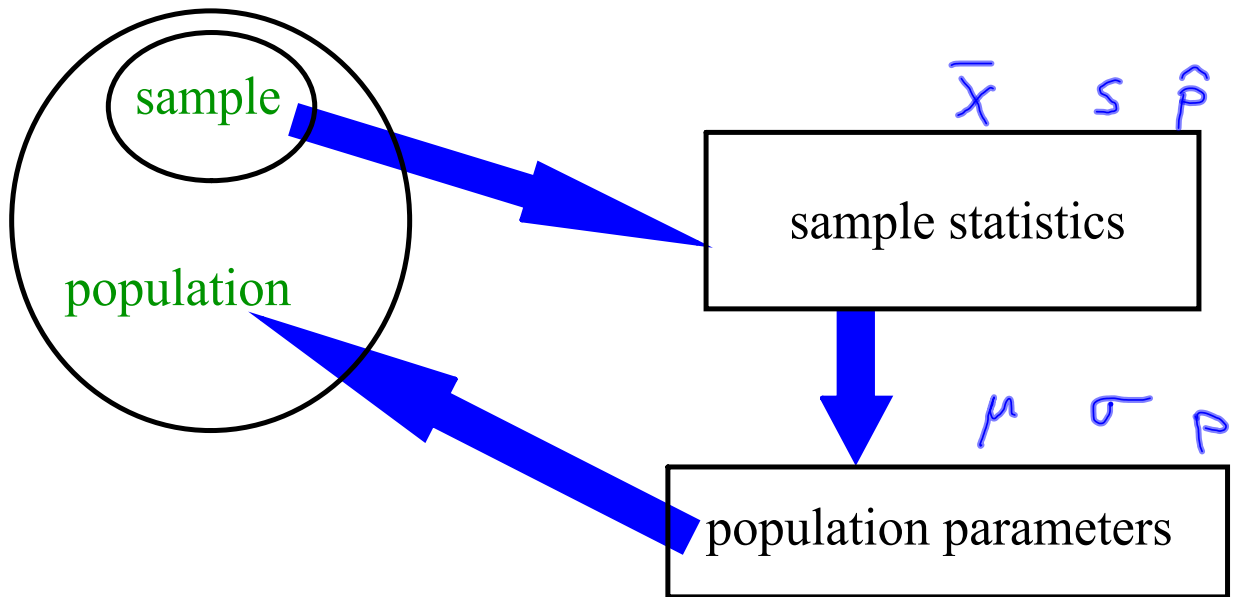


5.1 Designing samples



Entire group we want to know about: population

Part of population we actually examine: sample

Studying part of a population: sampling

Attempt to contact every individual in population: census

examples

corn field (**population**)

What insects?

Examine every plant (**census**) -- too long

Instead, examine a **sample** of 10 plants from various parts and inspect for insect damage.

No Child Left Behind (NCLB) requires that every student be tested in math once in high school.

Mo Alg. students (**population**) take the test, so this is a **census**. NCLB won't allow a **sample** because **sampling** would not measure every student.

ActivStats III.10.1 Sample

observational study observes individuals, measures variables, but does not attempt to influence the responses (no treatment assigned)

example

3 college students (1 at a time and briefly)
busy dorm restroom
observed individuals suspect nothing
monitor handwashing
observers do not influence behavior, just record it

experiment deliberately imposes a treatment on individuals in order to observe their responses

example

3 college students (1 at a time and briefly)
busy dorm restroom
observed individuals suspect nothing
monitor handwashing of certain individuals
next, post signs "Wash your hands!"
monitor handwashing of same individuals
Did signs affect behavior?

survey seeks responses from individuals knowingly responding to questions.

example

3 college students
outside busy dorm restroom
ask about handwashing habits upon exit

Sample **design**

method for choosing sample & getting information

biased design systematically favors certain outcomes

ActivStats III.10.2-3 Bias

Voluntary response sampling

people choose themselves

responding to an appeal

biased: strongest opinions most likely to respond

examples

Radio talk shows

Internet surveys

Warning: data reflect the views of only those motivated to respond, often with the most extreme opinions

Convenience sampling

choose individuals easiest to reach
biased: likely won't represent population

example

The Chicago Daily Tribune called prospective voters. (mostly the rich had phones)

example Selecting the first 25 shoppers at a store on Tuesday morning

A Cautionary Note: Data reflect the views of only those available at the time taken.



© Associated Press. TITLE: Dewey Defeats Truman
AP PHOTOGRAPHER: BYRON ROLLINS
11/4/1948

You can now do
5.1-5.7 on pg. 273

A **probability sample** has:

- no conscious choosing
- a definite procedure for selecting participants that involves use of probability or chance outcome

A **simple random sample (SRS)** of size n 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 selected.

