

STAT 121: Intro to Data Analysis

Icebreakers



About Me

- My Hometown(s)
- In my free time I...
- Something about me most people would never guess is...
- My favorite thing about teaching is...
- My favorite gospel principle is...



Your Turn (Talk to a Neighbor!)

Take 2 minutes to find out:

1. My name is _____, I'm from _____ and I'm majoring in _____.
2. My eventual dream job is _____.
3. My anxiety level for this class is _____ (1-10 with 10 = dreading it).
4. A game show that I would dominate is _____

What is Statistics?

- Statistics:
 - The science of solving data problems by collecting, organizing, analyzing, interpreting, and presenting data.
- Statisticians / Data Scientists:
 - Professionals who use data to solve complex problems.
For example...

Environmental Statistics

Statisticians help manage the earth's natural resources



Sports Statistics

Statisticians work for professional sports teams to make strategic draft picks



Statistical Genetics

Statisticians reduce disease and improve medicines

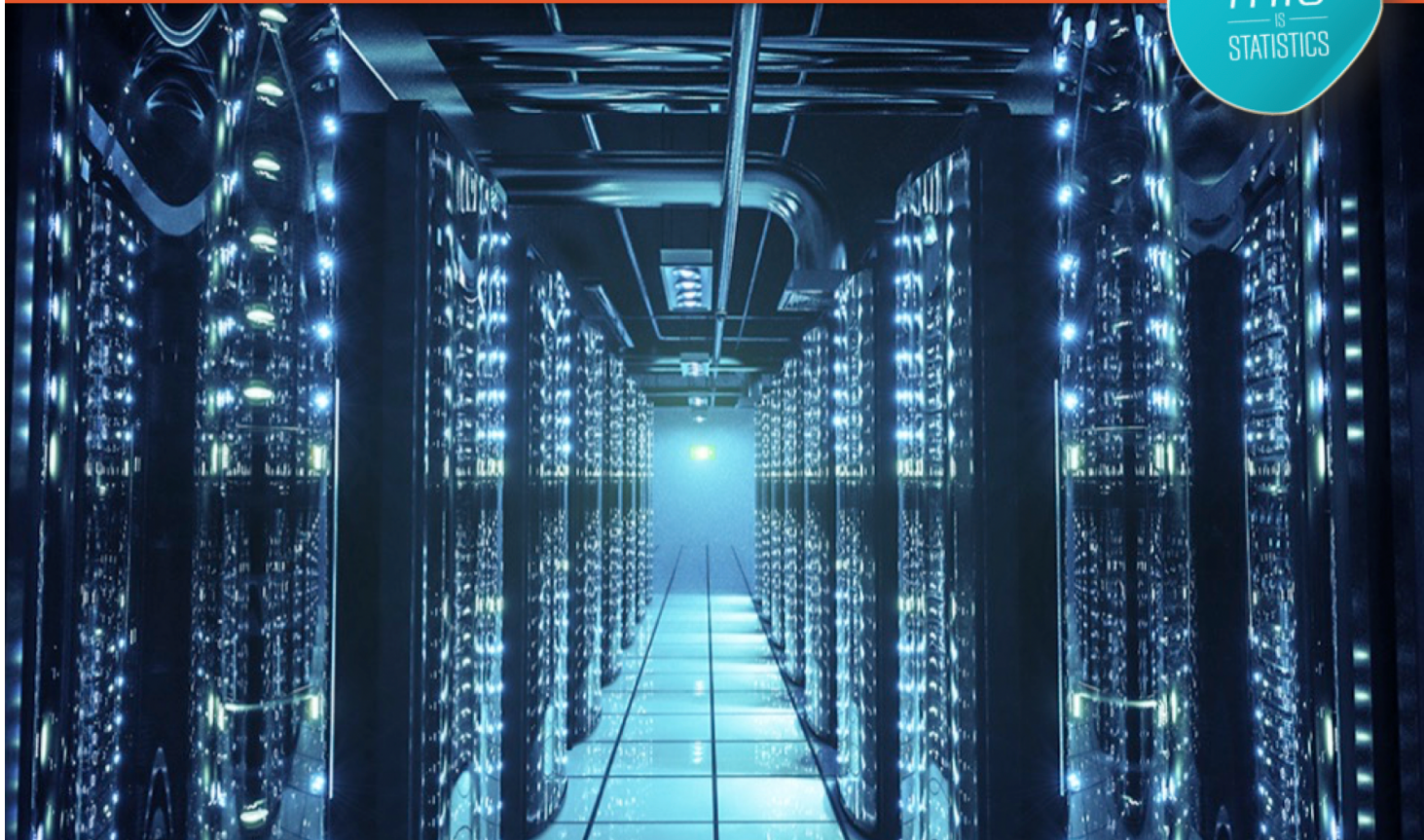
THIS
IS
STATISTICS



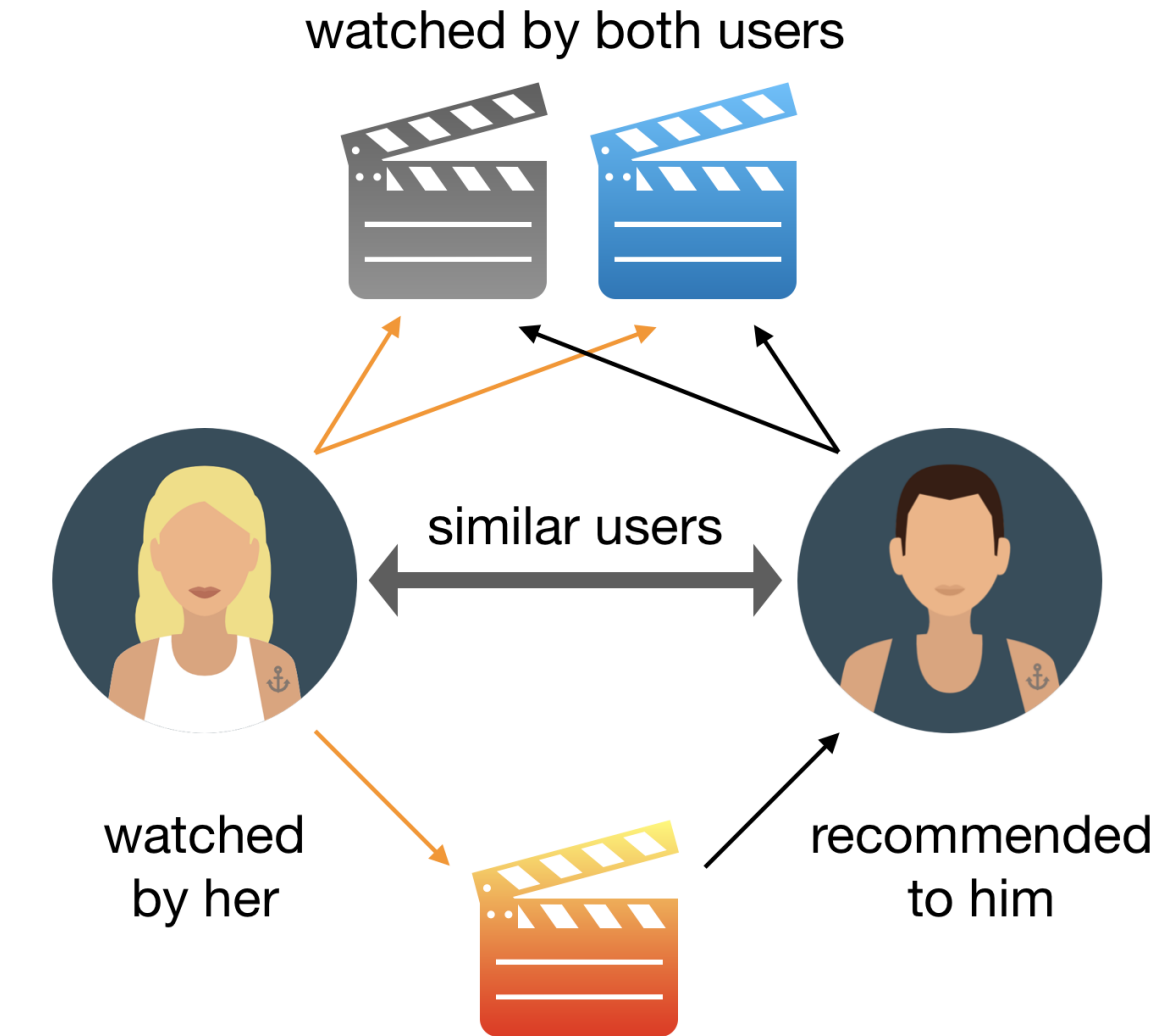
Artificial Intelligence

Statisticians advance computing through machine learning, speech recognition and artificial intelligence

