Chapter 2

AP STATISTICS
HW 2A

ANSWERS

Sampling and Experimentation

A Overview of Methods of Data Collection

1 EXAMPLE: The principal of a new school wants to know what mascot the incoming students would prefer. He summons each of the student body into his office and asks what the new mascot should be. What type of study is this?

CENSUS

2 EXAMPLE: The principal of the new school is also interested in what type of dress code the school should have. He telephones the parents of every tenth student summoned to his office and asks their opinions about school dress codes. What type of study is this?

Sample Survey (RANDOM)

REVISED 9/7

3A EXAMPLE: A pharmaceutical company wishes to test a new medication it thinks will reduce cholesterol. A group of 20 volunteers is formed and each has his or her cholesterol level measured. Half is randomly assigned to take the new drug and the other half is given a dummy pill, a pill with no active ingredients. After 6 months the volunteers' cholesterol is measured again and any change from the beginning of the study recorded. What type of study is this?

This is an experiment because the Company is imposing conditions (drug, placebo) to determine response to a condition (change in cholesterol)

3B EXAMPLE: A health studies research lab is interested in the effect of certain vegetables on cholesterol level. A group of 20 volunteers is formed and each keeps a diary of his or her food consumption for the next 6 months, after which time the diaries are collected and cholesterol level is measured. The researchers then examine the relationship between the rate of consumption of certain vegetables and cholesterol level. What type of study is this?

This is an OBSERVATIONAL STUDY because No conditions are imposed.

B Planning and Conducting Surveys

4 EXAMPLE: A journalism class prints a survey in their school newspaper. Readers are asked to clip the survey from the paper, complete it, and return it to a drop box in the school cafeteria. What type of sample is this?

This is a voluntary response survey

5 EXAMPLE: A journalism class stations pollsters in front of the stadium during a football game. They ask each student who enters his or her opinion of the quality of the school's athletics program. What type of sample is this?

This is a convenience study

6 EXAMPLE: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The voters' ID numbers are entered into a computer and 500 are chosen at random, without replacement, to comprise the sample. What type of sampling design is this?

SIMPLE RANDOM SAMPLE

7 EXAMPLE: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The voters' ID numbers and party affiliations (Democrat, Republican, Independent) are entered into a computer. The firm randomly selects, without replacement, 200 Democrats, 200 Republicans, and 100 Independents to comprise the sample. What type of sampling design is this?

COMPARE STRATA #7 VS CLUSTERS #8

STRATA ARE HOMOGENOUS GROUPS

STRATIFIED RANDOM SAMPLE

3 STRATA – Democrats 200 40%
Republicans 200 40%
Democrats 100 20%
500

* QUESTION ARE THE %'S FOR THE STRATA REPRESENTATIVE OF THE POPULATION?

HOW WOULD YOU FIND THIS OUT?

8 EXAMPLE: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The voters' ID numbers and voting precinct numbers (01–75) are entered into a computer. The firm first randomly selects 20 of the voting precincts. Then, from each of the chosen precincts, 25 voters are chosen with simple random samples. What type of sampling design is this?

CLUSTERS ARE HETEROGENOUS GROUPS (ie precincts)

- 15,000 voters in 75 precincts
- RANDOMLY select 20 voting precincts
  → THEN 25 voters are chosen with SRS.

CLUSTER SAMPLE

9 EXAMPLE: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration. The list is sorted by voter ID number. The firm randomly selects one person from the first 100 on the list; then it selects every 100th person after that. The sample size is 150. What type of sampling design is this?

SYSTEMATIC RANDOM SAMPLE

10 EXAMPLE: A journalism class prints a survey in their school newspaper about the school charging students for parking. Readers are asked to clip the survey from the paper, complete it, and return it to a drop box in the school cafeteria. Why is this sample biased?

SURVEY IS SUBJECT TO VOLUNTARY RESPONSE BIAS. STUDENTS WHO WANT OR ARE DRIVING TO SCHOOL ARE MORE LIKELY TO RESPOND AND SKEW THE RESULTS.

