

AP Stats Chapter 5 Notes: Producing Data

Write a question that you think we could ask of students at ACHS. How could we go about answering that question?

Suppose we want to answer a question about all adults in the United States, such as, "What percent of American adults think the Congress is doing an adequate job?"

How could we go about getting an answer to this question?

Kentucky just finished an election recently. Each year after an election the Attorney General of the state must conduct post election audits. This audit must be conducted in no fewer than 5% of Kentucky's 120 counties. How do you think the Attorney General goes about selecting the counties to be audited?

Refreshers:

Observational Study—

Experiment—

Population—

Sample—

Sampling versus a Census:

Sampling—

Census—

Types of Samples:

Voluntary Response Sample—

Examples:

http://www.huffingtonpost.com/2010/10/29/shape-editor-leann-rimes-_n_775830.html

Convenience Sampling—

Examples:

Bias—

Both of the sampling methods above are bias. Why?

Additional Examples:

http://www.huffingtonpost.com/akmuckraker/new-alaska-poll---democra_b_775174.html

<http://www.afa.net/>

<http://www.afa.net/Blogs/BlogPost.aspx?id=2147499796>

Assignment: p. 333-334 5.1, 5.2, 5.4, 5.6, 5.7

Simple Random Samples:

In a voluntary response sample who chooses whom responds?

In a convenience sample who chooses whom responds?

What is the problem with each of these sampling techniques?

A statistician solves this problem by allowing chance to select the sample. Chance allows neither favoritism by the sampler nor self selection by the respondents. Choosing a sample by chance attacks bias by giving all individuals an equal chance to be chosen. Rich and poor, young and old, black and white, all have the same chance to be in the sample.

Think of a simple way to use chance to determine a simple random sample:

Simple Random Sample (SRS):

Not only does the SRS give each individual an equal chance to be chosen (avoids bias) but it also gives every possible sample an equal chance to be chosen.

Drawing names from a hat may not always be feasible so another option would be a random number generator.

Table B: Random number table—Each number in the table has an equal chance of being any of the digits 0 to 9 and the entries are independent of one another.

Using the random number table:

Joan's small accounting firm serves 30 business clients. Joan wants to interview a sample of 5 clients in detail to find ways to improve client satisfaction. To avoid bias, she chooses an SRS of size 5.

Label each client with as few digits as possible.

	A-1 Plumbing		JL Records
	Accent Printing		Johnson Commodities
	Action Sports Shop		Keiser Construction
	Anderson Construction		Lui's Chinese Restaurant
	Bailey Trucking		Magic Tan
	Balloon's Inc		Peerless Machine
	Bennett Hardware		Photo Arts
	Best's Camera Shop		River City Books
	Blue Print Specialist		Riverside Tavern
	Central Tree Service		Rustic Boutique
	Classic Flowers		Satellite Services
	Computer Answers		Scotch Wash
	Darlene's Dolls		Sewer's Center
	Fleisch Realty		Tire Specialties
	Hernandez Electronics		Von's Video Store

Enter table B at line 130 and read two digit groups. The first 10 two digit groups are:

Which values should we ignore? Why?

Which values are left?

Continue along line 130 and on to 131 if necessary to finish choosing 5 clients. Identify the sample of clients by their number and name.

Guidelines for using Table B:

a. You may assign labels in the most convenient manner, such as alphabetical for names. Be certain all labels have the same number of digits. Why?

b. Use the shortest possible labels: one digit for a population of up to 10 members, two digits for 11 to 100 members...

c. Begin with label 1 or 01 or 001 as needed.

d. You can read from Table B in any order—down a column, across a row, and so on—because the table has no order, but general practice suggests reading across a row and starting at a different row each time you use the table.

AP Test—You will have to use the table on the AP test so that all answers will match. If everyone uses a different random number generator it would be nearly impossible to check answers for correctness.

Other Sampling Methods

Probability Sample:

Some probability sampling methods give each member of the population the same chance of being selected. This may or may not be true of other techniques that are more elaborate. In every case, however, ***the use of chance to select the sample is the essential principle of statistical sampling.***

Stratified Random Sample:

Example:

Cluster Sampling:

Example:

Describe the difference between stratified random sampling and cluster sampling.

Multi Stage Sampling Design: This is exactly what it sounds like. There are several stages to the sample.

Example: Current Population Survey, an office of the US government, survey about 50000 households each month about unemployment and employment. It's not practical to maintain a list of all households in the US for an SRS, so the Current Population Survey uses a multistage sampling design.

