# IDS 138
## Introduction to Applied Statistics

MWF, 1:50pm, Collins 323
Spring 2019


Jed Rembold, Ph.D

jjrembold@willamette.edu
http://willamette.edu/~jjrembold/classes/wu339
Collins 311
Lunch Hours: M–Th 12-1: I'll be eating but happy to help or chat!
Office Hours: MWTh 3-5, or if the door is open! (so pretty much always!)
Phone: (503) 370-6860

*This syllabus is subject to change or adaptation as the semester progresses. (Particularly the schedule...)*

**Course Description:** The world is becoming an ever more data-driven environment. The intent of this course is to supply students with the needed techniques and confidence to critique, judge, and find meaning in this sea of data. An emphasis will be placed on using modern statistical software to assist, visualize, and make tangible various statistical processes. Topics covered will focus on probability, regression, correlation, and inference. Particular attention will be focused on preparing students to compile solid statistical arguments of their own or to critique statistical arguments from others.
**Prerequisite(s):** No other stats course
**Note:** A minimum grade of C- is required for this course to count toward university credit.
**Credits:** 1.0

**Text:** *Introductory Statistics with Randomization and Simulation* (1st Ed.)
**Author:** David Diez, Christopher Barr, Mine Cetinkaya-Rundel
**ISBN-13:** 978-1500576691
*Note: This book is freely available online as a pdf copy from www.openintro.org! Or if you prefer a text copy it costs under $10.*

### Course Objectives:
Over the semester, students will gain a working knowledge in:

1. Understanding and working with different types of data
2. Drawing conclusions based on linear regression models
3. The basics of probability and the interplay between probability and statistics
4. Using inference to make and confirm or reject hypothesis tests
5. Using R to assist in data analysis, visualization, and simulation

The field of statistics is vastly interdisciplinary as it determines the interplay between raw data of any type and societal understanding or action based response to that data. By the end of the semester students should feel comfortable taking large data sets and making statistically sound arguments or conclusions based on that data set.

**Grade Weighting**

| | |
|---|---|
| Participation | 6% |
| Homework | 15% |
| In-class labs | 15% |
| Test 1 | 12% |
| Test 2 | 12% |
| Final Project | 20% |
| Final Exam | 20% |

**Letter Grade Distribution:**

| | | | |
|---|---|---|---|
| >= 92.00 | A | 72.00 - 77.99 | C |
| 90.00 - 91.99 | A- | 70.00 - 71.99 | C- |
| 88.00 - 89.99 | B+ | 68.00 - 69.99 | D+ |
| 82.00 - 87.99 | B | 62.00 - 67.99 | D |
| 80.00 - 81.99 | B- | 60.00 - 61.99 | D- |
| 78.00 - 79.99 | C+ | <= 59.99 | F |

**Student Learning Objectives (SLO):**

- To choose, describe, and interpret methods of sampling that will allow for robust statistical conclusions about populations. How a sample set is chosen can have far reaching implications on all subsequent statistical analysis.
- To design experimental or observational studies that address a question and minimize the effect of confounding variables. Real world problems often have immense complexity. How can a problem be addressed in a way to simplify or minimize as much of that extraneous complexity as possible?
- To choose appropriate and useful ways to visualize a data set and create those visuals either in R or by hand. Correct and clear visuals can be the difference in your conclusion being understood by others. Knowing what your options are and how to utilize them or crucial to clearly communicating your findings.
- To calculate and interpret summary statistics and confidence intervals. How robust of a conclusion can you really make from a particular data set?
- To understand the theory and be able to correctly choose and utilize the right statistical test for a question or data set. Students should be able to justify why they chose a particular statistical test or approach over another.
- To gain comfort and fluency in a modern statistical software language. R is incredibly popular in the data science world and throughout various disciplines. Gaining competancy using it opens many doors and methods of analysis that would remain closed or incredibly difficult otherwise.
- To accept responsibility for their own learning. Students need to evaluate what they do and don't understand about a topic and ask detailed questions to better articulate where they are struggling.