11 EXAMPLE: A journalism class stations pollsters in front of the stadium during a football game. They ask each student who enters his or her opinion of the quality of the school's athletics program. Why is this sample biased?

THIS IS A CONVENIENCE STUDY AND IS SUBJECT TO UNDER COVERAGE SINCE STUDY ONLY INCLUDES STUDENTS AT FOOTBALL GAME & LEAVES OUT LARGE PORTIONS OF THE POPULATION.

12 EXAMPLE: A media research firm is conducting a poll on an upcoming election for city council. The firm obtains a list of all 15,000 registered voters in the council ward under consideration and a simple random sample of 500 voters is chosen. Over a 24-hour period, telephone calls are placed to the voters, with follow-up calls made to those voters who do not answer the phone on the first attempt. Pollsters could not reach 37 of the 500 voters selected. What sources of bias could exist in this survey?

THIS IS NON RESPONSE BIAS.
→ THE 37 Voters could affect the results of the survey

13 EXAMPLE: A journalism class conducts a simple random sample of students at their school. They ask each student, "Given the fact that our school has won seven championships in the last five years, do you favor or oppose reducing funding for athletic programs?" What sources of bias could exist in this survey?

THIS IS A FORM OF RESPONSE BIAS.
THIS A LEADING QUESTION AND SHOULD BE REWORDED TO GET MORE ACCURATE RESULTS

D Sampling with Random Digit Tables

A random digit table can be used to select a random sample from a population. Each member of the population is assigned an identification number, with the length of the number in digits determined by the size of the population. For populations of 10 or less, one digit is needed. Populations of 11–100 require two-digit numbers. Three digits are required for populations of size 101–1,000, and so on. Note that if one has 100 individuals in a population, assigning 01–100 is *not* correct. The number 100 has three digits and all of the others from 01 to 99 have two. However, 00 is a two-digit number that will work for 100.

CONT →

Random numbers of the same length are selected from the random digit table by reading across the lines of the table. Each consecutive group of digits will either select one of the individuals in the population, or be ignored because it has already been selected, if sampling without replacement, or because it does not correspond with any member of the population.

EXAMPLE: Select a simple random sample of size $n = 6$ from the population of names below using the random digit table provided:

01 Anita ✓	02 Billy	03 Carol	04 Doug	05 Elmer
06 Francine	07 Glenda	08 Hector	09 Ivy	10 Jose ✓
11 Kelly	12 Lynn	13 Melvin ✓	14 Nicole ✓	15 Olive
16 Paul ✓	17 Quincy	18 Rae	19 Sue	20 Tom ✓

01401	49913	20134	96010
16290	33843	95945	04834
37520	93015	93615	03413

TIP

ANSWER: First, assign two-digit identification numbers to the members of the population.

THE POPULATION IS 20. THEREFORE NEED TO USE 2 DIGITS. NUMBER NAMES 1–20 alphabetically.

Next, take consecutive two-digit numbers from the random number table, excluding values 21–99 and 00, and those that have already been selected.

USE RANDOM NUMBERS AND GET THE FIRST SIX:

01 ANITA — ~~40~~ — 14 Nicole — ~~99~~ — 13 MELVIN — 20 TOM — Need 2 more

~~13~~ repeat # — ~~49~~ — ~~00~~ — 10 JOSE — 16 PAUL

The simple random sample of size $n = 6$ is ANITA, NICOLE, MELVIN, TOM, JOSE & PAUL.

E Planning and Conducting Experiments

115 EXAMPLE: A pharmaceutical company wishes to test a new medication it thinks will reduce cholesterol. A group of 20 volunteers is formed and each has his or her cholesterol level measured. Half is randomly assigned to take the new drug and the other half is given a placebo. After 6 months the volunteers' cholesterol is measured again and any change from the beginning of the study recorded. Identify the experimental units, factors and their levels, treatments, and response variable in this experiment.

