

STAT 2593

Lecture 001 - Populations, Samples, and Processes

Dylan Spicker

Populations, Samples, and Processes

Learning Objectives

1. What is statistics, and why do we care?
2. What are the components of statistical inference?
3. How are data categorized?
4. What are the different roles for statistics?
5. What are the different types of statistical studies?

The Big Question

What is **statistics** and why *should* you care?



ChatGPT



6σ



DOW JONES



THE LEGEND OF
ZELDA
TEARS OF THE KINGDOM



A Concrete Example

How might we be able to answer the
question:

Does pay discrimination exist?

A Concrete Example

- ▶ Can we survey every single person and ask them their wages, then look at discrepancies?

A Concrete Example

- ▶ Can we survey every single person and ask them their wages, then look at discrepancies?
- ▶ Can we look at all the wages at a single company, and see if there appears to be discrimination?

A Concrete Example

- ▶ Can we survey every single person and ask them their wages, then look at discrepancies?
- ▶ Can we look at all the wages at a single company, and see if there appears to be discrimination?
- ▶ Can we look at available internet data, on salary websites?

The Problem

None of these techniques will give us accurate information about the problem as a whole.

What is Statistics?

Statistics is the process through which **data** are collected and analyzed in order to derive (useful) **insight** regarding a **population** (or process) of interest.

Important Concepts: Data

- ▶ **Data:** A collection of facts.

Important Concepts: Data

- ▶ **Data:** A collection of facts.
- ▶ **Variables:** Characteristics which vary between individuals, processes, objects, etc.

Important Concepts: Data

- ▶ **Data:** A collection of facts.
- ▶ **Variables:** Characteristics which vary between individuals, processes, objects, etc.
- ▶ **Observation:** An individual piece of data.

Important Concepts: Data

- ▶ **Data:** A collection of facts.
- ▶ **Variables:** Characteristics which vary between individuals, processes, objects, etc.
- ▶ **Observation:** An individual piece of data.
- ▶ Data can be **univariate**, if only a single fact is collected for each observations, or **multivariate**, if more than one fact is collected for each observation.

Important Concepts: Populations

- ▶ **Population:** A well-defined collection of objects of interest.

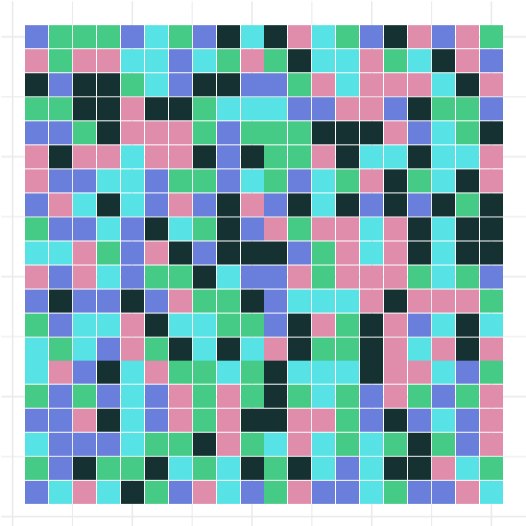
Important Concepts: Populations

- ▶ **Population:** A well-defined collection of objects of interest.
- ▶ **Census:** Collection of data for *all* members of a population.

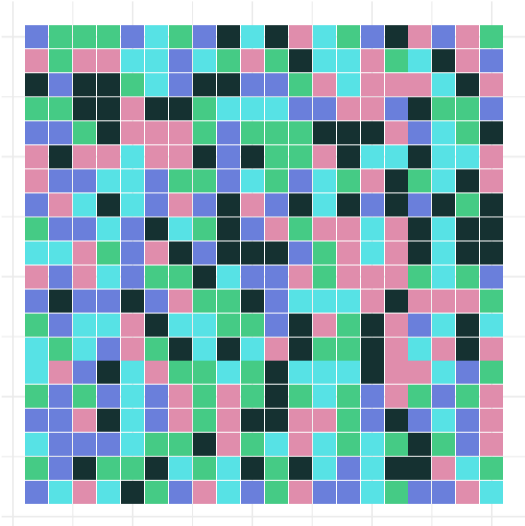
Important Concepts: Populations

- ▶ **Population:** A well-defined collection of objects of interest.
- ▶ **Census:** Collection of data for *all* members of a population.
- ▶ **Sample:** A subset of the population.

