STAT 2593

Lecture 021 - Statistics and their Distributions

Dylan Spicker

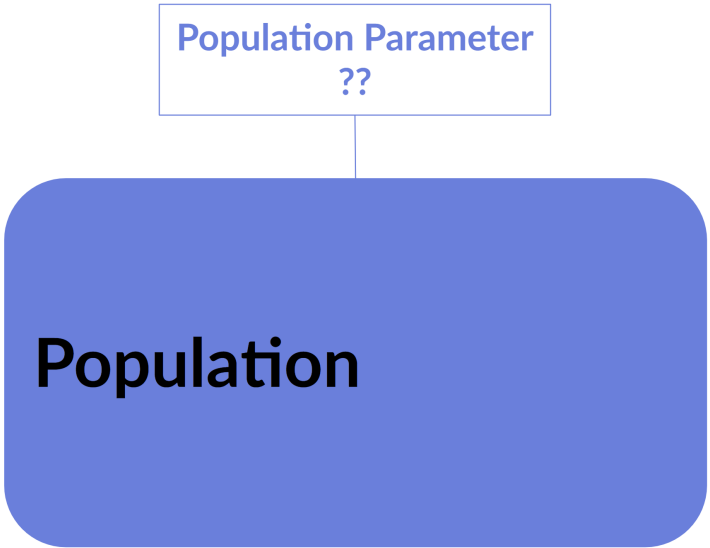
Statistics and their Distributions

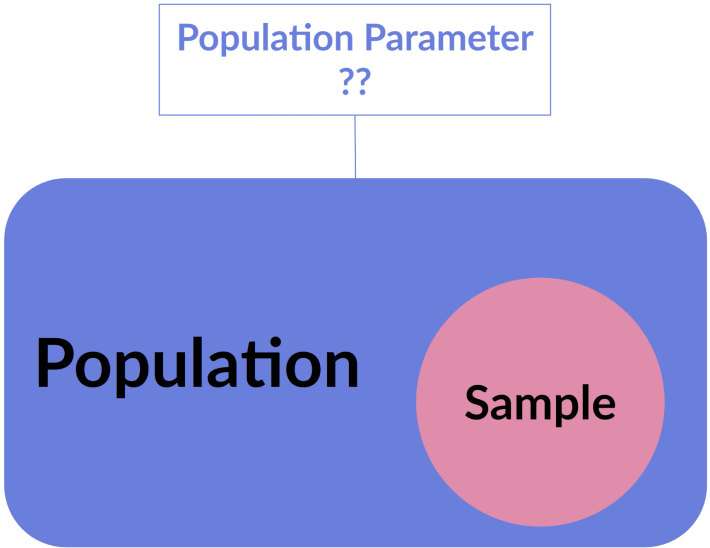
Learning Objectives

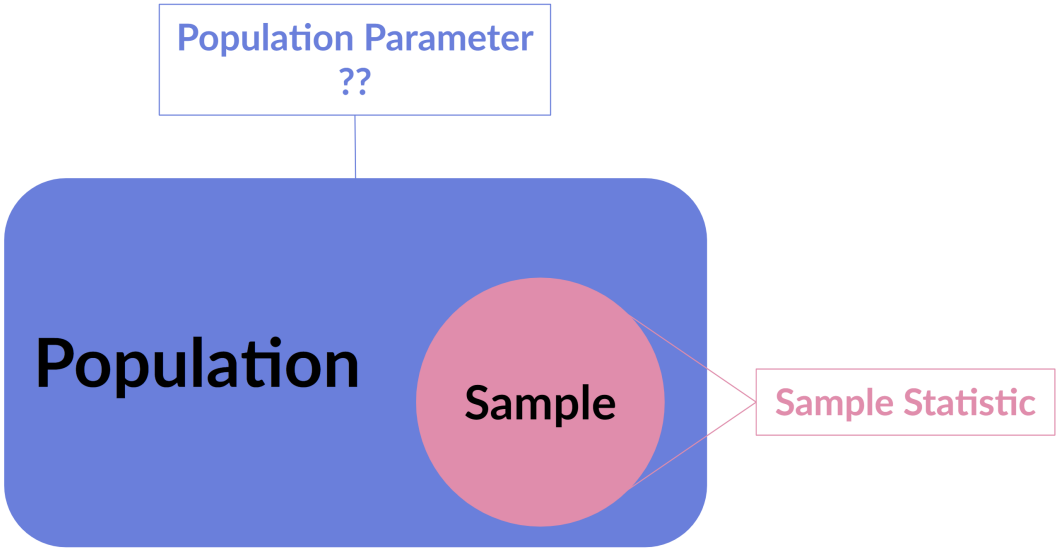
1. Understand and explain sampling distributions.

