Ch. 2: Generalization: How Broadly Do the Results Apply

Navigation

By Date

- February 4: start - end
- February 6: start - end
- February 11: start - end

By Section

- Section 2.1
- Section 2.2
- Section 2.3

Ch 2 Goals and Overview

- Explain when and why you can (sometimes) think of choosing your sample as a random process

    - When you can, Ch 1 applies
    - When you can't, use caution

- Sections:

    1. How sampling works in practice:
       Either from a process or a population

    2. Quantitative data:
       how processes work for means

    3. Errors and significance:
       Controlling how often we make a mistake

2.1: Sampling From A Finite Population

Review

population - The entire collection of observational units we are interested in

$$\text{population} \rightarrow \text{parameter}$$

sample - The subset of the population on which we record data

$$\text{sample} \rightarrow \text{statistic}$$

Up until now, we have sampled data using a **process** (coin flips, etc.).

Now, we're going to be sampling from a **population** of individuals; we'll record data about those individuals instead of the outcome of the process.

Example 2.1 - Sampling Students

Data from "College of the Midwest" for all students enrolled in Spring 2011.

Table 2.1: Data table of College of the Midwest

Student ID	Cumulative GPA	On campus?
1	3.92	Yes
2	2.80	Yes
3	3.08	Yes
4	2.71	No
5	3.31	Yes
6	3.83	Yes
7	3.80	No
8	3.58	Yes
...	...	...

What type of data is this?

Example 2.1 - Sampling Students

Data from "College of the Midwest" for **all students** enrolled in Spring 2011.
=> Population data

Student ID	Cumulative GPA	On campus?
	quantitative	*binary*
1	3.92	Yes
2	2.80	Yes
3	3.08	Yes
4	2.71	No
5	3.31	Yes
6	3.83	Yes
...	...	...

A **data table** stores each observation in one row, and each variable in a separate column.

Example 2.1 - Sampling Students

The entire **population** for College of the Midwest is 2919 students. We can use the data from these students to examine the distribution using an applet.

http://www.math.hope.edu/isi/applets.html

Histogram - a graph that summarizes the distribution of a quantitative variable

- separate the data into bins
- bar height = number of observations in the bin.
- bar width = limits of the bin

Parameters - numerical summaries of the population

Sampling Methods

We usually don't get to conduct a census (e.g. collect data from the whole population). Instead, statisticians sample from the population.

A **representative sample** is a sample collected that is similar to the results we would find in the entire population

---

Suppose the researcher only has time to collect data on a sample of students. The researcher asks the first 30 students he finds on campus one morning whether he or she lives on campus.

Is this sample **representative**? Discuss in your groups.

Sampling Methods

A **convenience sample** is a nonrandom sample of a population

What could the researcher have done instead to obtain a representative sample?

Why should we worry about whether or not the sample is not representative?

Bias: A sampling method is biased if the statistics from the samples consistently over-estimates or consistently under-estimates the population parameter of interest.

Biased Sampling

In 1948, major polling operations were using a technique called "quota sampling", where each interviewer was given a certain number of individuals in a demographic to interview (e.g. 7 white men living in rural areas). Interviewers were free to choose anyone who met the criteria to be part of the sample.

Data were collected until the amount of data collected matched statistics about the nation as a whole (race, gender, above/below 40, etc.).

Is this sampling method representative? Is it random?

Biased Sampling

Gallup conducted a poll on the presidential election using this method and determined that approximately 50% of the vote was for Thomas Dewey, 44% for Harry Truman, and 6% for third-party candidates Strom Thurmond and Henry Wallace.

The lead in the polls was so big, the Chicago Daily Tribune printed the headlines before the election results were in.
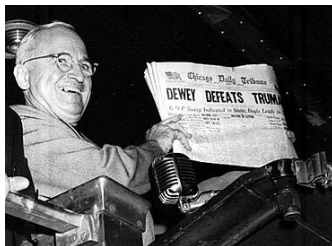
Biased Sampling

Later investigations of quota sampling found that there were two main problems with the method:

1. Interviewers could choose the people who they interviewed
   lack of randomness

2. You can always add more demographic characteristics, leading to ever more strict rules for who can be sampled.
   hard to get a truly representative sample

The biggest problem, however, was that quota sampling had a clear tendency to overestimate the Republican vote.
The quota sampling method was **biased** in favor of Republicans.

