EPBI 414 Syllabus - Fall 2016

Data Management and Statistical Programming

Instructor

Thomas A. Rehman, MA, MPH
Adjunct Instructor, Department of Epidemiology & Biostatistics
School of Medicine, Case Western Reserve University
tar9@case.edu

Course Website

http://sites.case.edu/a/case.edu/epbi414/

Course Description

This is an **online** course that offers no in-person meetings.

This course serves as a general introduction to the use of computer systems in epidemiologic investigations and biostatistical applications. Students will develop a conceptual understanding of data types, basic data structures, relational database systems and data normalization, data warehousing, control statements, and programming logic. Further, students will develop basic scripting skills and will learn to read in, manipulate, and perform basic descriptive analyses on research data using the SAS programming language. Primary emphasis in this course is on developing the knowledge and familiarity required to work with data in a statistical programming context. Basic familiarity with statistics is beneficial, as this course does not teach inferential statistical analysis in detail, but it is not vital to learning the course material.

Communications

Office Hours

This is an **online** course. The instructor will be available for meetings via Google Hangouts. In addition, a weekly Google Hangouts review session will be held for students to ask questions.

Course Mailing List

Students in this class will be enrolled in a mailing list at the start of the course. This list is intended to simulate asking questions in class, where everyone can hear the question, the answer, and any follow-up questions or comments. Please contribute to the list; good programming is a deeply creative endeavor, and discussions can provide everyone involved with new insights.

Submitting Assignments

Course submissions and online quizzes will be handled through the Blackboard platform. You must use Blackboard to submit your assignments; emailed submissions will not be accepted!

Computing Requirements

This course features assignments using multiple different technologies. Students will work with MySQL and SAS, and will be required to connect to a departmental Unix server to complete some assignments.

Most of the software needed for your assignments can be found on the computers in the departmental computer lab. If you desire, you can install **MySQL Workbench** on your personal computer; however, it has also been installed on the lab computers for your use. It is available for all major operating systems, and is available at the link below.

Students using Microsoft Windows can also download **SAS 9.4** from the Case Software Center for a nominal fee. Unfortunately, **SAS 9.4** is not available (in client version) for other operating systems; students using non-Windows personal computers will need to use the computer lab for SAS assignments.

Finally, students using Microsoft Windows will want to obtain an **SSH client** to use in connecting to departmental Unix servers. Some suggested options are given below, and will be discussed in the course materials.

- **MySQL Workbench:** Available for free at https://www.mysql.com/products/workbench/
- **SAS:** Available for a nominal fee through the Case Software Center at https://softwarecenter.case.edu/
- **Windows SSH Clients:**
    - **SSH Secure Shell** - Commercial software available through the Case Software Center at https://softwarecenter.case.edu/
    - **PuTTY** - Popular free and open-source terminal emulator available at http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html
    - **Cmder** - A drop-in replacement for the Windows command line that contains SSH, available at http://cmder.net/.
    - **Cygwin** - A more complex and robust Unix-like environment in Windows, available at https://www.cygwin.com/

Other Course Materials

By its nature, a course on statistical programming will make use of both traditional course materials as well as Internet and other resources. Useful course materials will be posted on the class website for all students to utilize. Students interested in some general-purpose books on this material might find the following of use, but none are required for this course.

SAS

Applied Statistics and the SAS Programming Language (5th Edition), Ron P. Cody (2005)

The Little SAS Book: A Primer, Fifth Edition, Lora Delwiche (2012)

Cody's Data Cleaning Techniques Using SAS, Second Edition (SAS Press), Ron P. Cody (2008)

General Data and Visualization Books

The Visual Display of Quantitative Information, Edward R. Tufte (2001)

R (for those students with an interest)

Using R for Data Management, Statistical Analysis, and Graphics (1st Edition), Nicholas J. Horton et al (2010)

Course Schedule

This is an **online** course that offers no in-person meetings. The course schedule is available on the course website and details when new lectures, homeworks, and quizzes will be posted, as well as when proctored exams must be completed. As we go through the semester, some modifications may be necessary - these will be posted to the course website and the email list.

