AP* Statistics Review

Designing a Study

Teacher Packet

Population vs. Sample

- The population is the entire group you would like to study or draw a conclusion about. Any numerical value that comes from the population is a parameter. Parameters are usually unknown. The study of an entire population is called a census.
- The sample is the part of the population that you take data from. Any numerical value that comes from a sample is called a statistic. The goal of this course is to use statistics to draw conclusions or make statements about (unknown) parameters.
- You should recognize which symbols represent parameters from the entire population and which represent statistics from the sample:

	Parameter	Statistic
Mean	μ	$\bar{x}$
Standard deviation	σ	s_x
Proportion	p	$\hat{p}$

Two kinds of studies

- In an observational study, people choose their own actions and scientists observe what they do. You cannot show a cause-and-effect relationship using an observational study.
- In an experiment, the scientist assigns treatments to subjects. A good experiment **controls** (or equalizes all the variables we are able to), **randomizes** the assignment of treatments to the subjects to hopefully equalize the variables we don't know about or cannot control, and **replicates** (uses a large sample size) to reduce variation.

Bias

- Bias is when a study systematically favors one outcome over another. It can occur when a certain group is over- or under-represented in your sample or when your measurement device affects the results. (For example, survey questions could be worded to encourage a certain answer, or a scale could be calibrated wrong so that everything weighs too much.)
- Bias changes the center (mean or median) of your distribution.
- Bias may not be done on purpose, but it should be avoided. There is no mathematical way to "fix" biased data. **Bad data leads to bad conclusions.**

Types of Bias

- Voluntary response bias occurs when subjects voluntarily choose to be in the sample, and people usually volunteer only if they have strong opinions.
- Undercoverage occurs when some groups of people are ignored when the sample is being chosen. A survey sent by e-mail ignores people without computers.
- Non-response bias occurs when some subjects chosen for the sample do not answer. (Only people with a lot of time on their hands respond.)
- Response bias occurs when subjects give incorrect answers, either because they have forgotten details or they lie about embarrassing or illegal activities.

Sampling Methods

- A simple random sample (SRS) of size n gives every individual and every group of size n an equal chance of being chosen. To carry out an SRS of size n, number the list of possible subjects. Take numbers off the random number table. For each chosen number, write down the name of the corresponding subject. Ignore repeated subjects. Continue until you have a list of n different subjects. In this process, if you have 10 or fewer possible subjects, you can use single digit numbers (including 0). If you have 11 to 100 possible subjects, you can use double digit numbers (including 00). Note: If you must describe how to take a random sample but you do not have to actually carry out your plan with random digits, then you may "put all of your subjects in a bag." Say that you will write the name of each subject on an identical piece of paper, put the papers in a bag, shake it up, and draw out slips of paper until you have a sample of the desired size.
- Choose a stratified random sample if you want to be sure to have some subjects from each subgroup in your sample. Split into subgroups, then take an SRS out of each subgroup.

  Note: All subjects in the subgroup should be similar (homogeneous). For example, the subgroups could be freshmen, sophomores, juniors, and seniors. Then your sample would be sure to have students from each grade level.
- Use a cluster sample if you have many groups that are similar to each other. Randomly choose one or more groups to be the sample.

  Note: The subjects in the groups do not have to be alike (heterogeneous), but each group should be similar to every other group. For example, if each fourth grade class in an elementary school has students of all ability levels and all socioeconomic groups, then randomly choosing one class as a sample would give an accurate representation of the fourth grade as a whole.
- Convenience samples are chosen by walking around and asking whoever you run into. Not a good idea!

- Systematic random sample: choosing every 4th name off a list or every 10th person through a door.

Vocabulary for experiments

Example: You want to know what combination of fertilizer brand and watering frequency results in the most growth for a wheat plant. You are interested in 3 brands (A, B, and C) of fertilizer and you plan to water either twice a week or four times a week.

