AP Statistics
Unit 03 – Sampling & Study Design
Day 04 Notes – Designing Experiments

Name Key 2018 – 2019
Period______

Observational Study: observes individuals and measures variables of interest but does not attempt to influence the response.

Experiment: deliberately imposes some treatments on individuals to measure their responses.

Example: Does Taking Hormones Reduce Heart Attack Risk after Menopause?

observational study

Should women take hormones such as estrogen after menopause, when natural production of these hormones ends? In 1992 several major medical organizations said "Yes." Women who took hormones seemed to reduce their risk of a heart attack by 35% to 50%. The risks of taking hormones appeared small compared with the benefits.

The evidence in favor of hormone replacement came from a number of observational studies that compared women who were taking hormones with others who were not. But the women who chose to take hormones were richer and better educated and saw doctors more often than women who didn't take hormones. Because the women who took hormones did many other things to better maintain their health, it isn't surprising that they had fewer heart attacks.

To get convincing data on the link between hormone replacement and heart attacks, we should do an *experiment*. Experiments don't let women decide what to do. They assign women to either hormone replacement pills or to placebo pills that look and taste the same as the hormone pills. The assignment is done by a coin toss, so that all kinds of women are equally likely to get either treatment. By 2002, several experiments with women of different ages agreed that hormone replacement does *not* reduce the risk of heart attacks. The National Institutes of Health, after reviewing the evidence, concluded that the first studies were wrong. Taking hormones after menopause quickly fell out of favor.

Why did the study fail? confounding! wealth was a factor (variable) that could lead to better healthcare, purchase of healthy meals, better health education, purchase of a gym membership, etc. These things could all cause a reduced risk of heart attack.

CHECK YOUR UNDERSTANDING:

1. Does reducing screen brightness increase battery life in laptop computers? To find out, researchers obtained 30 new laptops of the same brand. They chose 15 of the computers at random and adjusted their screens to the brightest setting. The other 15 laptop screens were left at the default setting – moderate brightness. Researchers then measured how long each machine's battery lasted. Was this an observational study or an experiment? Justify your answer.

   experiment! the researchers imposed a treatment (screen brightness) and measured the effects on battery life of the laptops.

2. Does eating dinner with their families improve students' academic performance? According to an ABC News article, "Teenagers who eat with their families at least five times a week are more likely to get better grades in school." This finding was based on a sample survey conducted by researchers at Columbia University.
   a. Was this an observational study or an experiment? Justify your answer.

      Observational study because researchers aren't imposing a treatment! students were not assigned to eat or not eat dinner w/ their families a specific # of times per week.

   b. What are the explanatory and response variables?

      if they eat dinner w/ family & frequency — academic performance (GPA)

   c. Explain clearly why such a study cannot establish a cause-and-effect relationship. Suggest a variable that may be confounded with whether families eat dinner together.

      There are probably other variables that influence the response variable (AND the explanatory variable). For example, students who participate in sports or had jobs may be busy during dinnertime. They may also have less time to study, resulting in lower academic performance.

The Language of Experiments

Treatment: a specific condition applied to the individuals in an experiment.

Factors: explanatory variables

Levels: magnitude or type(s) of value per factor

Experimental Units: the smallest collection of individuals to which treatments are applied (a treatment is applied).

Subjects: when experimental units are human beings, we call them subjects.

Example: TV Advertising

What are the effects of repeated exposure to an advertising message? The answer may depend on both the length of the ad and on how often it is repeated. An experiment investigated this question using 120 undergraduate students who volunteered to participate. All subjects viewed a 40-minute television program that included ads for a digital camera. Some subjects saw a 30-second commercial; others, a 90-second version. The same commercial was shown either 1, 3, or 5 times during the program. After viewing, all of the subjects answered questions about their recall of the ad, their attitude toward the camera, and their intention to purchase it.

For the advertising study, identify the experimental units or subjects, explanatory and response variables, and the treatments.

Experimental Units/Subjects: 120 undergraduate student volunteers

Explanatory Variable: 2 factors: length of ad, frequency of ad

Response Variable: recall of ad, attitude toward camera, intention to buy camera

Treatments:

frequency of ad (times)

length of ad	1	3	5
30 sec	T1	T2	T3
90 sec	T4	T5	T6

Bad Example: Are Online SAT Prep Courses Effective?