Stage 1: They first divide the US into 2007 geographical areas called Primary Sampling Units (PSUs). Select a sample of 756 PSUs. This sample includes the 428 PSUs with the largest population and a stratified sample of the 328 others.

Stage 2: Divide each PSU selected into smaller areas called "neighborhoods." Stratify the neighborhoods using ethnic and other information, and take a stratified sample of the neighborhoods in each PSU.

Stage 3: Sort the housing units in each neighborhood into clusters of four nearby units. Interview the households in a random sample of these clusters.

<http://www.census.gov/cps/>

<http://www.census.gov/>

http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=ACS&_submenuId=datasets_2&_lang=en

Assignment: p. 341-343 5.9 to 5.14

Cautions about Sample Surveys

Random sampling attempts to eliminate bias in the choice of the sample from a list of the population. However, when we are dealing with people, accurate information can be difficult to obtain (populations change, people move, die, etc.) There are two main problems that exist with surveying human beings:

Undercoverage:

Nonresponse:

Undercoverage in a phone sample will often miss about 7-8% of the population that does not have a residential phone line. This creates a bias in the sample if the people not sampled have a differing opinion than those that are sampled. What type of population would be most likely not to have a phone line?

Do you think their opinions may differ from those that do have a phone line?

How have recent times changed phone samples?

Undercoverage is a problem, but nonresponse is a greater source of bias. Sample surveys can have as much as 30% nonresponse, even with careful planning and call backs. Urban areas often have a greater nonresponse rate than the rural areas. So to help control the bias, people in urban areas are often replaced by others in the area to control the bias. Even with precautions, nonresponse is an issue since people who are rarely at home or who choose not to respond may still have a different opinion than those that do respond.

When conducting surveys, there are several factors that can influence the outcomes. These factors combine to make up **response bias** in sample results. Respondents may lie. The type of person asking the questions (race or sex) can influence the responses. Questions that ask respondents to recall past events can often times elicit incorrect responses. People tend to “telescope” the event further into the future than it should be.

Have you visited the dentist in the last six months?

Example 5.10 Did you vote? Response bias:

One of the most frequently observed survey measurement errors is the over reporting of voting behavior. People know that they should have voted, so those who did not vote tend to save face by saying that they did. Here are the data from a typical sample of 663 people after an election:

		What they Said	
		I voted	I didn't
What they Did	Voted	358	13
	Didn't vote	120	172

You can see that 478 people (72%) said they voted, but only 371 people (56%) actually did vote.

The **wording of questions** is the most important influence on the answers given to a sample survey. Confusing or leading questions can introduce strong bias, and even minor wording changes can alter a survey's outcome.

Example 5.11 Should we ban disposable diapers? Wording of questions

A survey paid for by makers of disposable diapers found that 84% of the sample opposed banning disposable diapers. Here is the actual question:

It is estimated that disposable diapers account for less than 2% of the trash in today's landfills. In contrast, beverage containers, third class mail and yard wastes are estimated to account for about 21% of the trash in landfills. Given this, in your opinion, would it be fair to ban disposable diapers?

This question gives information on only one side of an issue, then asks an opinion. That's a sure way to bias the responses. A different question that described how long disposable diapers take to decay and how many tons they contribute to landfills each year would draw a quite different response.

Example 5.12 Doubting the Holocaust? Wording of questions

An opinion poll conducted in 1992 for the American Jewish Committee asked: “Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?” When 22% of the sample said “possible,” the news media wondered how so many Americans could be uncertain that the Holocaust happened. Then a second poll asked the question in a different way. “Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?” Now only 1% of the sample said “possible.” The complicated wording of the first question confused many respondents.

Inference about the Population

Samples are just that, samples. They can give us an idea about the population, but they will be different from the results of a survey of the entire population. If we chose two samples at random from the same population, we will draw different individuals. So the sample results will almost certainly differ somewhat. **We can improve our results by knowing that larger random samples give more accurate information than smaller samples.**

Assignment: p. 347-348 5.15 to 5.20 Read the Section 5.1 Summary on pages 348 and 349 and be sure you understand all of those concepts

5.2 Designing Experiments

How would you develop an experiment to determine if people can tell the difference between Pepsi and Coke?

A study is an experiment when we actually do something to individuals in order to observe the response.

Experimental units:

Subjects:

Treatment:

Because the purpose of an experiment is to reveal the response of one variable in changes in other variables, the distinction between explanatory and response variables is important.