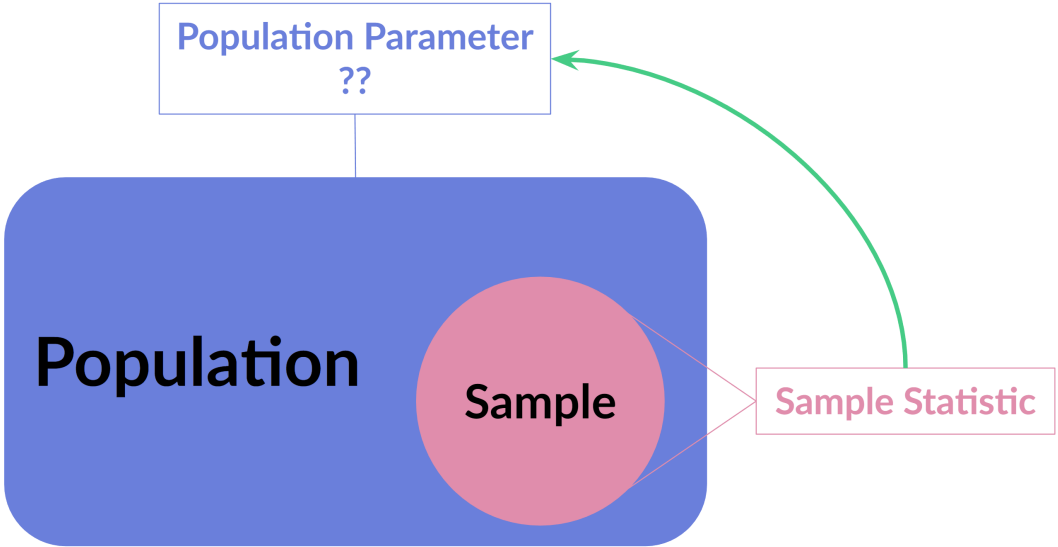
2. Differentiate between sampling distributions and population distributions.

Ultimately, the goal of statistics is to **learn** information about a **population** using observations from a **sample**.


Population Parameter
??
Population


Population Parameter
??
Population
Sample


Population Parameter
??
Population
Sample
Sample Statistic


Population Parameter
??
Population
Sample
Sample Statistic

The Fundamental Dilemma

- When we compute a **statistic** from a sample, this will **not** generally equal the underlying **parameter**.

The Fundamental Dilemma

- When we compute a **statistic** from a sample, this will **not** generally equal the underlying **parameter**.
- Instead, we wish to characterize how reliable our statistics are as proxies of the underlying parameters.

The Fundamental Dilemma

- When we compute a **statistic** from a sample, this will **not** generally equal the underlying **parameter**.
- Instead, we wish to characterize how reliable our statistics are as proxies of the underlying parameters.
- In order to do this, we need to understand the **distribution of the statistic**.

The Fundamental Dilemma

- When we compute a **statistic** from a sample, this will **not** generally equal the underlying **parameter**.
- Instead, we wish to characterize how reliable our statistics are as proxies of the underlying parameters.
- In order to do this, we need to understand the **distribution of the statistic**.
  - This is called the **sampling distribution**.

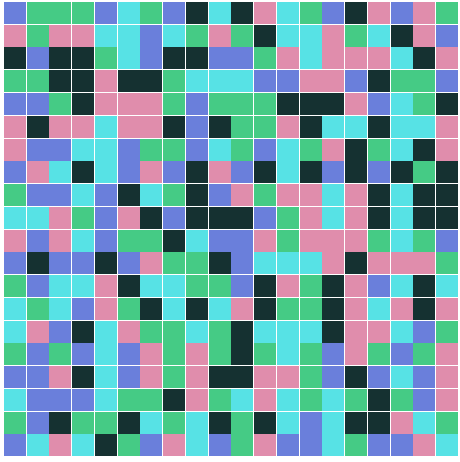
Sampling Distributions, Intuitively

Everytime that a sample is drawn from the population, and then a statistic is computed, we expect that there will be random variation. If you were to conduct this process again, you would expect to receive a different sample, and from this different sample, you'd compute a different value for the statistic.

What if we could run this experiment repeatedly?