Partial Course Options

Some students are interested in registering for only a portion of the material covered in this course. You have three options in taking EPBI 414, as described below.

- **Option 1 - Full Course:** This option is the "traditional" manner of taking EPBI 414. You complete the full course over the semester.
- **Option 2 - Data Principles Core:** Students interested in learning about data concepts, data management principles, and basic database and SQL information, may complete Weeks 1 - 7, with the final homework and examination submitted in Week 8.
- **Option 3 - SAS Core:** Those who are interested in only obtaining training in SAS may complete Weeks 8 - 14, with the final homework and examination submitted the following week.

Grading and Assessments

The full course features: 6 online, timed quizzes; 2 proctored examinations, one at the midterm and one at the end of the course, and; 14 graded homework assignments. Students enrolled in one of the two partial course options will complete only part of the full course requirements.

Assignments, quizzes, and exams are graded based on a point system, where the total number of points earned are divided by the total number possible to produce a percent score.

Your course grade is calculated from your average percent score on homeworks and quizzes, as well as your percent score on whatever examinations are required for your course option (full or partial). The weighted mean of these values is assigned as your percent score for the course, using the weighting described on the next page. A letter grade is assigned based on this percent score, as shown below.

In the past, students have found that the homework can be somewhat challenging. As this is an online course, and as you are being trained to be good programmers, you will find that you must apply the principles from the lecture to more complex homework questions. You are advised to budget your time accordingly and ask questions if you are confused!

Percent Score	90% - 100%	80% - 89%	70 - 79%	60% - 69%	0% - 59%
Letter Grade	A	B	C	D	F

Assignment	Option 1 % of grade	Option 2 % of grade	Option 3 % of grade
Average Homework Score	20%	20%	20%
Average Quiz Score	20%	30%	30%
Data Principles Exam Score	30%	50%	
SAS Programming Exam Score			50%
Comprehensive Exam Score	30%		

Submissions

Homework assignments will be submitted using Blackboard by the date and time specified in the Course Schedule. Please note that due dates and times are given in **Eastern** time.

Online quizzes are administered through Blackboard, and are covered in the orientation to the course. Be sure to leave yourself sufficient time to take your quiz before the due date!

Proctored exams may be supervised by any faculty or staff member in the Department of Epidemiology and Biostatistics. It is your responsibility to locate a proctor for your exam and to schedule your examinations by the due date. You must notify the Course Instructor of your selected proctor at least 24 hours before your scheduled exam date, so that the Course Instructor can provide your proctor with a copy of the exam.

Do not wait until the last minute to schedule your examinations.

Late Work

Late work will be accepted up to a week after the due date. The final percent score of all late work is reduced by 50%, after being graded (e.g. if you would have received a 90%, you will be given a 45%, or half of the original score).

If you anticipate being late, consider asking the Course Instructor for an extension. Much like the real world, you will need to negotiate the challenges of meeting your assignments while balancing many other demands.

Academic Integrity

Students in this course are fully expected to abide by the Academic Integrity Policy of the university, found at http://students.case.edu/groups/aiboard/policy.html. This is an online course and sometimes this can create new and unexpected situations that you may not have encountered in other classes. When you are in doubt, please seek out guidance from the Course Instructor.

You are expected to take proctored exams without glancing at other students papers, working in groups, or using any sort of notes or materials unless specifically allowed for that exam. Online quizzes must be completed ***alone***, though you may use materials to assist you. While you are encouraged to work in groups on your homework, the work you submit must be your own and must represent your honest intellectual efforts. If you make use of online resources, open-source code, or other such materials, you must disclose this when submitting your assignment (and you should share the materials with the Course Instructor, who will likely share them with the whole class!).

Evidence of academic dishonesty will be reported and investigated in accordance with university policies, and may result in judicial actions and a lowered or failing grade in the class.