Elementary Statistics: Math 080

Jordan Hanson

July 4, 2020

Whittier College Department of Physics and Astronomy

Course Introduction

1. ***What is statistical analysis?***
2. Math 080: Elementary Statistics
3. Read the syllabus for a roadmap
4. This is an online summer course that meets each day.
5. **Data science project and presentation**
6. Textbook: `https://openstax.org/details/books/introductory-statistics`
7. Download and install Excel, or LibreOffice Calc

Lecture format, with modifications

- Warm-up exercise, and solution (10-15 minutes)
- Lecture via Whiteboard and slides (10-20 minutes)
- Interactive questions or polls (10 minutes)
- Laboratory activity (20 minutes)
    1. Breakout rooms
    2. Offline
- Asynchronous content
    1. Homework clues
    2. Example problems
    3. Special topics

Unit 0 Outline

1. Topics from Chapter 1: 1.1, 1.2, 1.3
    - What is a statistic?
    - Probability examples
    - Data and sampling
2. Topics from Chapter 2: 2.1 - 2.4, 2.5 - 2.8
    - Data visualization
    - Location of the data in numerical space
3. Topics from Chapter 3: 3.1, 3.2, 3.3
    - Two rules of probability

Topics from Chapter 1

What is statistical analysis?

By tradition, we begin with Mark Twain.

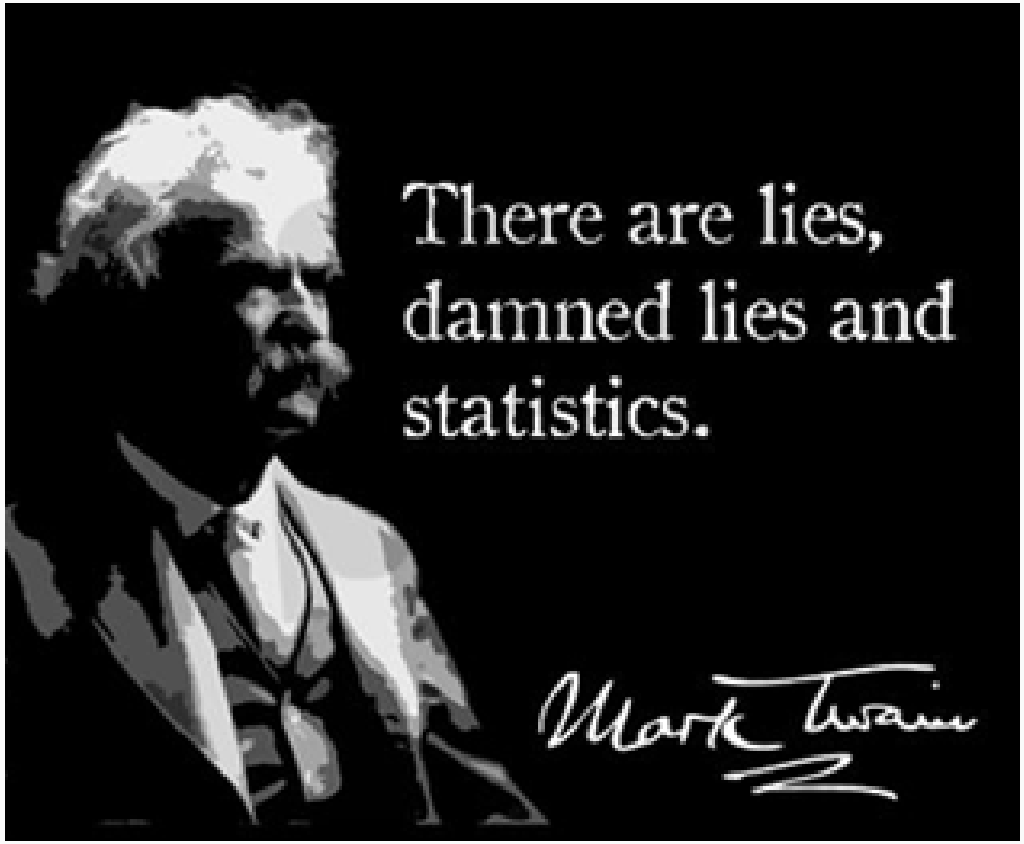

Figure 1: A famous quote from Mark Twain.

Warm-up exercises

COVID-19 data. In a March 2020 article in the magazine `wired.com`, Ferris Jabr points out that people were drawing comparisons between the influenza pandemic of 1918 and SARS-Cov-2 (COVID-19). The case fatality rate, or CFR, is the percentage of people who contract the disease that perish from it. In the 1918 outbreak, it is usually stated that there were approximately 500 million infections, 50-100 million fatalities, and an overall CFR of 2.5%. What is interesting is that the coronavirus seems to have a CFR (averaged over age) of ≈ 3 %, making it ... *higher*.