THIS
IS
STATISTICS

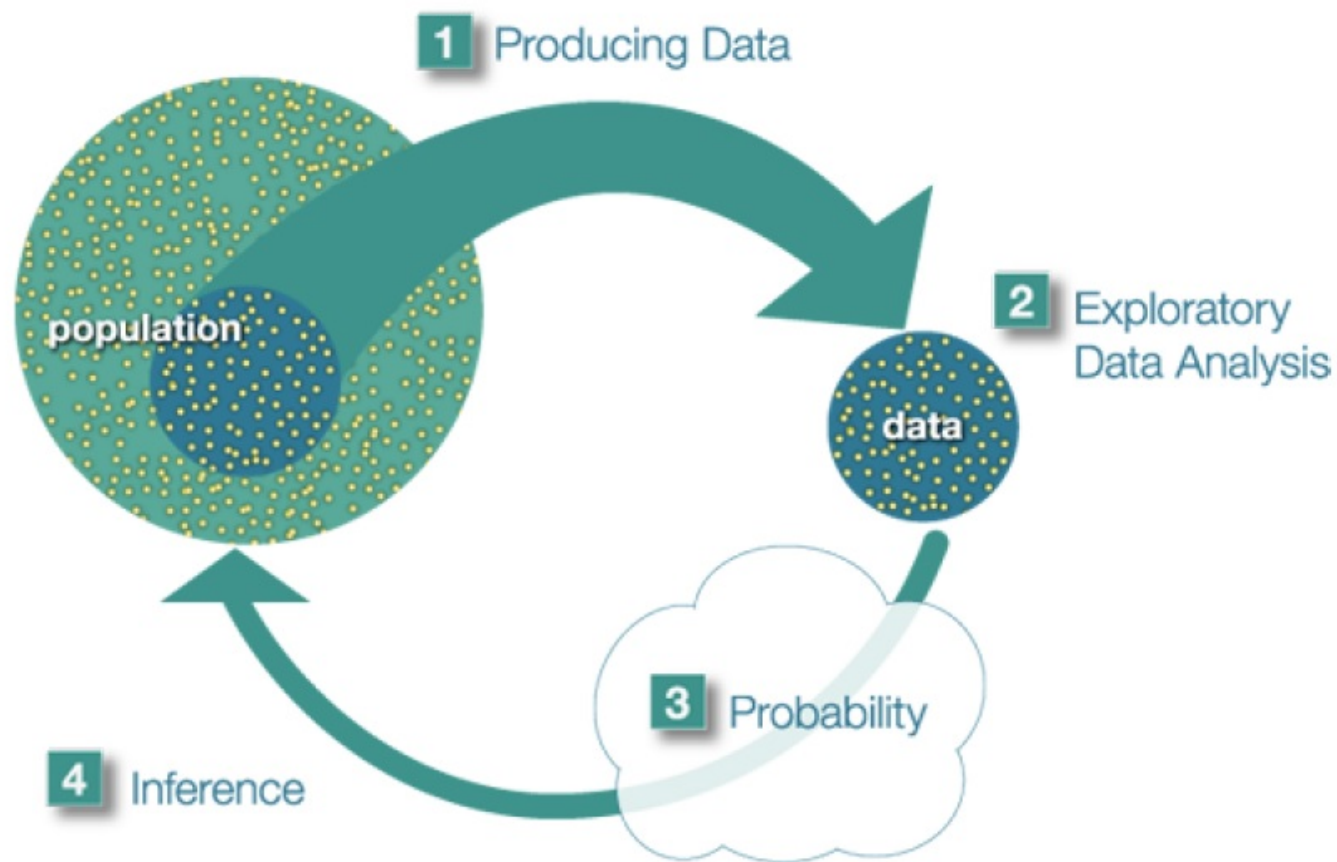


Recommender Systems



What's in Stat 121?

