

Overview

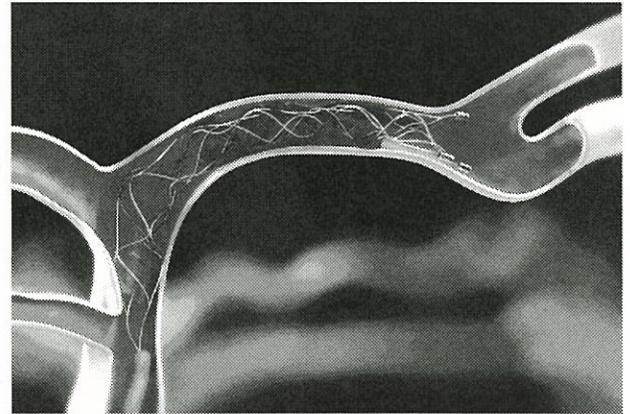
- Experimental Design and Statistical Significance
- Cases, Types of Variables
- Population, Sample, Parameter and Statistics

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

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

What is a stent?

A support placed inside a blood vessel.



who or what is being studied

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

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

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

Baseline to compare with

Experimental Design: Researchers randomly assigned participants 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?

They need a comparison group that did not get the treatment.

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, ethnicity

Data Table: Identify the cases and the variables

The cases are the patients (subjects)

Variables – what we're measuring

patient # group

0-30 days – did they have a stroke

0-365 days – did they have a stroke

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: *- decimal form*

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: $\frac{179}{224} = .7991$

Control group: $\frac{199}{227} = .8767$

*or 79.9% or 80%
or 87.7% or 88%*

results are greater than that due to chance

Important Questions:

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

It's about an 8% difference, which is large, but the results are not what we expected. Maybe the sample size was too small. Further study would be needed.

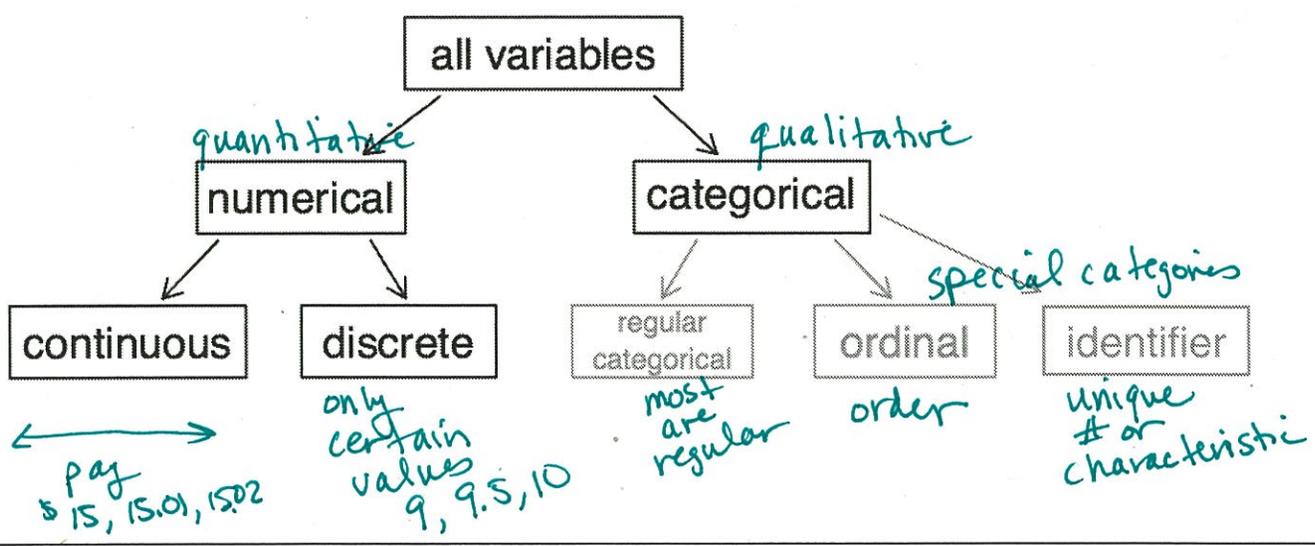
Can we generalize the results of this experiment to all patients at risk of stroke?