Question 1: The above paragraph listed four pieces of data. What are they?

Warm-up exercises

COVID-19 data. In a March 2020 article in the magazine `wired.com`, Ferris Jabr points out that people were drawing comparisons between the influenza pandemic of 1918 and SARS-Cov-2 (COVID-19). The case fatality rate, or CFR, is the percentage of people who contract the disease that perish from it. In the 1918 outbreak, it is usually stated that there were approximately 500 million infections, 50-100 million fatalities, and an overall CFR of 2.5%. What is interesting is that the coronavirus seems to have a CFR (averaged over age) of ≈ 3 %, making it ... *higher*.

Question 2: Which number, if any, seems to have a problem?

- A: The total number of infections in 1918
- B: The total number of deaths in 1918
- C: The 1918 influenza CFR
- D: The 2020 coronavirus CFR

Warm-up exercises

COVID-19 data. In a March 2020 article in the magazine `wired.com`, Ferris Jabr points out that people were drawing comparisons between the influenza pandemic of 1918 and SARS-Cov-2 (COVID-19). The case fatality rate, or CFR, is the percentage of people who contract the disease that perish from it. In the 1918 outbreak, it is usually stated that there were approximately 500 million infections, 50-100 million fatalities, and an overall CFR of 2.5%. What is interesting is that the coronavirus seems to have a CFR (averaged over age) of ≈ 3 %, making it ... *higher*.

Question 3: From the rest of the data in the paragraph, estimate the proper CFR of the 1918 influenza. Compare this number with the CFR of the 2020 coronavirus.

Topics from Chapter 1

Vocabulary:

1. **Probability**: The extend to which something is *likely* to occur, measured by the ratio of favorable cases to the whole number of cases possible.
2. **Population**: The total collection of people, objects, or cases under investigation.
3. **Sample**: A subset of the population for which statistical data is collected.
4. **Statistic**: *A statistic* is a number that represents a property of the sample. For example: the CFR of a *sample* of 2,500 coronavirus patients.
5. **Parameter**: Statistic measured from the *entire* population. A statistic attempts to reveal knowledge of a parameter.

Topics from Chapter 1

Vocabulary:

1. **Representative sample**: a sample that captures all of the properties of a population. Counter-example: psychological studies using undergraduate subjects.
2. **Variable**: A property of each member of the population that can be determined, either quantitative or categorical. **Data** are the actual values.

Mean: Definition 1

Let X represent a *variable* of a *population*, and x_i represent the actual value of the i-th member of a statistical *sample* of that *population*. The arithmetic mean $\bar{x}$ of the *sample* for that property is

$$\bar{x} = \frac{1}{N} \sum_{i}^{N} x_i \tag{1}$$

The mean of the variable X is the number $\bar{x}$ from the sample.

Topics from Chapter 1

Example 1: What's the average number of siblings in our community?

1. What is the population?
2. What is the sample? (Our class).
3. What is the variable?
4. What are the data?

Write in the chat area the number of siblings in your family, including yourself.

Topics from Chapter 1

Example 2: How many languages do you speak?

1. What is the population?
2. What is the sample? (Our class).
3. What is the variable?
4. What are the data?

Write in the chat area the number of languages that you can speak.

Topics from Chapter 1

Vocabulary:

1. **Proportion**: The total number of subjects in the sample that share a property, divided by the total number of subjects in the sample.
2. **Qualitative data**: Sometimes called categorical data, refers to non-numerical properties of subjects in sample (e.g. place of birth).
3. **Quantitative data**: Numerical values of variables for each subject in a sample (e.g. age).
    - Continuous quantitative data: average hours of sleep per night
    - Discrete quantitative data: average number of siblings

Topics from Chapter 1

Example 1: What fraction of Whittier College students live on campus?

1. What is the population?
2. What is the sample? (Our class).
3. What is the variable?
4. What are the data?

Write in the chat area the number 1 if you live on-campus, and the number 0 if you live off-campus or with your family.

Whittier College Factbook: 46.3% of undergraduates live on-campus.

Topics from Chapter 1

Example 2: What is the proportion of students to instructors here? (What is the student to faculty ratio of Whittier College)?

1. What is the population?
2. What is the sample? (Our class).
3. What is the variable?
4. What are the data?

Let's sum the students here, and then there is me.

