Chapter 4: Designing Studies

Objectives: Students will:
- Distinguish between, and discuss the advantages of, observational studies and experiments.
- Identify and give examples of different types of sampling methods, including a clear definition of a simple random sample.
- Identify and give examples of sources of bias in sample surveys.
- Identify and explain the three basic principles of experimental design.
- Explain what is meant by a complete randomized design.
- Distinguish between the purposes of randomization and blocking in an experimental design.
- Use random numbers from a table or technology to select a random sample.

AP Outline Fit:

II. Sampling and Experimentation: Planning and conducting a study (10%–15%)
- A. Overview of methods of data collection
  1. Census
  2. Sample survey
  3. Experiment
  4. Observational study
- B. Planning and conducting surveys
  1. Characteristics of a well-designed and well-conducted survey
  2. Populations, samples, and random selection
  3. Sources of bias in sampling and surveys
  4. Sampling methods, including simple random sampling, stratified random sampling, and cluster sampling
- C. Planning and conducting experiments
  1. Characteristics of a well-designed and well-conducted experiment
  2. Treatments, control groups, experimental units, random assignments, and replication
  3. Sources of bias and confounding, including placebo effect and blinding
  4. Completely randomized design
  5. Randomized block design, including matched pairs design

What you will learn:

A. SAMPLING
  1. Identify the population in a sampling situation.
  2. Recognize bias due to voluntary response sampling and other inferior sampling methods.
  3. Select a simple random sample (SRS) from a population.
  4. Recognize cluster sampling and how it differs from other sampling methods.
  5. Recognize the presence of undercoverage and nonresponse as sources of error in a sample survey. Recognize the effect of the wording of questions on the response.
  6. Use random digits to select a stratified random sample from a population when the strata are identified.

B. EXPERIMENTS
  1. Recognize whether a study is an observational study or an experiment.
  2. Recognize bias due to confounding of explanatory variables with lurking variables in either an observational study or an experiment.
  3. Identify the factors (explanatory variables), treatments, response variables, and experimental units or subjects in an experiment.
  4. Outline the design of a completely randomized experiment using a diagram like those in Examples 5.17 (page 360) and 5.19 (page 362). The diagram in a specific case should show the sizes of the groups, the specific treatments, and the response variable(s).
  5. Carry out the random assignment of subjects to groups in a completely randomized experiment.
  6. Recognize the placebo effect. Recognize when the double-blind technique should be used.
  7. Recognize a block design and when it would be appropriate. Know when a matched pairs design would be appropriate and how to design a matched pairs experiment.
  8. Explain why a randomized comparative experiment can give good evidence for cause-and-effect relationships.
Knowledge Objectives: Students will:
- Explain the difference between an observational study and an experiment.
- Give some examples of studies, and classify each as either an observational study or an experiment.

Vocabulary:
- Statistics – science of collecting, organizing, summarizing and analyzing information to draw conclusions or answer questions
- Information – data
- Data – fact or propositions used to draw a conclusion or make a decision
- Anecdotal – data based on casual observation, not scientific research
- Descriptive statistics – organizing and summarizing the information collected
- Inferential statistics – methods that take results obtained from a sample, extends them to the population, and measures the reliability of the results
- Population – the entire collection of individuals
- Sample – subset of population (used in the study)
- Placebo – innocuous drug such as a sugar tablet
- Experimental group – group receiving item being studied
- Control group – group receiving the placebo
- Double-blind – group receiving the placebo
- Variables – characteristics of individuals within the population

Key Concepts:

**Observational Studies**
- Draws inferences about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the control of the investigator
- Observational studies infer possible causes (basis of future experiments), but cannot determine cause and effect
- Experiments may be beyond the control of the investigator for a variety of reasons:
  - A randomized experiment would violate ethical standards
  - The investigator may simply lack the requisite influence
  - A randomized experiment may be impractical

**Experiments**
- Method of investigating causal relationships (cause and effect) among variables, or to test a hypothesis
- Often the experimenter is interested in the effect of some process or intervention (the “treatment”) on some objects (the “experimental units”), which may be people, parts of people, groups of people, plants, animals, etc.
- A controlled experiment generally compares the results obtained from an experimental sample against a control sample, which is practically identical to the experimental sample except for the one aspect whose effect is being tested (the independent variable).

