Math 243 Sections 1.1-1.2 Intro to Stats and Types of Variables

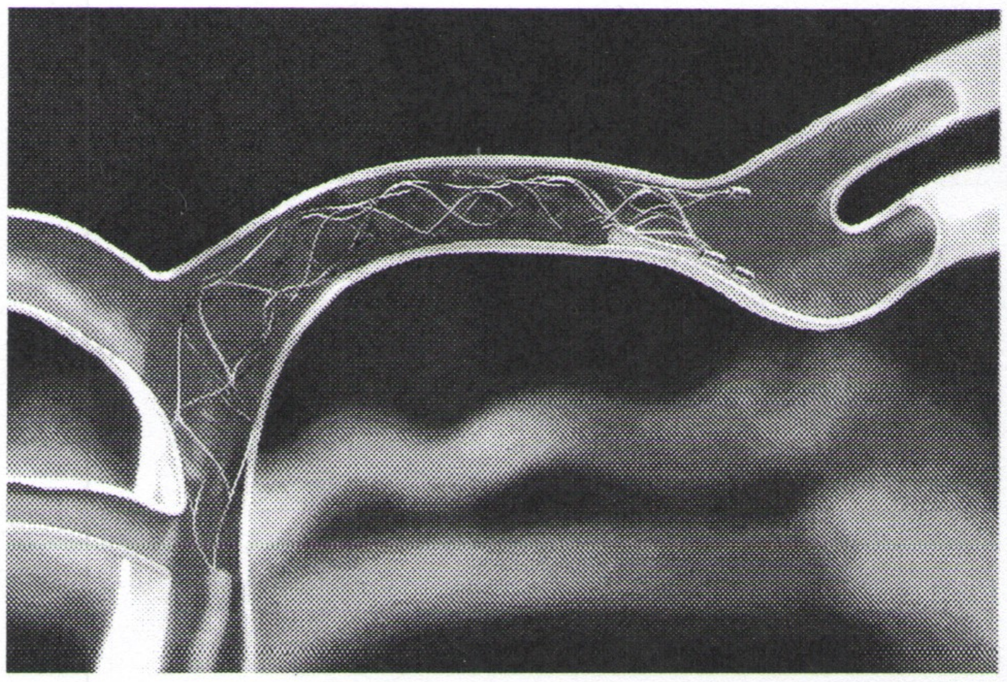
Overview

- Experimental Design and Statistical Significance
- Cases, Types of Variables

Section 1.1 A Case Study – Do Stents Help to Reduce the Risk of Stroke?

What is a stent? A wire support placed inside a blood vessel.

Research Question: Does the use of stents reduce the risk of stroke?



Cases: 451 at-risk patients volunteered for the study and were randomly assigned to two groups:

1 **Treatment Group**: Participants received a stent and medical management. The medical management included medications, management of risk factors and help in lifestyle modification.

2 **Control Group**: Participants received the same medical management as the treatment group, but they did not receive stents.

Experimental Design: Researchers randomly assigned 224 participants to the treatment group and 227 to the control group. They studied the effects at two time points: after 30 days and 365 days.

1. Why did the researchers use a control group?

(isolate the variable)

We need a baseline for comparison – otherwise we wouldn't know if the results were due to the stent or something else.

2. Why is it important to randomly assign participants to each group?

we want to spread out any additional variables between the 2 groups – age, gender, health pre-existing conditions, risk factors, ethnicity

Data Table: Identify the cases and the variables

cases → patients

variables in columns

Patient	group	0-30 days	0-365 days
1	treatment	no event	no event
2	treatment	stroke	stroke
3	treatment	no event	no event
⋮	⋮	⋮	
450	control	no event	no event
451	control	no event	no event

Table 1.1: Results for five patients from the stent study.

Results – Summary Statistics:

We want to compare the proportions (percentages) of patients who did not have a stroke in one year:

	Stroke in 0-365 days	No stroke in 0-365 days	Total
Treatment Group	45	179	224
Control Group	28	199	227
Total	73	378	451

Treatment group:

$\frac{179}{224} = .7991$ or 79.9%

Control group:

$\frac{199}{227} = .8767$ or 87.7%

Important Questions:

Are the results of this experiment due to random variation or are the results statistically significant?

