Statistics 112: Introductory Statistics (II)
Fall 2013

Richard Berk
564 McNeil
berkr@wharton.upenn.edu

August 28, 2013

Course Description This is a second course in statistics and data analysis. We will focus on the the general linear model (regression and analysis of variance) with some discussion of smoothers and logistic regression if there is time. With modern data analysis packages, actual computations have become trivial. In this course, therefore, the focus will be on what to compute and how to interpret the results. This is a course in applied statistics. The emphasis is on the intelligent use of statistics. This is not a math course, or a course in mathematical statistics. Underlying statistical proofs will be considered only as necessary and then primarily at an intuitive level. Students with lots of background in mathematics have no particular advantage.

Class Hours: Tuesday and Thursday, 9:00 to 10:30, F45-JMHH

Office Hours: By appointment. I have found that if I fix office hours, they do not work for a large number of students. The best way to get a quick response from me is through e-mail. This works especially well when you have specific questions that do not require a lengthy response. I check e-mail regularly.

TA Justin Bleich (jbleich@wharton.upenn.edu) — Justin will serve as a resource for you, but there will not be regular TA sections.

Text: There is no text as such. I will circulate notes based on my regression textbook as well as notes on other relevant material.

Computing: This is always an issue. Many of you have some working knowledge of SPSS, or JMP, or STATA, or something else. All are

at least optionally point-and-click, but differ somewhat in the user interface and what they can do. Moreover, there will be required computing that none do, or none do very easily. (Also, there is no way that I can be sufficiently familiar with each to provide help when you need it.) After experimenting for several years with different options, the majority of students came to favor a statistical programming language called R. R is *free*, easily downloaded and installed, and runs on all common platforms (http://www.r-project.org). But you will be working from the command line, which for many of you will be a new experience. I will provide lots of help in the form of complete examples and in the past, students rapidly caught on. The command line format will give you much greater control over exactly what is being computed. It will also give you a much better sense of what statistical computing is about.

It is perfectly feasible to work directly in R (usually with the help of an editor). But the *free* software R-Studio (http://www.rstudio.com) provides a user friendly interface that I strongly recommend. R-Studio is an example of an IDE (Integrated Development Environment). IDEs are becoming very common across a variety of computing environments (e.g., for JAVA). If you choose to use R-Studio, download and install R first. You will then download and install R studio.

Comunication: In the past, I have found both Blackboard and Canvas unnecessarily complicated. Students have agreed. We will communicate by e-mail (berkr@wharton.upenn.edu). I will circulate class materials as attachments.

Prerequisites: Stat 111 or the equivalent. This is a little tricky because it is difficult to know exactly what is "equivalent." AP courses, for example, are usually not adequate, but there have been exceptions. If in doubt, check with me.

Homework: There will be regular homework. The homework will not be graded by will only count if a serious effort is demonstrated. Late homework will not be accepted without a very good and credible excuse.

Grading: Grades will be based primarily on three to four short papers that will require an analysis of real data and a professionally done research report. Students who turn in all of their homework will be given extra credit toward their final grade on the course. The criteria

of each possible grade will be discussed in class. There is no curve. In the past, the average grade in the course has been a B. It is hard to get an A, but about a quarter of the students have done so. Of course, every class is different.