

Stat 154: Elementary Statistics

Jongyun Jung

Minnesota State University, Mankato

Ch 1: Introduction

jongyun.jung@mnsu.edu

January 14, 2019

Overview

1 Introduction

- What is Statistics?
- Variables vs. Data
- Descriptive statistics vs. Inferential statistics
- statistic vs. parameter

2 Population Parameters and Sample Statistics

3 Types of Data

- Levels of Measurements

4 Data Collection Method

5 Sampling Methods

- Random Sampling or Simple Random Sampling
- Stratified Random Sampling
- Cluster Sampling
- Systematic Sampling
- Multi-State Sampling and Convenience Sampling

What is Statistics?

- It is the discipline of how to collect, organize, analyze and draw conclusions from data.
- It is a science of gaining knowledge of a fact, often by using sample data, and often the decision is made with uncertainty.

Variables vs. Data

- **Variables** are the characteristics of an individual that are measured or observed.
- **Data** are the values that the variables can attain.

Descriptive statistics vs. Inferential statistics

- **Descriptive statistics** refers to the methods for organizing and summarizing the information in a data set.
 - Use numbers, graphs, and tables to describe the data set as a first step in the data analysis
 - Ch 1 and Ch 2 study for the descriptive statistical methods
- **Inferential statistics** refers to the methods of estimating and drawing conclusions about population characteristics based on the information contained in a sample using probability statements.

statistic vs. parameter

- **statistic** is a numerical measure that describes an aspect of a sample.
 - any function of a sample is a statistic.
 - It is a known quantity when a sample is selected.
 - It is a random quantity depends on the sample that got selected among all possible samples
- **parameter** is a numerical measure that describes an unknown population characteristic.

Population Parameters and Sample Statistics

- **Population** consists of all the subjects under study.
 - **Quantitative**: This is a numerical data, which are amounts or measurements. Height, weight, body temperature and number of items are some examples.
 - **Qualitative**: This can be separated into different categories according to some characteristics or attributes, such as gender, religious preference, geographical locations, quality of a product, etc.
- **parameter** is a population characteristic often identified by a constant, usually unknown.
- **sample** consists of only a subset of a population.
- **statistics** is a function of the sample measurements.
- **subjects** are the items on which the measurements are taken.

Types of Data

- **Qualitative:** This can be separated into different categories according to some characteristics or attributes, such as gender, religious preference, geographical locations, quality of a product, etc.
- **Quantitative:** This is a numerical data, which are amounts or measurements. Height, weight, body temperature and number of items are some examples.
 - **Discrete** data are the data that can take countable numbers (finite or infinite). Ex: the number of children in a family, the number of passengers on a plane, etc.
 - **Continuous** data are numbers having infinite possibilities even if they are represented by a whole number, such as, age, weight, height, income, distance, length, etc.
- **Non-numeric:** Any form of data that are presented using a verbal description and/or modes such as graphs, pictures, etc.
- **Numeric:** Data are presented using numeric symbols.

Levels of Measurements

- **Nominal level** is the measurement of data that have just labels or categories; no orderings are established, such as the name of person or a subject, different categories without any orderings, different colors such as red, blue, white, etc.
- **Ordinal level** is the measurement of data that can be ordered.
- **Interval level** is the difference between two measurements can be quantified, that is the difference can be measured exactly.
- **Ratio** is the interval measurements that have the natural zero, such as weights, prices, heights, counts, grade point average, etc.

Data Collection Method

- **Observational study:** the researchers often collect the data by observing the facts and then drawing conclusions based on the observed data.
- **Experimental study:** the researchers decide on the pertinent variables to find out how the variables under study have influenced the response or outcomes.

Random Sampling or Simple Random Sampling

- **Random Sampling or Simple Random Sampling:** the most common method of sampling. It is the best method of sampling when a population is homogeneous and each element in the population has an equal chance of being included in the sample.
 - **Picking Names Out of a Hat**
 - **Picking Numbers from a Random Generating Table**
 - Table 1: A Random Number Generating Table in Appendix II has 1,000 random digits. To generate the random number, follow this steps:
 - 1 Step 1: Arbitrarily select a position to start.
 - 2 Step 2: Consider the size of the population to be N with the maximum number of digits d . If the selected number containing d digits is less than or equals N , consider it a sample unit. If the selected number containing d digits is more than N , divide the number by N and consider the remainder of the division as a sample unit. If the remainder is zero, N is the sample unit. Continue the process for consecutive d digits until n such sample units are found. Ignore, if a number is repeated.
 - **Picking Numbers using Computer Software**

Stratified Random Sampling

- **Stratified Random Sampling:** when a population is divided into different groups to ensure proper representation of different groups in the sample, samples usually are selected proportionately from each group.
 - Ex: Taking samples from each gender group to ensure equal gender representation.

