Chapter 1

Designing studies

In this chapter we will learn two ways to acquire data: sampling and experiments. Both of them should be designed carefully to make sure the conclusions derived from them are persuasive.

1.1 Sampling

Basic concepts and ideas

Sampling is a kind of observational study. It doesn't interfere the subjects to be studied. For example, you want to know the average household income in China, you can sample 10,000 out of all the households and make a record of their yearly income. During this process, you just "record" without influencing their income.

All the households in China is called the **population**, and the 10,000 household sampled out is called a **sample**. If you recorded the data of all the households in China, then this process is called **census**.

The reason of sampling is to draw conclusions about the population, thus the sample should be representative. The first step in planning a **sampling survey** is to say exactly *what population* we want to describe. The second step is to say exactly *what we want to measure*, that is, to point out the variable we want to measure.

Bias and sources of bias

For a bad sample, it consistently overestimates or underestimates the value, and this phenomenon is called **bias**.

Good samples sometimes overestimate and sometimes underestimate the value of the variable to be estimated, and this variability is called **sampling errors**. But they never "consistently" overestimate or underestimate.

There are many sources of bias. Although each of the following sources of bias are defined separately, there is overlap, and many if not most examples of bias involve more than one of the following.

1. **Convenience sample** is a sample generated by choosing easy-to-reach individuals from a population. For example, to estimate the percentage of high school students who have graphing calculators, you draw a sample from our statistics class for convenience. Obviously, this sample is biased and it tends to overestimate the percentage of students with calculators.

2. **Voluntary response bias** is generated by letting the individuals participate the survey voluntarily. Often those with strong opinions are more likely to participate, leading to an over emphasis of those individuals. For example, a radio call-in program about whether should each one get a house for free.

3. **Undercoverage bias** happens when some individuals do not have equal chance to be selected. Telephone survey will ignore those who do not have telephones.

4. **Nonresponse bias** happens when some of the chosen individuals refuse to respond. Those unresponsive individuals may hold some particular opinions and they are ignored.

5. **Response bias** happens when participants give unfaithful answers. For example, no participants will say they are a criminal in a survey to estimate the percentage of criminals in the population.

6. **Wording bias** is caused by poorly worded question. Here is an interesting example.

   > It is estimated that disposable diapers account for less than 2% of the trash in todays 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?

   Do you think it has a wording bias?

In AP exam, if asked to describe how the design of a study leads to bias, you should **(1) identify the bias and (2) explain how the design leads to underestimation or overestimation**. For a design as "Draw a sample from your AP statistics class to estimate the percentage of students with calculators", you may say "It is an convenience sample and it tend to overestimate the percentage, because all the students in the AP statistics class are required to have a graphing calculator."

Good samples

1. **Simple Random Sample(SRS)** of size n is chosen in such a way that every group of n individuals in the population has an equal chance to be selected as the sample.

   Steps:

   (1) **Label.** Give each individual in the population a distinct number.
   (2) **Randomize.** Use random number number generator or *table of random digits* to generate n distinct numbers
   (3) **Choose.** Put the individuals corresponding to the selected number in the previous in the sample

2. **Systematic sample.** For example, you want sample 10 individuals out of a population of 100 systematically. The procedures will be:

   (1) Label the individuals in the population from 1 to 100.
   (2) Randomly choose a number from 1 to 10, let's say it is 6.
   (3) Pick every tenth number following 6. They are 16, 26, ... , 96
   (4) Select the individuals corresponding to the above numbers as a systematic sample.

3. **Stratified Random Sample** is generated first by dividing the population into *homogeneous* groups called *strata*, selecting SRS from each stratum, then combining all those SRS together.

   For example, you want investigate the hair length of all the students in a high school, you can divide the students into two strata by gender, since gender strongly influence the hair length, and the hair length in each group can be considered *homogeneous*. Then draw SRS from each group, and combine those two SRS together to form a stratified random sample.

Is *stratified random sample* simple random sample?