NO, not without further study.

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

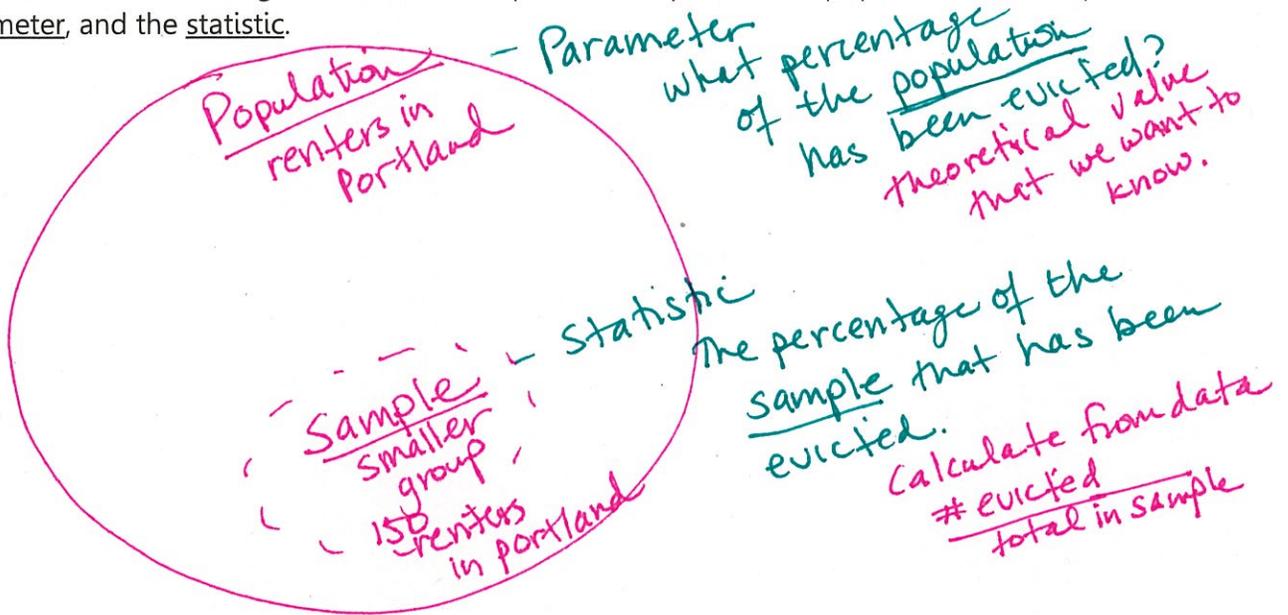
Write the type of variable represented by each question:

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 - $20, 21, 22$ discrete
 $20.1, 20.2$ 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
numbers can represent categories
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 or 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) categorical-ordinal
11. Which of the following pets do you have? Check all that apply.
 - Dog
 - Cat
 - Fish
 - Other: _____categorical

Section 1.3 Overview of Data Collection - Population, Sample, Parameter and Statistic

Example 1. (Categorical ^{7%} yes/no variable.) Researchers are interested in how many people have been evicted from their housing. Draw and label a picture to represent the population, the sample, the parameter, and the statistic.



Practice 1. (Numerical ^{average} variable.) Student Services is interested in how many hours students spend on campus per week when they are not in class. Draw and label a picture representing the population, the sample, the parameter, the statistic.

Notation

Type of Data	Quantity of Interest	Population Parameter	Sample Statistic
Categorical (yes/no)	Proportion	p	\hat{p}
Numerical	Mean	μ	\bar{x}

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.

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

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

Practice 3. 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?

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

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

(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?

Practice 4. Cats on YouTube (Problem 1.10). Suppose you want to estimate the percentage of videos on YouTube that are cat videos. It is impossible for you to watch all videos on YouTube so you use a random video picker to select 100 videos for you. You find that 2% of these videos are cat videos. Determine which of the following is an observation, a variable, a sample statistic or a population parameter.

(a) Percentage of all videos on YouTube that are cat videos

(b) 2%

(c) A video in your sample

(d) Whether or not a video is a cat video