

DEFINING POPULATION PARAMETERS AND SAMPLE STATISTICS.

An important idea when starting statistics is to understand the difference between *populations* and *samples*. Populations consist of everything or everybody you want to measure; for example, “all women in the Bay Area between the ages of 18 and 54.” “House cats raised as pets in American households.” “The light bulbs we are manufacturing in this plant.” “Every human being on planet earth.” These are all *populations*. Populations have parameters. If you survey everybody in a population, this is known as a *census*.

A *sample* consists of as random a selection of objects or persons in your population as possible (called a *sampling frame*). There are many ways to do randomization, but the most important constraint is to be as close to random as possible. To use the above examples, we can talk about “500 women in the Bay Area between the ages of 18 and 54.” “50 house cats drawn from households in San Francisco, CA.” (with which we have to assume has many of the same qualities as all of America.) “10 light bulbs manufactured at this plant in the last year.” “At least 3 people from every country in the world.”

The science of statistics is based on measuring these samples; in other words, *samples* have *statistics*. Hypothesis testing is done by testing the sample statistics against assumptions about the population parameters.

Here is a table of symbols for both population parameters and statistical samples:

NAME OF SYMBOL	STATISTIC (SAMPLE)	PARAMETER (POPULATION)
Mean (Average)	\bar{x}	μ
Variance, Standard Deviation	s^2, s	σ^2, σ
Correlation Coefficient	r	ρ
Slope of Least Squares Regression	b_1	β_1
Intercept of Least Squares Regression	b_0	β_0
Proportion	\hat{p} (alternatively, p)	P (alternatively, π)
Number (in sample, in population)	n	N

SOME BASIC TYPES OF RANDOMIZATION:

Simple random sample: Every object or person in the sample frame has an equal chance of being selected.

Systematic sampling: Lists the objects or persons in the sampling frame in an order, then selects the n th person or object on the list. Example: an alphabetical list of SFSU students where every 25th student is selected for a survey.

Stratified sampling: The sample frame is divided into parts or sections, and a sample is drawn from each part or section. Examples: 10 people chosen at random from every Southwest flight on a certain day. 20 people from each state in the United states. 5 students in each math class at SFSU.

Cluster sampling: The sample frame is also divided into parts or sections, but only certain parts or sections are used. However, all members of the parts or sections are sampled. Examples: All passengers on flights with even numbers. All students in the 8 AM and 2 PM MWF math classes,

Convenience sampling: The members of the sample frame volunteer or self-select. Examples: Internet polls, mailed surveys that need to be filled out and returned. This is considered the least reliable and therefore the least desirable method of sampling.