Quantitative Analysis Institute Pilot Course

Applied Data Analysis with R

Winter 2015

Instructor: Cassandra Pattanayak, Guthman Director of the QAI
cpattanayak@wellesley.edu

Dates: January 5-23, 2015

Description:
The Quantitative Analysis Institute is developing an online version of its applied data analysis course. This wintersession course is an informal pilot of the online materials prepared so far; the topics covered are a subset of QR 260 and the QAI Summer Course. There are no official grades, credit, or fees. Students will both learn the course content and provide feedback to help refine the online materials.

This is an intermediate statistics course focused on fundamentals of statistical inference and applied data analysis tools. Emphasis on thinking statistically, evaluating assumptions, and developing practical skills for real-life applications. Topics include t-tests and non-parametric alternatives, multiple comparisons, analysis of variance/linear regression, and missing data. Students can expect to gain a working knowledge of the statistical software R, which will be used for data analysis and for simulations designed to strengthen conceptual understanding.

Prerequisites: Any course that meets the QR Overlay requirement. Students who participated in the 2014 QAI Summer Course are not eligible.

Goals: This pilot course will enhance your ability to:
- Evaluate the strengths, weaknesses, and appropriateness of a variety of statistical techniques
- Given a data set: state hypotheses, explore the data using statistical software, identify and apply appropriate analysis methods, and assess assumptions
- Communicate statistical results graphically
- Handle common practical challenges of data analysis
- Use the statistical software R

Computing: This course involves learning the statistical software R. R is popular among statisticians and other researchers because it is free, downloadable, open source, field-neutral, and powerful. No matter where you are after college, R will always be available. R is downloadable from www.r-project.org (or, google the letter “R”). Instructions for downloading R and RStudio will be posted.
No previous experience with R is necessary. However, you should expect that it will take time to familiarize yourself with R. Computing questions (like conceptual questions) are usually easier to answer when we can discuss out loud rather than in writing, so plan ahead to attend virtual office hours when possible.

**Accommodations:** If you have documentation from Disability Services that will affect your work in this course, please let me know at the beginning of the session. I would also be interested in your feedback on how well the online materials accommodate your needs.

**Platforms:** We will primarily rely on a combination of the course website at edge.edx.org and google apps such as forms, groups, and hangouts.

- **Course Website:** Once you create your edge.edx.org account, you will automatically be enrolled in the course. The “Welcome” section provides guidance on how to interact with the website. The course is divided into six parts. Lessons will be posted under the “courseware.”

- **Google:** You will be invited to join the course google group, which I will use to email any announcements. Assignments will be submitted via a google drive folder that you share with me. We will use google forms for quizzes and feedback. We will experiment with google hangout office hours.

- **Skype:** If you don’t already have a Skype account, please create one. We will experiment with skype office hours.

**Expectations for Students:** Students should expect to spend 10-15 hours on the course for each of the three weeks. This includes time spent working through the online lessons, completing assignments, taking quizzes on course material, communicating with classmates, participating in office hours, and providing feedback.

- **Online Lessons:** The courseware tab of the course website is organized into six sections of videos, text, practice problems, discussion questions, and assignments. You should work through the lessons in order. The practice problems on the course website are for your benefit only – you should ignore references to grades that are built into the edge.edx website. I have not found a way to turn off the references to grades and points that are intended for massive open online courses. Similarly, ignore the “Progress” page that summarizes your scores on the practice problems.

- **Assignments:** An assignment will be posted at the end of each of the six sections. The assignment will include both conceptual questions and data analysis or simulations to be completed in R. You will submit your assignments by placing a google doc write-up and your R code in a google drive folder that you’ve shared with me.

- **Quizzes:** Each of the six sections of the course ends with a quiz, which you’ll complete via a google form survey. I expect that you will complete each quiz without looking at any course materials or other sources, but the quizzes are not timed.
Communicating: The course website includes both discussion boards and a wiki. Toward the beginning of the course, I will prompt you to practice using the boards. We also have the google group. As this is a pilot, we will work together to figure out which tool is best: please use the boards, the wiki, and/or the google group to ask questions or raise issues that could be of interest to others and to answer each other’s questions. Kily Wong, a Wellesley psychology major who completed my summer course, will spend time on the discussion boards, as will I. We will answer your questions as quickly as we can; we will not always be available in the hours just before a due date. You can also email me individually (though, I may suggest that we post the question and answer for all).