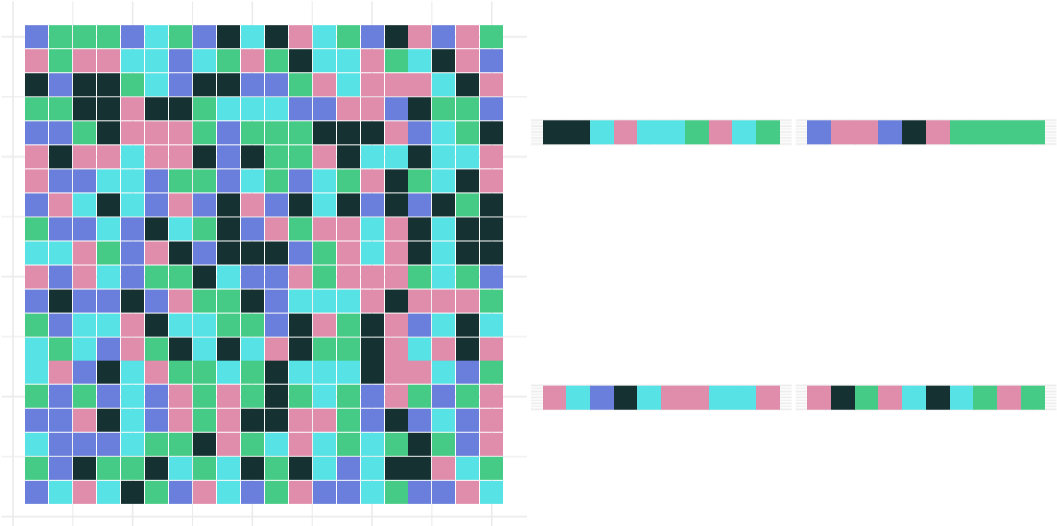
Population versus Sample, Visually



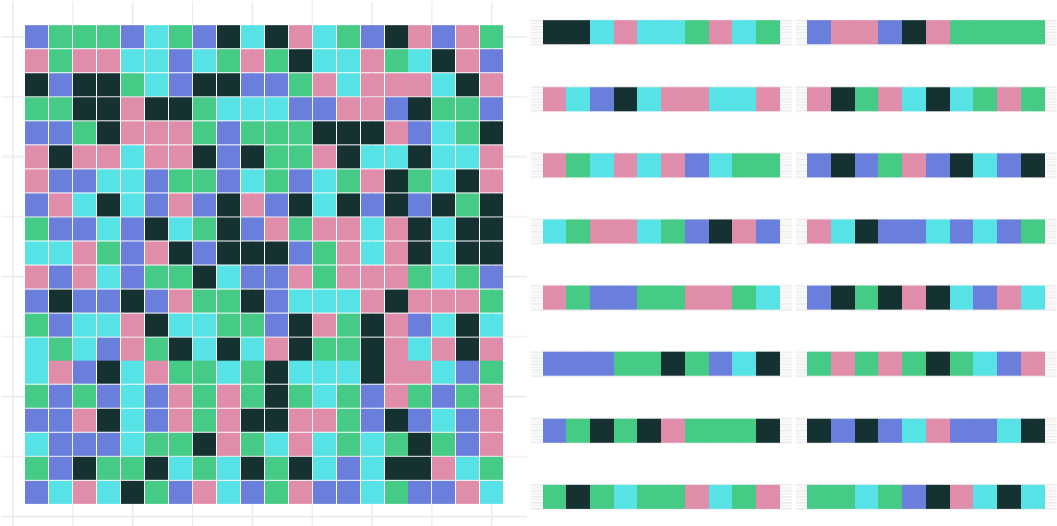
Population versus Sample, Visually



Population versus Sample, Visually



Population versus Sample, Visually



Important Concepts: Inferences

- ▶ **Parameters:** A descriptive measure for a population.

Important Concepts: Inferences

- ▶ **Parameters:** A descriptive measure for a population.

- ▶ **Statistic:** A descriptive measure for a sample.

The Process of Statistics

- ▶ In statistics we collect **observations** of **variables**, which constitute our **data**.

The Process of Statistics

- ▶ In statistics we collect **observations** of **variables**, which constitute our **data**.
- ▶ These data come from a **population** of interest, and make-up either a **census** or a **sample**.