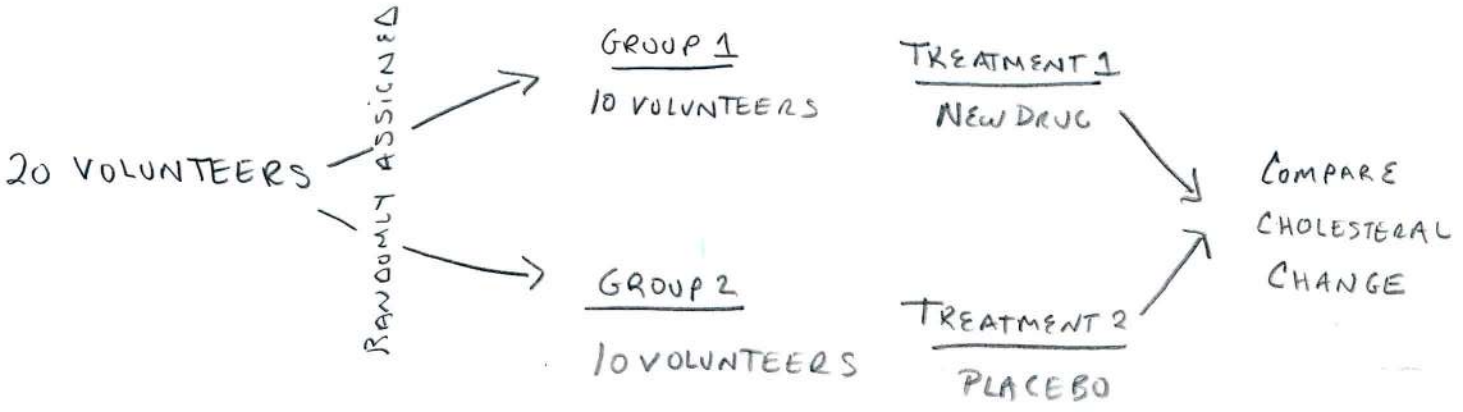
EXPERIMENTAL UNITS: 20 VOLUNTEERS

ONE FACTOR: MEDICATION WITH 2 LEVELS – NEW DRUG & PLACEBO

TWO TREATMENTS: ① NEW DRUG
② PLACEBO

RESPONSE VARIABLE: CHANGE IN CHOLESTEROL LEVEL DURING THE 6 MONTH PERIOD

DIAGRAM:



16 EXAMPLE: An agricultural researcher is interested in determining how much water and fertilizer are optimum for growing a certain plant. Twenty-four plots of land are available to grow the plant. The researcher will apply three different amounts of fertilizer (low, medium, and high) and two different amounts of water (light and heavy). These will be applied at random in equal combination to each of four plots. After 6 weeks, the plants' heights in each plot will be recorded. Identify the experimental units, factors and their levels, treatments, and response variable in this experiment.