**Observational Study versus Experiment**

<table>
<thead>
<tr>
<th>In an observational study, we observe individuals and measure variables of interest but do not attempt to influence the responses.</th>
</tr>
</thead>
<tbody>
<tr>
<td>In an experiment, we deliberately impose some treatment on (that is, do something to) individuals in order to observe their responses.</td>
</tr>
</tbody>
</table>

**Homework:** none
### Analyzing Experiments Template

<table>
<thead>
<tr>
<th><strong>Topic</strong></th>
<th><strong>Answers</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question:</td>
<td>What is the question the researchers are trying to answer?</td>
</tr>
<tr>
<td>Subjects / Experimental Units:</td>
<td>What are the experimental units?</td>
</tr>
<tr>
<td>Explanatory Variable(s) / Factor(s):</td>
<td>Type of variable: Quantitative or Categorical</td>
</tr>
<tr>
<td>Treatment(s):</td>
<td>What are the Factor(s) and their Levels?</td>
</tr>
<tr>
<td>Response Variable(s):</td>
<td>Type of variable: Quantitative or Categorical</td>
</tr>
<tr>
<td>Experimental Design Description:</td>
<td>Using words or diagrams describe the experimental design</td>
</tr>
<tr>
<td>Experimental Design Principles:</td>
<td>Explain how these design principles apply in this study</td>
</tr>
<tr>
<td>Control:</td>
<td>Eliminate confounding effects of extraneous variables</td>
</tr>
<tr>
<td>Randomization:</td>
<td>No systematic difference between the groups</td>
</tr>
<tr>
<td>Replication:</td>
<td>Reducing role of chance in results</td>
</tr>
<tr>
<td>Blocking:</td>
<td>Is blocking used? If so describe the blocking and why it was used.</td>
</tr>
<tr>
<td>Blinding:</td>
<td>Is blinding used? If so describe the blinding in context.</td>
</tr>
<tr>
<td>Concerns:</td>
<td>What concerns do you have about the experimental design?</td>
</tr>
<tr>
<td>Statistical Analysis Technique(s):</td>
<td>What statistical analysis techniques are appropriate?</td>
</tr>
<tr>
<td>Conclusions:</td>
<td>What conclusions can be drawn from the study?</td>
</tr>
</tbody>
</table>
Chapter 4: Designing Studies

Section 4.1: Samples and Surveys

Objectives: Students will:
- IDENTIFY the population and sample in a sample survey
- IDENTIFY voluntary response samples and convenience samples
- DESCRIBE how to use a table of random digits to select a simple random sample (SRS)
- DESCRIBE simple random samples, stratified random samples, and cluster samples
- EXPLAIN how undercoverage, nonresponse, and question wording can lead to bias in a sample survey

Vocabulary:
Population – the entire group of individuals that we want information about
Sample – a part of the population that is actually examined to gather the information
Sampling – the process of studying a part in order to gain information about the whole
Census – an attempt to contact every individual in the entire population
Voluntary response sample – consists of people who choose themselves by responding to a general appeal
Convenience sampling – contacting those individuals who are easiest to reach
Bias – the systematic favoring of a particular outcome
Simple Random Sample (SRS) – consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected
Random digits – A uniformly distributed random number set from 0 to 9 (from calculator or from Table B)
Probability sample – a sample chosen by chance
Stratum – a group of individuals, out of a population, that is similar in some way that is important to the response
Strata – plural of stratum
Stratified random sample – consists of one SRS for each stratum in the population
Cluster – a group within a population
Cluster sampling – a sampling method in which some clusters within a population are randomly selected, then all individuals within each chosen cluster are selected to be included in the sample
Response bias – survey responders for a variety of reasons do not relate accurate information (they lie!)
Undercoverage – occurs when some groups within the population are left out of the process of choosing the sample
Nonresponse – occurs when an individual chosen for the sample cannot be contacted or does not cooperate

Key Concepts:
Sampling Methods:
- Simple random sampling (SRS)
  - Everyone has an equal chance at selection
- Stratified sampling
  - Some of all
- Cluster sampling
  - All of some
- Systematic sampling
  - Using an algorithm to determine who to sample
- Multi-stage sampling
  - Dividing the sampling into stages
  - Perhaps using different techniques at different stages

- Sampling errors gives incomplete information about the population (bias)
  - Incomplete Frame – people you are most interested in aren’t included
  - Lack of Response – not enough people respond
- Poor sampling methods can produce misleading conclusions
  - Voluntary Response Sampling – people choose themselves by responding to a general appeal
  - Convenience Sampling – choosing individuals who are easiest to reach
Chapter 4: **Designing Studies**

In the following problems,

a) Determine is the survey design is flawed
b) If flawed, is it due to the sampling method of the survey itself
c) For flawed surveys, identify the cause of the error
d) Suggest a remedy to the problem

**Example 1:** MSHS wants to conduct a study regarding the achievement of its students. The principal selects the first 50 students who enter the building on a given day and administers the survey.