Factors:

Control

Laboratory experiments in science and engineering often have a simple design with only a single treatment, which is applied to all of the experimental units. The design of such an experiment can be outlined as

Treatment→Observed response

For example, we may place a heavy object on a support (treatment) and measure how much it bends (observation). We rely on the controlled environment of the laboratory to protect us from lurking variables. When experiments are conducted in the field or with living subjects, such simple designs can yield invalid data. That is, we cannot tell whether the response was due to the treatment or to the lurking variables.

Example 5.15: Treating ulcers placebo effect

Gastric freezing is a clever treatment for ulcers in the upper intestine. The patient swallows a deflated balloon with tubes attached, then refrigerated liquid is pumped through the balloon for an hour. The idea is that cooling the stomach will reduce its production of acid and so relieve ulcers. An experiment reported in the *Journal of American Medical Association* showed that gastric freezing did reduce acid production and relieve ulcer pain. The treatment was safe and easy and was widely used for several years. The design of the experiment was

Gastric freezing→Observe pain relief

The gastric freezing experiment was poorly designed. The patients' response may have been due to placebo effect. A placebo is a dummy treatment. Many patients responds favorably to any treatment, even placebo. This may be due to the trust in the doctor and expectations of a cure or simply to the fact that medical conditions often improve without treatment. The response to a dummy treatment is the placebo effect.

A later experiment divided ulcer patients into two groups. One group was treated by gastric freezing as before. The other group received a placebo treatment in which the liquid in the balloon was at body temperature rather than freezing. The results: 34% of the 82 patients in the treatment group improved, but so did 38% of the 78 patients in the placebo group. This and other properly designed experiments showed that gastric freezing was no better than a placebo, and thus its use was abandoned.

The first experiment results were misleading because the effects of the explanatory variable were confounded with the placebo effect. How can we overcome confounding?

Control Group:

Control:

Assignment: p. 357-358 5.33-5.38

Replication

In order to be sure that the difference between two groups is not simply a matter of chance, but an actual difference we use replication.

Replication:

Randomization:

It can be argued that randomization is the most important principle of experimental design since it is what allows us to assert that treatment groups are essentially similar—that there is no systematic difference between them before treatments are administered.

Two examples of randomization and replication:

1. Does talking on a hands free cell phone distract drivers? 20 students (control group) simply drove in a simulator. Another 20 (experimental group) talked on the phone while driving.

There is a single factor in this experiment, cell phone use, with two levels. The researchers placed the names of the 40 students into a hat and drew out 20 for the experiment group and the remaining 20 make up the control group. This is a completely unbiased way to decide the two groups.

2. Does regularly taking aspirin help protect people against heart attacks? The Physicians' Health Study looked at this and the effects of beta-carotene. The body converts beta-carotene into Vitamin A which may help reduce the risks of some cancer. The *subjects* were 21, 996 male physicians. There were two factors in this experiment, each having two levels: aspirin (yes or no) and beta-carotene (yes or no). The different levels of these factors form the four combinations of treatments. One fourth of the subjects were assigned each of these treatments.

Factor 2: Beta-Carotene

Factor 1: Aspirin	Yes		No	
	Yes			
	No			
	Aspirin	Beta-Carotene	Aspirin	Placebo
	Placebo	Beta-Carotene	Placebo	Placebo

On the odd numbered days the subjects either took the placebo or the aspirin, and on even numbered days they took the placebo or the beta-carotene. There were several response variables—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 that beta-carotene had any effect.

Randomized Comparative Experiments

The logic behind the randomized comparative design is:

- Randomization produces two groups of subjects that we expect to be similar in all respects before the treatments are applied.
- Comparative design helps ensure that influences other than the cell phone operate equally on both groups.
- Therefore differences in average break reaction time must be due either to talking on the cell phone or to the play of chance in the random assignment of subjects to the two groups.

Principles of Experimental Design

1. Control the effects of lurking variables on the response, most simply by comparing two or more treatments.
2. Replicate each treatment on many units to reduce chance variation in results.
3. Randomize—use impersonal chance to assign experimental units to treatments.

Our hope is to see a difference in the responses so large that it is unlikely to happen just because of chance variation. We can use the laws of probability, which give a mathematical description of chance behavior, to learn if the differences in treatment effects are larger than we would expect to see if only chance were operating. If they are, we call them **statistically significant**. An observed effect so large that it would rarely occur by chance is called statistically significant.

When all experimental units are allocated at random among all treatments, the experiment is said to have a **completely randomized design**. Both the cell phone and aspirin experiments were completely randomized designs.

TV Commercial: Completely randomized design

The figure below displays six treatments formed by the two factors in an experiment on response to a TV commercial. Suppose we have 150 students who are will to serve as subjects. We must assign 25 students to each group.