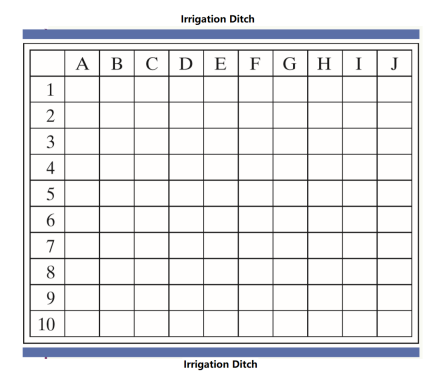
Is *cluster sample* simple random sample?

4. **Cluster sample** is generated by dividing the population into *heterogeneous* groups called *clusters*, then randomly select clusters from among all clusters.

   In previous example, instead of dividing the students by gender, we divide the students into classes. The hair length in each class is *heterogeneous* and can be considered as a miniature of the population. Each class is called a cluster. We randomly choose classes from among all the classes and combine the chosen classes to form a cluster sample.

Exercise

A British farmer grows sunflowers for making sunflower oil. Her field is arranged in a grid pattern, with 10 rows and 10 columns as shown in the figure on the next page. Irrigation ditches run along the top and bottom of the field. The farmer would like to estimate the number of healthy plants in the field so she can project how much money shell make from selling them. It would take too much time to count the plants in all 100 squares, so shell accept an estimate based on a sample of 10 squares.



Give the sampling method for

(a) simple random sample(SRS)

(b) systematic sample

(c) cluster sample

(d) stratified random sample

(e) In (c) why do you cluster the population that way?

(f) In (d) why do you stratify the population that way?

The advantages of stratified random sample

Let's use the previous exercise as an example. If we draw SRS many times, some of the samples may fall in the middle of the field(circle **L**), which is far away from the irrigation ditches, and consequently gives lower estimations of the production of sunflower seeds. Some samples may fall near the irrigation ditch(**circle H**) and give higher estimations of the production of sunflower seeds. Figure 1.1 gives the distribution of 100 samples of SRS and stratified random samples. Each dot represents a sample.

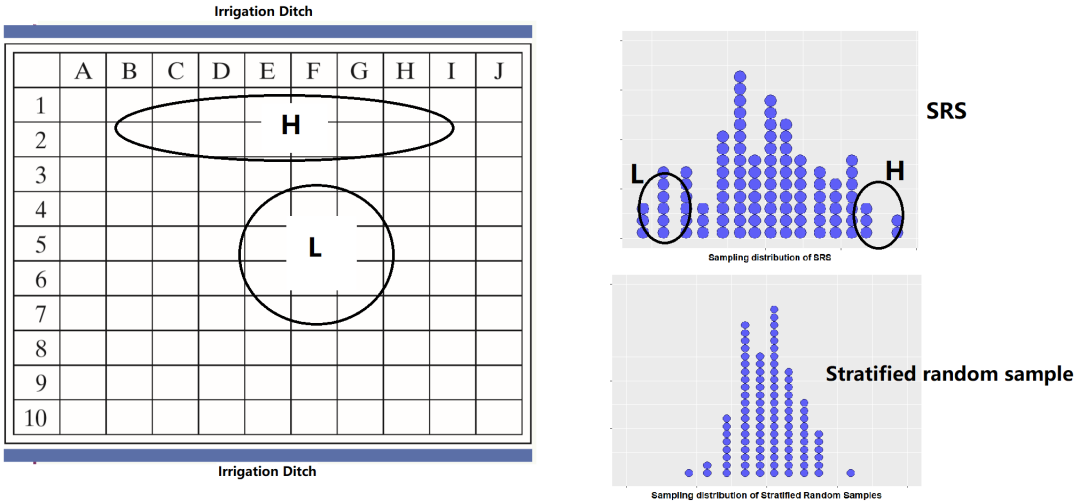


Figure 1.1: Compare SRS and stratified random sample

However, the extreme cases as circle **L** and circle **H** never happen for stratified random samples. The estimations based on stratified random samples have less variation than estimations based on simple random samples.

Stratified random sample can reduce the variation of the estimations than SRS.

1.2 Experiments

In an **observational study**, we can only come to the conclusion about *association*, not *cause and effect* or *causation* relation. For example, every day you get up, you see the rises. There is an *association* between "you get up" and the "sun rises", but you can not say the "sun rises" because "you get up".

