex:** Below is a diagram illustrating the block design. A clinical experiment to compare three therapies for cancer treatment might differ due to gender, so first the subjects are blocked by gender and THEN random assignment places both groups in the three therapies. Block allows us to draw separate conclusion about each block and allows more precise overall conclusions.

    Subjects

    Men → Random assignment → Group 1 → Therapy 1 / Group 2 → Therapy 2 / Group 3 → Therapy 3 → Compare survival

    Women → Random assignment → Group 1 → Therapy 1 / Group 2 → Therapy 2 / Group 3 → Therapy 3 → Compare survival

  - **Ex:** Automobile brake pads are either metallic or nonmetallic. An experiment is to be conducted to determine whether the stopping distance is the same for both types of brake pads. In previous studies, it was determined that car size (small, medium, large) is associated with stopping distance, but car type (sedan, wagon, coupe) is not associated with stopping distance. The experiment would be best done
    - (A) by blocking on car size
    - (B) by blocking on car type
    - (C) by blocking on stopping distance
    - (D) by blocking on brake pad type
    - (E) without blocking

    The answer is A. Since car size is associated with stopping distance, we should block on car size. Only block on variables that might affect the response.

- **Ex:** A dog food company wants to compare a new lower-calorie food with their standard dog food to see if it's effective in helping inactive dogs maintain a healthy weight. They have found several dog owners willing to participate in the trial. The 60 dogs have been classified as small, medium, or large breeds, and the company will supply the owners of each size of dog with one of the two foods. The owners have agreed not to feed their dogs anything else for a period of 6 months, after which the dogs' weights will be checked.

  It is okay to draw a diagram, but alone, it is not enough. The randomization must also be discussed. A diagram is not needed, however. The experimental design and randomization can be discussed in paragraph form.
  First, take the 60 dogs and divide them into blocks according to size. Let's say we have 20 dogs of each size, so there will be three groups of 20 dogs. The groups do not have to have the same amount, but it is desirable. The three groups will then be divided in half. Half of the dogs will take the new lower-calorie food and the other half of each group will eat the standard dog food. To randomize the dogs, label the dogs in each group 01-20 and using a random number table, ignoring repeats and numbers outside of our range, the first 10 numbers that are chosen will eat the lower-calorie food and the remaining 10 dogs will eat the standard dog food. This process is then repeated for the remaining two size groups. After 6 months, the dogs will be weighed and the differences of the before and after weights will be compared in each group and across the three groups.

- Another form of blocking is called the **matched pairs design.** Matched pairs design compares just two treatments. We choose blocks of two units that are as closely matched as possible (often these are twins) and we assign one of the treatments to each unit/subject randomly. Alternatively, each block in a matched pairs design may consist of just one subject, who gets BOTH treatments one after the other. Each subject serves as his or her own control. **The order of the treatments can influence a subject's response, so we must randomize the order for each subject!**
  - Matching subjects can reduce variation in the same way as blocking as subjects are paired because they are similar in ways NOT under study.
  - **Ex:** A researcher believes that students are able to concentrate better while listening to classical music. To test this theory, she plans to record the time it takes a student to complete a puzzle maze while listening to classical music and the time it takes him/her to complete another puzzle of the same difficulty level in a quiet room. Because there is so much variability in problem-solving abilities among students, a matched-pairs design will be used to reduce variability so that any differences recorded can be attributed to the conditions under which the student completed the puzzle. Each student will complete a puzzle under each of the conditions. A coin will be flipped to determine whether the task will first be done in a quiet room or while listening to classical music. The difference in the time it takes to complete each puzzle (Quiet – Music) is recorded for each student.
  - **Ex:** Another type of matched-pairs design involves two subjects. The two subjects are paired based on common characteristics that might affect the response variable. To test the durability of a new running shoe, marathon runners are matched by weight, physical build, and running times. A coin is tossed to determine which runner in each matched-pair will wear the new design of shoe and which runner will wear the manufacturer's old design. After the marathon, the difference in wear pattern for each pair of runners is then measured and recorded.

Syllabus Objectives: 2.10 – The student will analyze a study for sources of bias and confounding variables. 2.11 – The student will describe the need for blinding in an experiment.