**Example 2:** The town manager selects 10 homes in one neighborhood and sends an interviewer to the homes to determine household incomes.

**Example 3:** An anti-gun advocacy group wants to estimate the percentage of people who favor stricter gun laws. They conduct a nation-wide survey of 1,203 randomly selected adults 18 years old and older. The interviewer asks the respondents, “Do you favor harsher penalties for individuals who sell guns illegally?”

**Example 4:** Cold Stone Creamery is considering opening a new store in Marion. Before opening the store, the company would like to know the percentage of households in Marion that regularly visit an ice cream shop. The market researcher obtains a list of households in Marion and randomly selects 150 of them. He mails a questionnaire to the households that ask about their ice cream eating habits and favor preferences. Of the 150 questionnaires mailed, 14 are returned.

**Example 5:** The owner of shopping mail wishes to expand the number of shops available in the food court. She have a market researcher survey mall customers during weekday mornings to determine what types of food the shoppers would like to see added to the food court.

**Example 6:** The owner of radio station wants to know what their listeners think of the new format. He has the announcers invite the listeners to call in and voice their opinion.

**Homework:** 1, 3, 5, 7, 9, 11
Chapter 4: Designing Studies

Example 1: Describe how a university can conduct a survey regarding its campus safety. The registrar of the university has determined that the community of the university consists of 6,204 students in residence, 13,304 nonresident students, and 2,401 staff for a total of 21,909 individuals. The president has funds for only 1000 surveys to be given and then analyzed. How should she conduct the survey?

Example 2: Sociologists want to gather data regarding the household income within Smyth County. They have come to the high schools for assistance. Describe a method which would disrupt the fewest classes and still gather the data needed.

Example 3: The manager of Ingles wants to measure the satisfaction of the store’s customers. Design a sampling technique that can be used to obtain a sample of 40 customers.

Example 4: The Independent Organization of Political Activity, IOPA, wants to conduct a survey focusing on the dissatisfaction with the current political parties. Several state-wide businesses have agreed to help. IOPA has come to you for advice. Describe a multi-stage survey strategy that will help them.
Example 5: Describe how you would use the following sampling methods to select 80 students to complete a survey.

(a) Simple Random Sample

(b) Stratified Random Sample

(c) Cluster Sample

Summary:
- A sample survey selects a sample from the population of all individuals about which we desire information.
- Random sampling uses chance to select a sample.
- The basic random sampling method is a simple random sample (SRS).
- To choose a stratified random sample, divide the population into strata, then choose a separate SRS from each stratum.
- To choose a cluster sample, divide the population into groups, or clusters. Randomly select some of the clusters for your sample.
- Failure to use random sampling often results in bias, or systematic errors in the way the sample represents the population.
- Voluntary response samples and convenience samples are particularly prone to large bias.
- Sampling errors come from the act of choosing a sample. Random sampling error and undercoverage are common types of error.
- The most serious errors are nonsampling errors. Common types of sampling error include nonresponse, response bias, and wording of questions.

Homework: 17, 19, 21, 23, 25
Chapter 4: Designing Studies

Section 4.2: Experiments

Objectives: Students will:
- DISTINGUISH observational studies from experiments
- DESCRIBE the language of experiments
- APPLY the three principles of experimental design
- DESIGN comparative experiments utilizing completely randomized designs and randomized block designs, including matched pairs design

Vocabulary:
- Experimental unit – an individual upon which an experiment is performed
- Subject – a human experimental unit
- Treatment – a specific experimental condition applied to the experimental units
- Statistically significant – a term applied to an observed effect so large that it would rarely occur by chance
- Block – a group of experimental units that are known, prior to the experiment, to be similar in some way that is expected to systematically affect the response to the treatments
- Double-blind – neither the subjects nor the observers know which treatments any of the subjects had received in an experiment

Key Concepts:

Parts of an Experiment:
- Experimental units – individuals on which experiment is done
  - Subjects – experiment units that are human beings
- Treatment – specific experimental condition applied to units
  - Factors – the explanatory variables in the experiment
  - Level – the combination of specific values of each of the factors

Experimental Design Factors:
- Control
  - Overall effort to minimize variability in the way the experimental units are obtained and treated
  - Attempts to eliminate the confounding effects of extraneous variables (those not being measured or controlled in the experiment, aka lurking variables)
- Replication
  - Use enough subjects to reduce chance variation
  - Increases the sensitivity of the experiment to differences between treatments
- Randomization
  - Rules used to assign the experimental units to the treatments
  - Uses impersonal chance to assign experimental units to treatments
  - Increases chances that there are no systematic differences between treatment groups

Summary:
- We can produce data intended to answer specific questions by observational studies or experiments.
- In an experiment, we impose one or more treatments on a group of experimental units (sometimes called subjects if they are human).
- The design of an experiment describes the choice of treatments and the manner in which the subjects are assigned to the treatments.
- The basic principles of experimental design are control for lurking variables, random assignment of treatments, and replication (using enough experimental units).
- Many behavioral and medical experiments are double-blind.
- Some experiments give a placebo (fake treatment) to a control group that helps confounding due to the placebo-effect.
- In addition to comparison, a second form of control is to form blocks of individuals that are similar in some way that is important to the response. Randomization is carried out within each block.
- Matched pairs are a common form of blocking for comparing just two treatments. In some matched pairs designs, each subject receives both treatments in a random order.
Chapter 4: Designing Studies

Example 1: Two toothpastes are being studied for effectiveness in reducing the number of cavities in children. There are 100 children available for the study.

A) What are the test subjects?

B) What is the response variable?

C) What are the treatments?

D) What are the factors or levels?

E) What are the possible confounding variables?

Example 2: A baby-food producer claims that her product is superior to that of her leading competitor, in that babies gain weight faster with her product. As an experiment, 30 healthy babies are randomly selected. For two months, 15 are fed her product and 15 are feed the competitor’s product. Each baby’s weight gain (in ounces) was recorded.

A) What are the test subjects?

B) What is the response variable?

C) What are the treatments?

D) What are the factors or levels?

E) What are the possible confounding variables?

Example 3: A statistics class wants to know the effect of a certain fertilizer (0, 2, 4 oz) and water levels (2, 4, 6 oz) on tomato plants. They get 60 plants of the same type.

A) What are the test subjects?

B) What is the response variable?

C) What are the treatments?

D) What are the factors or levels?

E) What are the possible confounding variables?

Homework: Day 1: 37-42, 45, 47, 49, 51, 53
Chapter 4: Designing Studies

Example 1: Draw a picture detailing the following experiment:
A statistics class wants to know the effect of a certain fertilizer on tomato plants. They get 60 plants of the same type. They will have two levels of treatments, 2 and 4 teaspoons of fertilizer. Someone suggests that they should use a control group. The picture should include enough detail for someone unfamiliar with the problem to understand the problem and be able to duplicate the experiment. Picture must address the randomization in detail.

Example 2: We wish to determine whether or not a new type of fertilizer is more effective than the type currently in use. Researchers have subdivided a 20-acre farm into twenty 1-acre plots. Wheat will be planted on the farm, and at the end of the growing season the number of bushels harvested will be measured.

A) How do you assign the plots of land?

B) What is the explanatory variable?

C) What is the response variable?

D) How many treatments are there?

E) Are there any possible lurking variables that would confound the results?

Day Two Homework: 57, 63, 65, 67
When the objective is to compare more than two populations, the experimental design that decreases the variability within the samples is called a **randomized block design**.

Block designs in experiments are similar to stratified designs for sampling. Both are meant to reduce variation among the subjects. We use different names only because the idea developed separately for sampling and experiments. Blocks allow us to draw separate conclusions about each block; for example, about men and women are their response to a medication. Blocking also allows more precise overall conclusions, because the systematic differences due to gender or some other characteristic can be removed.

A block is a group of experimental units that are similar in some way that affects the outcome of the experiment. **In a block design, the random assignment of treatments to units is done separately within each block.** Rather than treating the subjects as if they were in a single pool we split the subject population.

Blocks are a form of control. They control the effects of some lurking variables (such as gender, weight, age, etc.) by bringing those variables into the experiment so they can be accounted/controlled for.