## Course Assessment:

- **Homework**
  - Written: Homework will be assigned each week and will be due at midnight each Monday. Homework will be a combination of book problems and problems of my own devising and will general have a few practice problems and 2 graded problems per assignment. Homework must be written up nicely and legibly, with different problems starting on new pages. Feel free to use scratch paper if you are worried about the trees. Graded homework problems should be photographed or scanned to a pdf and submitted through www.gradescope.com. The practice problems will either be odd numbered problems from the book so you can check yourself against the back or I'll provide a solution. The idea is that you can try the practice problems if you need before tackling the graded problems. Or if you are feeling confident, just jump straight to the graded problems!
- **In-class Labs**
  - Throughout the semester we will be setting aside some class days to work through in-class lab type assignments. This will focus heavily on using R to analyze and comprehend different data sets and reinforce what we have been talking about in the lecture. You will be expected to used Rmarkdown to output your results from each lab as a pdf and upload it to www.gradescope.com. You should be able to get most, if not all, of the lab work done during the class, but lab due dates will be several days after the class day to allow you some time to finish up if needed.
- **Tests**
  - There will be two midterms spaced throughout the semester and then a comprehensive final at the end. All tests will be in-class and thus limited to 1 hour. I will allow you a single-sided 3x5 index card to hand-write whatever you might find useful on and bring with you to tests. You should plan to bring a non-internet capable calculator (no phones!) with you to the test.
- **Attendance**
  - Attendance to lectures will be graded. Questions will be asked in class and students will respond via polling technology. Simply responding to each question will earn all your attendance points for the day, but answering correctly will earn you some extra credit. Over the course of the semester, that extra credit can really add up, so show up! In addition, I plan to have around a third of each class day devoted to actively solving problems in groups, some of which may also be homework problems. So show up!

## Course Policies:

### Late Work Policy

I understand that sometimes things come up and you are unable to get an assignment in on time, and I strive to be incredibly flexible and accepting of late work. However, there also comes a point when you get too far behind to realistically keep up with the class. In an effort to compromise between the two, my late policy allots you 10 cumulative days of late work throughout the entire semester. So you can turn 10 assignments in one day late, 2 assignments in 5 days late, etc. without penalty. Once you have used up your 10 days, any further late assignments will immediately be worth only 50% of their total possible points.

**Incomplete Policy**

An incomplete grade will only be granted in the case of prolonged illness or family emergencies that remove the student from the campus for an extended time period during the semester. Under no situations will an incomplete be granted due to a student falling behind through lack of motivation, understanding, or time management skills. If you are concerned about your progress and how you are doing in the class, please come visit me! We can sort out where you are struggling and work out a plan to get you back on track.

**Inclusive Classroom**

The practice of statistics in particular, and the scientific method in general, encourage the examination of deeply held assumptions and demand a willingness to set aside old ideas that are not consistent with the data. Despite this, mathematics and science have also often been used as tools to reinforce the power of some while marginalizing others. In this classroom we will strive to use statistics to question assumptions and liberate old ideas, and we will be particularly careful to use our knowledge to support rather than silence our classmates. We may well encounter datasets presenting complicated issues like gender, race, and mental or physical health that make us uncomfortable or cause pain or anger. Feel free to raise those concerns in a supportive manner in class, or come talk with me during office hours.

## Willamette Policies:

**Academic Honesty**

Cheating is defined as any form of intellectual dishonesty or misrepresentation of one's knowledge. Plagiarism, a form of cheating, consists of intentionally or unintentionally representing someone else's work as one's own. Integrity is of prime importance in a college setting, and thus cheating, plagiarism, theft, or assisting another to perform any of the previously listed acts is strictly prohibited. An instructor may imposed penalties for plagiarism or cheating ranging from a grade reduction on an assignment or exam to failing the course. An instructor can also involve the Office of the Dean of the College of Liberal Arts for further action. For further information, visit: http://www.willamette.edu/cla/catalog/resources/policies/plagiarism_cheating.php.