Sampling Methods

A **simple random sample** is a sample that ensures that every group of size n is equally likely to be the sample selected from the population

How do you take a simple random sample from a population?

- Sampling Frame: a list of all individuals (observational units) in a population
- Computer randomly selects n individuals in the sampling frame

If not all individuals in the population can be included in the sampling frame, the resulting random sampling can be biased.

Bias in random sampling methods:

- 2010 Pew Research Study: Landline only polls are biased
- 2017 Pew Research Study: Low response rates cause bias on some topics, but not others

---

In both cases, not all individuals in the population are equally likely to be included in the sample.

Example 2.1: Simple Random Sample

A simple random sample of 30 students gives us the following data:

ID	GPA	OnCampus	ID	GPA	OnCampus	ID	GPA	OnCampus
237	3.86	Y	595	3.55	Y	1600	2.26	N
2088	3.05	Y	2338	3.33	N	536	2.13	Y
2722	3.63	Y	184	3.71	Y	927	3.54	Y
118	2.99	N	510	3.72	Y	2136	3.63	Y
1772	2.70	N	737	2.93	Y	2194	3.09	Y
2109	2.53	Y	2755	3.18	Y	54	2.62	N
2373	3.40	N	11	3.42	Y	2461	2.49	Y
717	3.04	Y	1585	3.62	N	2048	3.57	N
1733	3.34	Y	444	3.95	Y	2160	3.93	Y
1497	3.36	Y	2882	2.22	Y	1855	3.54	Y

Statistics:

- Sample average of cumulative GPA $\overline{x} = 3.211$
- Sample proportion of on-campus students $\hat{p} = 0.733$

Sampling

Sampling Variability - variability from sample to sample

What if we take multiple samples of size 30 from the population?

- Each SRS will give us different values for the statistics

- What happens if we take 1000 different SRS? (long run pattern)

Exploration 2.1 - Sampling Words

Select a representative set of 10 words from the passage by circling them with your pen or pencil

Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we cannot dedicate, we cannot consecrate, we cannot hallow this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here.

It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion that we here highly resolve that these dead shall not have died in vain that this nation, under God, shall have a new birth of freedom and that government of the people, by the people, for the people, shall not perish from the earth.

Exploration 2.1 - Sampling Words

Record each word from your sample and indicate the length of the word (# letters) and whether or not it contains at least one letter e.

	Word	Length	Contains e? (Y/N)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Exploration 2.1 - Sampling Words

Individually, do questions 4-9 before Thursday. Upload your answers to Canvas by Thursday at 3:30 pm.

We will use google sheets on Thursday to collect and combine your results

---

Review

Review - Section 2.1

- **population** - entire collection of observational units
- **sample** - subset of data
- **parameters** - numerical summary of the population
- **statistics** - numerical summary of the sample

Review

- What is sampling?

- Why do we sample? What is the goal?

- What makes a sample good or bad?

Review

- **Representative** - data from a sample will be similar to results from the population

- **Bias** - statistics that consistently over/under estimate the population parameter, based on sampling method; occurs when not taking a representative sample

Review

- **Convenience sample** - a nonrandom sample of a population

- **Simple Random Sample** - every member of the population has equal probability of being chosen. To do this, you have to define (and list) the entire population.

- **Sampling frame** - List of individuals in a population, used when choosing a random sample

    - What might be the sampling frame to sample UNL students?

    - How might we select a SRS of everyone in the entire state?

Exploration 2.1

Answer Question 10 here: https://forms.gle/8vCzyVAxLGLx84hH6

Exploration 2.1

Answer Question 17 here: https://forms.gle/UxaVigrMn1gpVpaD7

Turn in answers to 10-19 on Canvas by Monday at 6pm. I will post the graphs shown in class on Canvas.

