Sampling: Summary

The Course So Far: Describe Observations in a Sample
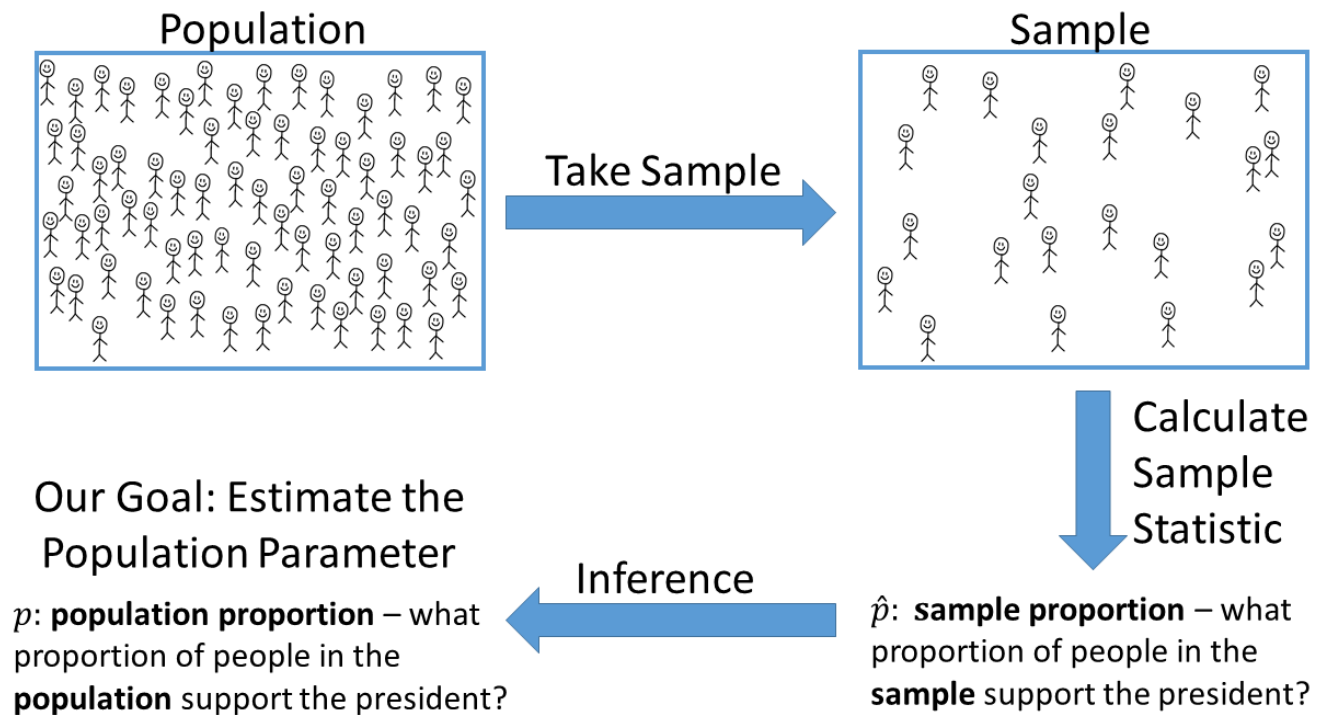
Variable Type(s)	Plot	Numeric Summary
1 Categorical	Bar	(Marginal) distribution
2 Categorical	Bar	Joint Distribution, Conditional Distribution
1 Quantitative	Histogram or Density	mean, median, quantiles, standard deviation, variance, IQR
1 Categorical, 1 Quantitative	Density Plot or Box Plot	summary statistics of the quantitative variable for each level of the categorical variable
2 Quantitative	Scatter Plot	correlation

From Now On: Use Observations in a Sample to Estimate Something about a Population

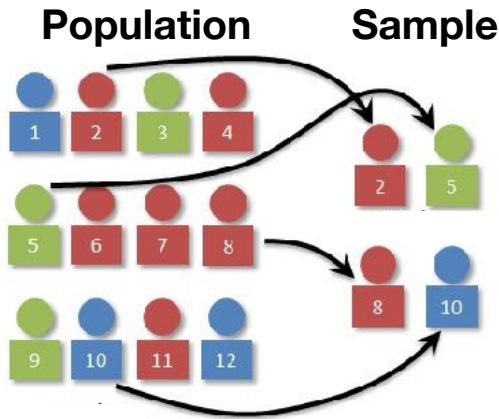**Population Parameter**: a number summarizing a variable across the whole population

