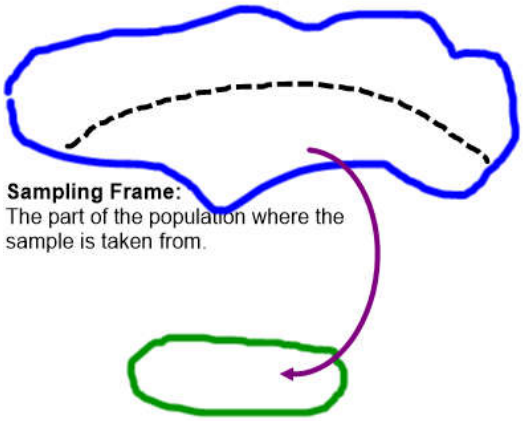
AP Statistics – Lesson Notes - Chapter 12: Sampling Techniques and Biases

Sample of a population - terminology

We have the ability to analyze data that has been collected. Often, this data is a **representative sample** of a larger **population**.



Sampling Frame:
The part of the population where the sample is taken from.

Population:
An entire group of individuals.

Parameters:
$examples: \mu, \sigma, r, \beta, p$

Sample:
A subset of the population.

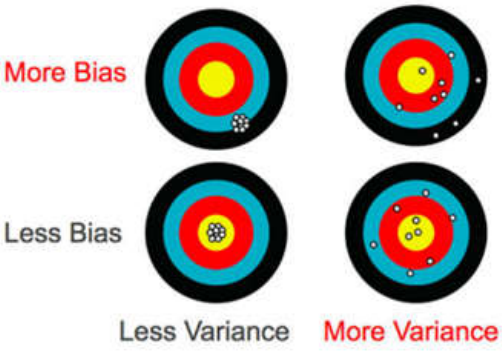
Statistics:
$examples: \bar{x}, s_x, r, b, \hat{p}$

Census:
If, for some reason, you can analyze the entire population, that is called taking a 'census'

It is more difficult than it sounds to take a representative sample. The goal in selecting a sample is to avoid the two 'enemies' of statistical analysis: **bias** and **variability**.

Bias = The sample is not representative of the entire population you are studying and to which your conclusions will apply.

Variability = If multiple samples are taken, variability is the amount that some measured statistic varies from sample to sample.



Both bias and variability make it more difficult for statistical analysis to reach conclusions and for those conclusions to apply generally.

Randomization produces the best sample

We saw in Ch11 in the Federalist Papers Activity why we shouldn't rely on human intuition to select a representative sample.

The best way to avoid human bias is to **use randomness** when selecting a sample.

Different types of sampling - Good Sampling Techniques

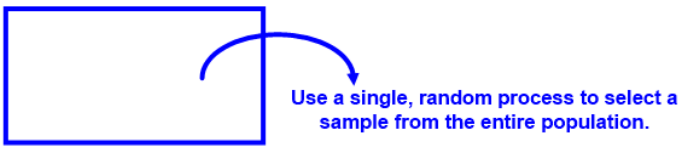
1) Simple Random Sample (SRS)

- The entire population is available, individually, for selection.
- Select *n* individuals from the entire population using a random process.

For an SRS, it must be true that...

- Every individual has the same probability of being selected.
- Every *combination* of individuals has the same probability of being selected.

Example: Number all the students in a school. Use randInt(1,3100,50) to randomly select 50 of these students.



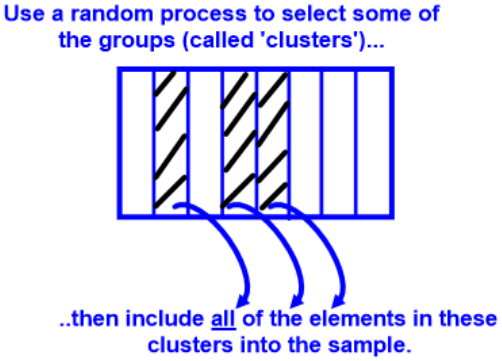
Another Example: We need to select a representative sample of 5 of the 40 athletes in a sports league to be featured in a promotional photo.

Write the names of the 40 athletes on identical slips of paper, put all the slips in a box and mix well, randomly draw out 5 slips of paper. The athletes whose names are on the slips are included in the sample for the photo.