The reason why we can not come to the conclusion of *cause and effect* is because we can not rule out the influence of other factors, such as the rotation of the earth. If you make another observational study, you may find the "rotation of the earth" and the "sun rises" are closely associated as well. When we can not distinguish the effect of more than one variable on a response variable, we say there is a **confounding**.

The observational studies can not come to conclusions about causations because of confounding. **Experiments can exclude confounding effect and give conclusions about cause and effect.**

Principles of Experimental design

Now, we developed a new painkiller. We want to do some experiments to test whether it is more effective than the existing one.

(1) **Comparison.** We have to compare the effectiveness of the new painkiller with with the existing one.

(2) **Random assignment.** We have to randomly assign the **subjects** into different **treatments:** one treatment is to take the new painkiller, the other treatment is to take the existing painkiller.

The purpose of *random assignment* is to make the subjects allocated to different treatments are roughly the same, and thus rule out *confounding*.

For example, if we allocate those with heavy body weight to the treatment of taking new painkiller, and the others to take the existing painkiller, we don't know whether the relief of pain is caused by body weight or the effectiveness of the new drug, and lead to *confounding*. Beside the observable factors, such as body weight, there are unobservable factors that may influence the outcome as well. Random assignment can also even out those unobservable factors.

(3) **Control** is to keep other variables that might affect the response the same. *Random assignment* is a way to control. Another way to control is **blocking**. A **control** group receives no treatments while all other variables are the same as the groups that receive treatments

(4) **Blocking** is to divide the subjects into different groups by the factors that may strongly influence the outcome, and design experiments within each group. Those different groups are called **blocks**.

(5) **Replication** is to use enough **experimental units** in each group so that any difference in the effects of the treatments can be distinguished from chance variance between groups.

For example, we only have two subjects, and each one is randomly assigned to take either the new or existing painkiller, and it happens that the new painkiller is more effective. People may argue that the one assigned to take the new painkiller happens to be more resistant to pains and the effectiveness if not caused by the new drug. In this case, we can not rule out the influence of chance variance.

Design experiments

(1) Completely randomized design

For example we have 80 subjects to test the effectiveness of the new painkiller. We can design the experiments in the following way.

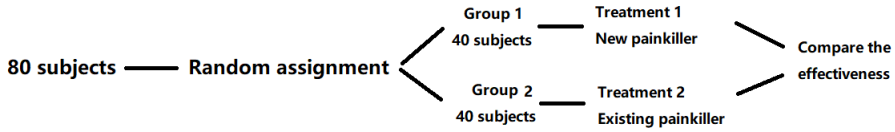


Figure 1.2: Completely randomized design

The procedures are given as follows:

(a) Assign numbers from 1 to 80 to those 80 subjects.

(b) Use random number generator to generate 40 distinct random numbers within 1 and 80

(c) The subjects corresponding to the 40 numbers in the above step go to group 1 and will take new painkillers. The other 40 subjects go to group 2 and will take existing painkiller.

(d) Compare the effectiveness between the two groups.

I. In AP exam, when you are asked to design an experiment, drawing the graph is not enough, you have to describe in detail how to carry out those procedures. Make sure people can carry out the experiment by following your design. You can not just say "random assignment", you have to describe how to do this "random assignment"

II. Here I say compare the effectiveness. It is not good. In AP example you have to describe exactly what variable to be measured. In this example, it could be like "After they take the painkillers, punch them with equal strength, and measure the loudness of their cry and lower loudness means the medicine is more effective.

(2) Randomized block design

Now we have 80 subjects, 40 of the are male and 40 of them are female. And we know gender can strongly influence the effectiveness of the medicine. We can do a *randomized block design* to rule out the confounding of gender in the following way.

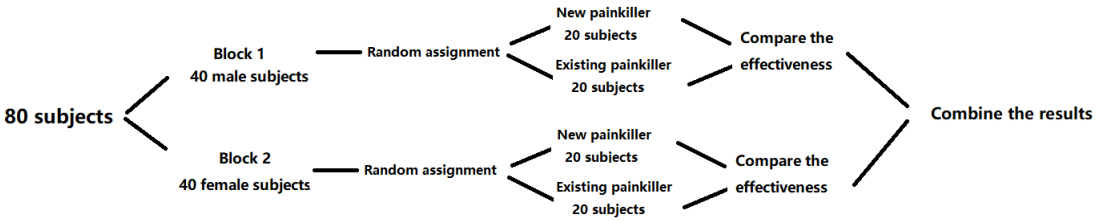


Figure 1.3: Randomized block design

Exercise
Describe in detail of the above *randomized block design*.