The control group did better, which is surprising. There is an 8% difference between the 2 groups so it may be significant. Further study is needed.

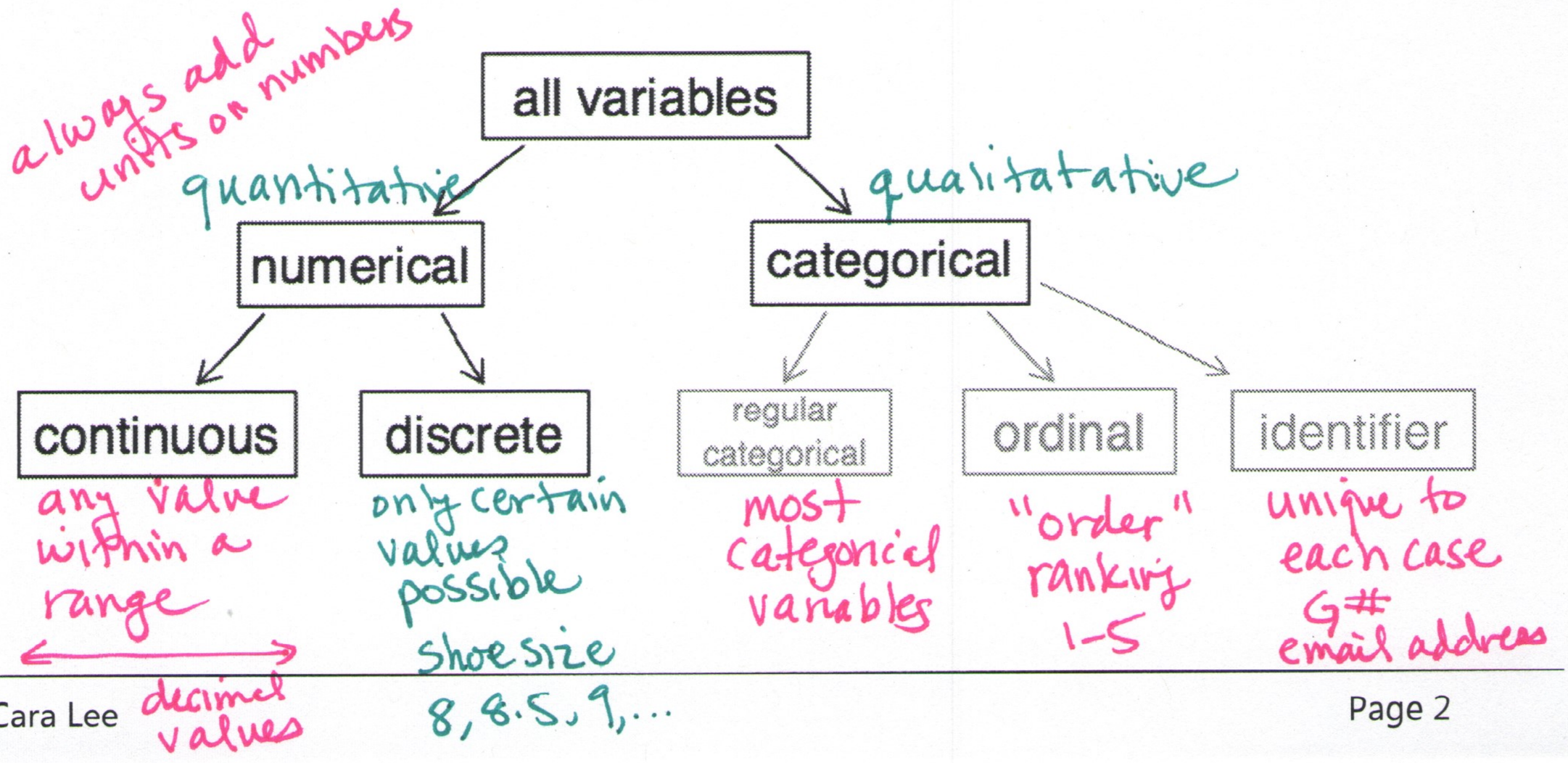
Can we generalize the results of this experiment to all patients at risk of stroke?