- Explanatory variable: x variable (called the independent variable in science class)
- Response variable: y variable (called the dependent variable in science class)
- Experimental units: What you are experimenting on (wheat plants)
- Factors: the explanatory variables (here, fertilizer brand and watering frequency)
- Levels: the choices you have for each factor (fertilizer brand has three levels and watering frequency has two levels)
- Treatments: the combinations you will test (there will be six treatments, which we find by multiplying the three levels for fertilizer by the two levels for watering)
- Completely Randomized Design (CRD): take the list of experimental subjects, number them, use the random number table to assign them to treatment groups, run the experiment for a given amount of time, then compare the groups on the response variable

Considerations in Experimental Design

- A control group (which either gets no treatment or gets the old, established treatment) must be used for comparison. For example, if you are testing a treatment for a sprained ankle, you must have a group that gets no treatment because sprained ankles naturally get better over time. You need to show that the treatment group gets better **faster** than the control group.
- You want the treatment group and the control group to be as equal as possible, so that the only difference between the groups is the treatment itself. A placebo, or fake treatment, is used to equalize psychological effects between the groups.
- In a blind experiment, the subjects do not know if they are getting the treatment or the placebo. In a double blind experiment, the person handing out the treatments or the person evaluating the results also doesn't know which subjects are in which group. This is to equalize psychological effects.
- At the end of an experiment, if you can say that the difference between the groups could be due to variables A or B, then A and B are confounded variables. For example, suppose a group of boys was taught to read with Method 1 and a group of girls was taught with Method 2. If one group has higher reading scores, you cannot tell if it is due to the difference in gender or the difference in teaching method. Gender and teaching method are confounded.

- If your subjects have differences that could affect the response variable, you need to make sure the control group and the treatment group are as equal as possible by using blocking. For example, block design is necessary if you have 50 female and 20 male volunteers to use in a medication study, and you think that the medication might affect men and women differently. Form blocks by gender by putting the 50 females on one list and the 20 males on another. Randomly choose 25 females and 10 males to be in the treatment group. The remaining 25 females and 10 males will be the control group. Since both groups have the same gender make-up, any difference between the groups can be attributed to the medicine, not to gender differences.
- A matched pair design is a special case of block design in which each block consists of only 2 subjects. Match up the subjects so that they are as similar as possible, and then randomly choose one from each pair to be in the treatment group and the other to be in the control group.

Multiple Choice Questions on Designing a Study

1. A human resources director of a large company is interested in how often employees use their computers during breaks. She watches a selected group of employees at their desks during the break times. This study would best be described as

(A) a census
(B) a survey
(C) an observational study
(D) an experiment
(E) a sample

2. A company wants to compare two washing detergents (Brands A and B) to see which best keeps colors from fading. Twenty new, identical red t-shirts will be used in the trials. Ten t-shirts are washed 15 times with Brand A in warm water. The other 10 t-shirts are washed with Brand B in cold water. The amount of fading is rated on a 0 to 100 scale, and the mean for the t-shirts washed in Brand A is compared to the mean for the others. Is this a good experimental design?

(A) No, because the means are not the proper statistics for comparison.
(B) No, because more than two brands of detergent should be used.
(C) No, because more temperatures of water should be used.
(D) No, because water temperature is confounded with brand of detergent.
(E) Yes.

3. The state would like to evaluate the usefulness of a program to randomly test high school athletes for steroid use. Initially, a state agency will test athletes in all 20 schools in Fort Worth, randomly selecting 3 athletes from each school. Is this a simple random sample of student athletes in Fort Worth?

(A) Yes, because athletes will be chosen at random.
(B) Yes, because each athlete is equally likely to be chosen.
(C) Yes, because stratified sampling is a special case of simple random sampling
(D) No, because not all possible groups of 60 athletes could be in the sample.
(E) No, because a random sample of Ft. Worth schools is not chosen.

4. Which of the following sample designs does **NOT** contain a source of bias?

(A) A legislator wishes to know how his district feels about a particular issue. As a result, his office e-mails a long, detailed survey about the issue to a random sample of adults in his district.
(B) A polling organization uses the telephone directory to randomly select adults for a telephone survey to obtain opinions on the current president.
(C) All 250 students at a review session are given numbered tickets. Five numbers are chosen randomly, and the individuals with the winning ticket numbers each win a $10 gift card.
(D) A news show asks viewers to call a toll-free number to express their opinions about their choice for president.
(E) A teacher asks high school students how often they drink alcohol.

5. A college counselor would like to select a simple random sample of all the 525 students in the college. She uses the numbers from 001 to 525 to number the students in a college database and then uses a random number table to choose her sample of 30. What numbers correspond to the first 5 students chosen?