Virtual Office Hours: I will hold an office hour four days each week at varying times, in an effort to catch a variety of time zones and schedules. Please see the “Virtual Office Hour Schedule” tab of the course website. The first week, we will use google hangout. The second week, we will use skype. The third week, we will use whichever tool seemed to work better for our needs. My experience is that office hours are the best way to learn and teach this material. I’ve never held virtual office hours before, but I expect that they will be productive – I strongly encourage you to plan your work such that you have questions to bring to office hours.

Feedback: I will frequently ask you for feedback on the course materials via anonymous google form surveys. I’m interested in how much time you spend on each component, which modes of instruction work best for you, which tools you prefer, etc. As you work through the course, please keep in mind that this is a pilot and make note of anything that you’d like to remember to mention on the next feedback survey. You can also email me individually with any thoughts. At the end of the course, you will complete a form that includes the usual SEQ questions.

Expectations for the Instructor:

Feedback: I will provide feedback on assignments that you submit via your google drive, and I will post solutions. I will also grade and post solutions on quizzes that you take through google forms. Though this course is not for credit, I will let you know roughly what your grade would have been if the course were for credit, at the end of wintersession.

Again, the problems on the course website are for your benefit; if this course were graded, they wouldn’t be part of your grade. I may look at your answers sometimes as I explore the capabilities of the website.

Communicating: I will participate in the discussion boards, wiki, and google group, and I will answer your individual emails. Though I cannot be available at all times, I will answer your questions as soon as I can. I will also create new course content as needed to address common questions or misconceptions.

Virtual Office Hours: As described above, I will hold video office hours four days a week.
Pace and Deadlines:
Each section of the course will be released with a due date. There are six sections and three weeks, so the due dates will be approximately half a week apart. These due dates assume that you’ll work approximately 2-3 hours each weekday, though you are not required to organize your time that way.

As you work through the materials in each section, keep in mind that the assignment posted at the end is meant to take at least half of the time you spend on the section, much as a problem set in a regular course takes at least as long as learning the material in class.

So, for example, if there are three weekdays between due dates, I expect that section to take 6-9 hours, including both the course materials and the assignment.

The EdX website refers to a universal time zone that is five hours ahead of Boston. I’m on Boston time, though: a Wednesday due date means that I will begin to grade when it’s Thursday in Boston.

What is the benefit of completing the course at the suggested pace?
I will grade assignments and quizzes shortly after the posted due dates. Then, I will post solutions and materials addressing frequently asked questions or common errors, in time to help with the next assignment.

What are the consequences of completing the course more slowly than suggested?
Assignments and quizzes submitted well after the posted due date may be graded late or not at all. So, you’ll be able see the solutions, but you may not receive timely feedback. Also, issues you encounter may not be addressed in the materials I post to help with the next assignment. That said, I don’t mind if you choose to work more slowly than suggested.

What if I would like to move ahead?
You are welcome to work faster than suggested. Not all course materials will be released immediately because I will revise later sections to incorporate feedback on earlier sections before posting. But, there is nothing stopping you from working on anything that is posted. I will still grade your assignments after the posted due dates, so you may not receive feedback in time if you’ve already moved on to subsequent assignments, and you won’t have a chance to benefit from my answers to other students’ questions.

Honor Code and Collaboration Policy: The Honor Code will be enforced. Sources must be cited, as appropriate. Unless specified otherwise, each student must write up assignments separately. However, you are strongly encouraged to discuss exercises with your classmates. Be sure that the answers you submit are the results of your own efforts. You may not copy another student’s R code, submit output from another student’s R session, or allow another student to share your code or output. A good rule of thumb: if a classmate asks if you would like to discuss a problem on an exercise, you are encouraged to say “yes”; if she asks to see your solutions or your code file, the answer is “no.”