Getting Started

Welcome to Stat 101! Information about the instructor:

- Name:
- Email:
- Office:

For sending emails: please put the phrase "STAT 101" and your lab section at the start of every email subject line and include a brief statement of your question or concern. Then please provide more detail in the body of the email. Your instructor will try to get back to you shortly.

Syllabus Highlights

Please read the syllabus, *I am begging you.* For due dates please check the syllabus or Canvas. In the unlikely event that an assignment date is changed, Canvas will have the most updated due date.

- Written homework is collected at the **beginning** of class or can be turned-in early via a folder in Snedecor 2418.
- late written homework, online homework, or online vocabulary quizzes will not be accepted and no extensions will be issued.
- Lab attendance is mandatory and there are no make-ups. If you have a valid conflict with a lab period please inform your instructor and your TA of your absence atleast 24 hours in advance.
- There will be weekly pop-quizzes administered during the lecture period.
- Two in-class midterms for Unit One and Unit Three
- One take-home midterm for Unit Two
- The final exam is cumulative with an emphasis on Unit Four Material

NO LAB THIS WEEK

Course Structure

- 17 Online Vocabulary Quizzes
- 11 Written Homework Assignments
- 13 Labs on provided paper and computers (**bring your notes to lab**)
- About 15 pop-quizzes (maybe more maybe less, tbd) in class on paper
- 2 in-class midterms, closed book, approximately an hour each
- 1 take-home midterm, open book
- 1 Final exam, closed book, approximately two hours

Grading

Assessment Name	Grade Percentage
Exam One	18%
Exam Two	9%
Exam Three	18 %
Final Exam	18 %
Labs	6 %
Online Homework	10 %
Written homework	10 %
Lecture Quizzes and Attendance	5 %
Vocab Quizzes	6 %

Online Homework Guidelines

- Online homework is submitted through Canvas
- Feel free to utilize your class notes, labs, fellow classmates, and TA for these assignments.
- You are only allowed one attempt at online homework assignments and there is no time limit. Be sure to save your progress as you complete your homework.
- Online homework assessments must be submitted by 8 AM on the due date to receive credit for the assessment. Homework assessments will not be reopened for any reason.

Written Homework Guidelines

- Assignments must be turned in at the **beginning** of lecture on the due date or can be turned in early to the folder in my office at Snedecor Hall 2418.
- No late homework will be accepted! If you know that you will not be able to attend class, you can email me a pdf of the assignment **before** the lecture time on the due date.
- All homework assignments are posted to Canvas and must be printed. Homework should be completed on the printed assignment sheets in the space provided.
- All homework assignments must be stapled. Do not use paperclips or binder clips, or tape, only staples. **5 points will be deducted for any homework assignment not stapled.**
- Write your Lab section at the top of the homework assignment by your name.

Chapter One

In the most simplistic terms, our goal for this course is to learn how to describe, quantify, explain, and analyze data! There will be a fair amount of computation and mathematics in this course but at the end of the day this course is about _____ and _____. But in order to facilitate our analysis we need to establish a common vocabulary between us to talk about data.

Components of Data

- **Obvservations or Cases:**
- **Variables:**
- **Data Inputs:**
- **Data Table:**

Example *At the beginning of the semester, STAT 101, students have been asked to complete a general survey to obtain information about demographic and other variables. ata were collected from the Spring 2004 through Spring 2007 semesters from 2,068 students.*

student	sex	age	year	eye color	height
1	female	18	first	hazel	68
2	male	20	third	brown	70
3	female	18	first	green	67
4	male	23	first	hazel	74
5	female	19	second	blue	62
6	male	19	second	green	67

Please highlight or label each component of the data table presented above.

Some takeaways from the example problem:

- Notice how each individual row of the data table is a single observation or case of our collected data
- Similarly, each column represents a different variable under consideration in our example.

The 5 Ws of Data

The main job of a statistician is to connect the analysis we learn from complex mathematical processes and our data to real-world conclusions. So context is incredibly important for talking about data. This context can be summarized as the 5 Ws of data:

Who

Who refers to who the data is collected from. So in our above example, the who is actually _____. One take-away from the last example is that the "who" of our data is almost always represented as the rows of our datatable.

Population vs. Sample

- **Population**:

- **Sample**:

What are some examples of the "who" in our data? Match the context of each study to the "who" might be the study participants (not strictly human), and determine whether the "who" is a sample or a population.

- An educator wants to study how attending statistics class impacts final course grades.
- An Agronomist wants to study whether a new type of fertilizer improves corn yields.
- THe U.S government wants to see how many people actually live in the Unites States.
- A couple of randomly selected corn fields
- The populace of the U.S.
- One section of STAT 101 Students

Before we move on, the total number of cases in a datatable is important information for hypothesis testing, we will discuss this in more depth later in the semester.

What

The what refers to characteristics of the "who". These are typically the columns of the datatable, more commonly referred to as _____.

Quantitative vs. Categorical Variables

- **Quantitative**:
  - Example: Age in years
  - Example:
  - Example:
- **Categorical**:
  - Example: Gender
  - Example:
  - Example:

The way a variable is used is also important for determining whether or note the variable in question is quantitative or categorical. Circle whether each variable is categorial or quantative given its usage.

Variable: Education

- The number of years in school: quantitative categorical
- Highest level of education: quantitative categorical

Variable: Annual Income

- The total amount of Annual Income: quantitative categorical
- Income as lower, middle, or upperclass: quantitative categorical

Where, When, How, and Why

This information is not necessarily found in the datatable, but provides additional context.

Questions to ask yourself about the data:

- Where did this data come from?
- When was this data collected?
- How was this data collected?
- Why Was this data collected?

Example *Data was collected on 54,000 certified diamonds sold in the United States during the year of 2008. Data were collected by taking information from sales records in order to study the relationship between "4Cs" of diamonds and their prices.*

diamond	carat	cut	color	clarity	price
1	0.23	Ideal	E	SI2	326
2	0.29	Premium	I	VS2	334
3	1.05	Very Good	J	SI2	2789
4	0.73	Very Good	H	VS2	2779
5	0.71	Premium	G	VS1	2825
6	0.23	Good	E	VS1	327
7	0.31	Good	J	SI2	335

In General	*In the context of our example*
Who is the data collected on	
What information is collected on the who	
When was the data collected	
Where was the data collected	
Why was the data collected	
How was the data collected	