Review

New Symbols

- Sample mean: _____
- Sample standard deviation: _____
- Population mean: _____
- Population standard deviation: _____

Review

New Symbols

- Sample mean: $\overline{x}$

- Sample standard deviation: s

- Population mean: μ

- Population standard deviation: σ

2.2: Inference on a Single Quantitative Variable

Introduction

In Ch 1, we used categorical/binary data...

Inference:

- Simulation (3 S's strategy)
- Standardized Statistics / p-values
- Theory based approximations

Now, we're going to shift to **quantitative** variables with the hopes of learning about the **population mean**

Testing with means

Variable of interest is now **quantitative**

- number of emails in a day
- amount of money spent Christmas shopping
- temperature

Parameter is the population mean μ

Null and Alternative Hypotheses:

$H_0 : \mu = a$
$H_A : \mu > a$, or $H_A : \mu < a$, or $H_A : \mu \neq a$

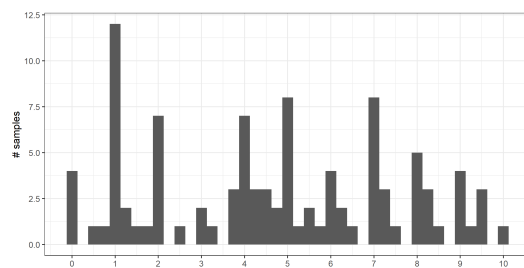Tests have the same interpretation and conclusions using p-values and standardized statistics

---

Describing Distributions

Quantitative variables can be distributed in more interesting ways...

Shape

Rounding effects (very common in human-recorded data)

Describing Distributions

Quantitative variables can be distributed in more interesting ways...

Shape
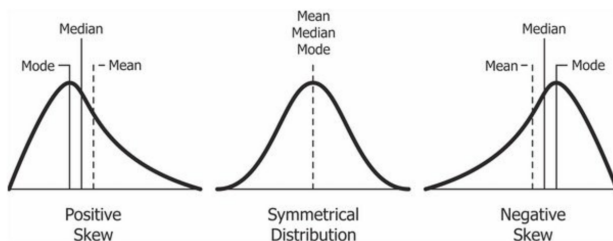
Describing Distributions

Center

If skew is shifting the mean, we might use another measure of center

- **median** - the middle number when the data values are sorted from smallest to largest
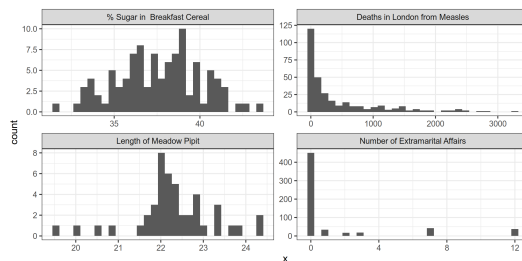- **mode** - the most common value

We usually use the median when the mean is not appropriate

Describing Distributions

Variability

The standard deviation of a binary variable is generally fixed; with quantitative variables, there is much more freedom.

# Describing Distributions

Unusual Observations

- Unusual observations are called **outliers**
- Some extreme values can impact our summary statistics (e.g. mean, standard deviation)
- A summary statistic is called **resistant** if its value does not change much when extreme observations are removed from the data set
- The median is a resistant summary statistic

Describing Distributions

Example
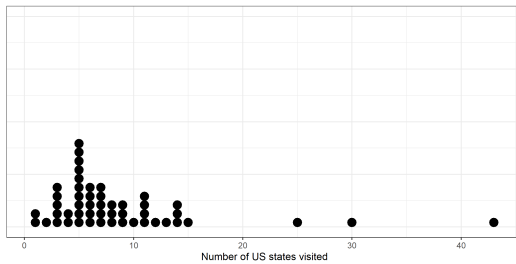
This distribution is skewed right

Median is ___

What if 43 is a typo and it should be 34?