Cautions about Experimentation

- If possible, an experiment should be **double-blind.** In a double-blind experiment, neither the subjects nor the people who have contact with them and who measure the response variable know which treatment a subject received. In a **single-blind** experiment, the subjects do not know which treatment they received, but the people in contact with them do know. This is not as good as double-blind.
- The most serious potential weakness of experiments is **lack of realism**. The subjects or treatments or setting of an experiment may not realistically duplicate the conditions we really want to study.
    - A study compares two television advertisements by showing TV programs to student subjects. The students know it's "just an experiment". We can't be sure that the results apply to everyday television viewers.
- **The Placebo Effect –** The placebo effect is the tendency in humans to show a response whenever they think a treatment is in effect. Well-designed experiments use a control group so that the placebo effect operates equally on both the treatment group and the control group thus allowing us to attribute changes in the response variable to the explanatory variable.
- **Confounding vs. Lurking Variables**
    - **Confounding –** An experiment is said to be confounded if we cannot separate the effect of a treatment (explanatory variable) from the effects of other influences (**confounding variables**) on the response variable.
        - **Ex:** Suppose we want to test the effectiveness of a meditation program in relieving workers' stress. We create a control group (a group who does not get any treatment, but whose stress levels are also monitored). If there are outside influences that reduce stress, they will also affect this group. By including a control group in our experiment, a confounding variable (like better weather, news of a bonus, a change in management) will act equally on both groups, and any change in stress levels can be attributed to the explanatory variable, meditation.
        - **Ex:** After many dogs and cats suffered health problems caused by contaminated foods, we're trying to find out whether a newly formulated pet food is safe. Our experiment will feed some animals the new food and others a food known to be safe, and a veterinarian will check the response. Why would it be a bad design to feed the test food to some dogs and the safe food to cats? *This would create confounding. We would not be able to tell whether any differences in animals' health were attributable to the food they had eaten or to differences in how the two species responded.*
    - **Lurking variable –** Confounding variables and lurking variables are alike in that they interfere with our ability to interpret our analyses simply. Each can mislead us, but the difference is where the confusion lies. A lurking variable creates an association between two other variables that tempts us to think that one may cause the other. A lurking variable, then, is usually thought of as a variable associated with both the response variable and the explanatory variable that makes it appear that the explanatory may be causing the response. A confounding variable is associated in a noncausal way with a factor and affects the response. Because of the confounding, we find that we can't tell whether any effect we see was caused by our factor or by the confounding variable – or even by both working together.

- **Ex:** A study of existing records of 27,000 automobile accidents involving children in Michigan found that about 10 percent of children who were wearing a seatbelt (group SB) were injured and that about 15 percent of children who were not wearing a seatbelt (group NSB) were injured. Which of the following statements should NOT be included in a summary report about this study?
  - (A) Driver behavior may be a potential confounding factor.
  - (B) The child's location in the car may be a potential confounding factor.
  - (C) This study was not an experiment, and cause-and-effect inferences are not warranted.
  - (D) This study demonstrates clearly that seat belts save children from injury.
  - (E) Concluding that seatbelts save children from injury is risky, at least until the study is independently replicated.

  The answer is D. This is an observational study and not an experiment, therefore no cause-and-effect relationship can be inferred.

Syllabus Objectives: 2.14 – The student will describe the types of conclusions that may be drawn from observational studies, experiments, and surveys. 2.15 – The student will determine the generalizability of results from observational studies, experiments, and surveys.

- **Statistical Significance –** When an observed difference is too large for us to believe that it is likely to have occurred by chance, we consider the difference to be statistically significant.

  - If the differences among the treatment groups are big enough, we'll attribute the differences to the treatments, but how can we decide what is big enough? The real question is whether the differences we observed are about as big as we might get just from the randomization alone, or whether they're bigger than that. If we decide that they're bigger, we'll attribute the differences to the treatments and say the differences are statistically significant.

  - **Ex:** Which of the following is a key distinction between well designed experiments and observational studies?
    - (A) More subjects are available for experiments than for observational studies.
    - (B) Ethical constraints prevent large-scale observational studies.
    - (C) Experiments are less costly to conduct than observational studies.
    - (D) An experiment can show a direct cause-and-effect relationship, whereas an observational study cannot.
    - (E) Tests of significance cannot be used on data collected from an observational study.

    The answer is D. Only a well-designed, controlled experiment can determine cause and effect between two variables.