The basic process of statistical analysis and modeling...



The Process of Statistical Modeling

1. Identify a **population** and **parameter** you are interested in studying.

- **Population** - the entire group of “individuals” to which we can make conclusions about
- **Individual** - an “entity” in the population (a rat, a person, a user)
- **Parameter** - a numerical characteristic of the population you want to learn about

2. Collect data by **sampling** from the population.

- **Sample** - the subgroup of individuals from the population that we have usable data on
- **Variable** - a feature that is measured and varies across individuals
- **Statistic** - a numerical characteristic of the sample (usually corresponds to the parameter)

The Process of Statistical Modeling

3. Posit a **model** for the population based on information from the sample.

- **Statistical Model** - a mathematical/probabilistic representation of the population

4. Draw inference about the population using your model.

- **Statistical Inference** - making a conclusion about the population based on information in the sample.

Example

A high school principal wishes to know the percentage of seniors at his high school who are planning on attending college. To figure this out, he randomly samples 50 seniors from his school and asks them if they plan to attend college after high school.

1. What is the population?
2. What is the individual?
3. What is the parameter?
4. What is the sample?
5. What is the variable?
6. What is the statistic?
7. What would constitute “statistical inference” in this example?

Example

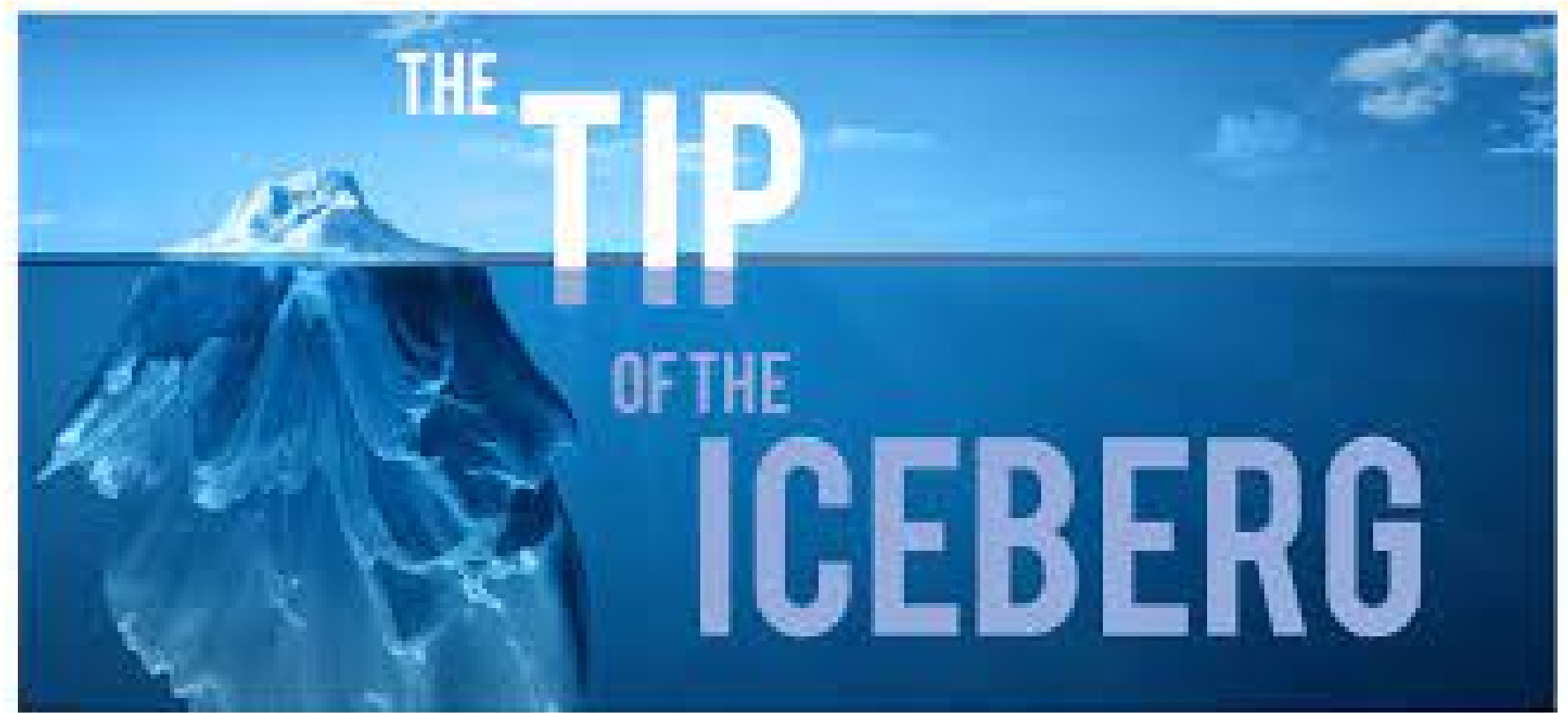
A high school principal wishes to know the percentage of seniors at his high school who are planning on attending college. To figure this out, he randomly samples 50 seniors from his school and asks them if they plan to attend college after high school.