(3) **Matched pairs design**

Matched pairs design is a special type of randomized block design. In each block there are either two *very similar* individuals or one individual with two measurements.

One individual with two measurements

Now let's do a *matched pairs design* to study whether the standing pulse rates generally higher than sitting pulse rates? Let's say the number of subjects is n.

(a) **Let each subject randomly decide which treatment comes first and apply the treatment**.

For example, by throwing a coin, if it is a head, then measure the standing heart rates first, otherwise measure the sitting heart rates first.

(b) **Apply the other treatment that is not applied on each subject.**

In this example, for those who measured standing heart rates first, measure sitting heart rates this time. And for those who measured sitting heart rates first, measure the standing heart rate this time.

(c) **Compare each pair of the data acquired in previous steps(*compare within block).***

Let $\{x_1, x_2, ..., x_n\}$ be the standing heart rates of the n subjects, and $\{y_1, y_2, ..., y_n\}$ be the corresponding sitting heart rates. We compare each pair of data by subtract y_i from x_i, then we get a new set of data $\{x_1 - y_1, x_2 - y_2, ..., x_n - y_n\}$.

(d) **Analyze the data generated from the previous step(*combine the results from each block*).**

We analyze data set $\{x_1 - y_1, x_2 - y_2, ..., x_n - y_n\}$ to see whether the standing heart rates is higher than the sitting heart rate(*We will learn how to analyze those data in chapter about statistical inferences*).

Remark:

I. Randomizing the order of the treatments is to eliminate the confounding that the order of the treatment may influence the response variable.

II. When you are asked to describe a matched pairs design, you are supposed to give details. The procedures given above are too general.

Two individuals in each block

Exercise

Nitrogen in tires—a lot of hot air?
Most automobile tires are inflated with compressed air, which consists of about 78% nitrogen. Aircraft tires are filled with pure nitrogen, which is safer than air in case of fire. Could filling automobile tires with nitrogen improve safety, performance, or both?
Consumers Union designed a study to test whether nitrogen-filled tires would maintain pressure better than air-filled tires. They obtained two tires from each of several brands and then filled one tire in each pair with air and one with nitrogen. All tires were inflated to a pressure of 30 pounds per square inch and then placed outside for a year. At the end of the year, Consumers Union measured the pressure in each tire. The amount of pressure lost (in pounds per square inch) during the year for the air-filled and nitrogen-filled tires of each brand is shown in the table below.

Brand	Air	Nitrogen	Brand	Air	Nitrogen
BF Goodrich Traction T/A HR	7.6	7.2	Pirelli P6 Four Seasons	4.4	4.2
Bridgestone HP50 (Sears)	3.8	2.5	Sumitomo HTR H4	1.4	2.1
Bridgestone Potenza G009	3.7	1.6	Yokohama Avid H4S	4.3	3.0
Bridgestone Potenza RE950	4.7	1.5	BF Goodrich Traction T/A V	5.5	3.4
Bridgestone Potenza EL400	2.1	1.0	Bridgestone Potenza RE950	4.1	2.8
Continental Premier Contact H	4.9	3.1	Continental ContiExtreme Contact	5.0	3.4
Cooper Lifeliner Touring SLE	5.2	3.5	Continental ContiProContact	4.8	3.3
Dayton Daytona HR	3.4	3.2	Cooper Lifeliner Touring SLE	3.2	2.5
Falken Ziex ZE-512	4.1	3.3	General Exclaim UHP	6.8	2.7
Fuzion Hrl	2.7	2.2	Hankook Ventus V4 H105	3.1	1.4
General Exclaim	3.1	3.4	Michelin Energy MXV4 Plus	2.5	1.5
Goodyear Assurance Tripletred	3.8	3.2	Michelin Pilot Exalto A/S	6.6	2.2
Hankook Optimo H418	3.0	0.9	Michelin Pilot HX MXM4	2.2	2.0
Kumho Solus KH16	6.2	3.4	Pirelli P6 Four Seasons	2.5	2.7
Michelin Energy MXV4 Plus	2.0	1.8	Sumitomo HTR$^+$	4.4	3.7
Michelin Pilot XGT H4	1.1	0.7			

Figure 1.4: Matched pairs design

Does filling tires with nitrogen instead of compressed air reduce pressure loss? Give appropriate graphical and numerical evidence to support your answer.