A high school regularly offers a review course to prepare students for the SAT. This year, budget cuts will allow the school to offer only an online version of the course. Suppose the group of students who take the online course earn an average increase of 45 points in their math scores from a pre-test to the actual SAT test. Can we conclude that the online course is effective?

This experiment has a very simple design. A group of subjects (the students) were exposed to a treatment (the online course), and the outcome (increase in math scores) was observed. Here is the design:

Students → Online course → increase in math scores

A closer look showed that many of the students in the online review course were taking advanced math classes in school. Maybe the students in the online course improved their math scores because of what they were learning in their school math classes, not because of the online course. This confounding prevents us from concluding that the online course is effective.

Outside the lab, badly designed experiments often yield worthless results because of confounding.

Comparison: experiments compare 2 or more treatments.

Random Assignment: experimental units are assigned to treatments using a chance process.

Good Example: SAT Prep – Online versus Classroom

This year, the high school has enough budget money to compare the online SAT course with the classroom SAT course. Fifty students have agreed to participate in an experiment comparing the two instructional methods.

alternative: 25 online, 25 classroom. students pull out one slip (one at a time)

Describe how you would randomly assign 25 students to each of the two methods:

a) Using 50 identical slips of paper hat method. write each student's name on an identical slip of paper. Put slips in a hat and mix thoroughly. Draw out slips one at a time until you have 25 slips of paper. The students whose names are on the slips are assigned to the online course. The remaining students are assigned to the classroom course.

b) Using technology
Assign subjects a unique # 1-50 alphabetically by last name. Use the random # generator on your calculator to produce 25 unique/nonrepeated #s and assign the corresponding students to the online course. The remaining students are assigned to the classroom course.

c) Using Table D (or B)
assign subjects a unique 2-digit # 01-50, alphabetically by last name. Choose line 135 on table B, going left to right. Choose the first 25 unique/nonrepeated 2-digit #s between 01-50 and assign the corresponding students to the online course. The remaining students are assigned to the classroom course.

PRINCIPLES OF EXPERIMENTAL DESIGN

The basic principles for designing experiments are as follows:

1. **Comparison:** describe the 2+ treatments you're comparing & what you will measure. comparing ___ to ___ on ___.

2. **Random Assignment:** use chance to assign experimental units to treatments. Doing so helps create roughly equivalent groups of experimental units by balancing the effects of other variables among the treatment groups.

3. **Control:** keep other variables that might affect the response the same for all groups (limit confounding).

4. **Replication:** use enough experimental units in each group so that any differences in the effects of the treatments can be distinguished from chance differences between the groups.

Placebo: a sham or dummy treatment (and the resulting response). placebo effect

Example: The Physicians' Health Study

Does regularly taking aspirin help protect people against heart attacks? The Physicians' Health Study was a medical experiment that helped answer this question. In fact, the Physicians' Health Study looked at the effects of two drugs: aspirin and beta-carotene. Researchers wondered whether beta-carotene would help prevent some forms of cancer. The subjects in this experiment were 21996 male physicians. There were two explanatory variables (factors), each having two levels: aspirin (yes or no) and beta-carotene (yes or no). Combinations of the levels of these factors form the four treatments shown in the figure below. One-fourth of the subjects were assigned at random to each of these treatments.

		Factor 2: Beta-carotene	
		Yes	No
Factor 1: Aspirin	Yes	Aspirin & Beta-carotene T1	Aspirin & Placebo T2
	No	Beta-carotene & Placebo T3	Placebo & Placebo T4

On odd-numbered days, the subjects took a tablet that contained aspirin or a dummy pill that looked and tasted like aspirin but had no active ingredient (a placebo). On even-numbered days, they took either a capsule containing beta-carotene or a placebo. There were several response variables – the study looked for heart attacks, several kinds of cancer, and other medical outcomes. After several years, 239 of the placebo group but only 139 of the aspirin group had suffered heart attacks. This difference is large enough to give good evidence that taking aspirin does reduce heart attacks. It did not appear, however, that beta-carotene had any effect on prevent(ing) cancer.