1. What is the population? All seniors at this high school.
2. What is the individual? A senior
3. What is the parameter? The percentage of ALL seniors who plan to attend college.
4. What is the sample? The group of 50 seniors
5. What is the variable? Whether a senior plans to attend college or not
6. What is the statistic? The percentage of the 50 sampled seniors who plan to attend college.
7. What would constitute “statistical inference” in this example? Using the percentage of the 50 sampled seniors to plan to attend college to understand the percentage of ALL seniors who plan to attend college.

What's in Stat 121?

1. Data Collection
2. Univariate Quantitative Data Analysis
3. Univariate Categorical Data Analysis
4. Analysis of Multiple Means
5. Analysis of Multiple Proportions
6. Simple Linear Regression
7. Multiple Linear Regression

What's in Stat 121?



A Note on Computing

All statistical analysis generally requires the use of a computer. Different disciplines use different statistical software:

- Social Sciences: Stata, SPSS, SAS
- Life Sciences: SAS, R
- Mathematical Sciences: R, Python
- Business: Excel, Minitab, R

For this class, we are going to use an [online R tool](#) that was developed specifically for this class.

A Note on AI

There is a now famous study that came out in 2025 that looked at the effect of using AI tools (like ChatGPT) on learning. The study used fMRI brain scans to measure brain connectivity (a proxy for learning) in students who were asked to complete tasks either with or without AI assistance:

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task

[Nataliya Kosmyna](#), [Eugene Hauptmann](#), [Ye Tong Yuan](#), [Jessica Situ](#), [Xian-Hao Liao](#), [Ashly Vivian Beresnitzky](#), [Iris Braunstein](#), [Pattie Maes](#)

the main conclusions of the study were:

- those who initially used AI to do tasks had substantially lower brain connectivity (i.e. learning)
- those who initially did NOT use AI to do tasks had substantially *higher* brain connectivity (i.e. learning)
- those who initially did NOT use AI but did learn to use AI after an initial learning period still had high brain connectivity

A Note on AI

In this course, you may not use AI tools (e.g., ChatGPT, Copilot, Gemini) to complete assignments, projects or exams.

This is an introductory course, and our goal is to develop your own analytical thinking skills. While AI can be a valuable tool in professional data analysis, it can also shortcut the learning process before you've built a solid foundation. We want you to fully understand the principles and processes of data analysis so that, later, you can use AI effectively and critically — rather than relying on it blindly.

Course Learning Activities

- In-class Practice Problems (5% of your grade): Near daily class quizzes during lecture. We will drop your lowest 5 scores.
- Homework (10% of your grade): online in LearningSuite covering about 1-2 lectures and part of an analysis. Mostly multiple choice, numerical answers, etc.
 - 3 HW versions, keep the high score
- Mini-projects (15% of your grade): online in LearningSuite covering about 3-5 lectures. Full analyses of datasets. Mostly multiple choice and numeric answers.
- In-class Analyses (5% of your grade): “flipped” classroom days where you analyze data in class (due 10 minutes after the end of your class period).
- Exams (65% of your grade): online in LearningSuite covering several miniprojects
 - 2 Versions, keep the high score
 - 20% midterm 1, 20% midterm 2 and 25% final. Exams are OPEN note but CLOSED other people.
- Extra Credit (0.5%): fill out the course evaluation survey at the end of the semester

Grade Breakdown

- A: 94.0-100; A-: 90.0-93.99; B+: 87.00-89.99; B: 83.00-86.99 (and so on, see LearningSuite)
- Note: I don't "curve" or "adjust" grades at the end of the semester. I think 3 possible HW options for each HW and 2 versions of each exam where I keep the high score is sufficiently merciful.