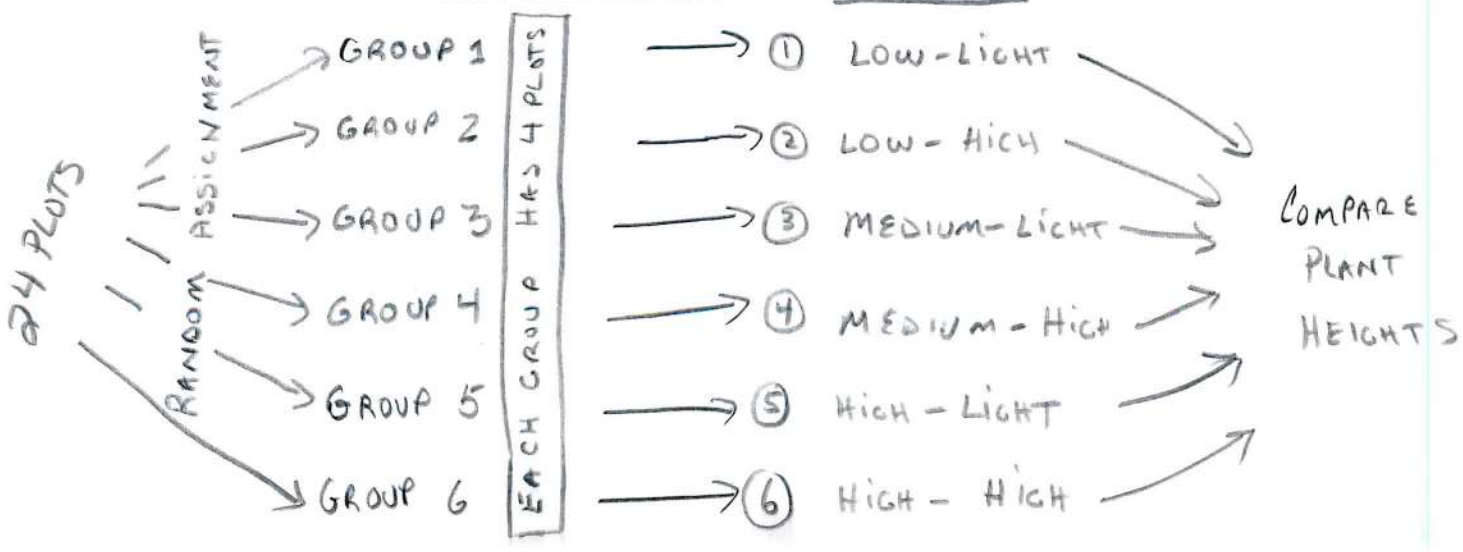
Experimental Units – 24 plots of land

Two Factors – ① Fertilizer w/ 3 levels (low, medium, high)
② Water w/ 2 levels (light + heavy)

Six Treatments – ① Low Fertilizer – Light Water
② Low " – High
③ Medium " – Light
④ Medium " – High
⑤ High " – Light
⑥ High " – High

Response Variable – Height of plants

DIAGRAM



F

Characteristics of a Well-Designed Experiment

17 **EXAMPLE:** A pharmaceutical company wishes to test a new medication it thinks will reduce cholesterol. A group of 20 volunteers is formed and each has his or her cholesterol level measured. Half is randomly assigned to take the new drug and the other half is given a placebo. Neither group knows which pill it is taking. After 6 months the volunteers' cholesterol is measured again and any change from the beginning of the study recorded. Explain where control, randomization, and replication are present in this study.

CONTROL: ① USE OF PLACEBO ② BLIND STUDY

RANDOMIZATION: GROUPS ARE RANDOMLY ASSIGNED TO 2 Treatments

REPLICATION: EACH TREATMENT IS IMPOSED 10 Times (more would be better)

18 **EXAMPLE:** A health studies research lab is interested in the effect of certain vegetables on cholesterol level. A group of 20 volunteers is formed and each keeps a diary of his food consumption for the next 6 months, after which time the diaries are collected and cholesterol level is measured. The researchers then examine the relationship between the rate of consumption of certain vegetables and cholesterol level. What confounding variables could be present in this observational study?

CONFOUNDING VARIABLES:

- Exercise
- Diet outside of vegetables in the study
- Other health-related variables
- Those that eat vegetables may eat lower fat diets
- all variables that affect cholesterol levels