**Example 1:** An agronomist wishes to compare the yield of five corn varieties. The field, in which the experiment will be carried out, increases in fertility from north to south. Outline an appropriate design for this experiment. Identify the explanatory and response variables, the experimental units, and the treatments. If it is a block design, identify the blocks.

**Example 2:** You are participating in the design of a medical experiment to investigate whether a calcium supplement in the diet will reduce the blood pressure of middle-aged men. Preliminary work suggests that calcium may be effective and that the effect may be greater for African-American men than for white or Hispanic men. Forty randomly selected men from each ethnic category are available for the study. Outline the design of an appropriate experiment. What kind of design is this? Can this experiment be blinded?
Example 3: An educational psychologist wants to test two different memorization methods to compare their effectiveness to increase memorization skills. There are 120 subjects available ranging in age from 18 to 71. The psychologist is concerned that differences in memorization capacity due to age will mask (confound) the differences in the two methods. What would the design look like?

Example 4: Men and women experience different physiological differences among diseases. In a study of blood pressure three different methods (a drug, yoga, and meditation) will be tried on both men and women randomly selected from a large company to see which is most effective in reducing high blood pressure. Construct an appropriate design diagram. Would a control group be necessary? Explain. Can this experiment be blinded? What is the parameter of interest in this experiment? What is the population of interest in this problem?

Homework: Day 3: 77, 79, 81, 85
Section 4.3: Using Studies Wisely

Objectives: Students will:
- DESCRIBE the challenges of establishing causation
- DEFINE the scope of inference
- DESCRIBE data ethics in designing studies

Vocabulary:
Statistics – xxxx

Key Concepts:

Summary:
- Inference about the population requires that the individuals taking part in a study be randomly selected from the larger population. A well-designed experiment that randomly assigns treatments to experimental units allows inference about cause-and-effect.
- Lack of realism in an experiment can prevent us from generalizing its results.
- In the absence of an experiment, good evidence of causation requires a strong association that appears consistently in many studies, a clear explanation for the alleged causal link, and careful examination of possible lurking variables.
- Studies involving humans must be screened in advance by an institutional review board. All participants must give their informed consent, and any information about the individuals must be kept confidential.
Chapter 4: Review

Objectives: Students will be able to:
- Summarize the chapter
- Define the vocabulary used
- Know and be able to discuss all sectional knowledge objectives
- Complete all sectional construction objectives
- Successfully answer any of the review exercises

Vocabulary: None new

Review Questions:

1. What is one reason for using random allocation to assign units to treatments in an experiment?
   a. to produce the placebo effect
   b. to produce experimental groups that are similar
   c. to eliminate lack of realism.
   d. to produce the blocks in a block design.

2. What is a specific experimental condition applied to the subjects or units in an experiment called?
   a. an observation
   b. the placebo effect
   c. a treatment
   d. the control

Control groups are used in experiments in order to - - -
   a. control the effects of extraneous variables on the response
   b. control the subjects of a study so as to insure all participate equally
   c. guarantee that someone other than the investigators, who have a vested interest in the outcome, control how the experiment is conducted
   d. achieve a proper and uniform level of randomization

4. A study was conducted to determine whether a football filled with helium would travel farther when kicked than one filled with air. Though there was a slight difference, it was not statistically significant. What are the treatments?
   a. the gas (air or helium) with which the football is filled.
   b. the kickers.
   c. whether or not the football was kicked.
   d. the distance that the football traveled.

5. (a) _______________________________ bias occurs when a representative sample is chosen for a survey, but a subset cannot be contacted or does not respond.
   (b) _______________________________ bias occurs when participants respond differently from how they feel, perhaps because of the way questions are worded or the way the interviewer behaves.

6. A large medical organization with membership consisting of doctors, nurses, and other medical employees wants to know how its members feel about health maintenance organizations (HMOs). Name the type of sampling plan they would use in each of the following scenarios.
   (a) They randomly sample 500 members from each of the lists of all doctors, all nurses, and all other employees and survey these 1500 members. _______________________________________
   (b) They randomly choose a starting point from the first 50 names in an alphabetical list of members, then choose every 50th member in the list, starting at that point. _____________________________
   (c) They select a random sample of hospitals where their members work and survey all members of the organization who work in each hospital. _______________________________________

7. If a sample is selected so that it systematically favors certain groups of the population, we say it is ___________________________.

