1. Introduction

STAT*2060: Statistics for Business Decisions

Nishan Mudalige

Department of Mathematics and Statistics
University of Guelph

Fall 2014

Table of Contents

Table of Contents

Basics

Definition (Statistics)

Statistics is the science of collecting, classifying, summarizing, analyzing and interpreting data.

Definition (Descriptive Statistics)

Numerical and graphical methods used to analyze, interpret and represent data.

Definition (Inferential Statistics)

Use information from a sample to make generalizations about a population.

Table of Contents

Types of Data

Definition (Quantitative data)

Numerical data that can be measured.

- Examples:
  - Number of hours you studied this week
  - Distance from your house to the university
  - Area (in ft^2) of the floor of a concourse

- Quantitative data can also be divided into:
  - Discrete data
  - Continuous data

Types of Data Ctd...

Definition (Discrete data)

Measurements can only take specific vales.

- Examples:
  - The number of heads you get when you toss a coin 5 times
  - The number of rooms in a residence

Definition (Continuous data)

Measurements can take any value within a specific range

- Examples:
  - Height
  - Weight
  - The time taken to complete a task

Types of Data Ctd...

Note

Note that quantitative data can also be categorized as:

- *Interval data*
- *Ratio data*

However we will not go into this much detail in this course.

Types of Data Ctd...

Definition (Qualitative data)

Data can not be measured on a numerical scale. Instead the data falls into categories.

- Examples:
  - Favourite flavour of ice-cream
  - Day of the week (Mon, Tue, ...) that an event occurred
  - City you live in

Types of Data Ctd...

Note

Note that qualitative data can also be categorized as:

- *Nominal data*
- *Ordinal data*

However we will not go into this much detail in this course.

Exercise

	Quantitative data	Qualitative data
Age		
Age cohort		
Weight		
Gender		
Time taken to run a lap		
Cost of a textbook		
Major at university		
Classroom seating capacity		
Interest rate		
Average annual return of a stock		
Shoe size		
Year of university (According to the # of credits completed)		

Table of Contents

Types of Studies

Definition (Observational Studies)

We observe units and take measurements without assigning treatments

- Examples:
  - The effect of smoking on lung capacity.
  - The effect of heroin on brain function.

Types of Studies Ctd...

Definition (Experimental Studies)

Treatments are assigned to units and then the effects are observed and measured

- Examples:
  - Testing whether the packaging of a product is appealing to consumers before sending it out to the market.
  - Providing a group with windows pc's and another similar group with mac's and measuring the time taken to complete certain tasks.

Types of Studies Ctd...

Definition (Sample surveys)

Data is obtained from a selected part of the population using a survey instrument.

- Similar to an experimental designed study.
- Examples:
  - Satisfaction survey of guests at a resort.
  - CSA O-week concert survey.

Types of Studies Ctd...

- Other types of studies include:
  - Case control studies
  - Cohort studies
  - Cross sectional
  - Prospective studies
  - Retrospective studies
- These will not be discussed but are mentioned for completeness

Table of Contents

Introduction to Inferential Statistics

Definition (Unit)

Objects we are interested in and which measurements are recorded.

Definition (Population)

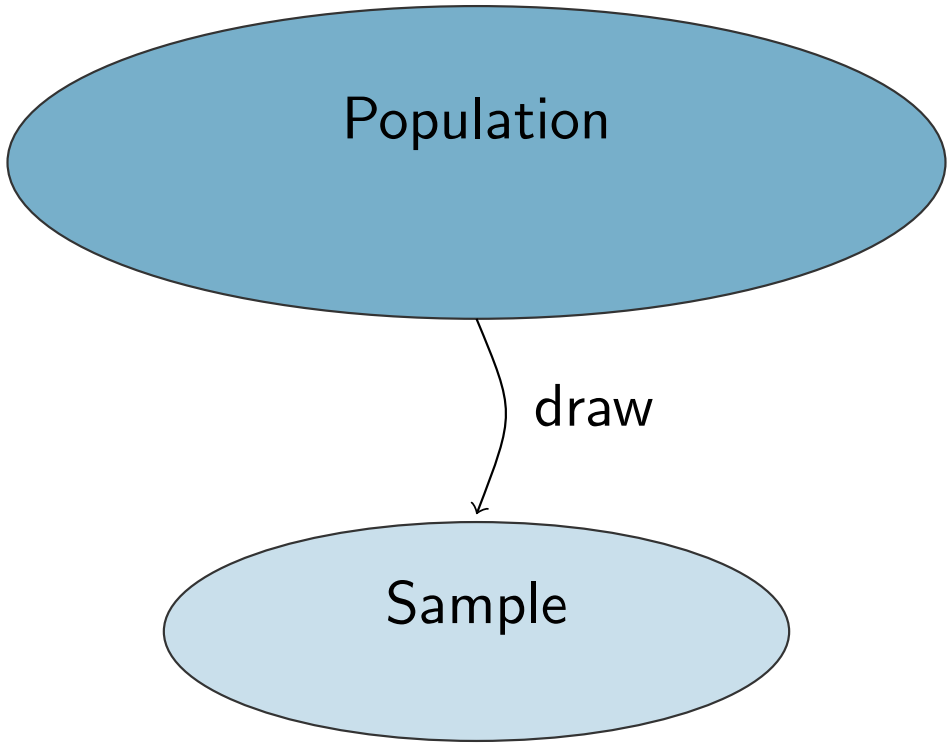
A (large) group of units that we are interested in studying.