06385 61327 51790 63618 23145 46124 20031

(A) 06 38 56 13 24
(B) 63 85 61 32 45
(C) 063 132 517 361 145
(D) 063 132 182 314 124
(E) 063 327 361 145 242

6. Twenty men and 20 women with migraine headaches were subjects in an experiment to determine the effectiveness of a new pain medication. Ten of the 20 men and 10 of the 20 women were chosen at random to receive the new drug. The remaining 10 men and 10 women received a placebo. The decrease in pain was measured for each subject. The design of this experiment is

(A) completely randomized with one factor, gender
(B) completely randomized with one factor, drug
(C) randomized block, blocked by drug and gender
(D) randomized block, blocked by gender
(E) randomized block, blocked by drug

7. A student organization wants to assess the attitudes of students toward a proposed change in the hours the library is open. They randomly select 50 freshmen, 50 sophomores, 50 juniors, and 50 seniors to survey. This situation is described as

(A) a stratified random sample
(B) a simple random sample
(C) a convenience sample
(D) a systematic random sample
(E) an observational study

8. A group of 420 college students are enrolled in a blind taste test. The school's food service wants to see if they can improve the taste of their lattes. They decide to try two types of coffee beans (Arabica and Robusta); three types of syrup (vanilla, hazelnut, and mocha); and two types of milk (soy and low fat). The best combination of ingredients is sought. The latte experiment will have

(A) 2 factors, 7 levels, and 420 treatments
(B) 2 factors, 3 levels, and 12 treatments
(C) 3 factors, 7 levels, and 420 treatments
(D) 3 factors, 12 levels, and 420 treatments
(E) 3 factors, 7 levels, and 12 treatments

9. The primary reason for using blocking when designing an experiment is to reduce

(A) variation
(B) the need for randomization
(C) bias
(D) confounding
(E) the sensitivity of the experiment

10. The head of the admissions office at a small college wants to understand why minority students who visit her school do not eventually enroll. The college holds a preview weekend for students who have been admitted. Two months later, after the students have decided while college to attend, a survey is sent out to all minority students who attended the weekend visit but who did not choose to attend this college. About a third of them returned the survey, with 48% of those indicating that they received a larger scholarship offer elsewhere. Which is true?

I. The population of interest is all potential college students.
II. This survey design suffered from non-response bias.
III. Because it comes from a sample, 48% is a parameter, not a statistic.

(A) I only
(B) II only
(C) I and II only
(D) II and III only
(E) I, II, and III

Free Response Questions on Designing a Study

1. A veterinary school would like to test a food supplement that is meant to increase joint flexibility in dogs. Dogs lose flexibility as they age, with many older dogs (ages 8 and above) exhibiting joint problems. In addition, the joints of larger breeds of dogs are subject to more stress, so joint problems appear more frequently and more severely in larger dogs.

Twelve dogs have been volunteered for this study by their owners. They have been categorized by age and size, where small breeds are under 35 pounds, medium breeds are 35 to 70 pounds, and large breeds are over 71 pounds. Owners will administer the food supplement to the six dogs chosen for the treatment group, while the other six dogs will be given a placebo by their owners. Before the study, the owners will complete a questionnaire about their dog's comfort and activity level. They will answer the same questions after three months of treatment, and the results will be compared.

Dog number	Age	Breed Size
1	9	Large
2	6	Small
3	3	Large
4	5	Large
5	4	Small
6	6	Medium
7	5	Large
8	10	Medium
9	11	Large
10	2	Large
11	9	Medium
12	5	Medium

a) A matched-pairs block design will be used for this study. Using the dog numbers, tell which dogs will be in each of the six blocks **and** describe the criteria used to form the blocks.

b) Describe how the dogs will be assigned to the treatment or control group.

c) Describe why a placebo is necessary in this study.

2. A new type of soy-based baby formula has been developed to help babies who have trouble gaining weight in their first six months of life. The parents of 150 babies have volunteered their children to participate in this study. All of the babies are between 8 and 9 weeks of age (a time when they should be growing very rapidly) and have been identified by their doctors as gaining weight more slowly than normal.

The babies will be weighed before the study begins. Half of them will be given the new formula and the other half will be given the type of formula that is currently recommended for slowly growing babies. After three months, all the babies will be weighed again.

a) Describe an appropriate method for assigning the subjects to the two groups so that each group will have an equal number of subjects.

b) In this study, the researchers chose to include a group who used the currently recommended brand of formula. Why is it important to include a control group in this study even though weights will be measured at the beginning and at the end of the study?

c) Many babies continue to drink formula until they are 12 to 15 months of age. Why would the researchers choose to use only babies who are about 2 months old (about 8 to 9 weeks old) in the study?