Inference of Quantitative Data

- Simulations
    - Get many samples of the same sample size from the population
    - Create a null distribution using the average of each simulated sample
- Find p-value by evaluating how our sample statistic relates to the null distribution
- Null distribution is centered at the value of the population mean according to the null hypothesis
- Review how the sample size affected the null distribution
- Simulations are difficult

Theory-Based Inference for Quantitative Data

- The principles from categorical data carry over to quantitative data
- Using theory-based methods, the standard deviation of the sample mean is calculated to be

$$s = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Remember, σ is the standard deviation of the population

σ is a parameter

Standardized Statistic

Remember, the general template is

$$z = \frac{statistic - parameter_{H_0}}{\sigma}$$

Oftentimes, we don't know σ, which is fine because we can rely on s

We can use the standard error in place of σ

$\frac{s}{\sqrt{n}}$ is the **standard error** (measure of the accuracy of an estimate) of the sample mean

Standardized Statistic for Chapter 2

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

T-statistics vs. Z-statistics

- As with normal approximation, we need a reference distribution
  - Categorical data: find z and compare to a normal distribution
  - Quantitative: use s in place of σ -> we can't compare to a normal distribution
    Instead, we use the t distribution, which is very similar
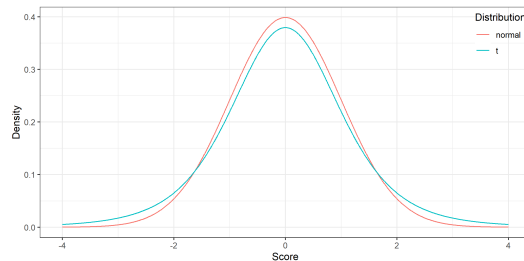
Interpretations and conclusions are still the same

Distributions

Z-Distribution	T-Distribution
Categorical data - falls into specific groups; corresponding probabilities	Quantitative Data - numerical values, with mean/averages
Test estimates of population proportions	Test estimates of population mean
Theory Validity Conditions 10 successes and 10 failures	**Theory Validity Conditions** $n \geq 20$ (sample size)
Can simulate because we have finite groups and/or a known process	Cannot simulate

Significance test - Example 2.2 (Sleep)

From the Theory Based Inference Applet:

- Sample size: $n = 21$
- Sample mean: $\overline{x} = 6.69$
- Sample standard deviation: $s = 1.295$

On a quiz or exam, values or data will be given

Hypotheses:

- To see if the population average hours of sleep is 8 or is less than 8

Using our sample, we need to find a p-value and a standardized statistic

P-value from Theory based inference Applet: _____

Significance test - Example 2.2 (Sleep)

Interpretation With our p-value of _____, we (reject/fail to reject), because the p-value is (less/greater) than 0.05.

Conclusion From our p-value we (do/do not) have enough evidence to conclude that the average hours slept per night for UNL students is significantly below 8 hours.

Standardized Statistic - Example 2.2 (Sleep)

Calculation with Applet values:

- Sample size: $n = 21$
- Sample mean: $\overline{x} = 6.69$
- Sample standard deviation: $s = 1.295$

1. Find the standard error:

$$\frac{\text{sample std.dev.}}{\sqrt{\text{sample size}}} =$$

We are trying to use the **standard deviation from our one sample** to estimate the standard deviation of the mean for **many samples** of size n

2. Input values into S.S. formula

$$t = \frac{\overline{x} - \mu_0}{\text{standard error}} =$$

Standardized Statistic - Example 2.2 (Sleep)

1. Find the standard error

2. Input values into S.S. formula

3. **Interpretation**: We will _____ the null because our standardized statistic is _____, which is _____ and provides _____
   to suggest that UNL students get _____ 8 hours of sleep per night.

4. **Conclusion**: We can use the same conclusion as before, but with the standardized statistic as our reasoning.

Example: Sunscreen

Most dermatologists recommend using sunscreens that have a sun protection factor (SPF) of at least 30. One of the authors wanted to find out whether the SPF of sunscreens used by students at her school (which is in a very sunny part of the US) exceeds this value, on average?

- Identify the variable of interest and whether the variable is categorical or quantitative.

- Describe the author's parameter of interest and assign an appropriate symbol to denote it.

- Write the appropriate hypothesis using symbols

Example: Sunscreen

Most dermatologists recommend using sunscreens that have a sun protection factor (SPF) of at least 30. One of the authors wanted to find out whether the SPF of sunscreens used by students at her school (which is in a very sunny part of the US) exceeds this value, on average?

