

Some Questions to Ask Before Producing Data

- What is the group of interest?
- What information about the group are we interested in?
- How do we collect this information?

The Population

- The entire group of individuals about which we want to get information.
- If we can quickly and/or cheaply do so, we perform a **census**:
 - The collection of data from *every* member of a population.
- Often, it is too time-consuming or expensive to obtain data for the entire population.
- How do we get around these constraints?

The Sample

- The subset of the population from which we actually collect data.
- Selected according to some prescribed manner.
- For this sample data, we:
 - Organize
 - Describe
 - Analyze



Example

- The 2000 Census tried to gather basic information from every household in the United States. A "long form" requesting much more information was sent to about 17% of households.
- Population =
- Sample =



How do we describe objects in a data set?

- A **variable** is any characteristic whose value may change from one object to another in a population.
- Univariate data
 - Observations on a single variable
- Bivariate data
 - Observations on each of two variables
- Multivariate data
 - Observations made on more than one variable



- Descriptive Statistics:
 - Summarize and describe key features of the data
 - Includes graphical methods and calculation of numerical summary measures
 - Discussed in Ch. 1
- Inferential Statistics:
 - Generalize and draw conclusions about the population from the sample data
 - Discussed in Ch. 6 16





Simple Random Samples (SRS)

- Every possible sample of a specified size has an equal chance of being selected.
- Every individual also *has an equal chance* of being selected for the sample.

Example							
Researchers wish to test a promising weight loss medication on a group of overweight persons. The list below needs to be randomly divided into two groups: a treatment and control group.							
[01	Birnbaum	35	11	Moses	25	
	02	Brown	34	12	Nevesky	39	
	03	Brunk	30	13	Obrach	30	
	04	Cruz	34	14	Rodriguez	30	
	05	Deng	24	15	Santiago	27	
	06	Hernandez	25	16	Smith	29	
	07	Jackson	33	17	Stall	33	
	08	Kendall	28	18	Tran	35	
	09	Loren	32	19	Wilansky	42	
	10	Mann	28	20	Williams	22	

Stratified Sampling

- First, divide the population into at least two subgroups.
 - All of the individuals in each subgroup should share some common characteristic.
 - The subgroups are called **strata**.
- Then draw a SRS from each subgroup.