To carry out the random assignment, label the 150 students 001 to 150. Enter table B and read three digit groups until you have selected 25 students to receive Treatment 1 (a 30 second ad shown once). Continue in table B to select 25 more students for Treatment 2 and so on until you have assigned 125 of the students to Treatment 1 through Treatment 5. The 25 students who remain get Treatment 6. The randomization is straight forward but very tedious to do by hand.

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 systematically affect the response to the treatment. In a block design, the random assignment of units to treatments is carried out separately within each block.

Block designs can have blocks of any size. A block design combines the idea of creating treatment groups as similar as possible with the principle of forming treatment groups at random. Blocks are another form of control. They control the effects of some outside variables by bringing those variables into the experiment from the blocks. Here are some typical examples of block designs.

1. The progress of a type of cancer differs in women and men. A clinical experiment to compare three therapies for this cancer therefore treats gender as a blocking variable. Two separate randomizations are done, one assigning the female subjects to the treatment and the other assigning the male subjects. Note that there is no randomization involved in making up the blocks. They are groups of subjects who differ in some way (gender in this case) that is apparent before the experiment begins.

2. The soil type and fertility of farmland differ by location. Because of this, a test of the effect of tillage type (two types) and pesticide application (three application schedules) on soybean yields uses small fields as blocks. Each block is divided into six plots, and the six treatments are randomly assigned to plots separately within each block.

3. The Tennessee STAR class size experiment used a block design. It was important to compare different class types in the same school because the children in a school come from the same neighborhood, follow the same curriculum, and have the same school environment outside class. In all, 79 schools across Tennessee participated in the program. That is, there were 79 blocks. New kindergarten students were randomly placed into the three types of class separately within each school.

Blocks allow us to draw separate conclusions about each block. Blocking also allows more precise overall conclusions. Randomization will then average out the effects of the remaining variation and allow an unbiased comparison of the treatments. *Control what you can, block on what you can't control, and randomize the rest.*

Matched Pairs Designs

The simplest use of blocking is a matched pairs design, which compares just two treatments. The subjects are matched in pairs. For example, an experiment to compare two advertisements for the same product might use pairs of subjects with the same age, sex, and income. The idea is that matched subjects are more similar than unmatched subjects, so comparing the responses within a number of pairs is more efficient than comparing the responses of groups of randomly assigned subjects.

Example 5.24 Matched Pairs Design: Cereal Leaf Beetles

Are cereal leaf beetles more strongly scattered by the color yellow or by the color green? Agriculture researchers want to know because they detect the presence of the pests in farm fields by mounting sticky boards to trap insects that land on them. The board color should attract beetles as strongly as possible. We must design an experiment to compare yellow and green mounting boards on poles in a large field of oats.

The experimental units are locations within the field far enough apart to represent independent observations. We erect a pole at each location to hold the boards. We might employ a completely randomized designing which we randomly select half the poles to receive a yellow board while the remaining poles receive green. The locations vary widely in the number of beetles present, however. For example, the alfalfa will have extra beetles. This variation among experimental units can hide the systemic effect of the board color.

It is more efficient to use a matched pairs design in which we mount boards of both colors on each pole. The observations (number of beetles trapped) are matched in pairs from the same poles. We compare the number trapped by the green board to the number trapped by the yellow board on the same pole. Because the boards are mounted one above the other, we select the color of the top board at random. Just toss a coin for each board—heads, yellow board is mounted above the green.

Matched pairs designs compare just two treatments. We chose blocks of two units that are as closely matched as possible. The two boards on the same pole form a block. We assign the treatment by tossing a coin. Alternatively, each block in a matched pairs design may consist of one subject, who gets both treatments one after the other. Each subject serves as his or her own control. *Matched pairs are an example of block design.*

Double Blind Experiment

There are many limitations to experiments. For example, the environment of an experiment can limit or influence the results in unpredictable ways. So before experiments are believed, they are often performed many times in many different environments.

Another problem is a lack of realism. The subjects or environment may not actually mirror the conditions we really want to study.

Example 1: A study of marijuana effects. A group of marijuana users was recruited for the study. Some were randomly assigned to smoke marijuana, others were given a placebo cigarette. The control group recognized they did not have actual marijuana and complained. The blindness of the experiment failed because the subjects could tell what treatment they were receiving.

Example 2: 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. Many behavioral science experiments use as subjects students who know they are subjects in an experiment. That is not a realistic setting.

This lack of realism can make it difficult to apply the conclusions of an experiment to the settings of greater interest.

Assignment: p. 371-373—5.45, 5.46, 5.48, 5.49 Read through the section 5.2 summary
