Terminology and Introduction

Math 122 - Introduction to Statistics and Probability

January 14, 2012

Terminology

In statistics, we study DATA from a particular POPULATION.

Terminology

- The POPULATION is the collection of all individuals being studied.
- Data are collections of observations (measurements, genders, survey responses, etc).

Terminology

Data are collected in EXPERIMENTS and OBSERVATIONAL STUDIES

- In an OBSERVATIONAL STUDY we observe and measure characteristics without trying to affect the subjects being studied.

  Example:
  Polls/surveys are observational studies.
  If I want information about how many CU students eat breakfast in Janzow, I can count them or ask students.

- In an EXPERIMENT we apply some treatment to subjects and then observe its effects.

  Example: To test a medication, you can give the medication to sick patients and see what happens.

Terminology

Usually, a population is too big to collect data from every single member, so data is collected from a smaller part of the population.

- A CENSUS is a collection of data from every member of a population.
- A PARAMETER is a measurement based on a census of the entire population.
- A SAMPLE is a collection of some (but not all) members of a population.
- A STATISTIC is a measurement based on a sample from the population.

Parameter vs. Statistic

Parameter $\leftrightarrow$ Population

Statistic $\leftrightarrow$ Sample

A basic assumption of this class is that under the right assumptions, a statistic can be a good approximation to a parameter.

Parameter ↔ Population

A parameter associated with the population of all full-time undergraduate students at Concordia is their average height.

- To find this parameter, I would need to measure every full-time undergraduate student on campus and average the results.
- This measurement is a parameter.

Statistic ↔ Sample

To approximate the average height of a full-time undergraduate student on campus at Concordia I can collect 50 random students and measure their heights.

- This measurement is a statistic.
- Depending on how I select the population, this statistic may be a good approximation of the actual parameter.

Types of Data

There are two main types of data that we will look at

- QUANTITATIVE DATA - numbers representing actual counts or measurements
- CATEGORICAL DATA - names or labels that do not come from actual counts or measurements. These usually separate subjects into groups.

Quantitative or Categorical?

- Gender
- Height
- Weight
- Number of siblings
- Hours of sleep
- Home state
- Eye/hair color
- Credit hours
- Right/left handed
- Four random digits
- Pulse

Types of Quantitative Data

Quantitative date comes in two forms

- CONTINUOUS DATA comes from a set of possible values that covers a range without gaps or jumps

  The amount of water used to wash dishes in Janzow each day.

- DISCRETE DATA comes from a set of possible values which are isolated from each other by gaps or jumps

  The number of cats in Ruby nebraska.

Continuous or Discrete?

- Height
- Weight
- Number of siblings

- Hours of sleep
- Credit hours
- Pulse

Statistical Thinking

There are five factors that should be considered in any statistical analysis

- Context of the data
  - Affects what type of statistical analysis can be used.
- Source of the data
  - Bias? Corruption of data?
- Sampling method
  - **Bad sampling techniques render statistics useless!**
  - Worst mistake - VOLUNTARY RESPONSE SAMPLE (or SELF-SELECTED SAMPLE)
- Conclusions
- Practical implications
  - Results may be statistically significant but not practically significant.

Sampling

- A SIMPLE RANDOM SAMPLE of n subjects is a sample which is collected in such a way that every possible sample of the same size n has the same chance of being selected.
- This is the IDEAL.
- Most statistical tools require a simple random sample.
- But...

Sampling

A RANDOM SAMPLE is a sample which is selected in such a way that every individual member of the population has an equal chance of being selected.

Sample

- SIMPLE RANDOM SAMPLE - all samples equally likely

  To select a simple random sample of 50 students at CU, I could randomly select names from a directory.

- RANDOM SAMPLE - all individuals equally likely

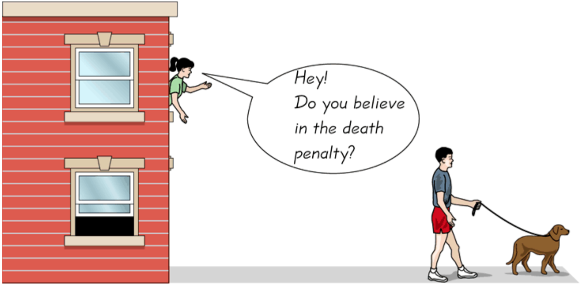
  To select a random sample (not simple) of 50 students at CU, I could randomly select 5 dorm floors and then randomly select 10 people from each floor. Everyone is equally likely to be selected, but some combinations cannot happen.

Systematic Sampling

Select some starting point in a list and then select every k^{th} subject for some k.

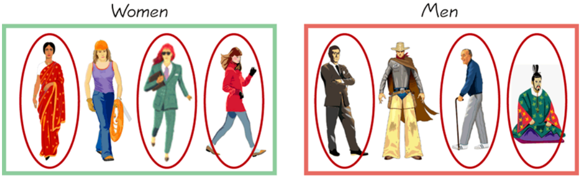
Convenience Sampling

Use results that are easy to get.



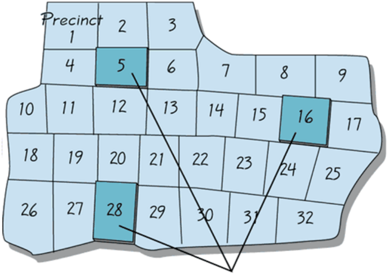
Stratified Sampling

Divide the population into groups with similar characteristics and then select a sample from each group.



Cluster Sampling

Divide the population into sections or clusters. Randomly select some of the clusters. Choose all of the members of the selected clusters.



Interview all voters in shaded precincts.

Critical (Skeptical) Thinking

- Bad samples
- Small samples
- Voluntary response
- Correlation vs. Causality
- Incorrectly reported responses
- Wording of questions
- Non-response/missing data
- Motivation
- Precise numbers
- Deliberate distortions