

STAT 121: Intro to Data Analysis

About Me

- Where from?
- Hobbies and interests
- How long at BYU?
- Weird fact about myself

What is Statistics?

- Statistics:
 - The **science** focused on the collection, organization, analysis, interpretation, and presentation of 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

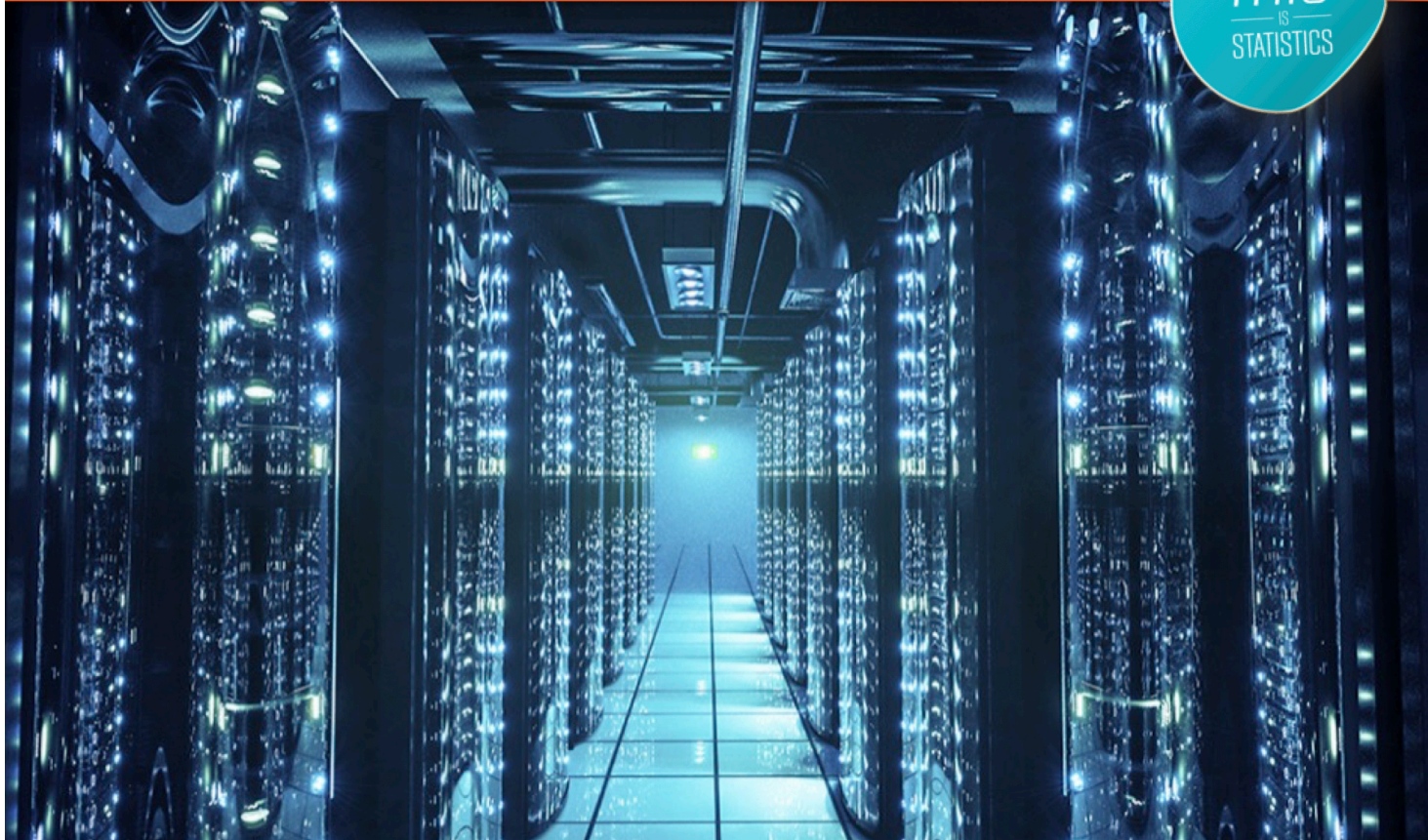
THIS
IS
STATISTICS



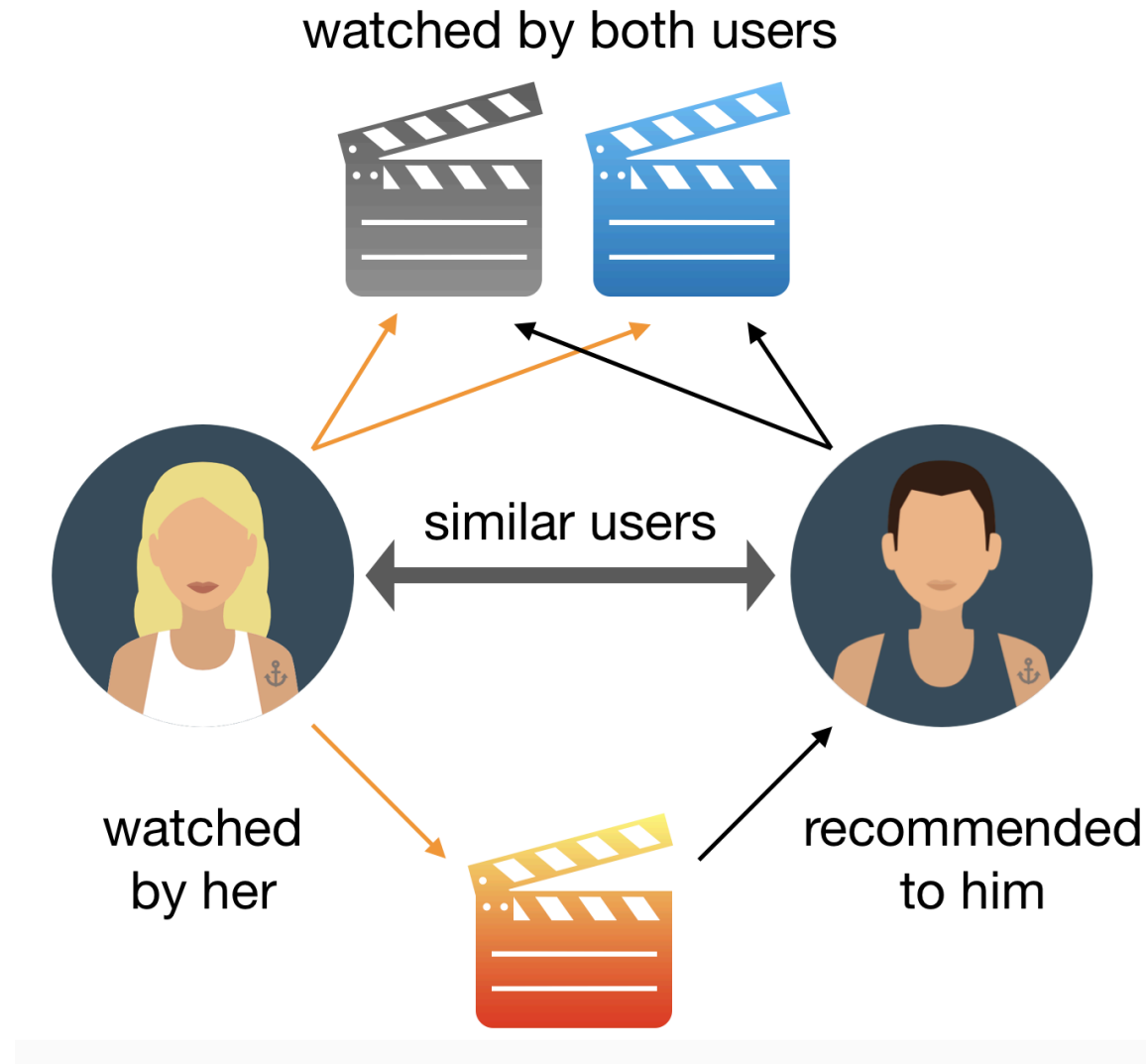
Artificial Intelligence

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

THIS
IS
STATISTICS



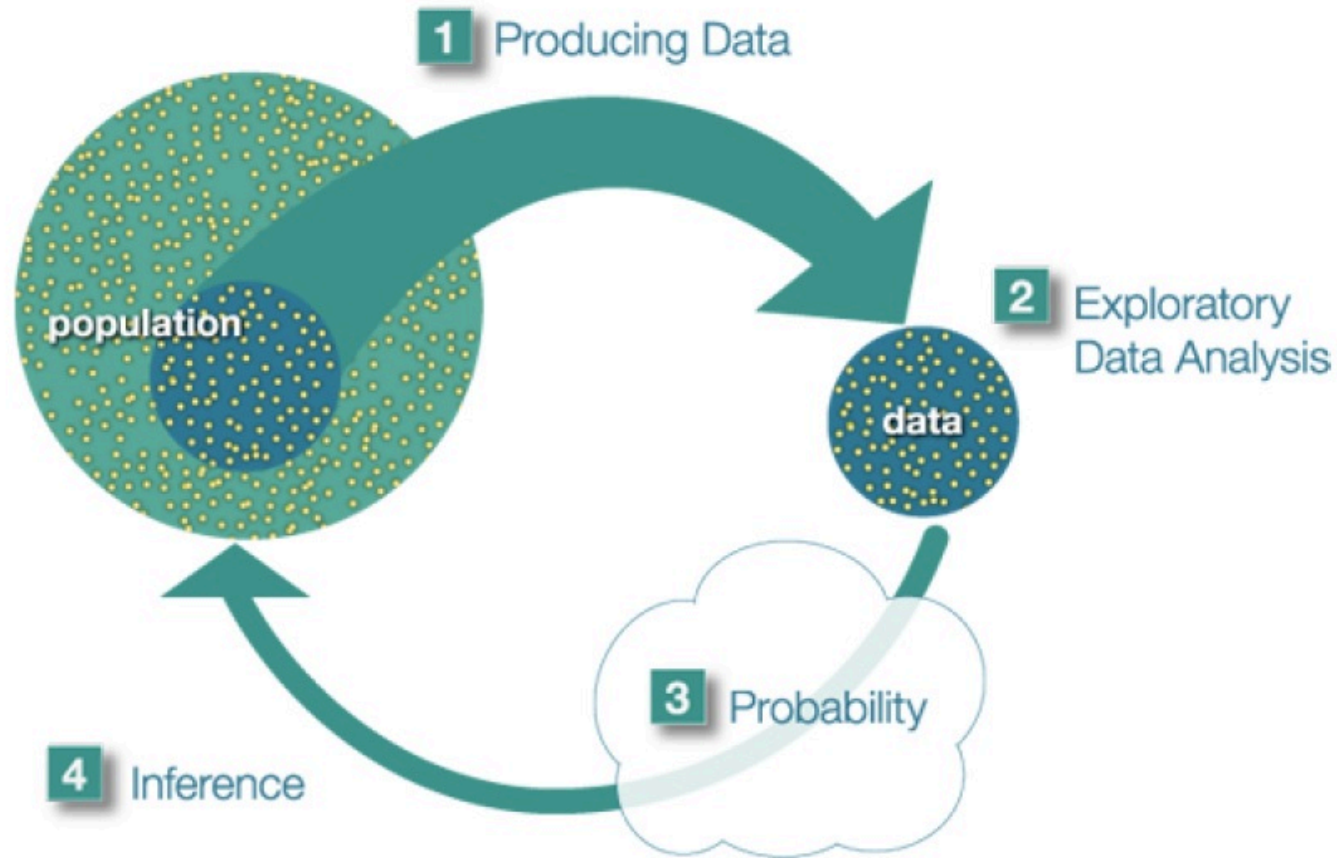
Recommender Systems



Statistics in YOUR Career

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” you are interested in
- **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 you have 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 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