3. At a certain neighborhood public library, the librarian was interested in finding out if the users of the library would like more children's activities to be scheduled. Based on her conversations with parents, she believed that many library users put a high priority on children's activities. She decided to do a survey to estimate the proportion of adult library users who would like the number of children's activities to be increased. During month of July, the librarian asked the first twenty adults entering the library each morning to answer the following question:

> Many library users think that more children's activities should be provided at this library. Do you think that the number of children's activities should be increased, even though this would take away resources for other programs such as adult education or public internet access?
>
> ____ Yes _____ No _____ No opinion

(a) This survey used a convenience sample. In this situation, explain how bias may have been introduced based on the way this convenience sample was selected and suggest how the sample could have been selected differently to avoid that bias.

(b) In this situation, explain how bias may have been introduced based on the way the question was worded and suggest how it could have been worded differently to avoid that bias.

Key to Designing a Study Multiple Choice

1. C The workers themselves choose to use their computers or not, so this is an observational study.
2. D The difference in fading could be due to either detergent brand or water temperature, so this is an example of confounding.
3. D This is stratified, not simple.
4. C Distractors: A is undercoverage and non-response bias; B is undercoverage; D is voluntary response bias; E is response bias
5. D Use three digit numbers, and ignore numbers over 525.
6. D This is not a completely randomized design because the treatment group was not chosen randomly from a list of all 40 volunteers. Since the treatment group was chosen randomly from 2 separate lists (males and females), this study is blocked by gender.
7. A Since students are chosen from each group, this is a stratified random sample.
8. E Three factors (coffee, syrup, milk), 7 levels (2 choices of coffee, 3 choices of syrup, 2 choices of milk), and $2 \times 3 \times 2 = 12$ combinations
9. A Blocking groups together subjects that are similar and thus reduces variability that is attributable to the differences between the blocks.
10. B I is not true because the population of interest is minority students who have visited this college. II is true because only a third of the students returned the survey. III is false; 48% is a statistic because it comes from a sample.

Rubric for Designing a Study Free Response

1. **Solution**

Part (a):
The blocks (by number) are

2 and 5	(small, younger)
6 and 12	(medium, younger)
8 and 11	(medium, older)
3 and 10	(large, quite young)
4 and 7	(large, young)
1 and 9	(large, older)

Note: students must describe the blocks by number. They do not have to add the descriptions in parentheses.

Dogs of ***similar*** size and age were assigned to the same block.

Part (b):
One dog from each block must be randomly assigned to the treatment group, and the other dog must go to the control group. A process must be described clearly enough that another person could carry it out. For example: a coin is tossed for each block. If it is heads, the lower-numbered dog goes to the treatment group and the higher-numbered dog goes to the control group. If it is tails, do the reverse. Or put the names of the two dogs in each block on slips of paper in a bag. Draw out one name. That dog goes to the treatment group and the other dog in the block goes to the control group. Repeat for each block.

Part (c):
Since the owners are the ones evaluating the results of the study, it is important that they not know if their dogs are getting the supplement or not. The use of a placebo allows the study to be blind. This prevents an owner from answering the questions differently or from changing other variables (such as the amount of exercise they give their dogs) that could affect the results of the study based on their knowledge of which group their dog is in.

Scoring
Parts (a) and (b) can be essentially correct (E), partially correct (P), or incorrect (I). Part (c) can be essentially correct (E) or incorrect (I).

Part (a) is correct if the student correctly identifies the blocks **and** describes how blocks were assigned.
Part (a) is partially correct if the student does one of the two tasks above.
Part (a) is incorrect if the student does neither.

Part (b) is correct if the method assigns one dog from each block to the control group and one to the treatment group **and** describes a clear method that results in a random assignment.

Part (b) is partially correct if the process assigns one dog from each block to the treatment group but the process is not random, or a random process is used but it does not necessarily assign one dog from each block to the treatment group, or if the random process is not described clearly enough for someone else to follow it.

Part (c) is correct if the student identifies the placebo as needed to make the study blind or to equalize psychological effects or knowledge in the owners.

4 Complete Response

All parts essentially correct.

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and two parts partially correct

1 Minimal Response

One part essentially correct and either zero or one part partially correct

OR

No parts essentially correct and two parts partially correct

2. **Solution**

Part (a):
The student must clearly describe a random process that will choose 75 babies for the treatment group and 75 for the control group. For example, put the names of all of the volunteers on identical slips of paper, put them in a bag, shake them up, and pull out 75 slips. Those babies form the treatment group. The remaining 75 are the control group.