Explain how each of the four principles of experimental design was used in the Physician's Health Study.

Comparison: researchers compared the effects of aspirin and beta-carotene treatments on heart attack rates, cancer rates, and other health outcomes. to a placebo

Random Assignment: random assignment was used but the method was not specified. this helps ensure that the treatment groups were roughly equivalent to begin with.

Control: everyone took a pill every day. All subjects were male physicians.

Replication: each treatment group had 1/4 of the nearly 22000 (21996) participants, which is sufficiently large. This large # of subjects helped ensure that the difference in heart attacks was due to aspirin and not chance variation in the random assignment.

Completely Randomized Design:

the experimental units are assigned to the treatments completely by chance.

Notice that the definition of a completely randomized design does not require that each treatment be assigned to an equal number of experimental units. It does specific that the assignment of treatments must occur completely at random.

Think about it: Does using chance to assign treatments in an experiment guarantee a completely randomized design?

Actually, no. Let's return to the SAT prep course experiment. Another way to randomly assign the 50 students to the two treatments is by tossing a coin. If it's heads, then the student will take the course online. If it's tails, then the student will take the classroom course.

As long as all 50 students toss a coin, this is still a completely randomized design. Of course, the two experimental groups are unlikely to contain exactly 25 students each due to the chance variation in coin tosses.

The problem comes if we try to force the two groups to have equal sizes. Suppose we let the coin tossing continue until one of the groups has 25 students and then place the remaining students in the other group. This is no longer a completely randomized design, because the last few students aren't being assigned to one of the treatment groups by chance. In fact, these students will all end up in the same group, which could lead to bias if these individuals share some characteristic that would systematically affect the response variable. For example, if the students came to toss the coin last because they're lazier than the other students who volunteered, then the SAT prep class they're in will seem less effective than it really is.

* avoid this by using the hat method!

Example: Tomatoes

An ad for OptiGro plant fertilizer claims that with this product you will grow "juicier, tastier" tomatoes. You'd like to test this claim, and wonder whether you might be able to get by with half the specified does. How can you set up an experiment to check out the claim?

Of course, you'll have to get some tomatoes, try growing some plants with the product and some without, and see what happens. But you'll need a clearer plan than that. How should you design your experiment? Let's say you have access to 24 tomato plants from a local garden store, 24 identical plots of land to grow these plants, and you want to know how the different levels of the fertilizer affect the tomato juiciness and tastiness, evaluated on a 1-7 scale.

QUESTION: How would you design an experiment to best OptiGro fertilizer?

PLAN State what you want to know	*I want to know whether tomato plants grown with OptiGro yield juicier, tastier tomatoes than plants raised in otherwise similar circumstances but without the fertilizer.*
RESPONSE Specify the response variable	*I'll evaluate the juiciness and taste of the tomatoes by asking a panel of judges to rate them on a scale from 1-7 in juiciness and in taste.*
EXPERIMENTAL UNITS Specify the experimental units	*I'll obtain 24 tomato plants of the same variety from a local garden store.*

EXPERIMENTAL DESIGN

Address the principles of design:

Compare the factor levels and the treatment	We will compare the effects of different amounts of OptiGro plant fertilizer on tomato tastiness & juiciness (on a 1-7 scale).
Randomly assign experimental units to treatments, to equalize the effects of unknown or uncontrollable sources of variation	I will randomly assign each plant a unique # 1-24. I will write the #s 1-24 on identical slips of paper, place in a hat, and mix thoroughly. I will pull out slips one at a time, until I have 8, and I w/o replacement will assign the corresponding plants to T1 (full fertilizer). ~~I will then pull~~ Out the next 8 slips and assign the corresponding plants to T2 (½ fert.). The remaining 8 plants will be assigned to T3 (no fert.).
Control any sources of variability you know of and can control	Identical plots of land will be used, all tomatoes are from the same store.
Replicate results by placing more than one plant in each treatment group	8 plants will be in each treatment group.
Make a picture: a diagram of your design can help you think about it clearly.	

24 tomato plants → rand. ass. → T1 (full fert.) 8 plants / T2 (½ fert.) 8 plants / T3 (no fert.) 8 plants } compare juiciness & tastiness

*you can add a spot in the diagram for groups if you want.