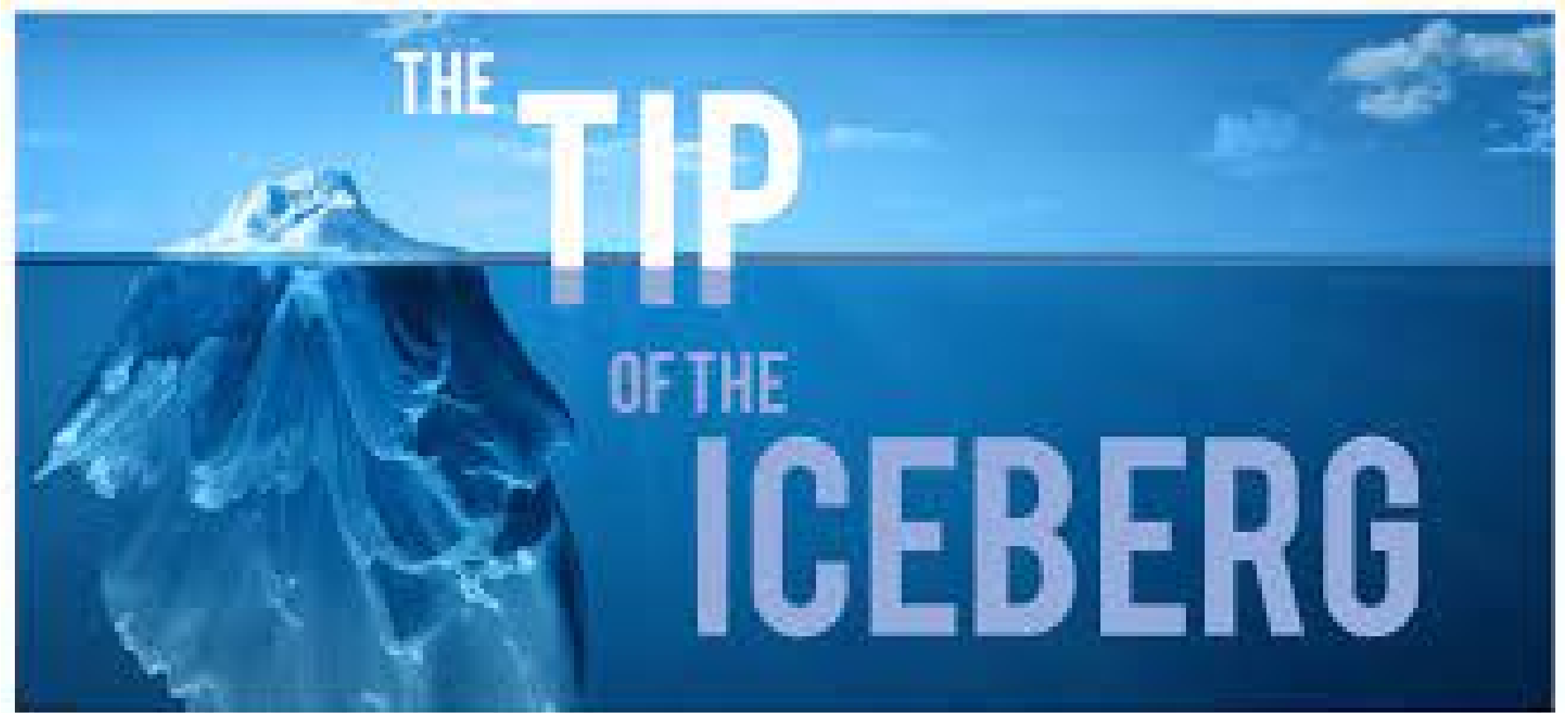
A high school principal wishes to know the percentage of seniors at his high school who are planning on attending college. To figure this out, he randomly samples 50 seniors and asks them if they plan to attend college after high school.

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.

Course Learning Activities

- YPoll (5% of your grade): Near daily class quizzes during lecture (80% participation, 20% correctness). 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.
 - 4 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
- 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 4 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 or Path (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](#)

Getting Help

1. Email stat121-hwhelp@byu.edu and we'll get back to you (between hours of 8am-5pm M-F).
2. Come to the open lab in 1151 WVB between 8am and 5pm. There will always be a 121 TA there to help.
3. Email me directly to schedule an appointment if you want to talk with me directly.

Zoom Policy

Please plan on attending in person but I realize that you will likely miss a day or two for one reason or another during the semester. We do record the lecture which you can request as follows:

1. Email stat121-zoom@byu.edu with your name, NetID, Professor name and date you want the recording.
2. You can request up to 3 zoom recordings during the semester (yes, we are going to keep track of how many you ask for).
3. If you are gone for an in-class analysis day, you must request the recording no later than **THE DAY OF** to be able to do the in-class assignment.

FAQs

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

- Everything is online so either get it done early or do it away from Provo. There is no late HW accepted in this class.

2. 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