

Advanced Statistical Inference

ITAO70200

Office: 321 Mendoza

Email: seth.berry@nd.edu

Office Hours

Mondays & Wednesdays – 1:00 to 3:00

If you need time other than those office hours, I am in my office everyday from ~8:30 to ~4:30 – just let me know if you are dropping by.

Class Days and Time

Monday and Wednesday – 3:30 to 5:20

Location – L068

Course Objectives

By the end of this class, I want for you to be able use the general linear model, generalized linear models, and interpret results from both. While mastery is not required, you will have a high-level understanding of these methods and you should be able to appropriately apply them to problems.

Attendance

Attendance in this course is not required, in that I will not be taking attendance; attendance is certainly recommended (and even encouraged). While the lecture presentations will be available, they are not verbatim recitations of what was covered and you may not rely on them to make it through everything in the course. Learning statistics takes effort and attending class is but one small part of that effort.

Although attendance is not required, we will have a weekly short comprehension check question. Each question is worth 5 points and is essentially a participation credit.

Readings

There is no official textbook for this course, but there are going to be a few assorted readings and resources for many topics. I will also share resources for topics. We (read: you) are absolutely not committed to the readings; however, they will give you some very helpful background and I encourage you to read them; in other words, they are suggested readings, not required readings.

Homework

The bulk of your grade will be comprised of topical homework assignments – you will always have 1 week to complete the assignments. These are intended to help you learn how to implement and explain the techniques that we cover. For each homework assignment, you will be provided with data and a question to answer. You are the analyst here, so you decide how to tackle the issue and offer an explanation. With the exception of obvious nonsense, there are no wrong answers, so long as you can defend your decisions and they are statistically defensible.

Homework assignments do not need to be fancy... I do not expect you to bring me beautifully-formatted Word documents. This is a course in statistics, not typesetting. Text files of output and your explanations will be more than sufficient. I do, however, expect them to be legible, sensible, and relatively free of spelling/grammatical errors.

Please feel free to work together on homework, but each assignment needs to be your own work. Putting your heads together to formulate an analytic attack plan is perfect (we all stand on the shoulders of giants), but copying and pasting text from each other is unacceptable.

Homework will be scored on an increasing scale. The first homework assignment will be worth 10 points, the second worth 15, and the third worth 25.

Exam

There will be one exam during the course of this semester and it will be a takehome test. You cannot work together on this. You will have one week to complete the test.

Presentations

I promise you that these will be the easiest presentations you will ever give. Select your coursework that you liked best and prepare a 2 minute overview of what you did and what you found (bonus points if you can explain it with one visualization!). These will be given during the designated final day and time. These will be graded on the content of your presentation, not necessarily on the scientific merit of your work! I really want you to be able to speak about your results.

R

There are many statistical programs available and you have likely had some exposure to many of them. In this course, we will be using R exclusively. R is a free and open-source statistical computing language and it is *lingua franca* for modern statistics. We are going to dedicate some time to learning R and working through examples together; our arrangement will be very similar to the science labs that we all remember from our undergrad days. If you have never done any object-oriented programming, it will take a little work – I am only ever an email or visit away.

Grade Breakdown

Comprehension Checks – 30 points (15%)

Homework – 50 points (25%)

Test – 100 points (50%)

Presentation – 20 points (10%)

Total – 200 points

A = 186+ points

A- = 180-185 points

B+ = 174-179 points

B = 166-173 points

B- = 160-165 points

C+ = 154-159 points

Topic 1 – Breaking The Point-And-Click Chains

Mon., 03-19 & Wed. 03-21

R and RStudio

Reproducible Research

Probability, Data Types, Distributions, and Samples

Model Selection

Topic 2 – A Picture Is Worth ~1000 Words

Mon., 03-26 & Wed., 03-28

Homework #1 assigned on Mon., 03-26

Data Exploration

Data Visualization

Visual Inference

Topic 3 – The General Linear Model

Mon., 04-02 through Mon., 04-09

Homework #2 assigned on Mon., 04-02

Linear regression

T-tests

ANOVA and its variations

Topic 4 – Generalized Linear Models

Wed., 04-11 & Mon., 04-16

Families and Link Functions

Logistic regression

Poisson regression

Mon., 04-16 – TEST 1 ASSIGNED

Wed., 04-18 – Review Day

Topic 5 – Mixed Models

Homework #3 assigned on Mon., 04-23

Mon., 04-23 & Wed., 04-25

Fixed and random effects

Topic 6 – Students' Choice

Mon., 04-30 & Wed., 05-02

Regression's Wacky Variants

Robust Models

Quantile Regression

Feature Selection (Ridge -> LASSO -> Elastic Net)

Non-linear Models

Non-linear models

Generalized Additive Models

The Great Statistics Civil War

Bayesian Data Analysis

Presentations

Monday, May 7, 3:30 to 5:30

Mendoza 159