Using Your Data

Lecture notes in this class will be using your data from a class survey. By staying in this class, you are consenting to me showing your data in class. All data will be **deidentified** and kept **confidential**.

Course Materials

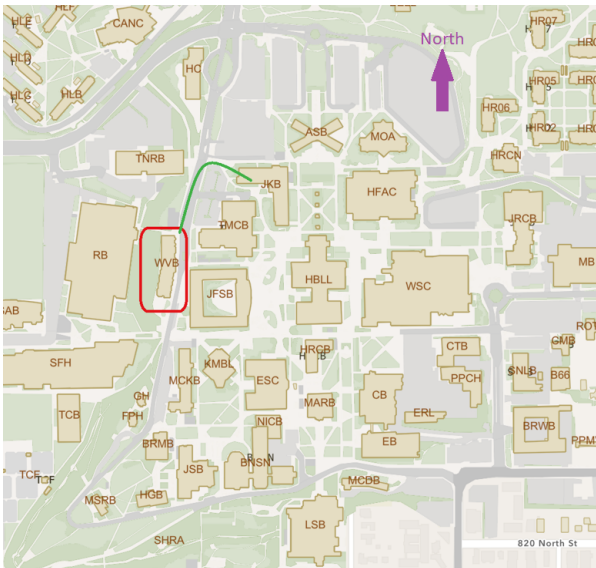
- Lectures notes will be posted on LS via Schedule Tab (wait till after the add/drop deadline to download Units 2+).
- Supplemental reading will be at [Introduction to Modern Statistics](#).
- All HW, projects and exams are through LearningSuite.
- Bookmark the [Analysis Webpage](#)

Lecture Recordings

- Lectures will be recorded and the recordings posted on Learning Suite approximately **1-3 days after** the lecture.
- In-class practice problems and assignments will **not** be available after the class. You have to be present to earn credit for those assignments. **But**, we drop 5 practice problem sets so if you miss a few, it won't hurt your grade.
- You can also find the access codes for In-Class Analyses where the lectures are uploaded on Learning Suite, in case you miss an In-Class Analysis. These codes will be posted after the in-class analysis day as well.

Getting Help

1. Come to the open lab in 1152 WVB between 8am and 5pm M-F (its a bright green room). There will always be a 121 TA there to help.



2. Email me directly to schedule an appointment if you want to talk with me.

FAQs

1. What if I know I am going to be gone?

- We are generally willing to work with you *if* you plan ahead. Email stat121-exceptions@byu.edu before anything is due and we'll work with you.

2. I forgot to do an assignment (or I didn't contact you *before* an assignment was due), do you have a late policy?

- Because they are worth a bigger chunk of your grade, In-class Analyses and Miniprojects can be turned in late at a 25% penalty per day for a max penalty of 75% (so you can turn in assignments anytime for 25% credit).
- Exams and HW are not accepted late. If you have a legitimate reason for missing an exam, we will work with you if you contact us *before* the exam.

3. What if something goes really wrong in life?

- Notify me and stat121-exceptions@byu.edu and we will work with you.

Accommodations

Many of you have accommodations through the disability office. To use those accommodations, simply email stat121-accommodations@byu.edu with your request and we'll check it against the letter we receive from the disability office.

Some Final Notes

1. Do the course survey and associated LearningSuite “quiz.”
2. Do the course policy quiz to make sure you know course policies.
3. Use schedule and path on LearningSuite to follow the class progression.
4. I fully expect you to adhere to the honor code.
5. Information on campus resources (mental health, Title 9, etc.) is available on LearningSuite.

Key Terms to Remember

- Population
- Parameter
- Individual
- Statistical Model
- Statistic
- Sample
- Statistical Inference