OR

Number a list of volunteers from 001 to 150. Look at three-digit numbers from a random number table. The first 75 non-repeated numbers between 001 and 150 will be the treatment group. The remaining volunteers will be the control group.

OR

Flip a coin for each participant. If it is heads, put that baby in the treatment group. Other wise, put the baby in the control group. Continue flipping coins until 75 babies have been chosen for the treatment group. The remaining babies will go in the control group.

Part (b):
All babies should gain some weight in a three month period beginning at the age of 8 to 9 weeks old. We must have a control group for comparison purposes so we can see if the treatment group gained ***more*** weight than the control group. Students may identify some other variable leading to weight gain (other than the natural weight gain as the babies get older), but they do not have to identify any other variable.

Part (c):
The babies chosen for the study are all about the same age in order to reduce variation. Babies of different ages are expected to grow at different rates, so choosing babies of the same approximate age helps insure that the groups are similar in growth rate, eliminating age as a confounding variable and allowing more direct comparison between the groups.

Scoring
Parts (a) and (b) can be essentially correct (E), partially correct (P), or incorrect (I). Part (c) can be essentially correct (E) or incorrect (I).

Part (a) is essentially correct if the student describes a method that is based on random assignment and that will result in 75 babies in each group.
Part (a) is partially correct if the student

- describes a random method that will not result in 75 babies in each group, or
- attempts to describe a method that will assign 75 babies to each group, but the explanation of random assignment is unclear or incomplete.

Part (a) is incorrect if the method described is not random or if the student refers to "random assignment" without describing a method of assignment.

Part (b) is essentially correct if the student states that the control group is needed for comparison purposes in order to see how much of the growth can be attributed to the new formula and how much would generally be expected with the old formula.
Part (b) is partially correct if the student identifies that the control group is needed in order to compare the two groups but does not identify the need for comparison to control the other variables (such as the natural growth of the babies or another variable they have identified).
Part (b) is incorrect if the student fails to see the need for a comparison group.

Part (c) is correct if the student explains that age is related to growth and expresses the desire for more homogeneous groups or says that choosing babies of the same age will reduce variability.
Part (c) is incorrect if the student only says that age is associated with growth or that age is a confounding variable.

4 Complete Response

All parts essentially correct.

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct
OR
One part essentially correct and two parts partially correct

1 Minimal Response

One part essentially correct and either zero or one part partially correct
OR
No parts essentially correct and two parts partially correct

3. **Solution**

Part (a):
Bias may have been introduced by this convenience sample because the first people entering the library in the morning during the summer may differ in their opinions from those who use the library in the evenings or at other times. For example, many people going to the library during normal business hours during the summer may be parents with children who are out of school for summer break. These people would be more likely to support an increase in children's activities.

Instead of gathering data using a convenience sample, surveys should be sent to a simple random sample of adults holding library cards at this branch or a simple random sample of adults who have used this library at any time during the past year.

Part (b):
The wording of the question may be leading. Since the question says that other people support an increase in children's activities, respondents may be more likely to agree with the statement.

Since the question states that increasing children's activities may take resources away from other programs, some people may NOT support the increase in children's activities in order to maintain support of other programs they value.

A better way to word the question would be to simply ask, "Should the number of children's activities at this library be increased?"

Scoring
Each part can be essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct if the student indicates that
- choosing the first 20 people entering the library each July morning could introduce bias because those people's opinions might differ from those of other library users, and
- a random sample should be taken of library users (either an SRS or a systematic sample with a random starting point)

Part (a) is partially correct if the student correctly does one of the above.

Part (a) is incorrect if the student does neither.

Part (b) is essentially correct if the student
- points out at least one of the wording problems with the survey, and
- addresses that concern raised with a reasonable alternative wording.

Note: the student did not need to point out both wording problems. However, if they pointed out both problems, their alternative wording must correct both problems.

Part (b) is partially correct if the student identifies one or both of the wording problems, but does not propose a reasonable alternative to fix the question.
Part (b) is incorrect if the student points out both problems, but argues that one will increase the proportion in favor and one will decrease the proportion in favor, so that the two effects cancel each other out.

4 Complete Response

Both parts essentially correct.

3 Substantial Response

One part essentially correct and one part partially correct

2 Developing Response

One part essentially correct and one part incorrect
OR
Both parts partially correct

1 Minimal Response

One part partially correct