The Process of Statistics

- ▶ In statistics we collect **observations** of **variables**, which constitute our **data**.
- ▶ These data come from a **population** of interest, and make-up either a **census** or a **sample**.
- ▶ Our goal is to learn information about a **parameter** of interest, using **statistics** which we can compute.

Common Types of Variables

- ▶ Categorical versus quantitative:

Common Types of Variables

- ▶ Categorical versus quantitative:
 - ▶ Occupation (categorical) compared to years of experience (quantitative)

Common Types of Variables

- ▶ Categorical versus quantitative:
 - ▶ Occupation (categorical) compared to years of experience (quantitative)
- ▶ Discrete versus continuous:

Common Types of Variables

- ▶ Categorical versus quantitative:
 - ▶ Occupation (categorical) compared to years of experience (quantitative)
- ▶ Discrete versus continuous:
 - ▶ Age in years (discrete) compared to actual age

Common Types of Variables

- ▶ Categorical versus quantitative:
 - ▶ Occupation (categorical) compared to years of experience (quantitative)
- ▶ Discrete versus continuous:
 - ▶ Age in years (discrete) compared to actual age
- ▶ Is **education** categorical or quantitative?

Common Types of Variables

- ▶ Categorical versus quantitative:
 - ▶ Occupation (categorical) compared to years of experience (quantitative)
- ▶ Discrete versus continuous:
 - ▶ Age in years (discrete) compared to actual age
- ▶ Is **education** categorical or quantitative?
- ▶ We care since our analysis will depend on variable types!

The Four Roles the Statistics Can Play

1. **Descriptive Statistics:** “What has happened?”

The Four Roles the Statistics Can Play

1. **Descriptive Statistics:** “What has happened?”
2. **Inferential Statistics:** “What conclusions can be drawn?”

The Four Roles the Statistics Can Play

1. **Descriptive Statistics:** “What has happened?”
2. **Inferential Statistics:** “What conclusions can be drawn?”
3. **Predictive Statistics:** “What is going to happen in the future?”

The Four Roles the Statistics Can Play

1. **Descriptive Statistics:** “What has happened?”
2. **Inferential Statistics:** “What conclusions can be drawn?”
3. **Predictive Statistics:** “What is going to happen in the future?”
4. **Prescriptive Statistics:** “What should be done?”

Statistical Studies

- ▶ Studies in statistics can be divided into **observational studies** versus **designed experiments**.

Statistical Studies

- ▶ Studies in statistics can be divided into **observational studies** versus **designed experiments**.
- ▶ **Observational studies** occur when the data being collected have not been directly manipulated by the researchers.

Statistical Studies

- ▶ Studies in statistics can be divided into **observational studies** versus **designed experiments**.
- ▶ **Observational studies** occur when the data being collected have not been directly manipulated by the researchers.
- ▶ **Designed experiments** use researcher intervention, and the process of randomization, to compare populations or processes.

Observational Studies

- ▶ Harder to discern causation (but not impossible!).

Observational Studies

- ▶ Harder to discern causation (but not impossible!).
- ▶ Have to watch-out for **confounding**.

Observational Studies

- ▶ Harder to discern causation (but not impossible!).
- ▶ Have to watch-out for **confounding**.
- ▶ Samples need to be **representative**, typically.

Sample Techniques

- ▶ **Convenience sampling:** select those individuals who are available.

Sample Techniques

- ▶ **Convenience sampling:** select those individuals who are available.
- ▶ **Simple random sampling:** any observation is equally likely to be included.

Sample Techniques

- ▶ **Convenience sampling:** select those individuals who are available.
- ▶ **Simple random sampling:** any observation is equally likely to be included.
- ▶ **Stratified random sampling:** A simple random sample is performed in different groups (or strata).

Designed Experiments

- ▶ **Randomization** helps to eliminate confounding.

Designed Experiments

- ▶ **Randomization** helps to eliminate confounding.
- ▶ Normally focused on a very specific intervention.

A Note on Probability

- ▶ **Probability** represents the opposite process of statistics.
- ▶ You assume that you know something about a population and then ask what is expected to be observed in samples.

Summary

- ▶ Statistics is the process of using data to make inferences about a population from a sample.
- ▶ Parameters represent population quantities of interests, statistics represent sample quantities of interest.
- ▶ Variables can be quantitative or categorical, discrete or continuous.
- ▶ Statistics can be used to describe, infer, predict, or prescribe.
- ▶ Studies in statistics are either observational or experimental.
- ▶ Probability performs the opposite process of statistics