Topic 1: Introduction and Basics of Data

STOR 155: Introduction to Data Models and Inference

Dr. Teressa Bergland

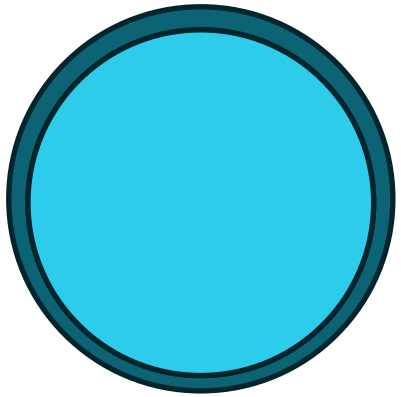
Fall 2025

Introductions

Welcome to STAT 155! Your first question: Who IS Dr. Bergland?

About me (T/F):

- I am an introvert.
- I have two cats.
- I am a knitter.
- I am a pianist.
- I am a Carolina Hurricanes fan.
- I prefer science fiction to fantasy.



Answers:

- True
- True! I can bring photos
- False… but I do quilt!
- True, mostly in musicals
- False – Go Penguins!
- False. Dragons any day.

Important Syllabus Items: Logistics

Textbook: Open-source and therefore free! Available on Canvas.

Homework: STOR 155 uses **WebAssign** for homework, which costs money ☹
You can purchase semester-long access via the bookstore or online

- Use the syllabus link to enroll in our WebAssign section by **Tuesday!**
  - (That way we have a bit of time to solve any problems that may occur)
- First assignment will be due **Thursday, January 15, 11:59pm!**

Attendance: Attendance is mandatory and part of your grade. I will start using **PollEverywhere** for attendance on **Thursday, January 15**

- Take the **syllabus quiz** on Canvas by Wednesday for a bonus attendance point!

Important Syllabus Items: Resources

Dr. Bergland's Office Hours: Hanes 310, Mondays 10-1 and Wednesdays 1-4

- Drop by to ask questions, get advice, or just to say hello!

Piazza: A full-class discussion board, shared by my two sections of STOR 155

- Piazza **(not email)** is the best way to reach me outside office hours!

Tutorial sessions: They're free! They're open to all! They're led by experienced instructors! They start **Monday, January 19** and run every night Sunday-Thursday!

- Held virtually – Zoom link will be posted on Canvas

Need accommodations? Let me know ASAP! The sooner I know what you need, the better I can be prepared to help.

What to Expect in Lecture

Blank lecture notes: These PDFs are provided in advance for you to print or access on the device of your choice. They include lots of blank space for notes.

- I will mostly lecture using the chalkboard – taking notes is up to you.
- **Where to find them:** on Canvas, go to the Modules tab. Each module is divided into Topics. The blank notes can be found on each Topic page before lecture.

Slideshows: Slide decks will mostly be used for images, interactive elements like PollEverywhere, and exercises where we pause to think, discuss, and synthesize.

- Slide decks will also display headings **from the blank notes** to help guide you during the lecture.
- **Slide decks will NOT contain detailed notes!**

Let's see an example of how this will work!

Course Overview

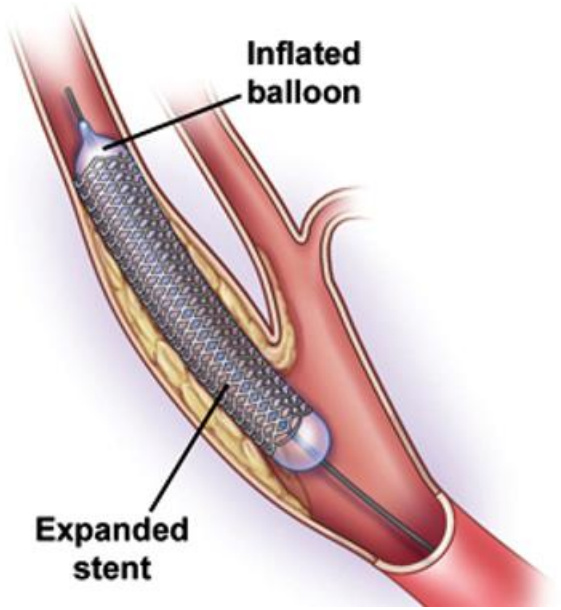
What will we learn in STOR 155?

Steps of inquiry:

1. Identify a question or problem of interest
2. Collect relevant data
3. Analyze the data
4. Form a conclusion

Case Study: Stents

Do stents reduce the risk of stroke in at-risk patients?



	Stroke	No Event
Treatment	45	179
Control	28	199

Let’s try a poll!

Do you find this study data surprising?

	Stroke	No Event
Treatment	45	179
Control	28	199

For future polls, you’ll need a **PollEverywhere account!** If you don’t have one yet, go to https://edtech.unc.edu/service/poll-everywhere/ for instructions.

- I will start running polls this way next week! We’ll do a test run on Tuesday before using it for attendance.

Case Study: Stents

The Big Question: Do we have enough information to conclude that stents change the risk of stroke in this type of patient?

Test poll!

Flip a coin – physically or virtually – and report the result.

(To flip a virtual coin, just Google “Flip a coin”)

Case Study: Stents

The Big Question: Do we have enough information to conclude that stents change the risk of stroke in this type of patient?

We use statistical tools to determine if the difference is so large that we should reject the notion that it was due to chance.

Data Sets: Vocabulary

	A	B	C	D	E	F	G
1	loan_amount	interest_rate	term	grade	state	total_income	homeownership
2	22000	10.9	60	B	NJ	59000	rent
3	6000	9.92	36	B	CA	60000	rent
4	25000	26.3	36	E	SC	75000	mortgage
5	6000	9.92	36	B	CA	75000	rent
6	25000	9.43	60	B	OH	254000	mortgage
7	6400	9.92	36	B	IN	67000	mortgage
8	3000	17.09	36	D	NY	28800	rent
9	14500	6.08	36	A	MO	80000	mortgage
10	10000	7.97	60	A	FL	34000	rent
11	18500	12.62	60	C	FL	192000	mortgage
12	17000	17.09	36	D	MD	73000	rent
13	12000	5.31	36	A	HI	120000	mortgage
14	16000	7.35	36	A	CT	100000	rent
15	16500	5.31	36	A	NE	105000	mortgage
16	3000	7.96	36	A	CA	34000	rent

Data Sets: Vocabulary

Two types of variables:

- Numerical
  - Discrete
  - Continuous
- Categorical
  - Ordinal
  - Nominal

Let's Practice! Variable Types

	A	B	C	D	E	F	G
1	loan_amount	interest_rate	term	grade	state	total_income	homeownership
2	22000	10.9	60	B	NJ	59000	rent
3	6000	9.92	36	B	CA	60000	rent
4	25000	26.3	36	E	SC	75000	mortgage
5	6000	9.92	36	B	CA	75000	rent
6	25000	9.43	60	B	OH	254000	mortgage
7	6400	9.92	36	B	IN	67000	mortgage
8	3000	17.09	36	D	NY	28800	rent
9	14500	6.08	36	A	MO	80000	mortgage
10	10000	7.97	60	A	FL	34000	rent
11	18500	12.62	60	C	FL	192000	mortgage
12	17000	17.09	36	D	MD	73000	rent
13	12000	5.31	36	A	HI	120000	mortgage
14	16000	7.35	36	A	CT	100000	rent
15	16500	5.31	36	A	NE	105000	mortgage
16	3000	7.96	36	A	CA	34000	rent