Cluster Sampling

- **Cluster Sampling:** when a population is divided into a large number of groups, a random sample of a few groups is selected, and a complete survey is performed within each sampled group.
 - Ex: on a large college students there are 25 different elementary statistics sections. To sample elementary statistics students, an investigator decides to randomly sample 5 sections and survey those 5 sections to obtain their responses about the survey questions.

Systematic Sampling

- **Systematic Sampling:** when the complete frame (the list of the population individuals) is not available, an investigator may decide to sample periodically.
 - Ex: an investigator decides to consider every 17th item in the production line for every thousand items produced and obtain 25 such items to ensure the quality of the product.

Multi-Stage Sampling and Convenience Sampling

- **Multi-Stage Sampling:** this is similar to cluster sampling, but each sampled group is subdivided into groups, and samples of subgroups are obtained.
- **Convenience Sampling:** researchers collect the data from the subjects that are convenient.

Exercises

- 1 Determine whether the following statements indicate a **parameter** or a **statistic**:
- a A researcher determines that of all 20-year-olds in her city, 19 % are married.
 - b A researcher examines the records of all the registered voters in one city and finds that 45 % are registered Democrats.
 - c A sample of 40 patients is selected from the patients admitted to the emergency room at a hospital, and it is found that 32 % have no health insurance.

Exercises

- ① Identify whether the measurement scale in the following statements are **continuous** or **discrete**:
 - a The number of automobile accidents per year in Montana, a state in the USA
 - b The length of time to complete a phone call
 - c The number of telephone calls made from a telephone booth
 - d The weights of truckloads of factory products
- ② Identify whether the measurement scales in the following statements are **Nominal**, **Ordinal**, **Interval** or **Ratio**:
 - a Nationalities of survey respondents
 - b Test grades as A, B, C, D or F
 - c Year of a historical event
 - d The colors of cars driven by college students

Exercises

- 1 Identify whether the following statements indicate the data are **experimental** or **observational**:
- a A sample of fish is taken from a lake to measure the effect of pollution from a nearby factory on the fish
 - b A study of female college graduates who go to graduate school
 - c A clinic gives a drug to a group of 10 patients and a placebo to another group of 10 patients to find out if the drug has an effect on the patients illness
 - d A marketing firm does a survey to find out how many people use a product. Of the 100 people contacted, 15 say they use the product.

Exercises

- 1 Classify the following sampling methods indicated by the respective statements as either **random**, **stratified**, **systematic**, **cluster** or **convenience**:
- a A complete list of all stockholders is compiled, and every 500th name is selected from each brokerage house
 - b A pollster uses a computer to generate 420 random numbers and interviews the voters corresponding to those numbers
 - c All of the stockholders' zip codes are collected, and five stockholders are randomly selected from each zip code
 - d At the annual stockholders' meeting, a survey is conducted of all who attend

Exercises

- 1 Select a random sample of 12 households from 150 neighborhood households using the random number table, Table 1 in Appendix II. Start with a random starting point and do not forget to mention the starting point.

Solution.

- 1 We need to sample 12 households from a population of 150 neighborhood. This population consists of 3 digits.
- 2 Let us start from the 2nd row and the 6th column of Table 1 of Appendix II.
- 3 It starts with the number of 796.
- 4 Since the 796 is a greater than our population of 150, we will divide 796 by 150. Then the remainder is 46.
- 5 Thus, our first random sample data is 46. Similarly, the next three digits of number is 685. Then the remainder is 85.
- 6 By doing so, the randomly selected data is 46, 85, 125, 134, 69, 48, 147, 23, 36, 106, 29, 112

References



Mezbahur Rahman, Deepak Sanjel, Han Wu. Statistics Introduction, Revised Printing

Kendall Hunt