

# Syllabus slides

*DJM*

*January 10, 2016*

## Course info

### Instructor:

Daniel McDonald

Office: Ballantine Hall 669

Website: <http://mypage.iu.edu/~dajmcdon>

Slack: @dajmcdon

Email: [dajmcdon@indiana.edu](mailto:dajmcdon@indiana.edu)

### Office hours:

Me: TBA

Gunchang: TBA

### Course webpage:

WWW: <https://github.com/STATS-432Sp2017>

Slack: <https://stat-s432-sp17.slack.com/>

### Lectures:

TR 9:30 - 10:45, BH 148

### Textbook:

Required: *Advanced data analysis from an elementary point of view*

Optional: *Introduction to Statistical Learning*

### Prerequisite:

STAT-S 431 or equivalent or permission of the instructor.

## Course objective

This is a course in advanced data analysis. Based on the theory of linear models covered in S431, this course will focus on applying many types of data analysis methods to interesting datasets. The focus will be dealing with and describing data rather than on particular methods per se.

The course combines analysis with methodology and computational aspects. It treats both the “art” of understanding unfamiliar data and the “science” of analyzing that data in terms of statistical properties. The focus will be on practical aspects of methodology and intuition to help students develop tools for selecting appropriate methods and approaches to problems in their own lives.

I expect that the material should be appropriate and interesting to students coming from either statistics or informatics/computer science. I will make an effort to provide necessary background so that lectures are accessible to both populations.

We will put special emphasis on learning to use certain tools common to companies which actually *do* data science. This includes Slack, GitHub, Rstudio, and Rmarkdown. As such, we will require the use of these technologies.

## Lectures

Class time will consist of a combination of lecture, discussion, questions and answers, and problem solving, with a focus on problem solving. You are strongly encouraged to bring a laptop to class.

## Textbook

The required textbook for this class is still in progress, but is excellent. Download it somewhat regularly, as it is still being updated.

## Course communication

This class will use Slack for class communication. Slack is a Team messaging app which is frequently adopted by companies and research groups. We will use it for announcements, questions, homework communication, etc. You are required to use it.

Since homework and the project are team-based, you need a way to communicate with your team. Slack is that way. You can create a special channel for your team. You can direct message individuals. You can message me or the TA. You can post pictures of cats. The sky is the limit. By using slack, we will avoid sending lots of email back and forth.

Slack has apps for iPhone or Android. Or you can use the web (as I do).

If you send me email instead, I may ignore it.

## Grading

- 10 In-class exercises (5 points each)
- 6 Homeworks (30 points each)
- 1 Midterm project report (100 points)
- 1 Final project report (200 points)
- 1 Midterm exam (200 points)
- 1 Final exam (300 points)

In all cases, a fixed number of points may be **earned** for each exercise. Points are not deducted for poor performance in this class. Points are awarded for excellent performance.

## In-class exercises

We will regularly have short programming or theoretical exercises during class time. These will always occur on Thursdays. At the end of class, you will push your results to GitHub for grading. Complete submissions will receive full credit. Partial submissions will receive 2 points. There will be at least 10 of these during the semester. Any submissions you make in excess of 10 will be extra credit.

## Homeworks

Homework assignments will be completed in groups which I assign. These groups will rotate for each assignment. Each team will have their own private repository. All assignments will be submitted on Github. Homework grades will be awarded based on complete and accurate analyses according to the rubric provided with each assignment. Grades will be adjusted up or down based on participation according to the Github commit messages. You are expected to consult only with me, the TA, or your assigned group, with the

exception of conversations which occur during class meetings or office hours in the presence of me or the TA. This means that you and your friend (who is in a different group) may not do the homework together.

## Project

You will be required to complete a team project (2-4 people). You may choose this group. The goal is to analyze a dataset in depth using everything that you have learned.

A Midterm Report is due on **Thursday, March 2** at 11:59pm. This report must include your group members, the questions you are interested in, and evidence that you have successfully loaded the data onto your computer in processable form.

The entire class will give brief presentations during the last day of class. You will also be graded on your completed project. You will also submit a final report due on **Thursday, April 20** at 11:59pm.

## Exams

There will be two take home exams. Both will be completed entirely individually. The midterm will be made available on **Friday, March 3** for submission by **Friday, March 10**.

The final exam will be available **Friday, April 21** for submission by **Friday, April 28**. Note that this exam occurs during “dead week”. There will be no course meeting during the scheduled final exam time.

## Academic Integrity

As a student at IU, you are expected to adhere to the standards and policies detailed in the Code of Student Rights, Responsibilities, and Conduct. When you submit an assignment with your name on it, you are signifying that the work contained therein is yours, unless otherwise cited or referenced. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged. If you are unsure about the expectations for completing an assignment or taking a test or exam, be sure to seek clarification beforehand. All suspected violations of the Code will be handled according to University policies. Sanctions for academic misconduct may include a failing grade on the assignment, reduction in your final course grade, a failing grade in the course, among other possibilities, and must include a report to the Dean of Students, who may impose additional disciplinary sanctions.

## Note Selling

Several commercial services have approached students regarding selling class notes/study guides to their classmates. Please be advised that selling a faculty member’s notes/study guides individually or on behalf of one of these services using IU email, Canvas, or Oncourse violates both IU information technology and IU intellectual property policy. Selling the faculty member’s notes/study guides to fellow students in this course is not permitted. Violations of this policy will be considered violations of the Code of Student Rights, Responsibilities, and Conduct and will be reported to the Dean of Students as a violation of course rules (academic misconduct). Sanctions for academic misconduct may include a failing grade on the assignment for which the notes/study guides are being sold, a reduction in your final course grade, a failing grade in the course, among other possibilities.

## Solutions

Some of the problems that are assigned are similar or identical to those assigned in previous years by me or other instructors for this or other courses. Using proofs or code from anywhere other than the textbooks

(with attribution), this year's course notes (with attribution), or the course website is not only considered cheating (as described above), it is easily detectable cheating. Such behavior is strictly forbidden.