Sample Statistic: a number summarizing a variable for the sample

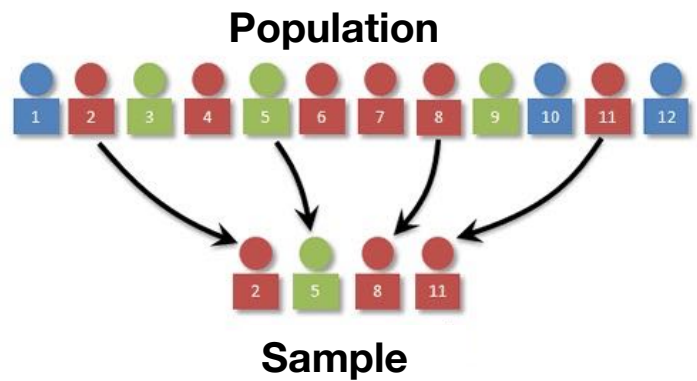Our Hope: The sample statistic will be a good estimate of the population parameter.

Population

Sample

Take Sample

Calculate Sample Statistic

Our Goal: Estimate the Population Parameter

Inference

p: **population proportion** – what proportion of people in the **population** support the president?

$\hat{p}$: **sample proportion** – what proportion of people in the **sample** support the president?

How Do We Get Our Sample?

Simple Random Sample

Population **Sample**

Systematic Sample (every 3rd)

Population

Sample

Stratified Sample

Population **Sample**

Strata
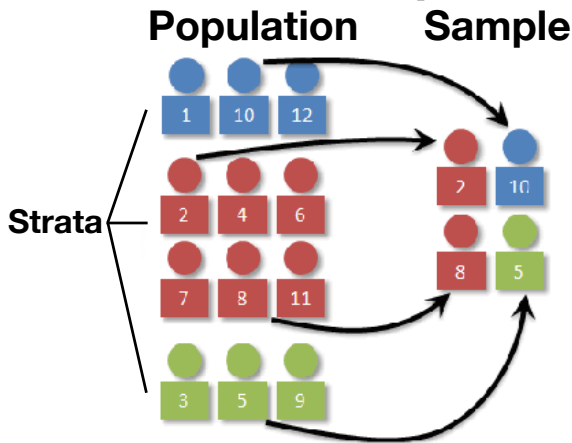
Cluster Sample
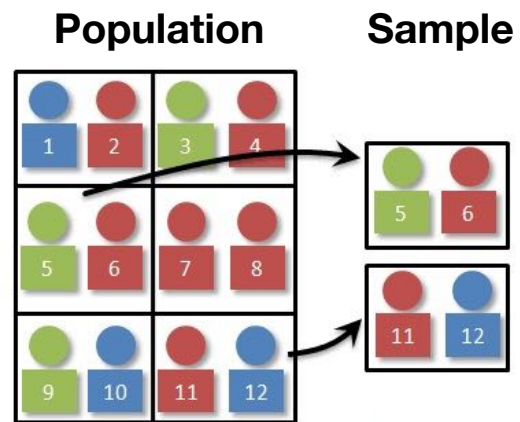
Population **Sample**

Bias

- For the sample statistic to be a good estimate of the population parameter, the sample needs to be **representative** of the population.
- Definition: Sampling methods that tend to over-emphasize or under-emphasize some characteristics of the population are **biased**.
- Common sources of bias:
    - **Sample Volunteers/Convenience Sampling**: just include people in the sample who are easy to recruit
    - **Bad Sampling Frame/Undercoverage**: only choose your sample from among a subset of the population
    - **Nonresponse**: some people selected for your sample choose not to respond
    - **Response bias**: your phrasing or survey design encourages people to answer a certain way

Sampling Variabilty

- Every sample you take is different!
- Imagine taking 10 different samples of people in the US
- Each group of people you select will have different numbers who support the president
- So each sample will have a different sample statistic (different proportion of the sample who support the president).
- Definition: The **sampling distribution** is the distribution of values of the sample statistic that you would get from all possible samples of a given size. (We will explore this more in the lab today and in Chapter 17.)