No, not without further study. Did the surgery increase the risk of stroke? Check the medical management of both groups.

Section 1.2 Data Basics - Types of Variables

Activity 1. Class Data Collection. Login to MyPCC and D2L. On the Content tab you will find a link to the Math 243 class survey. Please complete the anonymous survey. No identifying variables are collected.

Variables – The characteristics being recorded or measured



Activity 2. Class Data – Identify the Cases and Types of Variables

Define the cases:

Students in math 243

Write the type of variable represented by each question:
Categorical, Numerical Discrete, Numerical Continuous, Ordinal and/or Identifier

1. How do you identify your gender? (*female, male, non-binary, other*) categorical

2. About how many credits do you take each term? numerical discrete

3. About how many hours are you employed per week? numerical continuous

4. How many siblings do you have (not including yourself)? numerical discrete

5. Choose one of these numbers at random: 1 2 3 4. categorical (zip code)

6. Which award would you rather win:
Academy Award, Olympic Gold, Nobel Prize, none? Categorical

7. How many piercings (ear, nose, ...) do you have? (2 ears count as 2) Numerical Discrete

8. How many tattoos do you have? Numerical – Discrete / Continuous

9. Do you have children? (yes or no) Categorical

10. Overall, how satisfied are you with PCC?
(rate on a scale from 1=low to 5=high) ordinal

11. Which of the following pets do you have? Check all that apply. Categorical

- ☐ Dog
- ☐ Cat
- ☐ Fish
- ☐ Other:_____

Cases, Variables and the Research Question

Practice 2. For the study described below, identify (i) the cases, (ii) the variables and their types, and (iii) the main research question in the study.

In performing research for an ecology class, students at a college in upstate New York collect data on streams each year. To monitor the health of the local streams, they record a number of biological, chemical, and physical variables, including the stream name, the substrate of the stream (limestone, shale or mixed), the acidity of the water (pH), the temperature (°C), and the BCI (a numerical measure of biological diversity).

(i) The Cases – The respondents, participants, subjects or experimental units that are being asked or measured.

The Streams

(ii) The Variables – The characteristics or measurements that are being recorded or measured. Include units whenever the variable is numerical/quantitative.

Stream name – identifier
substrate of the stream (limestone, shale, mixed) – categorical
acidity (pH) – numerical continuous
BCI – numerical (we don't know whether it's discrete or continuous)
Temperature (°C) – numerical continuous

(iii) The main research question – The purpose of the data collection. What are researchers trying to figure out?

They want to monitor the health of the streams over time. Reporting to agencies. Educational purposes.

Practice 2. Sinusitis and Antibiotics, Part I. (Problem 1.2). Researchers studying the effect of antibiotic treatment for acute sinusitis compared to symptomatic treatments randomly assigned 166 adults diagnosed with acute sinusitis to one of two groups: treatment or control. Study participants received either a 10-day course of amoxicillin (an antibiotic) or a placebo similar in appearance and taste. The placebo consisted of symptomatic treatments such as acetaminophen, nasal decongestants, etc. At the end of the 10-day period patients were asked if they experienced significant improvement in symptoms. The distribution of responses is summarized below.

(a) What percent of patients in the treatment group experienced a significant improvement in symptoms? What percent in the control group?

		Self-reported significant improvement in symptoms		
		Yes	No	Total
Group	Treatment	66	19	85
	Control	65	16	81
	Total	131	35	166

Treatment Group: $\frac{66}{85} = .7765$ or about 78%

Control Group: $\frac{65}{81} = .8025$ or about 80%

(b) Based on your findings in part (a), which treatment appears to be more effective for sinusitis?

The control group had a higher percentage, which is the opposite of what we would expect.

(c) Do the data provide convincing evidence that there is a difference in the improvement rates of sinusitis symptoms? Or do you think that the observed difference might just be due to chance?

The difference between the groups is small, only about 2%, so it seems like this difference is due to chance. The resuets do not seem statistically significant.