Skip Chap 9

Chapter 10: Sample surveys p290

See this kind of thing all the time:

- a survey asking "if there were a provincial election tomorrow, which party would you vote for?"

How is that done, and why?

What would happen if we tried to survey everybody?

Examine a part of the whole p290

- The entire group of individuals that we want information about is called the **population**.

- We'd like to know about an entire population of individuals, but examining all of them is usually impractical, if not impossible.

- We settle for examining a smaller group of individuals—**a sample—selected from the population**.

- Opinion polls are examples of sample surveys, designed to ask questions of a small group of people in the hope of learning something about the entire population.
  - Professional pollsters work quite hard to ensure that the sample they take is representative of the population.
  - If not, the sample can give misleading information about the population.

How might we select a representative sample?

1. Carefully select individuals to *match* the population in every way we can think of:
    - Males and females
    - the right mix of ages
    - right number of people living in each city/rural area
    - right mix of political opinions
    - etc, etc.
   - Difficult to do.
   - Might miss important way of matching population.
2. Select individuals **at random**.
    - Easy to do
    - A*pproximately* represents population in all ways, *including ones you didn't think of.*

Why does randomization work?

- Short term unpredictable, long term predictable

- *Cannot predict* which individuals are going to end up in sample

- With a large sample, sample will have approximately right proportion of males/females, urban/rural, old/young, etc., *and anything else we didn't think of*.

Three keys for sampling:

1. Examine a part of the whole (sample)
2. Randomize (to obtain the sample)
3. Sample size

It's the sample size p292

- How large a random sample do we need for the sample to be reasonably representative of the population?

- It’s the size of the sample, not the size of the population, that makes the difference in sampling.

  - Exception: If the population is small enough and the sample is more than 10% of the whole population, the population size *can* matter.

- The *fraction* of the population that you’ve sampled doesn’t matter. It’s the *sample size* itself that’s important.

Does a Census Make Sense? P293

- Wouldn't it be better to just include everyone and "sample" the entire population?
  - A survey with all individuals in the population is called a census.

It can be difficult to complete a census—there always seem to be some individuals who are hard to locate or hard to measure, high cost etc.

Sometimes the population changes while you work, and your findings might not be relevant to the current population by the time your complete your census.

Populations and parameters, samples and statistics p293

- A **parameter** is a number that describes the population.
-
- A parameter is a fixed number, but in practice we do not know its value.
-
- A **statistic** is a number that describes a sample.
-
- The value of a statistic is known when we have taken a sample, but it can change from sample to sample.
-
- We often use a statistic to estimate an unknown parameter.

- **Simple Random Sample** p294

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 actually selected.

This requires a list of whole population (sampling frame).

Drawing a simple random sample using random digit table p295

Suppose we have a population of 80 students, numbered 01—80, want a simple random sample of 6 of them. Use these random digits: 43623 33434 94776 15780 95603 64962 46971 95188.

43, 62, 33, 34 all ok

34 is a repeat: reject

94, too big, reject

77, 61: ok

so sample is students numbered 43, 62, 33, 34, 77, 61.

Stratified Random Sampling p296

- Divide the population into groups of similar individuals, called **strata**.
- Then choose a separate SRS in each stratum and combine these SRSs to form the full sample.

Example

Why is this good?

- Stratified random sampling can reduce bias.

- Stratifying can also reduce the variability of our results.

- Therefore sample statistic should be closer to population parameter.

Drawing a stratified sample

Population of 100 students: 01-60 female, 61-99 and 00 male.

Use random digits: 18406 28903 75909 66389 28937 46983 49652 37406 .

Draw stratified sample of 6 females and 4 males.

18, 40 females (2 so far)

62, 89 male (2 so far)

03 female (3 so far)

75, 90 male (now 4 males, don't sample any more)

96, 63, 89 reject (would be more males)

28 (4^{th} female), 93, 74, 69, 83 (reject)

49, …, 23 (last 2 females).

Cluster and multistage sampling p297

- Sometimes stratifying isn’t practical and simple random sampling is difficult.
- Splitting the population into similar parts or clusters can make sampling more practical.
  - Then we could select one or a few clusters at random and perform a census (or take a sample if the clusters are large) within each of them.

This sampling design is called cluster sampling

Example

How would you randomly sample 100 words from the textbook?

- simple random sampling: number every single word and then sample from them.
- easier: randomly sample 10 *pages* first, then randomly sample 10 words on each page. Why is that easier to do?
- not same as simple random sampling: if you select a particular page, other words on the same page *more likely* to be in sample.
- *cluster sampling.*

Multistage sampling p299

- Often hierarchy of clusters eg. chapter – section – sentence – word, and could choose:
  - chapters
  - section within chosen chapter
  - sentence within chosen section
  - word within chosen sentence

Called *multistage sampling.* At each stage, choice made by simple random sampling.

- Sometimes we use a variety of sample methods

Choose cluster/multistage sampling for *convenience*, choose stratified sampling for *accuracy*.

Systematic Samples p300

- Sometimes we draw a sample by selecting individuals systematically.

- For example, you might survey every 10th person on an alphabetical list of students.

- To make it random, you must still start the systematic selection from a randomly selected individual.

- When there is no reason to believe that the order of the list could be associated in any way with the responses sought, systematic sampling can give a representative sample

Things that can go wrong with sample surveys 306

- not getting Who you want (nonresponse)
  - *call back later*
- getting the question(s) right
  - *avoid favouring a certain answer in way question is asked.*
- not giving choices for answer (ie. getting open-ended response)
  - *eg. use strongly agree – strongly disagree*
- sampling volunteers
  - *don't rely on people who choose to respond, eg. callers to radio show*
- sampling badly but conveniently
  - *see above*
- undercoverage
  - *not being able to sample certain parts of population.*