Statistics for Data Science

UC Berkeley, School of Information

2021-11-23

Contents

Live Session Introduction

This is the live session work space for the course. Our goal with this repository, is that we're able to communicate *ahead of time* our aims for each week, and that you can prepare accordingly.

Bloom's Taxonomy

An effective rubric for student understanding is attributed to Bloom (1956). Referred to as *Bloom's Taxonomy*, this proposes that there is a hierarchy of student understanding; that a student may have one *level* of reasoning skill with a concept, but not another. The taxonomy proposes to be ordered: some levels of reasoning build upon other levels of reasoning.

In the learning objective that we present in for each live session, we will also identify the level of reasoning that we hope students will achieve at the conclusion of the live session.

1. **Remember** A student can remember that the concept exists. This might require the student to define, duplicate, or memorize a set of concepts or facts.
2. **Understand** A student can understand the concept, and can produce a working technical and non-technical statement of the concept. The student can explain why the concept *is*, or why the concept works in the way that it does.
3. **Apply** A student can use the concept as it is intended to be used against a novel problem.
4. **Analyze** A student can assess whether the concept has worked as it should have. This requires both an understanding of the intended goal, an application against a novel problem, and then the ability to introspect or reflect on whether the result is as it should be.
5. **Evaluate** A student can analyze multiple approaches, and from this analysis evaluate whether one or another approach has better succeeded at achieving its goals.

6. **Create** A student can create a new or novel method from axioms or experience, and can evaluate the performance of this new method against existing approaches or methods.

Chapter 1

Probability Spaces

1.1 Learning Objectives

At the end of this week's learning, student will be able to:

1. *Find* and *access* all of the course materials
2. *Develop* a course of study that is builds toward success
3. *Apply* ...

1.2 What can you learn in this class?

- Probability theory
- Statistics
- Applying theory to data
- How statistics relates to other fields
- Composing a persuasive statistical analysis.

1.3 Instructor Introduction

-

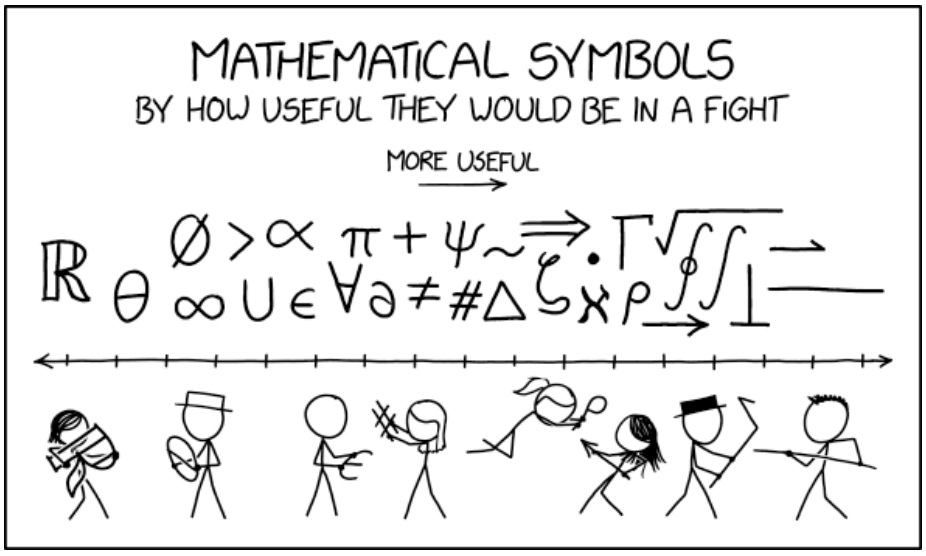

Figure 1.1: xkcd

1.4 What does a statistician look like?

1.4.1 A statistician looks like YOU!

- It doesn't matter if you've never taken a stats class before, or if you're reviewing using this class.
- It doesn't matter how old, young, tall, or short you are.
- It doesn't matter if you identify as a woman or a man or trans or non-binary.
- It does not matter what color your skin is.
- It does not matter what your sexual orientation is.

1.5 Student Introductions

Please take 90 seconds to tell us:

- Where you dial in from
- What kind of work you do / are interested in

Unfortunately, it's important to keep these intros to just 90 seconds, so there is time for everything we want to cover!