2) Cluster sampling

For large populations it may be difficult to 'number' all the individuals for random selection. We can divide the population into groups called **clusters** which are each representative of the entire population, and then randomly select one or more clusters and then **include all the elements in the selected clusters as the sample.**

Example: If we want to randomly select households in a city, we may not want to number each household for selection. Instead, we could number the neighborhoods, randomly select some of the neighborhoods, and for those selected neighborhoods, include all of the houses in the sample.



(Note that even though we believe that each cluster is representative of the entire population, it is a subgroup which could have unique properties so **cluster sampling is not considered an SRS of the population.)**

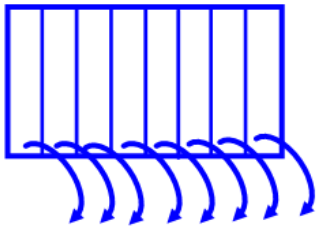
Different types of sampling - Good Sampling Techniques

3) Stratified Random Sample

Sometimes, we suspect that there might be differences between subgroups in a population and we want to make sure some of each subgroup is included in the sample.

We can divide the population into groups (called **strata**) so that the elements within each group are similar to each other, but there are differences between the groups that are meaningful to the analysis. We then use a separate, random process to select a Simple Random Sample (SRS) within each statrum to be included in the sample.

To include some from each stratum...



... we take separate SRSs from each of the strata to include into the sample.

(Again, **stratified random sampling is not considered an SRS of the population.)**

Cluster sampling is just a way of reducing the number of things we have to individually identify compared to Simple Random Sample (SRS).

To see why we might want to use Stratified Random Sampling, we will do the River Problem Activity in your Ch11,12,13 Required Practice Packet together.

You'll need a calculator - and need to 'seed' your random number generator first:

1) <type any 4 digit number>

2) press the 'STO' (store) key

3) press MATH, right arrow to PRB, rand, enter twice

<The River Problem Activity is in the Ch11,12,13 Required Practice Packet>

4) Systematic sampling

Using some systematic method, not randomness, to select items in a population.

Example: To select sentences to measure in a book, instead of numbering each sentence and selecting randomly, you could start with the first sentence and count sentence, systematically selecting every 100th sentence (selecting the 100th, 200th, 300th, etc. sentence).

Systematic sampling is not an SRS because not all samples have equal likelihood of being selected.

Even though randomness is not being used, this is still considered a valid sampling technique because a system (rather than human intuition) is being used.

Different types of sampling - Good Sampling Techniques

5) Multistage sampling

Stratified, cluster, and SRS can be combined in **multistage sampling.**

Example: If you want to select a sample of the sentences in a book...
*- **Stratify** a book into portions: 'beginning', 'middle', 'end'. Then randomly select a chapter within each stratum.*
*- **Cluster**: we could then divide each selected chapter from each stratum into pages(cluster = a page) and randomly select a page (cluster) for each stratum.*
- ***SRS***: finally, we could number and randomly select the sentences on the selected pages.

Different types of sampling - Bad Sampling Techniques

6) Random Sample

- Randomization is used in some way to select from the population.

In a random sample...

- Every individual has the same probability of being selected.
- ***Not*** every *combination* of individuals has the same probability of being selected.

Example: Flip a coin. If heads, use randInt(1,#boys,50) to randomly select 50 boys. If tails, use randInt(1,#girls,50) to randomly select 50 girls.

7) Convenience sampling

Selecting individuals in some way that is simply convenient to perform.

Example: Asking survey questions of the next 100 people who exit a grocery store.

Convenience sampling is not an SRS and is almost guaranteed to produce a biased result.

Different types of sampling

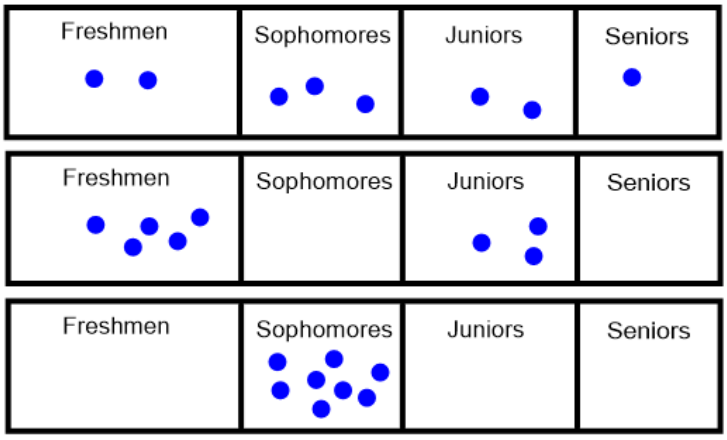
Example: Select a sample of students from a high school

Simple Random Sample

Slips of paper, each containing ID number of a student, for all students, into a box, randomly draw out students.