Whittier College Factbook: average student to faculty ratio: 11

Topics from Chapter 1

Example 3: You go to the supermarket and purchase three cans of soup:

- 19 ounces tomato bisque
- 14.1 ounces lentil
- 19 ounces Italian wedding

...and two desserts:

- 16 ounces pistachio ice cream
- 32 ounces chocolate chip cookies

Create three data sets: one quantitative discrete, one quantitative continuous, and one categorical.

Interactive Questions

Interactive Questions

Almost always, we will give multiple-choice questions with answers A-D. If you are lost, or need extra explanation, or just feel we are going to fast, select the letter E. E stands for WAT...

After 1 round, we examine the *answer distribution*, and if 70% get it right, we move on. Otherwise, we discuss via chat with each other, explaining why we picked our answer. Then we have round 2. Remember to hit E if you are confused.

Interactive Questions

To battle the pandemic, backup health care workers were called in to work in hospitals A, B, and C. Hospital A began with 50, hospital B began with 40, and hospital C began with 60. Hospital A received an additional 10, B received an additional 25, and C received an additional 5. What is the average number of workers at hospitals in this sample (A, B, and C)?

- A: 53
- B: 63
- C: 42
- D: 32

Interactive Questions

Suppose a sample of students record the duration of their sleep each night for a week, and gather the data at the end. What kind of data is this?

- A: Quantitative discrete
- B: Qualitative or categorical
- C: Quantitative continuous
- D: Variable

Interactive Questions

Fall 2019 Country of Citizenship and Student Count					
Argentina	0	Hungary	0	Saudi Arabia	8
Australia	3	India	3	South Africa	1
Bhutan	1	Ireland	1	Spain	0
Brazil	1	Italy	3	Sweden	1
Cambodia	1	Japan	1	Turkey	1
Canada	2	Kosovo	1	Ukraine	1
China	12	Netherlands	1	United Kingdom	5
Egypt	0	New Zealand	1	Zimbabwe	1
France	1	Russia	1		
Germany	0	Rwanda	1		

What kind of data is represented in the population above? (There may be more than one answer).

- A: Quantitative discrete
- B: Qualitative or categorical
- C: Quantitative continuous
- D: Variable

Interactive Questions

Fall 2019 Country of Citizenship and Student Count					
Argentina	0	Hungary	0	Saudi Arabia	8
Australia	3	India	3	South Africa	1
Bhutan	1	Ireland	1	Spain	0
Brazil	1	Italy	3	Sweden	1
Cambodia	1	Japan	1	Turkey	1
Canada	2	Kosovo	1	Ukraine	1
China	12	Netherlands	1	United Kingdom	5
Egypt	0	New Zealand	1	Zimbabwe	1
France	1	Russia	1		
Germany	0	Rwanda	1		

The total number of international students is 52 in the above table. What proportion of international students are from China?

- A: 12
- B: 12%
- C: 23
- D: 23%

Interactive Questions

Fall 2019 Country of Citizenship and Student Count					
Argentina	0	Hungary	0	Saudi Arabia	8
Australia	3	India	3	South Africa	1
Bhutan	1	Ireland	1	Spain	0
Brazil	1	Italy	3	Sweden	1
Cambodia	1	Japan	1	Turkey	1
Canada	2	Kosovo	1	Ukraine	1
China	12	Netherlands	1	United Kingdom	5
Egypt	0	New Zealand	1	Zimbabwe	1
France	1	Russia	1		
Germany	0	Rwanda	1		

The total number of international students is 52 in the above table. What proportion of international students are from Europe?

- A: 15%
- B: 23%
- C: 50%
- D: 12

Laboratory Activity

Laboratory Activity

Go to the following link and watch the interesting TED talk by Steven Levitt from 2005 about driving safety.

`https://www.ted.com/talks/steven_levitt_surprising_stats_about_child_carseats?utm_campaign=tedspread&utm_medium=referral&utm_source=tedcomshare`

Answer the questions on the form entitled Laboratory Exercise 1 on Moodle for this week, and submit them via email: jhanson2@whittier.edu. (This is part of your warm-ups grade...see syllabus).

Conclusion

Unit 0 Outline

1. Topics from Chapter 1: 1.1, 1.2, 1.3
    - What is a statistic?
    - Probability examples
    - Data and sampling
2. Topics from Chapter 2: 2.1 - 2.4, 2.5 - 2.8
    - Data visualization
    - Location of the data in numerical space
3. Topics from Chapter 3: 3.1, 3.2, 3.3
    - Two rules of probability