SRS

- every individual has equal chance
- every sample of size n has equal chance
- chosen randomly

Some ways:

- a) choose names from hat
- b) computer choose randomly
- c) assign #s to individuals & use table of random digits to select #s

examples

- lottery
- drawing student names from hat

A **table of random digits**:

- Each digit is equally likely to be any digit 0 through 9
- digits are independent of each other (knowledge about one part of the table tells you nothing about any other part)

<http://bcs.whfreeman.com/yates2e/pages/bcs-main.asp?v=category&s=00020&n=99000&i=99020.01&o=>



Pick a SRS of 6 names from the 12 students here using the random number table. Start at line 110.

1. I assign each name a number 01 through 12.
2. I select 2 digits at a time from the table and see if it matches a name and ignore all other 2 digit numbers. I ignore repeats. I write on the table.
3. Identify the SRS.

110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	50626	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776

stratified random sample

1. Divide population into strata (groups) of *similar* individuals
2. Choose a separate **SRS** in each stratum
3. Combine these SRS's to form the full sample.

examples

Strata are often formed around race, gender, residence, or economic status. **The strata must be selected so members of any particular stratum are similar/homogeneous.**

A farmer with 4 chicken breeds samples egg yield for each.

In planning to start a PSP program, a principal expects students who are in sports, work program, or full day may differ on what is a desirable parking space. She assembles lists of these various groups and interviews a couple students from each group.

A reporter randomly selects 5 freshmen, 5 sophomores, 5 juniors, and 5 seniors for a survey.

Cluster Sample - a population has naturally occurring groups, each well mixed/heterogeneous and we randomly choose an entire group/cluster as our sample.

examples

Marketing researchers randomly choose a subdivision from a town and survey every household there about their shampoo.

An agriculture researcher randomly chooses a farm from the county and measures the soil temperature in each field there.

Multistage Sample - a series of SRS's.

examples

To predict an election, a political scientist takes a SRS of electoral sub-divisions from the state, then picks a SRS of blocks of houses, then selects a SRS of individual houses to be polled.

To survey students, a principal randomly picks 1 of the 4 classes, selects a SRS of homerooms, then in each of those homerooms, selects a SRS of students.

Caution: methods to analyze a multi-stage sample differ from the methods used to analyze a SRS.

Don't confuse these!

stratified	cluster sampling
we ÷ population into groups	groups naturally occur
randomly select from each strata	randomly choose entire cluster(s)

Systematic Sample: start at some random location in a list of members of population, then select every n^{th} member

examples

A farmer's hogs are tagged with ID numbers. He randomly chooses one hog out of the first 10 and then every 20th hog thereafter. He ended up weighing the 4th hog, then the 24th, 44th, 64th, etc..

For a poll, we could select one of the first 100 student numbers from an ordered list & then every 100th student number after that one.

Be careful that the selection method isn't associated with the way the population is organized (it's not appropriate to choose a day of the week at random and then select every 7th day after that).

Quota Sampling: we build a sample that is similar to the population in proportions of race, gender, residence, income, politics, economic status, etc..

examples

Restaurant marketing researchers may see how a new sandwich appeals to a sample selected to mirror the U.S. population in terms gender, age, or other factors they believe most influence food choice.

Surveys with screening questions to select who is or is not included.

Danger: surveyor chooses who to interview within the constraints of the quotas set.

Suppose 24 students are seated in rows of 6 and I want to select a sample of 6. What kind of sample is each:

I assign numbers 01-24 and choose 6 at random (ignoring repeats)

Answer

I pick 3 boys and 3 girls

Answer

I put all the names in a hat and draw out 6 names without replacement

Answer

I pick the 6 closest to me

Answer

I pick a whole row at random

Answer

I rolled a 4-sided die and got a 2 so I chose the 2nd, 6th, 10th, and 14th, 18th, and 22nd students.

Answer

You can now do
5.9 and 5.11 on pg. 279

Sources of bias and cautions about sample surveys

Population ⇒ **sampling frame** ⇒ sample

- list of individuals available to be chosen
- incomplete sampling frame can cause bias

undercoverage: some groups are left out of the process of choosing the sample

examples

Surveys of households won't represent the homeless, inmates, & students in dorms.

Phone surveys may miss those without a landline.

Most sample surveys also suffer from **nonresponse**-- this happens when someone is unavailable for selection or refuses to cooperate. Non-respondents tend to differ from those who are readily available.

examples

Some voters refuse to participate in election exit polls.

Some people sign up for the no-call list or are not at home when a pollster calls.

The surveyors at the mall miss those who do not shop at the mall or who refuse to participate.

Household bias occurs when a sample includes only one member of any given household. This underrepresents the members of large households.

response bias: when someone answers falsely.

examples

voter might lie in an exit poll, hoping that early returns may motivate some voters to get to the polls or to stay home.

One may *think* he saw 3 movies in the last 6 months, when it has actually been 8 months.

wording effects:

questions are confusing, leading, or put in a particular order.

example

Some surveys have leading questions and so, participants may answer in hopes of "getting it right".

You can now do
5.13 - 5.17 on pg. 283