- Identify the variable of interest and whether the variable is categorical or quantitative.
  SPF of sunscreen used by students at the school

- Describe the author's parameter of interest and assign an appropriate symbol to denote it.
  average SPF of sunscreen used by all students at the school; (\mu)

- Write the appropriate hypotheses using symbols

$$H_0 : \mu = 30$$
$$H_A : \mu > 30$$

Example: Sunscreen

To collect data, the author surveyed students in her introductory statistics course and found that in a sample of 48 students the average SPF was 35.29 and the standard deviation of the SPF was 17.19

- Assign appropriate symbols to the numbers:

$$= 48 \qquad = 35.29 \qquad = 17.19$$

- Is the theory-based approach (one-sample t-test) valid to find a p-value to investigate whether the data provide evidence that the average SPF of sunscreens used by the students at this school exceeds 30? Explain

- Find and report a standardized statistic (t-statistic)

- Use the Theory-Based Inference applet to conduct a one-sample t-test and find a p-value

- State your conclusion from this study based on the p-value, being sure to provide justification for your conclusion.

Example: Sunscreen

To collect data, the author surveyed students in her introductory statistics course and found that in a sample of 48 students the average SPF was 35.29 and the standard deviation of the SPF was 17.19

- Assign appropriate symbols to the numbers:

$$n = 48 \qquad \bar{x} = 35.29 \qquad s = 17.19$$

- Is the theory-based approach (one-sample t-test) valid to find a p-value to investigate whether the data provide evidence that the average SPF of sunscreens used by the students at this school exceeds 30? Explain
  Yes, the theory-based approach is valid. We have more than 20 observations in our sample, which means that we can expect our sample mean to be normally distributed thanks to the central limit theorem.

- Find and report a standardized statistic (t-statistic)

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{35.29 - 30}{\frac{17.19}{\sqrt{48}}} = 2.132$$

Example: Sunscreen

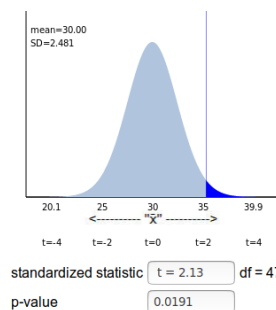- Use the Theory-Based Inference applet to conduct a one-sample t-test and find a p-value

Example: Sunscreen

- State your conclusion from this study based on the p-value, being sure to provide justification for your conclusion.

We reject the null hypothesis, because p-value = 0.0191 is less than 0.05. We have evidence that the average SPF of sunscreen used by students at this college is higher than 30.

2.3: Errors and Significance

Introduction

So far:
Significance tests are for examining the **strength of evidence** against H_0

Another approach:
Significance tests are used to **make a decision** about whether the data provide enough evidence to reject H_0

But decisions imply making mistakes...

Significance and Errors

- How extreme is extreme enough?

  - Changes with the study: the consequences may be different for different types of errors

    Criminal justice system:

    > It is better that ten guilty persons escape than that one innocent suffer - William Blackstone

- In statistics, we control "extreme enough" with the value where we reject the null, aka. **significance level** or **alpha** (α)

- Changing α has consequences:

  - if we increase it (lowering the level of "extreme" necessary to reject), we might find evidence of an effect even though there is only chance.
  - if we decrease it (we need more evidence to reject), we might fail to reject an effect that is actually there

Errors

- Any time we make a decision based on data, a mistake is possible

- In significance tests, there are two types of mistakes we can make:

  - (The effect is really there) and we fail to reject H_0
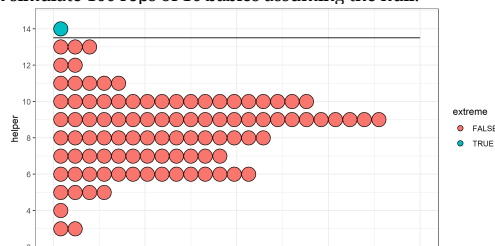 - (The effect is really not there) and we reject H_0

		Reality	
		`$H_0$` is true	`$H_A$` is true
What we say	Fail to reject `$H_0$`	Correct decision!	Missed opportunity (Type 2 error)
	Reject `$H_0$`	False alarm (Type 1 error)	Correct decision!