Definition (Sample)

A subset of the population.

Intro to Inferential Statistics Ctd...

Our picture so far:



Intro to Inferential Statistics Ctd...

Definition (Parameter)

A numerical measure of a population.

- Usually not known since it is very hard to take measurements on every unit in a population.
- Parameters of interest for us:
  - μ : Population mean
  - σ : Population standard deviation
  - p : Population proportion

Intro to Inferential Statistics Ctd...

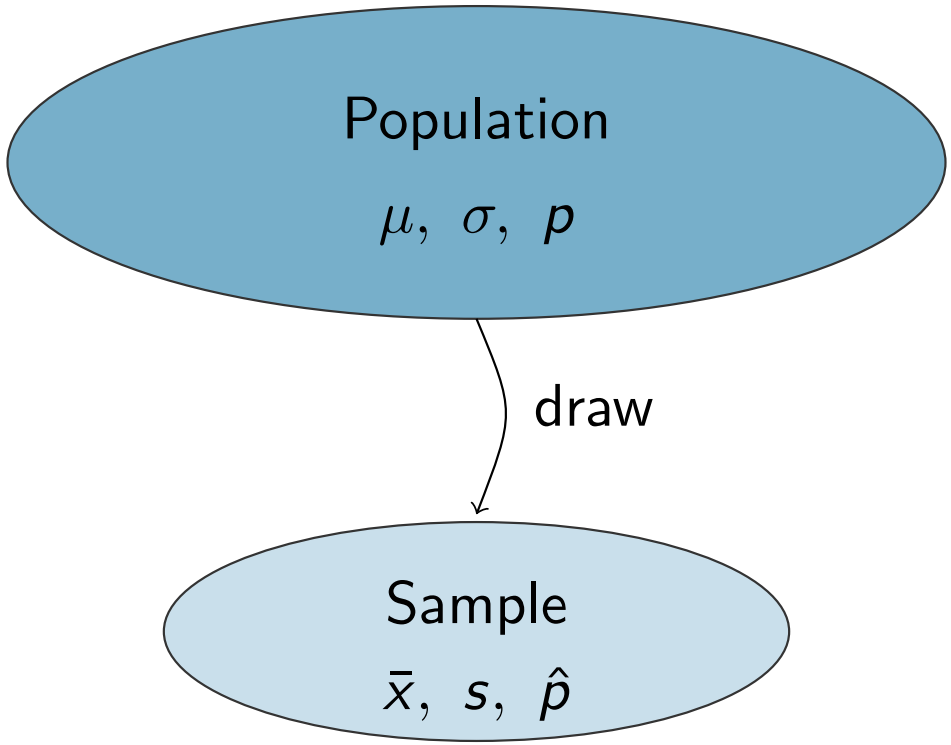
Definition (Statistic)

A numerical measure of a sample.

- Known as it is much easier to take measurements on a sample and calculate statistics.
- Statistics of interest for us:
  - $\bar{x}$: Sample mean
  - s : Sample standard deviation
  - $\hat{p}$: Sample proportion
- A statistic is also often referred to as an estimator.

Intro to Inferential Statistics Ctd...

Updated picture:



Intro to Inferential Statistics Ctd...