Things that could happen:

Every possible combination of students is equally likely to be selected.

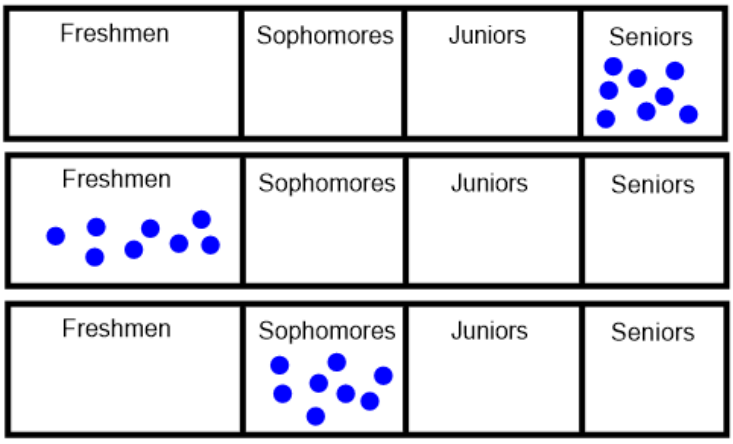


Cluster Sampling

Randomly select a class, and use the entire class as the sample.

Things that could happen:

Every student is equally likely to be selected, but you would only want to use cluster sampling if you believed there is no difference between the clusters.

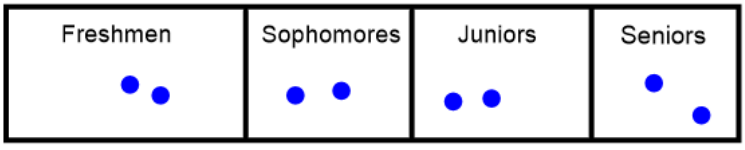


Stratified Random Sampling

Randomly select part of the sample from each class.

Things that could happen:

You are guaranteed to include members of each class. Especially useful if you believe there are differences between the strata.



Kinds of Bias...

Undercoverage bias: Some portion of the population is not sampled at all.

Example: 1936 presidential campaign (Alf Landon, F.D.Roosevelt) a poll was conducted asking, "Who will you support for president?" Alf Landon projected the winner 57% to 43%. Actually, F.D.R. won 62% to 37%. The poll was conducted by telephone, but in 1936 only wealthy households had phones, so a large portion of the country was not sampled.

Response bias: Anything in the survey design which influences the responses of the subjects.

Examples:

- *A survey may be designed which includes language or contextual references which are unfamiliar to some of the population.*
- *If a survey is conducted face-to-face and the interviewer is physically attractive, a participate may try to give the response they feel is what the interviewer 'wishes to hear'.*
- *If a participant's responses are not confidential, they may not respond honestly (due to fear of retaliation, for example from his/her boss).*

Voluntary response bias: If sampling people, and you ask for volunteers to participate you may get more participation from subgroups.

Example: Ann Landers asks parents to write in an answer, "If you had it to do over again, would you have children?". 70% of 10,000 respondents said no. But a more careful verbal telephone survey showed that 90% of parents are happy with their decision to have children. When people needed to volunteer to write in, mostly parents with a particular point of view were motivated to respond.

Nonresponse bias: Usually, if people are polled, participation in the survey is not mandatory. Some people usually decide not to participate (nonresponse). But it may be that subsets of the population are more likely to be nonresponsive.

Voluntary Response Bias: Researchers don't choose the sample - subjects volunteer to be in the sample.

Nonresponse Bias: Researchers select the sample, but then some subjects opt-out.

Judgment bias (not in book): If a sample is created by allowing the researchers to 'judge' who should and should not be included.
Example: A researchers including 2 people from each of many religions. (They will not be represented in the sample in the same proportions they appear in the population).

This last phrase is not really a bias in the same sense...

Confirmation bias (not in book): People evaluate all information based upon their current 'mindset' based upon previous experience. There are psychological reasons why humans tend to accept things that match their pre-conceived notions and reject as invalid things which do not.

...but is about how people are or are not convinced by results.