G Experiment Designs

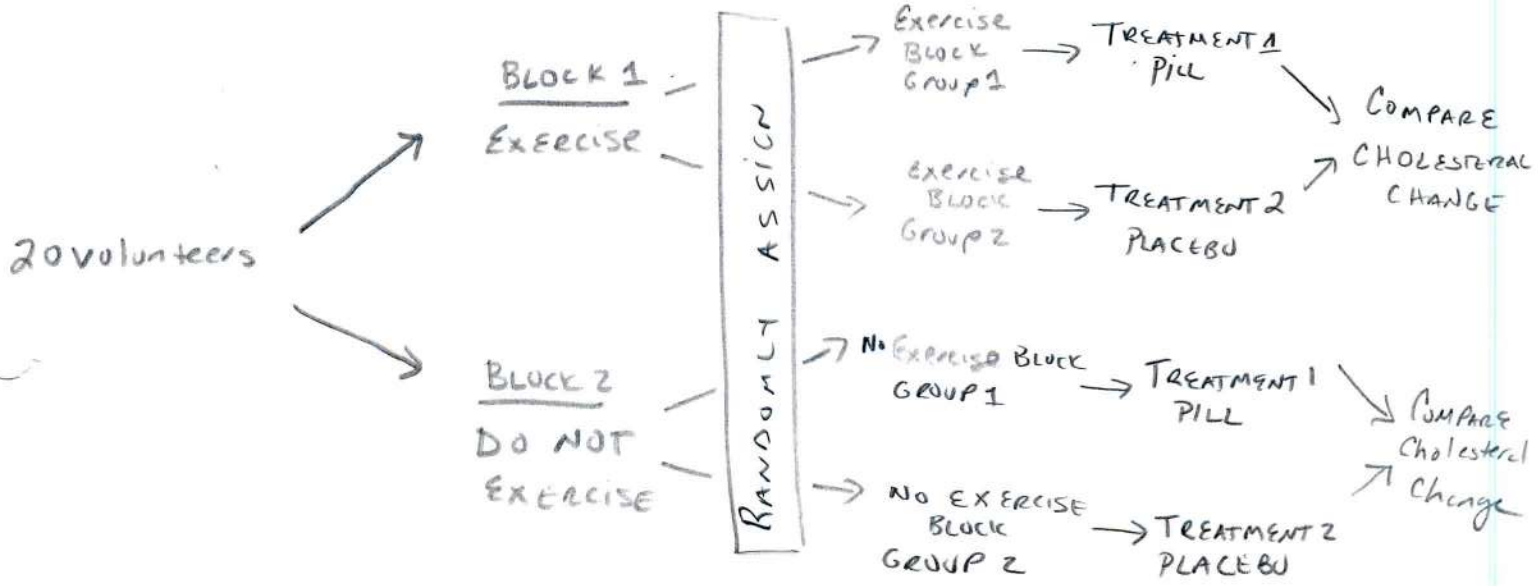
19 EXAMPLE: A pharmaceutical company wishes to test a new medication it thinks will reduce cholesterol. A group of 20 volunteers is formed and each has his or her cholesterol level measured. Half is randomly assigned to take the new drug and the other half is given a placebo. Neither group knows which pill it is taking. After 6 months the volunteers' cholesterol is measured again and any change from the beginning of the study recorded. What type of experimental design is this?

This is a Completely randomized design

No blocking done such as gender, age, etc

20 EXAMPLE: A pharmaceutical company wishes to test a new medication it thinks will reduce cholesterol. A group of 20 volunteers is formed and each has his or her cholesterol level measured. After 6 months the volunteers' cholesterol is measured again and any change from the beginning of the study recorded. The researcher believes that regular exercise may influence the change in cholesterol level. Create a randomized block design that takes account of subjects who exercise regularly.

2 BLOCKS – EXERCISE REGULARLY + THOSE THAT DO NOT



21 EXAMPLE: A pharmaceutical company wishes to test a new medication it thinks will reduce cholesterol. A group of 20 volunteers is formed and each has his or her cholesterol level measured. After 6 months the volunteers' cholesterol is measured again and any change from the beginning of the study recorded. The researcher believes that initial cholesterol level may influence the change in cholesterol level. Create a matched pairs design that takes account of subjects' initial cholesterol level.

PAIR THE subjects by initial cholesterol level.

GROUP 1 The two highest initial cholesterol level

2 The next 2 highest initial cholesterol level

3 and so...

4

5

6

7

8

9

10