8. A random sample of 1001 University of California faculty members was asked, “Do you favor or oppose using race, religion, sex, color, ethnicity, or national origin as a criterion for admission to the University of California?” 52% responded “favor.” What was the population for this survey?
Chapter 4: Designing Studies

9. List the two characteristics necessary for a sample to be a simple random sample.
   1.
   2.

10. A popular magazine often presents readers with the opportunity to answer a survey question by mailing in their response to the magazine. A typical question might be, “Do you think there is too much violence on television?” This type sample is called a/an __________________________ sample.

11. (a) Explain briefly the difference between an observational study and an experiment.
   
   (b) In which one of these is it safer to conclude that the difference in response was caused by the effect of the explanatory variable? __________________________

12. List the three basic principles of experimental design (key words are sufficient):
   (a) _______________________
   (b) _______________________
   (c) _______________________

13. Sometimes researchers think that experimental units are different enough in regard to an important variable that they should be grouped on that variable and then randomly assigned to treatments. These groups are called __________________________.

14. To prevent bias, experimenters try to assign subjects to a group so that neither the subjects nor the people who evaluate them know which treatment group the subject is in. An experiment of this type is described as __________________________.

15. Doctors investigated the relationship between a person’s heart rate and the frequency at which that person stepped up and down on steps of various heights. There were 3 rates of stepping and 2 different step heights used. A subject performed the activity (stepping at one of the 3 stepping rates at one of the 2 possible heights) for three minutes. His heart rate was then measured.
   
   (a) State what the factors are in this experiment.
   Next to each factor state its number of levels.
   (b) How many treatments are in this experiment? ______________
   (c) Identify one of the treatments. _____________________________
   (d) What is the response variable for this study? ______________
Names of 12 subjects are listed followed by a line of random digits. 
Ahbel   Barnes   Calhoun   Dancer   Freda   Keller   
Magee   Marge   McCullion   Stevens   Meier   Winokur 
41842   81068   09001   03367   49497   54580   
81507   27102   56027   55892   33063   71035 

e) Demonstrate your understanding of simple random sampling by using the random digits to determine which subjects would be randomly assigned to the first treatment. List these names:
___________________________________________________________________________
f) Describe how your selections were made. Be sufficiently clear in your description that I can duplicate your work.
g) Demonstrate your understanding of random blocked sampling by using the random digits to determine which subjects would be randomly assigned to the first treatment. List these names:
___________________________________________________________________________
h) Describe how your selections were made. Be sufficiently clear in your description that I can duplicate your work.

16. A 1994 article in Science magazine discussed a study comparing the health of 6000 vegetarians and a similar number of people who were not vegetarians. The vegetarians had a 28% lower death rate from heart attacks.

(a) Is this an observational study or an experiment? ________________________________

(b) Give an example of a potential confounding variable and explain what it means to say that this is a confounding variable.

(c) Give an example of an extraneous variable that you would not expect to be a confounding variable. Explain why you think this variable would not be confounding.

Homework: pg 380-3 problems 5.61-3, 66, 68, 70-72
Chapter 4: Designing Studies

From the AP Stats Listserv a response from one of the “fathers” of AP Statistics:
Chris and Peter,

I have always liked the way George Cobb presents this idea in experimental design.

In Introduction to the Design and Analysis of Experiments, George Cobb (1998) describes the variability inherent an experiment in the following way:

Any experiment is likely to involve three kinds of variability:

(1) Planned, systematic variability. This is the kind we want since it includes the differences due to the treatments.

(2) Chance-like variability. This is the kind our probability models allow us to live with. We can estimate the size of this variability if we plan our experiment correctly.

(3) Unplanned, systematic variability. This kind threatens disaster! We deal with this variability in two ways, by randomization and by blocking. Randomization turns unplanned, systematic variation into planned, chance-like variation, while blocking turns unplanned, systematic variation into planned, systematic variation.

The management of these three sources of variation is the essence of experimental design.

The management of the unplanned, systematic variability (3) is the issue between you and Peter. One important goal of randomization is to turn this systematic variability into chance-like variation that adds into our standard error. If that randomization "fails", then all bets are off. Randomization is never a guarantee against confounding variables, but it is our best defense.

This is an example I created to try to clarify the situation. (Linda Young bears no responsibility in what follows, but, nevertheless, it is all her fault.)