Errors

- Helper toy example: Do more than half the babies pick the helper toy?

$$H_0 : \pi = 0.5 \quad H_A : \pi > 0.5$$

- We can simulate 100 reps of 16 babies assuming the null:

- Is this strong evidence against the null?

- Enough evidence to reject?

Errors

- Usual threshold is $\alpha = 0.05$
  What that actually means: we are ok with making a type 1 error 5% of the time

- We can choose something other than 5% ($\alpha = 0.05$)

- What if $\alpha = .10$?

  - 10% of the time we will reject a true null hypothesis (in the long run)

- The only thing that changes in our procedure is that we're deciding the cutoff for rejecting; it might be something other than 5%.

- Where do we draw the line?

  - How small does p have to be?
  - Field dependent
  - Consequences dependent: Would you want a lower p-value for
    - deciding whether someone should be executed,
      or for
    - deciding whether someone should get a sticker?

Types of Error

Type 1 Error - reject the null hypothesis when H_0 is actually true

- False alarm - believe we have discovered a difference when no difference exists
- i.e. Students at UNL use sunscreen that is on average 30 spf, but we said it was more.

Type 2 Error - fail to reject H_0, but in reality, H_0 is actually false

- Missed opportunity - failing to detect a difference that is there
- i.e. UNL students use sunscreen that is over 30 spf, but we did not reject the null hypothesis of $\mu = 30$

Reality

- We never actually know whether we've made an error

- We can control what we want to happen (depending on discipline)

- **Significance level** is the probability of Type 1 error

  - Smaller α = less likely to make a Type 1 error
  - Trade off: more likely to make a Type 2 error

- **Power** is the probability of rejecting a false null hypothesis

  - We want to reject when the null is not true
  - We want our tests to be as powerful as possible

Reality

		Reality	
		H_0 is true	H_A is true
What we say	Fail to reject H_0	Correct decision!	Missed opportunity (Type 2 error)
	Reject H_0	False alarm **(Type 1 error)**	Correct decision! **Power**

Example

You are sitting on a jury and must decide whether the defendant is innocent or guilty. If they are guilty, they will go to jail.

- Hypothesis (in words):

  - H_0 :
  - H_A :

- Errors:

  - Type 1:
  - Type 2:

- Which error would be worse?

- Significance level:

- How should we adjust our significance level?

Example

You are sitting on a jury and must decide whether the defendant is innocent or guilty. If they are guilty, they will go to jail.

- Hypothesis (in words):

  - H_0 : The defendant is innocent
  - H_A : The defendant is guilty

- Errors:

  - Type 1: Send the defendant to jail when they are innocent
  - Type 2: Let a guilty person go free

- Which error would be worse? In common law jurisdictions, it is seen as worse to send an innocent person to jail.

- Significance level: Percent of time that you send an innocent person to jail

- How should we adjust our significance level? (If it's worse to send an innocent person to jail) we would want to decrease our significance level α so that the probability of sending an innocent person to jail is lower.

Example

You need to decide whether to bring your umbrella with you tomorrow

- Hypothesis (in words):

    - H_0 :
    - H_A :

- Errors:

    - Type 1:
    - Type 2:

- Which error would be worse?

- Significance level:

- How should we adjust our significance level?

Example

You need to decide whether to bring your umbrella with you tomorrow

- Hypothesis (in words):

    - H_0 : It will not rain tomorrow (I should leave the umbrella at home)
    - H_A : It will rain tomorrow (I should bring my umbrella)

- Errors:

    - Type 1: Bringing your umbrella even though you don't need it
    - Type 2: Getting rained on with no umbrella

- Which error would be worse?

- Significance level: Proportion of the time you are stuck carrying an umbrella when it's not raining?

- How should we adjust our significance level?

Errors

The most important thing about errors is when they occur

- Type 1 errors only occur if you reject H_0
- Type 2 errors only occur if you fail to reject H_0

Common quiz/test question:

Draw a conclusion for a study and determine which type of error you may have made.