

AP Statistics

2019-01-08 Section 4.1 Sampling & Surveys

Notes taken by: **Noah Overcash**

A **population** consists of the entire group of individuals about which we want information. These individuals can be people, animals, objects, etc.

A **sample** is the part of the population from which we collect information. The information from this sample is used to draw conclusions about the entire population.

Planning a sample survey

State what population you wish to draw conclusions about

State what you want to measure and all variables

Describe how the sample will be chosen

A **convenience sample** is one where the members of the population who are easiest to reach are sampled.

This type of sampling often causes biases to form.

A **bias** occurs when a value is consistently over or under-estimated

A **voluntary sample** (wherein people choose themselves to vote/take a survey) also causes extreme bias.

Very opinionated people typically are the only ones who participate in such samples.

For example, a survey on a right-wing news site is likely to be biased towards right-wing responses.

Random sampling removes favouritism by the sampler and self-selection bias. This uses chance to select a sample and is a central principle of statistical sampling.

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 have an equal chance to be the sample actually selected.

Random sampling using a table:

Assign each member of the population a unique number of the same length (of digits)

Read consecutive groups of the length of digits

If the group is outside of the bounds of the numbers assigned to the population, continue to the next. If it is a repeat of a previous number, continue to the next. Continue iterating until the appropriate number of items are selected.

Example:

5 subjects in the population, numbered 1-5: 1 A, 2 B, 3 C, 4 D, 5 E

Using the numbers in the table, iterate and find subjects. Example random numbers: "04834 29665"

We want three members of the population.

Zero is not in the range 1-5; skip it.

Four is in the range, so subject 4 (D) will become part of our sample

Eight is not in the range; skip it

Three is in the range, so subject 3 (C) will become part of our sample

Four has already been selected; skip it

Two is in the range, so subject 2 (B) will become part of our sample.

Subjects D, C, and B have been selected.

A **stratum** is a group, **strata** is the plural

An alternative to SRS (simple random sample) is a **Stratified Random Sample**, where important groups are sampled.

Important groups (strata) are made from the population.

From there, a SRS is conducted of each stratum. Each stratum is then combined to form the full sample.

Strata can be chosen based on defining facts **known before the sample**. For example: male/female, freshman/sophomore/junior/senior, etc.

Cluster samples is another sampling method which samples individuals who are near others.

First, the population must be divided into clusters that mirror the characteristics of the population.

Then, an SRS of the clusters overall (not within each cluster) is taken, and the ENTIRETY of the selected clusters are sampled.

For example, clusters could be formed in a large arena based on row numbers. A SRS of row numbers is chosen

(say, rows 3, 4, and 7). The entirety of these rows are fully sampled.

Clusters and strata differ based on what they reflect of the population.

A cluster has to represent an entire population (e.g. all types of students, both political parties, every ticket type)

A stratum is just a way of dividing the population to make it easier administratively to sample.

Random samples help obey the laws of probability and make trustworthy inferences about an entire population.

Larger random samples help decrease margins of error and make better generalizations.

THIS WILL BE CONTINUED NEXT LESSON

2019-01-08