**Time Commitments**

Willamette's Credit Hour Policy holds that for every hour of class time there is an expectation of 2-3 hours work outside of class. Thus, for a class meeting three days a week you should anticipate spending 6-9 hours outside of class engaged in course-related activities. Examples include study time, reading and homework, assignments, research projects, and group work.

**Diversity and Disability**

Willamette University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. Our goal is to create learning environments that are usable, equitable, inclusive and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, please notify the professor as soon as possible. Students with disabilities are also encouraged to contact the Accessible Education Services office in Matthews 103 at 503-370-6737 or accessible-info@willamette.edu to discuss a range of options to removing barriers in the course, including accommodations.

## Tentative Course Outline:

The weekly coverage might change as it depends on the progress of the class. However, I highly recommend you follow along with the reading, as it makes a large difference!

| Week | Date | Reading | Description | Due |
|---|---|---|---|---|
| 1 | Jan 23 | | Motivation and Syllabus | |
| | Jan 25 | | Into the Streets (No Class) | |
| 2 | Jan 28 | Ch 1.1–1.4 | Intro to Data | |
| | Jan 30 | | Introduction to R Lab | |
| | Feb 1 | Ch 1.5–1.7 | | |
| 3 | Feb 4 | | Intro to Data Lab | HW 1 |
| | Feb 6 | A.1 1st half | Basic Probability | |
| | Feb 8 | A.1 2nd half | | |
| 4 | Feb 11 | | Probability Lab | HW 2 |
| | Feb 13 | Ch 5.1 | Intro to Regression | |
| | Feb 15 | Ch 5.2 | Fitting a Line | |
| 5 | Feb 18 | Ch 5.3 | Regression outliers | HW 3 |
| | Feb 20 | | Linear Regression Lab | |
| | Feb 22 | Ch 6.1 | Intro to Multiple Regression | |
| 6 | Feb 25 | Ch 6.1, 6.3 | Checking Multiple Regression Visually | HW 4 |
| | Feb 27 | A.2 | Conditional Probability | |
| | Mar 1 | Ch 2.1, 2.2 | Foundations for Inference | |
| 7 | Mar 4 | | Sampling Lab | HW 5 |
| | Mar 6 | Ch 2.3, 2.4 | Hypothesis Testing and Simulation | |
| | Mar 8 | | **Test 1: (Ch 1,5,6,A.1)** | |
| 8 | Mar 11 | A.3 | Random Variables | HW 6 |
| | Mar 13 | Ch 2.5, 2.6 | CLT and Normal Distributions | |
| | Mar 15 | Ch 2.7, 2.8 | Confidence Intervals | |
| 9 | Mar 18 | | Confidence Interval Lab | HW 7 |
| | Mar 20 | Ch 3.1 | Intro to Categorical Inference | |
| | Mar 22 | Ch 3.2, 3.3, 3.4 | Categorical Inference Testing | |
| 10 | Mar 25 | | | |
| | Mar 27 | | Spring Break! | |
| | Mar 29 | | | |
| 11 | Apr 1 | | Categorical Inference Lab | |
| | Apr 3 | Ch 4.1, 4.2 | Intro to Numerical Inference | |
| | Apr 5 | Ch 4.3 | Differences Between Means | |
| 12 | Apr 8 | Ch 4.4, 4.5? | ANOVA Testing | HW 8 |
| | Apr 10 | | Numerical Inference Lab | |
| | Apr 12 | Ch 5.4 | Inference for Regression | |
| 13 | Apr 15 | Ch 5.4 | Continuing Inference for Regression | HW 9 |
| | Apr 17 | Ch 6.2 | Inference for Multiple Regression | |
| | Apr 19 | | **Test 2: (Ch 2,3,4,A.2,A.3)** | |
| 14 | Apr 22 | Ch 6.4 | Prediction | HW 10 |
| | Apr 24 | | SSRD (No Classes!) | |
| | Apr 26 | Ch 6.4 | More Prediction | |
| 15 | Apr 29 | | Open Day | |
| | May 1 | | Presentations A | |
| | May 3 | | Presentations B | |
| 16 | May 6 | | Review and Final Prep | Projects |
| | May 10 | | **Final: (Comprehensive)** | |