Lecture Times

Lectures: 1 hour pre-recorded lectures twice per week

Instructors and Office Hours

Professor: Donald P. Taylor, PhD, MS, MBA
don@cmu.edu

Teaching Assistant: Mr. Rahul Ladhania
rladhania@cmu.edu

Out-of-classroom Assistance: Email Professor Taylor or Mr. Ladhania at least 48 hours in advance to schedule a Skype session. Please also use the Blackboard forums created according to general questions, each week’s topic and homework, and the final project.

Course Description

This distance-learning course covers the fundamentals of statistical exploration and visualization of data. We will fit models and produce specialized graphs to explore data in a detailed and statistics-oriented manner. This course also serves as an introduction to R, a widely used statistical programming language.

In this class, students will learn:

1. How to use R to perform basic data manipulation such as filtering, aggregating, and organizing data sets
2. How to produce graphics in R
3. How transformations, model fits, and residuals can be used to explore and check assumptions about data

Course policies

Prerequisites
A first course in statistics is required, such as either 95-796 or 90-711.

Computer policy
You will need a computer capable of installing R and RStudio

Textbooks
There are no required textbooks. Helpful references:
If you are interested in a deeper discussion of the methods presented in class, you may want to reference:


### Coursework and Grading

Your grade in this course will be based on 5 homework assignments and a mini-project.

- **Homework (65%)**
  - There are 5 homeworks, equally weighted towards your grade.
  - Homework should be submitted online via blackboard, by 11:59 PM EST on the scheduled due date.
  - Each student has 5 late days. You may use them at your discretion, to cover travel for interviews, illness, or general business. Otherwise, late homework will not be accepted. Note – late days cannot be applied to the Mini-Project.

- **Mini-Project (35%)**
  - We will give you a data set and one or more questions to explore.
  - The project will cover one or more aspects of the course (so not everything you have learned will necessarily be relevant).
  - You will be responsible for structuring the analysis yourself and deciding what tools should be used.
  - You will also need to present your analysis and results in a clear and concise manner in the form of a final report.
  - No late days may be applied to the Mini-Project.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Release Date</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hw 1</td>
<td>Wed May 24</td>
<td>Wed May 31</td>
</tr>
<tr>
<td>Hw 2</td>
<td>Wed May 31</td>
<td>Wed Jun 7</td>
</tr>
<tr>
<td>Hw 3</td>
<td>Wed Jun 7</td>
<td>Wed Jun 14</td>
</tr>
<tr>
<td>Hw 4</td>
<td>Wed Jun 14</td>
<td>Wed Jun 21</td>
</tr>
<tr>
<td>Hw 5</td>
<td>Wed Jun 21</td>
<td>Wed Jun 28</td>
</tr>
<tr>
<td>Mini-Project</td>
<td>Wed Jun 16</td>
<td>Fri Jun 30</td>
</tr>
</tbody>
</table>
Homework rubric (12 points total)

- 1 pts: if the code compiles
  - 1/1 if the grader runs “knit HTML” and an HTML document is returned instead of errors
  - 0/1 if we ask you to resubmit because the homework did not compile the first time
  - if you do not resubmit in 36 hours: homework is LATE.

Assuming the HW eventually is compile-able:

- 10 pts: correct solution (2 point per problem)
  - 2/2 for each problem that is 100% correct or has inconsequential errors
  - 1/2 for each problem with significant errors
  - 0/2 for each problem that is not attempted

- 1 pt: readability
  - 1/1 if the homework is well written:
    - variables are well named
    - avoids needless replication
    - good commenting within R code
    - reasonable coding style
    - non-code writing is clear and to the point
  - 0.5/1 if exposition is pretty good but doesn’t deserve full credit
  - 0/1 if there are significant coding or writing issues, and we don’t understand what you are doing

Grading

Students’ numeric grades for deliverables, and their final numeric grades, will not be curved. However, the instructor will apply a curve when deciding on the final assignment of letter grades based on the numeric grades, in order to meet Heinz College standards for mean student GPA in core courses. Therefore, I do not pre-specify the grading scale. Below are how the cutoffs came out for a previous semester, but this year’s cutoffs will depend on the overall distribution of scores.

A+ 96-100%  B+ 80-83%  C+ 68-71%
A 88-95%  B 76-79%  C 64-67%
A- 84-87%  B- 72-75%  C- 56-63%

Professor & Teaching Assistant Communication

Questions should be sent at least 48 hours before any deadline to assure a timely response. Please include the course code 95-868 in the subject line of your emails. There is also a Blackboard forum created for Q/A that will be moderated to ensure only clarifications to the assignments are discussed and not answers to questions. We encourage you to use the Blackboard forum as the primary medium to ask questions.
Please be conscientious about not posting any answers to assignments on the forum (or anywhere) in order to maintain academic integrity.

**Collaboration**

You are encouraged to discuss general approaches and clarification questions with your fellow students. However, you should do your homework yourself.

- Do not look at (or copy) another student's homework.
- Do not copy from another student's homework.

If you receive any help from another student or outside the class (such as stackexchange or other forums or websites), you must clearly identify where you received help. The expectation is that your grade must reflect the work that you alone did.

**Tentative Schedule**

Week 1 (May 22): Introduction
  - Intro to R, Rstudio, and RMarkdown
  - Data cleaning and aggregation

Week 2 (May 29): Graphics

Week 3 (Jun 5): Analyzing means and sample sizes

Week 4 (Jun 12): Univariate distributions
  - Quantiles, residuals, transformations
  - Simulation and permutation tests

Week 5 (Jun 19): Model fitting
  - 1-D functions, linear regression, splines
  - Identically distributed residuals

Week 6 (Jun 26): Higher dimensional data
  - Stepwise variable selection, interaction terms
  - Generalized additive models, coplots

**Student Health & Safety**

**Take care of yourself.** Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support.
Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

CaPS: 412-268-2922  
Re:resolve Crisis Network: 888-796-8226  
If the situation is life threatening, call the police:  
   On campus: CMU Police: 412-268-2323  
   Off campus: 911