Welcome

I am happy that you've chosen to take Regression Models, part of the Johns Hopkins Data Science Specialization on Coursera! This course presents the fundamentals of regression modeling that you will need for the rest of the specialization and ultimately for your work in the field of data science.

We believe that the key word in Data Science is "science". Our course track is focused on providing you with three things: (1) an introduction to the key ideas behind working with data in a scientific way that will produce new and reproducible insight, (2) an introduction to the tools that will allow you to execute on a data analytic strategy, from raw data in a database to a completed report with interactive graphics, and (3) on giving you plenty of hands on practice so you can learn the techniques for yourself.

Regression Models represents a both fundamental and foundational component of the series, and it presents the single most practical data analysis toolset. Using only a bare minimum of mathematics, we will attempt to provide you with the fundamentals for the application and practice of regression.

We are excited about the opportunity to attempt to scale Data Science education. We intend for the courses to be self-contained, fast-paced, and interactive, and we intend to run them frequently to give people with busy schedules the opportunity to work on material at their own pace.

Some Basics

A couple of first week housekeeping items. First, make sure that you've had R Programming, the Data Scientist's Toolbox, Reproducible Research and Statistical Inference before taking this class. At a minimum you must know: very basic git, basic R and most of the Statistical Inference Coursera class. The small amount of knitr that you need for the project you can pick up quickly.

An important aspect of this class is to peruse the materials in the github repository. All of the most up to date material can be found here: https://github.com/bcaffo/courses/tree/master/07_RegressionModels

You should clone this repository as your first step in this class and make sure to fetch updates periodically. (Please send pull requests too!) It is one of the most essential components of the Specialization that you start to use Git frequently. We're practicing what we preach as well by using the tools in the series to create the series, especially git.

You can clone the whole repo with (http)

git clone https://github.com/bcaffo/courses.git

or (ssh)

git clone git@github.com:bcaffo/courses.git

The lectures are in the index.Rmd lecture files. In Developing Data Products, we'll cover how to create these sorts of slides. However, for the time being, you should be able to open them in R Studio and look at their contents. You will see all of the R code to recreate the lectures. Going through the R code is the best way to familiarize yourself with the lecture materials.

If you'd prefer to watch the videos on YouTube, you can find them here: https://www.youtube.com/playlist?list=PLpl-gQkQivXjqHAJd2t-J_One_fYE55tC

If you'd like to keep up with the instructors I'm @bcaffo on twitter, Roger is @rdpeng and Jeff is @jtleek. The Department of Biostat here is @jhubiostat.