1.6 Probability Theory

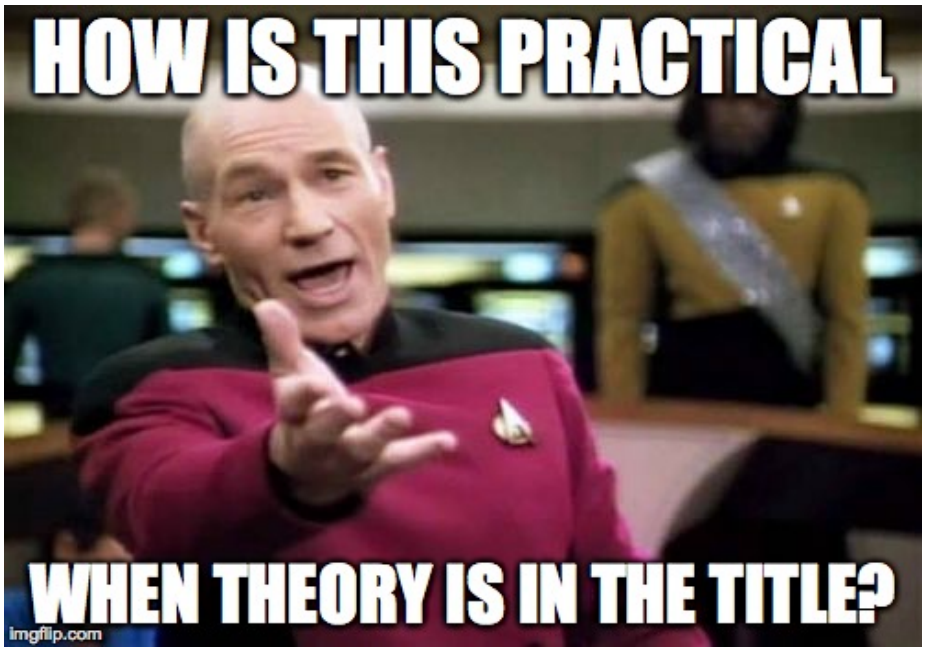

Figure 1.2: picard

1.7 Probability

- Probability is a system of reasoning that we use to model the world under incomplete information
- This model underlies virtually *every* other model you'll ever use

1.8 Working with a Sample Space, Part I

1. **You roll two six-sided dice**:
    1. How would you define an appropriate sample space, Ω?
    2. How many elements exist in Ω?
    3. What is an appropriate event space, and how many elements does it have?
    4. Give an example of an event.

1.9 Working with a Sample Space, Part II

2. **For a random sample of 1,000 Berkeley students**:
    1. How would you define an appropriate sample space, Ω?
    2. How big is Ω? How many elements does it contain?
    3. What is an example of an event for this scenario?
    4. Can a single person be represented in the space twice? Why or why not?

1.10 Working with a Sample Sapce, Part III

3. **Suppose that you're sitting in a surf lineup, and you have to pick a wave that is the right height. Too small, and you won't get anywhere, too large and you'll get crushed.**
    1. What sample space is appropriate to represent the height of a single wave, Ω?
    2. How big is Ω? How many elements does it contain?
    3. What is an example of an event that could be part of the event space?
    4. What sample space is appropriate to represent the height of the next 10 waves? How large is this sample space?

To represent 10 waves, you should use $\mathbb{R}^{10}$. It is an interesting mathematical fact that $\mathbb{R}$ and $\mathbb{R}^{10}$ actually have the same cardinality. There exists a 1-to-1 function between these sets.

1.11 Proofs: Style Counts

In each week of a class, you are either caught up or behind.

- The probability that you are caught up in Week 1 is 0.7.
- If you are caught up in a given week, the probability that you will be caught up in the next week is 0.7.
- If you are behind in a given week, the probability that you will be caught up in the next week is 0.4.
- **What is the probability that you are caught up in week 3?**

1.12 Proofs: Style Counts (cont.)

Identify as many ways to improve this proof as you can:

If you are caught up in a week, there are two possibilities for the previous week: caught up and behind. Let $P(C)$ be the probability of being caught up. In week 1, $P(C) = .7$. The probability of being behind is $P(B) = 1 - .7 = .3$. We first break down the probability for week 2:

$$P(C) = .7 \cdot .7 + .3 \cdot .4 = .65$$

Now we can repeat the process for week 3:

$$P(C) = .65 * .7 + .35 * .4 = .595$$

1.13 Reminders

1. Welcome!
2. Before next live session:
    1. Complete the homework that builds on this unit
    2. Complete all videos and reading for unit 2

1.14 Good luck getting started!

Chapter 2

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

2.1 Learning Objectives

1.
2.
3.

Chapter 3

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

3.1 Learning Objectives

1.
2.
3.

Chapter 4

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

4.1 Learning Objectives

1.
2.
3.

Chapter 5

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

5.1 Learning Objectives

1.
2.
3.

Chapter 6

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

6.1 Learning Objectives

1.
2.
3.

Chapter 7

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

7.1 Learning Objectives

1.
2.
3.

Chapter 8

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

8.1 Learning Objectives

1.
2.
3.

Chapter 9

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

9.1 Learning Objectives

1.
2.
3.

Chapter 10

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

10.1 Learning Objectives

1.
2.
3.

Chapter 11

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

11.1 Learning Objectives

1.
2.
3.

Chapter 12

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

12.1 Learning Objectives

1.
2.
3.

Chapter 13

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

13.1 Learning Objectives

1.
2.
3.

Chapter 14

{WEEK TITLE, REPLACE WITH ACTUAL TITLE}

14.1 Learning Objectives

1.
2.
3.