Course Title

Regression Models

Course Instructor(s)

The primary instructor of this class is Brian Caffo. Brian is a professor at Johns Hopkins Biostatistics and co-directs the SMART working group.

This class is co-taught by Roger Peng and Jeff Leek. In addition, Sean Kross and Nick Carchedi have been helping greatly.

Course Description

Linear models, as their name implies, relates an outcome to a set of predictors of interest using linear assumptions. Regression models, a subset of linear models, are the most important statistical analysis tool in a data scientist's toolkit. This course covers regression analysis, least squares and inference using regression models. Special cases of the regression model, ANOVA and ANCOVA will be covered as well. Analysis of residuals and variability will be investigated. The course will cover modern thinking on model selection and novel uses of regression models including scatterplot smoothing.

Course Content

This class has three main components

1. Least squares and linear regression
2. Multivariable regression
3. Generalized linear models

The full list of topics are as follows:

Module 1, least squares and linear regression

- 01_01 Introduction
- 01_02 Notation
- 01_03 Ordinary least squares
- 01_04 Regression to the mean
- 01_05 Linear regression
- 01_06 Residuals
- 01_07 Regression inference

Module 2, Multivariable regression

- 02_01 Multivariate regression

- 02_02 Multivariate examples
- 02_03 Adjustment
- 02_04 Residual variation and diagnostics
- 02_05 Multiple variables

Module 3, Generalized linear models

- 03_01 GLMs
- 03_02 Binary outcomes
- 03_03 Count outcomes
- 03_04 Olio

Module 4, Logistic Regression and Poisson Regression

- Logistic Regression
- Poisson Regression
- Hodgepodge

Github repository

The most up to date information on the course lecture notes will always be in the Github repository. Please issue pull requests so that we may improve the materials. Note my GitHub repo will generally be more up to date than the Data Science Specialization Repo.

Youtube videos

If you'd prefer to watch the videos on youtube, they can be found here:

https://www.youtube.com/playlist?list=PLpl-gQkQivXhdgUCdaUQcdb31CRe8Mm2y

Book: Regression Models for Data Science in R

A companion book is available here: https://leanpub.com/regmods

The book is published via leanpub, and the suggested price is $14.99. You can get it for free or pay what you feel it is worth.

Quizzes

- There are four weekly quizzes.
- You must earn a grade of at least 80% to pass a quiz
- You may attempt each quiz up to 3 times in 8 hours.
- The score from your most successful attempt will count toward your final grade.

Course Project

The Course Project is an opportunity to demonstrate the skills you have learned during the course. It is graded through peer assessment. You must earn a grade of at least 80% to pass the peer assessment.

Grading Policy

You must score at least 80% on all assignments (Quizzes & Project) to pass the course.

Your final grade will be calculated as follows:

- Quiz 1 = 15%
- Quiz 2 = 15%
- Quiz 3 = 15%
- Quiz 4 = 15%
- Course Project = 40%

swirl Programming Assignment (optional)

In this course, you have the option to use the swirl R package to practice some of the concepts we cover in lectures.

While these lessons will give you valuable practice and you are encouraged to complete as many as possible, please note that they are **completely optional** and you can get full marks in the class without completing them.

Differences of opinion

Keep in mind that currently data analysis is as much art as it is science - so we may have a difference of opinion - and that is ok! Please refrain from angry, sarcastic, or abusive comments on the message boards. Our goal is to create a supportive community that helps the learning of all students, from the most advanced to those who are just seeing this material for the first time.

Plagiarism

Johns Hopkins University defines plagiarism as "...taking for one's own use the words, ideas, concepts or data of another without proper attribution. Plagiarism includes both direct use or paraphrasing of the words, thoughts, or concepts of another without proper attribution." We take plagiarism very seriously, as does Johns Hopkins University.

We recognize that many students may not have a clear understanding of what plagiarism is or why it is wrong. Please see the following guide for more information on plagiarism:

http://www.jhsph.edu/academics/degree-programs/master-of-public-health/current-students/JHSPH-ReferencingHandbook.pdf

It is critically important that you give people/sources credit when you use their words or ideas. If you do not give proper credit -- particularly when quoting directly from a source -- you violate the trust of your fellow students.

The Coursera Honor code includes an explicit statement about plagiarism:

I will register for only one account. My answers to homework, quizzes and exams will be my own work (except for assignments that explicitly permit collaboration). I will not make solutions to homework, quizzes or exams available to anyone else. This includes both solutions written by me, as well as any official solutions provided by the course staff. I will not engage in any other activities that will dishonestly improve my results or dishonestly improve/hurt the results of others.

Reporting plagiarism on course projects

One of the criteria in the project rubric focuses on plagiarism. Keep in mind that some components of the projects will be very similar across terms and so answers that appear similar may be honest coincidences. However, we would appreciate if you do a basic check for obvious plagiarism and report it during your peer assessment phase.

It is currently very difficult to prove or disprove a charge of plagiarism in the MOOC peer assessment setting. We are not in a position to evaluate whether or not a submission actually constitutes plagiarism, and we will not be able to entertain appeals or to alter any grades that have been assigned through the peer evaluation system.

But if you take the time to report suspected plagiarism, this will help us to understand the extent of the problem and work with Coursera to address critical issues with the current system.