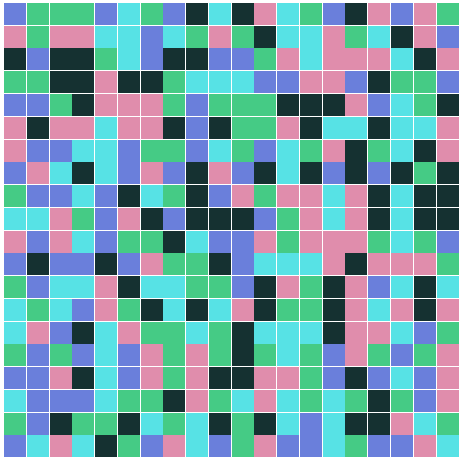
Sampling Distributions, Visually

Population: P=0.195



Sampling Distributions, Visually

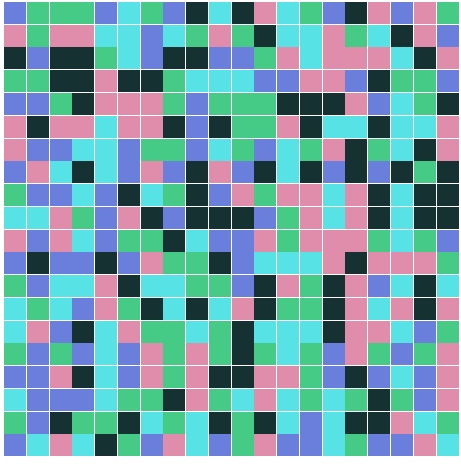
Population: P=0.195



Sample 1: P=0.2

Sampling Distributions, Visually

Population: P=0.195



Sample 1: P=0.2

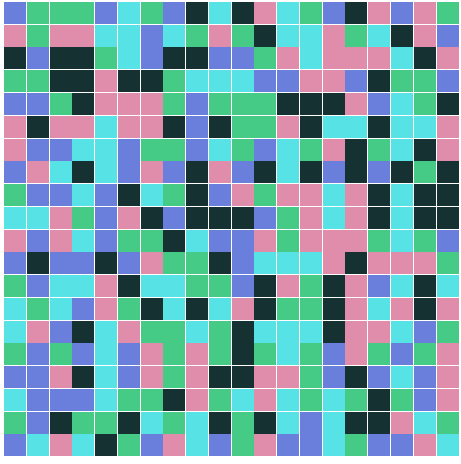
Sample 2: P=0.1

Sample 3: P=0.1

Sample 4: P=0.2

Sampling Distributions, Visually

Population: P=0.195



Sample 1: P=0.2

Sample 2: P=0.1

Sample 3: P=0.1

Sample 4: P=0.2

Sample 5: P=0

Sample 6: P=0.3

Sample 7: P=0.1

Sample 8: P=0.1

Sample 9: P=0

Sample 10: P=0.3

Sample 11: P=0.2

Sample 12: P=0.1

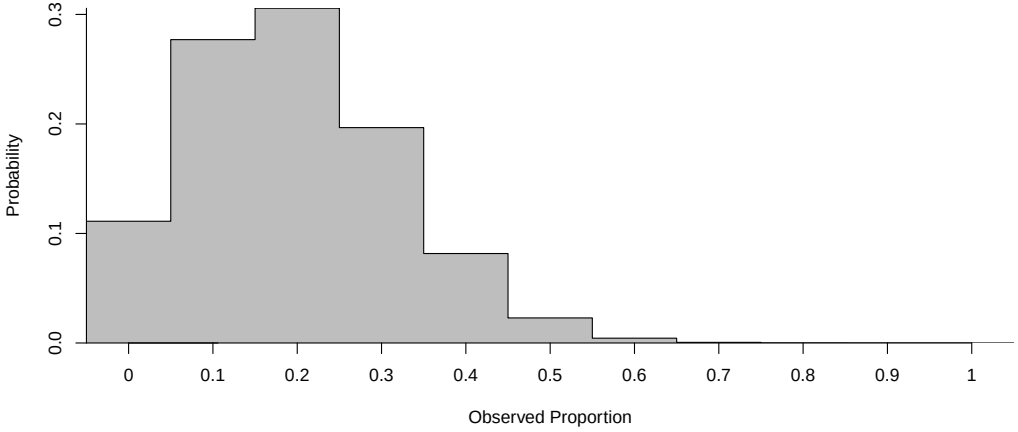
Sample 13: P=0.3

Sample 14: P=0.3

Sample 15: P=0.1

Sample 16: P=0.2

Sampling Distributions, Visually



Sampling Distributions

- Because statistics are computed based on data which is random, a statistic is also a random variable.

Sampling Distributions

- Because statistics are computed based on data which is random, a statistic is also a random variable.
- The sampling distribution is the distribution of the statistic.

Sampling Distributions

- Because statistics are computed based on data which is random, a statistic is also a random variable.
- The sampling distribution is the distribution of the statistic.
- It can be thought of as arising from repeated experiments, many times over.

Sampling Distributions

- Because statistics are computed based on data which is random, a statistic is also a random variable.
- The sampling distribution is the distribution of the statistic.
- It can be thought of as arising from repeated experiments, many times over.
  - This can be determined through simulation.

Summary

- Statistics are random variables and correspondingly have a distribution.
- The distribution of a statistic is called the sampling distribution.
- Sampling distributions can be assessed to quantify the reliability of estimates, and are an important component of statistical inference.