Goal (Statistical Inference)

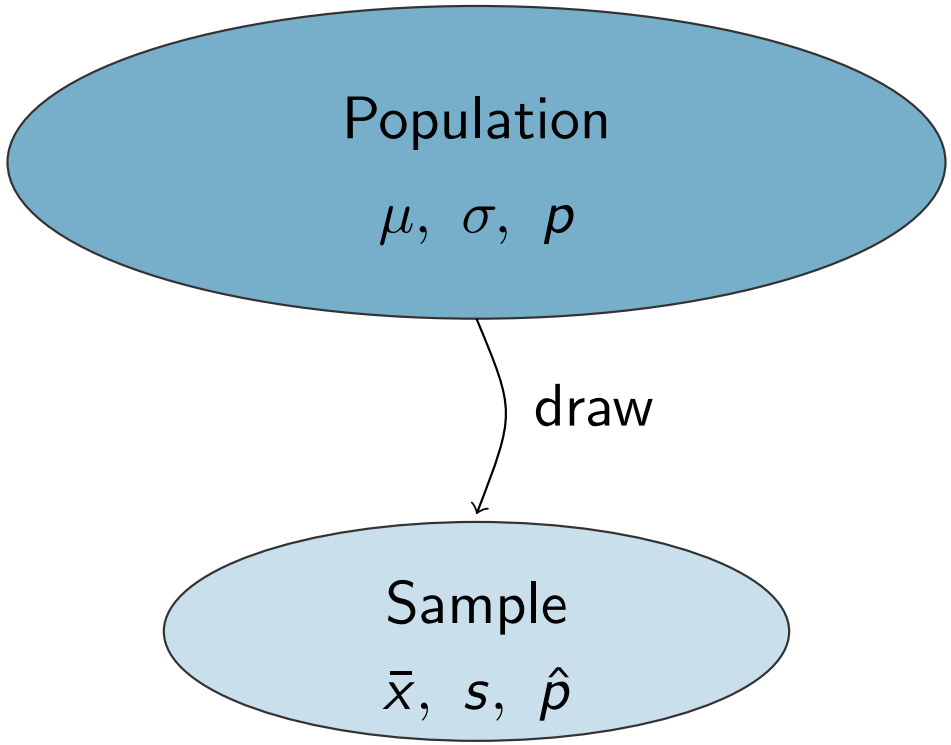
To calculate estimators of population parameters, and to quantify the accuracy of these estimators with probabilities.

- Interested in parameters in a population.
- Calculate statistic from a sample.

$$\text{Statistics} \xrightarrow{\text{Estimate}} \text{Parameters}$$

- Quantify how confident we are about a statistic estimating its corresponding parameter.

Intro to Inferential Statistics Ctd...



$$\bar{x} \xrightarrow{\text{Estimate}} \mu$$
$$s \xrightarrow{\text{Estimate}} \sigma$$
$$\hat{p} \xrightarrow{\text{Estimate}} p$$

Table of Contents

Information on Sample Data

- It is usually very difficult (or impossible) to measure every unit in a population.
- It is more feasible and practical to draw a sample and take measurements on all units in this sample.
- We would like our sample to be representative of the population it is drawn from.
- This is because we would like to make generalizations about the population based on our sample.

$$\text{Sample} \xrightarrow{\text{Generalize}} \text{Population}$$

Types of samples

Definition (Random Sample)

Each unit in a population has an equal chance of being selected. Also each of combination of size n has an equal chance of selection.

- Other sampling techniques include:
  - Stratified sampling
  - Cluster sampling
  - Multistage sampling
- However these will not be covered in this course

Issues with Collecting Sample data

- Collecting sample data is expensive (in terms of time and money).
- Data may be biased:

Selection bias	:	Sample is not representative of the population since a subset of the population has no chance of being selected for the sample
Nonresponse bias	:	There may be a reason that certain respondents refuse to participate. As such we lose information.
Measurement error bias	:	The response measured and recorded for an individual unit is not correct.

- Ideally we would like to draw a random sample.

Table of Contents

Homework

Readings

- Introduction to Probability and Statistics
  Read 1.1 — 1.4.1 (Pages 2 — 15);
  Read 1.6 (Pages 21 — 22)

Exercises

- Custom Edition of Statistics for Business and Economics (McClave, Benson, Sincich):
  Exercises 1.15 — 1.17;
  Exercises 1.25 — 1.31

- 11^{th} Edition of Statistics for Business and Economics (McClave, Benson, Sincich):
  Exercises 1.14 — 1.17;
  Exercises 1.21 — 1.28