To focus the discussion, consider the variation inherent in the following experimental setting. To keep the computations simple and clear, the sample size is unrealistically small. I hope the gain in simplicity and clarity from this example outweighs the obvious problem with sample size.

Example Experiment:

Compare two kinds of rabbit food on weight gain (in ounces) from the age of two weeks to the age of six weeks of life. We want to know if the rabbits will gain more weight on one diet than on the other. We have space to house eight rabbits for this experiment.

Sources of Variation:

The most obvious is, perhaps, that the rabbits are all different rabbits, and so they will all grow at different rates. If different breeds of rabbit are used, then we will have an additional source of variation. Young California rabbits do not grow at the same rate as young Florida White rabbits, for example. The environment in which the rabbits live will not be exactly the same. They cannot all live in the same location; some will be in slightly warmer areas while others will be in areas with more light. They will not all have exactly the same amount of exercise or sleep. The food will be carefully weighed before it is given to the rabbits, but there will inevitably be measurement error in the amount of food given to each rabbit. Similarly, the rabbits will be weighed before the experiment begins and after the experiment ends. Measurement error (hopefully small) will occur in both these weighings. All of the aspects of the experimental setting mentioned so far can be considered natural chance-like variation.

There is another source of essential variation: the systematic difference in the rate of growth that is a result of the different diets. This is a variation we want to investigate. One way to think about the goal of the experiment is that we want to know if the variation that is a result of the diet is larger than the variation that is due to all the natural variation inherent in rabbit growth. In designing our experiment, we want to accentuate this planned, systematic variation, while reducing the natural chance-like variation.

This chapter considers some of the methods the experimenter has for managing these three sources of variation in the example experiment. The methods are Control, Randomization, Replication, and Blocking.
Chapter 4: **Designing Studies**

**Control:**

We control the experiment by organizing the structural components of the experiment to remove as many sources of chance-like variation as possible. We want to keep the experimental units (in this example, the rabbits in their cage) as similar as possible. We would like to use only one breed of rabbit. We might also prefer to use just one gender of rabbit, since male and female growth patterns may differ. We certainly want to keep the cages in a single location so that the effects of heat, light, air flow, and other unknowable affects will be as consistent as possible. As much as possible, we want the only difference among the rabbits to be the diet they are receiving.

To control the measurement error, we want to use the same scale when measuring the food each day. Similarly, it is important to use the same scales to measure the weights of the rabbits before and after the experimental regimen. We would prefer to use the same technician as well. In the end, no matter how much control we have in our experiment, some chance-like variation remains. It is the natural variation in average weight gain for our rabbits. By control, we try to make sure that our estimate of this variation is as accurate as possible and is not inflated by being combined with other extraneous sources of variation.

In the end, we want to do a calculation like a t-test, and the standard error in the denominator of this t-test will be our estimate of the natural chance-like variation in average weight gain for rabbits. The smaller we make our estimate of this chance-like variation, the more likely we are to detect a difference in the two varieties of rabbit food if a difference really exists. This is called increasing the power our test. We can gain power through controlling the experiment.

**The Price of Control: Scope of Inference**

The scope of inference refers to the population to which inference can reasonably be drawn based on the study. This population is the population from which the random sample used in the study was drawn. If only one breed of rabbit and one gender (male) is used, we might consider the results a random sample of results possible with this breed of male rabbit. We can comment only on this breed of male rabbit. If someone suggests that other breeds or females would behave differently with the diets, we have no counter-argument. We only have information about the breed of male rabbits we considered.

If we had taken a random sample of rabbits from several breeds, we introduce the variation inherent in those breeds. Some breeds are smaller and more active than others, while others are larger and more sedate. This variation makes it more difficult for us to find a difference in weight gain due to diet if one exists. However, if we do find a significant difference in weight gain, we can say something about rabbits of different breeds, not just one special breed. Similarly, if we used males and females, our population of inference would be rabbits of either gender.

Through control, the experimenter attempts to accentuate or make as visible as possible the planned, systematic variation between treatment groups, while at the same time reducing or removing as much chance-like variability as possible. The smaller the population of inference, often, the greater the control we have.

**Randomization:**

A second approach to handling the chance-like variability is through randomization. Clearly, it is not possible to remove all chance-like variation through our methods of control. Rabbits are still different, even if they are the same breed and gender; some will grow faster than others, regardless of the diet. Measurement error is always present even if the same scales and technicians are used.

