# Concepts of Statistics

STAT 135

Instructor: Rebecca Barter

Kelvin Lee

# **Contents**

1	Intr	oduction to Inference	3
	1.1	Parameters, populations, and estimates	3
		1.1.1 Common parameters of interest in statistics	3
		1.1.2 Inference	3
	1.2	Bias in data	3
	1.3	Random Sampling	4

## 1 Introduction to Inference

### 1.1 Parameters, populations, and estimates

**Definition 1.1.1** (Population). A *population* is the complete set of individuals or entities that we are interested in. We usually only have data on a subset of them.

**Definition 1.1.2** (Parameter). A parameter is any quantifiable feature of a population.

#### 1.1.1 Common parameters of interest in statistics

The most common population parameters we are interested in are:

- 1. Mean
- 2. Proportions (averages of binary data)

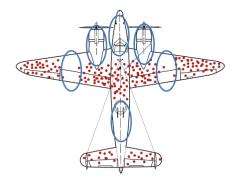
#### 1.1.2 Inference

**Definition 1.1.3** (Inference). *Inference* involves using data to compute an estimate of a population parameter of interest.

**Remark.** The population should always be defined in the context of where the results will be applied. Accurate inference is only possible when the data is representative of the population (i.e., the data is **unbiased**).

#### 1.2 Bias in data

**Example 1.2.1** (Survivorship bias). In the figure below, each dot corresponds to a place that a returning plane has been hit. Where should you reinforce the plane's armor?



**Figure 1.1:** If the bullets hit the top circled area, the plane goes down and does not return. The data is a biased representation of where the planes are getting hit.

**Definition 1.2.2** (Biased). Data is *biased* if it does not reflect the population it was designed to represent. **Biased data leads to biased results**.

**Example 1.2.3.** If AI-driven skin cancer detection is built only using patients with light skin tones but is used to detect skin cancer in racially diverse patients, the algorithm might be biased.

### 1.3 Random Sampling