Populations and Samples

STAT-UB.0001 Statistics for Business Control

Ningshan Zhang

IOMS Department
nzhang@stern.nyu.edu

July 3, 2018

Course Logistics

See the syllabus on NYU Classes.

Introduction to Statistics

The word 'Statistics' means

- In ordinary conversations, "statistics" means a collection of numbers. (Sport statistics, census statistics, etc.)
- In this course, "statistics" is an analytical discipline.

Wikipedia: Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data.

"Statistics is using a *sample* to make a statement about a *population*."

Introduction to Statistics

Why statistics? They say...[1]

- Apply to various fields.
- Use math to solve problems.
- Not a major, but a set of skills.
- Open so many opportunities.

[1] https://youtu.be/d5urHxg_KSs

Introduction to Statistics

Statistics is used in a variety of fields:

- Finance (Banking, Risk Management)
- Marketing (Targeted Advertising)
- Health (Drug Development)
- Politics (Election Polls)
- Education (College Admissions Data, Teacher Evaluations)
- Many more...

Course Topics

- Populations and Samples
- Descriptive Statistics
- Probability
  - Basic Probability
  - Conditional Probability
  - Discrete Random Variables
  - Continuous Random Variables
- Inferential Statistics
  - Sampling Distributions
  - Confidence Intervals
  - Hypothesis Tests

Populations vs Samples

"Statistics is using a *sample* to make a statement about a *population*."

Population

The set of items or individuals that we are interested in studying and drawing conclusions about.

- What we care about. The world.

Sample

A subset of items or individuals from the population, i.e. the data.

- What we have access to. A measurement of the world.

Populations vs Samples

Example

You want to estimate the average amount of time that NYU undergraduates spend on social networking websites per day.

- What is the population?
- How to collect a sample?

Representative

A representative sample is a sample whose properties accurately reflect those of the population.

- Not possible to guarantee a representative sample in practice.
- If we know the population, take the sample to be equal to the population, which is representative.
- Otherwise, take an *unbiased* sample.

Unbiased sample

A sample is unbiased when every member of the population has an equal chance of being included in the sample.

- An unbiased sample is not guaranteed to be representative, but it will be representative on average.
- A large unbiased sample will be approximately representative.

If some member of the population is more likely to be in the sample, then the sample is biased.

Sources of Bias

Selection Bias

There is a systematic tendency for one group to be overrepresented or underrepresented. In other words, some type of individuals are more (or less) likely to be surveyed.

Example: The Wall Street Journal wants to estimate its popularity amongst college students. To do this, an employee surveys a random sample of Stern undergraduates.

Any problem?

Sources of Bias

Nonresponse Bias

The researcher is unable to obtain data on all experimental units selected for the sample. In other words, some type of individuals are more (or less) likely to respond.

Example: A university uses course faculty evaluations to assess the quality of its instructors. There is no consequence if a student does not participate.

Any problem?

Sources of Bias

Example: During WWII, statistician Abraham Wald[2] was asked to help the British decide where to add armor to their bombers. Wald recommended adding more armor to the places where there was no damage!

Why?

[2] Known for the Wald test.

Summary

- Populations and samples
- Representative sample
- Unbiased sample
- Sources of bias