

Sheldon M Ross - Statistics
Notes and Exercises

Anirudh Krishnan

July 28, 2021

Contents

1	Introduction to Statistics	2
2	Descriptive Statistics	5

Chapter 1

Introduction to Statistics

“Sorry, your sampling is biased.”

Descriptive Statistics : gather and describe data using summary measures. Covered by school-level classes.

Inferential Statistics : drawing conclusions from statistics using statistical models. Is the observed trend mere chance or is it a significant result worthy of more scrutiny. The right set of assumptions making up a *probability model*, are not always apparent given the dataset.

Population : the set of all possible elements that can be scrutinized. In real-world examples, the population is usually too large to be analyzed in full.

Sample : a small subsection of the population that is supposedly a good representation of the properties of the population itself. Random sampling is crucial for the inferred population statistics to be useful.

Applications throughout history : Surveying populations in order to calculate men available for military enlistment, taxable populations, and insurance contracts. Astronomy, physics, anthropology, demography, sociology and many other contemporary fields of study depend on the advancements in statistics over the 18th and 19th centuries.

Relation to probability theory : probability theory was largely divorced from statistics until the late 19th century. The study of probability was necessary to develop inferential statistics and begin to apply statistics to business, medicine, and politics.

Exercises

1. Sample A is biased in favor of young people.
Sample B is biased in favor of wealthy people.
Sample C is likely to be representative.
Sample D is biased in favor of people watching the particular channel.
Sample E involves the user choosing names and is therefore bad.
2. The proportion of voters in 1936 owning automobiles or telephones would have been extremely biased in favor of the wealthy. This sampling bias led to the incorrect prediction.
The prevalence of phone and automobile ownership is much closer to 100% today, making this a much better sampling mechanism.
3. No, because obituaries are typically not inclusive of child mortality. The obituary sample will be biased in favor of old people.
4. The age bias is likely to be lowest at the shopping mall. This makes it the best polling location.
5. No. It would be an overestimate. Unemployed graduates might be more likely not to return a filled questionnaire out of a sense of shame.
6. How likely is a person to be out on the streets at night as a function of clothing color?
7. Graunt surveys only certain neighborhoods in the city. Possible wealth bias in his sampling. All of England is not the same demographics as London.
8. Let x be the total population

$$0.02 \times x = 12246 \quad (1.1a)$$

$$x = 612300 \quad (1.1b)$$

9. Find the remaining lifetime of each person based on their current age. Charge some additional percentage as the annuity down-payment to earn a profit.
10. Simple probability calculations

(a)

$$\frac{\text{died after age 6}}{\text{total number dead}} = \frac{24 + 15 + 9 + 6 + 4 + 3 + 2 + 1}{36 + 24 + 15 + 9 + 6 + 4 + 3 + 2 + 1} = 64\% \quad (1.2a)$$

(b)

$$\frac{\text{died after age 46}}{\text{total number dead}} = \frac{4 + 3 + 2 + 1}{36 + 24 + 15 + 9 + 6 + 4 + 3 + 2 + 1} = 10\% \quad (1.2b)$$

(c)

$$\frac{\text{died at age 6-36}}{\text{total number dead}} = \frac{24 + 15 + 9}{36 + 24 + 15 + 9 + 6 + 4 + 3 + 2 + 1} = 48\% \quad (1.2c)$$

Chapter 2

Descriptive Statistics

“I will skim through these topics since you’ve already covered them in school.”

Descriptive Statistics : gather and describe data using summary measures. Covered by school-level classes.

Inferential Statistics : drawing conclusions from statistics using statistical models. Is the observed trend mere chance or is it a significant result worthy of more scrutiny. The right set of assumptions making up a *probability model*, are not always apparent given the dataset.

Population : the set of all possible elements that can be scrutinized. In real-world examples, the population is usually too large to be analyzed in full.

Sample : a small subsection of the population that is supposedly a good representation of the properties of the population itself. Random sampling is crucial for the inferred population statistics to be useful.

Applications throughout history : Surveying populations in order to calculate men available for military enlistment, taxable populations, and insurance contracts. Astronomy, physics, anthropology, demography, sociology and many other contemporary fields of study depend on the advancements in statistics over the 18th and 19th centuries.

Relation to probability theory : probability theory was largely divorced from statistics until the late 19th century. The study of probability was necessary to develop inferential statistics and begin to apply statistics to business, medicine, and politics.

Exercises

1. Sample A is biased in favor of young people.
 Sample B is biased in favor of wealthy people.
 Sample C is likely to be representative.
 Sample D is biased in favor of people watching the particular channel.
 Sample E involves the user choosing names and is therefore bad.
2. The proportion of voters in 1936 owning automobiles or telephones would have been extremely biased in favor of the wealthy. This sampling bias led to the incorrect prediction.
 The prevalence of phone and automobile ownership is much closer to 100% today, making this a much better sampling mechanism.
3. No, because obituaries are typically not inclusive of child mortality. The obituary sample will be biased in favor of old people.
4. The age bias is likely to be lowest at the shopping mall. This makes it the best polling location.
5. No. It would be an overestimate. Unemployed graduates might be more likely not to return a filled questionnaire out of a sense of shame.
6. How likely is a person to be out on the streets at night as a function of clothing color?
7. Graunt surveys only certain neighborhoods in the city. Possible wealth bias in his sampling. All of England is not the same demographics as London.
8. Let x be the total population

$$0.02 \times x = 12246 \quad (2.1a)$$

$$x = 612300 \quad (2.1b)$$

9. Find the remaining lifetime of each person based on their current age. Charge some additional percentage as the annuity down-payment to earn a profit.
10. Simple probability calculations

(a)

$$\frac{\text{died after age 6}}{\text{total number dead}} = \frac{24 + 15 + 9 + 6 + 4 + 3 + 2 + 1}{36 + 24 + 15 + 9 + 6 + 4 + 3 + 2 + 1} = 64\% \quad (2.2a)$$

(b)

$$\frac{\text{died after age 46}}{\text{total number dead}} = \frac{4 + 3 + 2 + 1}{36 + 24 + 15 + 9 + 6 + 4 + 3 + 2 + 1} = 10\% \quad (2.2b)$$

(c)

$$\frac{\text{died at age 6-36}}{\text{total number dead}} = \frac{24 + 15 + 9}{36 + 24 + 15 + 9 + 6 + 4 + 3 + 2 + 1} = 48\% \quad (2.2c)$$