Stats starts here

Al Nosedal
University of Toronto

Summer 2019

My momma always said: "Life was like a box of chocolates. You never know what you're gonna get."

Forrest Gump.

Exercise

- Take a box of *SMARTIES*® and open it. Don't eat them...yet!
- Count the number of candies inside your box.
- Now you can eat them.
- Wait for your turn.
- Tell me your name and how many candies you got.

Definitions

- Statistics is the science of collecting, describing, and interpreting data.
- Individuals are the objects described by a set of data. Individuals may be people, but they may also be animals or things.
- A variable is any characteristic of an individual. A variable can take different values for different individuals.

Descriptive Statistics

Most of the statistical information in newspapers, magazines, company reports, and other publications consists of data that are summarized and presented in a form that is easy for the reader to understand. Such summaries of data, which may be tabular, graphical, or numerical, are referred to as descriptive statistics.

Statistical Inference

Many situations require information about a large group of elements. But, because of time, cost, and other considerations, data can be collected from only a small portion of the group. The larger group of elements in a particular study is called the population, and the smaller group is called the sample. As one of its major contributions, Statistics uses data from a sample to make estimates and test hypotheses about the characteristics of a population through a process referred to as statistical inference.

Key Statistical Concepts

- A **population** is the group of **all** items of interest to a Statistics practitioner. It is frequently very large and may, in fact, be infinitely large.
- A descriptive measure of a population is called a **parameter**.
- A **sample** is a set of data drawn from the studied population. A descriptive measure of a sample is called a **statistic**.
- **Statistical inference** is the process of making an estimate, prediction, or decision about a population based on sample data.

Example

According to "50 Amazing and Strange Astronomy Facts", an article posted on astroceanomy.com, only 55% of all Americans know that the sun is a star.
An elementary school student thinks that his fifth grade classmates are smarter than the average American. He asks a random sample of 50 of his classmates this question, "Is the sun a star?" Thirty-eight of the students say "Yes." He reports that he is 95% confident that the true proportion of fifth graders at this school that would answer "Yes" to the question, "Is the sun a star?" is between 0.642 and 0.878.
Determine whether the scenario is an example of descriptive or inferential statistics.

Homework

This course uses R. R is an open-source computing package which has seen a huge growth in popularity in the last few years. R can be downloaded from https://cran.r-project.org

Please, download R and bring your laptop next time.

Homework

Please read:

- 1.1 Distinguish the difference between descriptive and inferential statistics
- 1.2 Given a study scenario, identify the population and sample