By randomly assigning the rabbits to the treatment groups, we will spread the chance-like variation among the treatment groups. This adds to the variation in each group, but it removes the bias that would otherwise doom the experiment. This random assignment of experimental unit to treatment group is essential for an experiment and distinguishes it from an observational study. In an observational study, the experimental units or subjects are not randomly assigned to the treatment groups. As an example, if we looked at rabbits at Farm A who are presently being fed Diet A and rabbits at Farm B who are presently being fed Diet B, we might find that the average weight gain for the rabbits at Farm A (on Diet A) had a greater average weight gain. However, the diet and farm are inextricably confounded. We could never say that Diet A was better than Diet B, since any weight gain could be a result of other aspects of rabbit life at the two farms. If, in the end, we want to say that one diet produces greater average weight gain in rabbits than the other, then randomization is essential.

Randomization plays other roles in our experiment. We should place the rabbit cages in our designated locations through a random process. This keeps the lurking variables of heat, light, air flow, and other unknowable variables from biasing the results. These unknown variables can produce systematic, unplanned variation if randomization is not used. The effects of these variables, if any, is distributed to the two treatment groups by this randomization process. This form of randomization is our protection against bias (unplanned, systematic variation) in the experiment.

Finally, a randomization process creates the probability models we use for the basis of hypothesis tests. If the null hypothesis is true (diet has no effect on average weight gain), then the variation we see between treatment groups must all be of the chance-like variety. We have estimated the size of this variation, and we build our hypothesis tests around it.
Chapter 4: **Designing Studies**

Both of these types of randomization are essential to a good experimental design. A third type of randomization, random sample of experimental units from the population of inference, is not essential and is often not possible. However, if a random sample is taken, we can make inferences to the population once we have our result.

**Replication:**

Replication in an experiment just means using more than one experimental unit. In this context, it does not mean repeating the whole experiment multiple times. Each additional rabbit is called a replicate. We must have a way to estimate the size of the chance variation and we need at least two values to compute a standard deviation. Without replication, there is no way for the experimenter to estimate the chance-like variation to compare to the systematic, planned variation between treatment groups. The more rabbits used for each diet, the more accurate is the estimate of the natural variation in weight gain. There is a second benefit of using more rabbits; the greater the number of replicates in each treatment group, the smaller the standard error used in the t-test, since the estimated variance of the mean weight gain of n rabbits is the estimated variance for single rabbits divided by n. This corresponds to a reduction in the estimate for the size of the chance-like variation in the mean increase in weight.

**Blocking:**

The final method for managing the variability inherent in an experiment is through blocking. Blocking is more complicated than control, randomization, and replication. Suppose, due to availability, we were forced to use two different breeds of rabbits instead of just one. We have four Californian and four Florida White rabbits for use in this experiment. We believe that the Californian will grow faster than the Florida White and so the weight gains for these four rabbits will be larger than that of the other four, regardless of the diet. For example, we might expect the average weight gain for Florida White rabbits to be about 10 ounces less than the average gain for the larger Californian rabbits. However, we don't think there will be an interaction between breed and diet. This means that the effect of each diet on the rabbits' growth will be the same additive amount. We might suspect that, for example, Diet A will add 6 ounces to the weight gain for both California and Florida White rabbits. The variability due to the two breeds is not chance-like; it is systematic, unplanned variation. We can turn this variation into chance-like variation by our random assignment process, but the variation caused by having two different breeds will be included in our estimate of chance variation, inflating it and reducing the power of our test.

A better solution comes from the process called blocking. We will use the breed as a blocking variable. We are not really interested in the effect of breed, so we think of breed as a nuisance variable. We want to estimate the amount of variation added by having two different breeds, and remove it from our estimate of the chance-like variation (reducing our standard error) by the process of ANOVA.

In AP Statistics students are taught that a good experiment requires Control, Replication, and Randomization. Each of these attributes offers the experimenter a way to manage the inescapable variability inherent in the experimental process.

The planned, systematic variability is emphasized by controlling extraneous sources of variation, while the chance-like variability is managed by randomization and reduced by replication and control. Finally, the unplanned, systematic variability that an destroy our results can be managed by blocking when its causes are recognized prior to the experiment and through randomization in any event.

The experimenter must always pay careful attention to the design of the experiment, since the method of analysis is determined by the manner in which the experimental units are randomized to treatments. The way you randomize is the way you analyze.