(4) **Double blind**

The **Rosenthal effect** and **Self-fulfilling prophecy** tell us that what you think may influence the outcome. This phenomenon happens in experiments when the subjects are people.

Give a patient a sugar and tell him this is a painkiller, he may really feel that his pain is relieved after taking it. This is called **placebo effect**, and the sugar is a **placebo**.

Double-blind experiment can exclude *placebo effect*. In a *double-blind experiment* neither the subjects nor those who interact with them and measure the response variable know which treatment a subject received.

Who knows which treatment a subject received? People who design the experiment knows! Once a person knows which treatment a subject received, he should not participate in the procedure of measuring the response variable.

For example, there is a new drug to decrease blood pressure. To test whether it works or not we can do a *double-blind experiment*. Randomly divide the subjects into two groups, one group is treated with real drug, the other is treated with placebo. But all are told they are treated with drug. After they have taken the drug or placebo for a period of time, the blood pressure is measured by those who don't know whether the subject was fed with drug or placebos.

Sometimes double-blind experiment is impossible, we have to do single blind experiment.

For example we want to know whether diet or exercise is more effective in losing weight. No matter how the experiment is designed, the subjects will know what type of treatments they receive. But the people who measure the subjects' weight can be blind about the treatment each subject received. This type of experiment is called **single-blind experiment**.

(5) **Conclusions and their generalization**

In order to draw conclusions, an observed effect need to be **statistically significant**, which means that the effect can rarely be explained by chance effect.

Once the effect is statistically significant, what how far can we go about the inferences of the populations? What type of inference can we make?

Previously, we mentioned that observational study can not come to conclusions about *cause and effect* relation because of confounding. When the experiment is designed without *random assignment*, confounding can not be ruled out either, and *cause and effect* relation can not be reached.

When subjects are not *randomly selected*, they are not representative of the population, the conclusion drawn from those subjects can not be generalized to the population.

The following table gives the scopes of inference on different occasions.

		Were individuals randomly assigned to groups?	
		Yes	**No**
Were individuals randomly selected?	**Yes**	Inference about the population: YES Inference about cause and effect: YES	Inference about the population: YES Inference about cause and effect: NO
	No	Inference about the population: NO Inference about cause and effect: YES	Inference about the population: NO Inference about cause and effect: NO

Vitamin C and Canker Sores

A small-town dentist wants to know if a daily dose of 500 milligrams (mg) of vitamin C will result in fewer canker sores in the mouth than taking no vitamin C.

The dentist is considering the following four study designs:

Design 1: Get all dental patients in town with appointments in the next two weeks to take part in a study. Give each patient a survey with two questions: (1) Do you take at least 500 mg of vitamin C each day? (2) Do you frequently have canker sores? Based on patients answers to Question 1, divide them into two groups: those who take at least 500 mg of vitamin C daily and those who dont.

Design 2: Get all dental patients in town with appointments in the next two weeks to take part in a study. Randomly assign half of them to take 500 mg of vitamin C each day and the other half to abstain from taking vitamin C for three months.

Design 3: Select a random sample of dental patients in town and get them to take part in a study. Divide the patients into two groups as in Design 1.

Design 4: Select a random sample of dental patients in town and get them to take part in a study. Randomly assign half of them to take 500 mg of vitamin C each day and the other half to abstain from taking vitamin C for three months.

For whichever design the dentist chooses, suppose she compares the proportion of patients in each group who complain of canker sores. Also suppose that she finds a statistically significant difference, with a smaller proportion of those taking vitamin C having canker sores.

What can the dentist conclude for each design?