Regression and Multivariate Data Analysis

B90.2301 / C22.0017

Gary Simon
KMC 8-53
212-998-0451
gsimon@stern.nyu.edu

Office hours

Monday	noon – 1:30 p.m.
Wednesday	4:00 p.m. – 5:30 p.m.

Web information
Course materials will appear in the Web location www.stern.nyu.edu/~gsimon. Click on the entry for B90.2301.

Textbooks:

Samprit Chatterjee and Ali S. Hadi, *Regression Analysis By Example*, 4th edition, John Wiley and Sons (2006). This is highly recommended, but it is not required. It is an excellent easy-to-read general reference on linear regression.

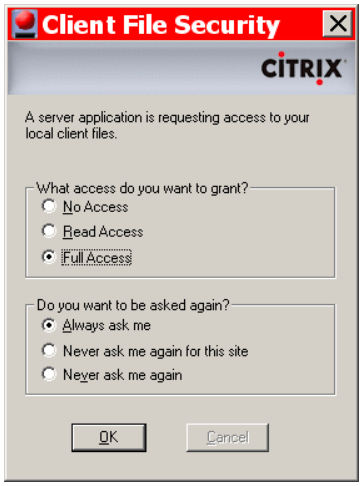
Samprit Chatterjee, Mark S. Handcock, and Jeffrey S. Simonoff, *A Casebook for a First Course in Statistics and Data Analysis*, John Wiley and Sons (1995). This book is optional. You may find it useful in terms of the case report style. These reports are over-explanatory, and reports that you write will want a more concise style.

Our course software will be Minitab, release 15.1. There are several ways in which you can get Minitab 15.

1. Purchase the CD version from the Professional Bookstore, about $115.
2. Rent the program for $30 for six months from www.e-academy.com/minitab. Navigate to the site's eStore.
3. Rent the program for $50 for twelve months from www.e-academy.com/minitab.
4. Use the program for no charge through the CITRIX system.

CITRIX is reached through http://apps.stern.nyu.edu. CITRIX will permit you to connect to a server accessing Minitab 15, as well as other software. Connection difficulties are not uncommon; Stern's Information Technology group can help if you have connection difficulties.

When Minitab is launched through CITRIX, you will get the panel shown at the right. Click the radio button for **Full Access** so that you can read and write files to the C:\ drive of your computer. When Minitab opens, you will want to read a worksheet through **File** ⇒ **Open Worksheet** or perhaps read a project thorough **File** ⇒ **Open Project**.



In the list on the **Open** panel, click on "My Computer." This will show drives available to you. The C:\ drive on your local machine will be identified as "C$ on 'Client' (V:)" .

The latest release number of Minitab for Windows is *at least* 15.1.30.0. Within Minitab, use **Help** ⇒ **About Minitab** to see the release number. You can also do **Help** ⇒ **Check for Updates**.

This course will use computers very heavily. The class illustrations will be with Minitab release 15. You may use any package you wish, on any machine that you wish, as long as it performs the necessary calculations. It is however the responsibility of the student to check whether the other packages are flexible. Plausible other packages are R, S-Plus, BMDP, and SAS. Excel is not rich enough to handle the material for this course.

Minitab 15 has critical features not found in Minitab 14 or in Student Minitab.

Our class only meets for 12 weeks, and there is no extra final exam session. It is important to review basic regression material before the first class. Please see the material related to "It is assumed that . . ." below on page 5.

There will be handouts in many of the class sessions during the semester. These will end up on the B90.2301 tab at Web location www.stern.nyu.edu/~gsimon. You should make every effort not to miss classes, however, since the material covered in class will be somewhat different from the material in the textbook.

Course grade will be based on homework projects *only*.

You will need to write statistical reports for this class. The Chatterjee, Handcock and Simonoff (CHS) Casebook gives many examples of such reports, though your reports will probably be less exhausting than these. You can read some of the cases that appear in the book to see what such reports look like if you have concerns about this.

There will be five projects for data analysis. You will obtain your own data for three of those five.

Do not take data from a textbook. You should obtain your data from original data sources.

Please provide complete source information for your data. If the data come from a printed source, include a photocopy, as well as the book name and page information. If the data come from the World Wide Web, provide the complete URL and the date of download. Many Web sites are complicated and a reference like "espn.com" is not enough to locate the data.

Data sets should not be taken from university archives or statistical libraries. Data sets should not be taken from a source that provides a similar analysis.

You will have about two weeks to complete each assignment. Assignments should be word processed, and the printed forms should be submitted.

Assignments will be scored up to 10 points. Yes, there are "adjustments" for late submissions. An assignment loses points as follows:

loses one point if submitted on days 1-3 after due date
loses two points if submitted on days 4-7 after due date
loses three points if submitted on days 8-10 after due date
loses four points if submitted on days 11-14 after due date

Assignments will not be accepted after 14 days.

If professional obligations keep you away from class, assignments should be sent in electronically to gsimon@stern.nyu.edu.

Don't wait until the last minute to do an assignment, as you might run into last-minute obstacles. For example, computing facilities might be down, your laptop's hard drive might crash, or your printer might run out of ink.

Please keep in mind that you will be graded on the quality of your analysis, and *not* on how exciting your data are. Don't waste time trying to find the perfect data set; you're working on a homework assignment, not a Congressional white paper. If you find that you're spending more time finding data than you are on analyzing it and writing up the analysis, you are allocating your time ineffectively.

With regard to the projects,

* I will be happy to answer questions about your analyses, either in person or by e-mail. As the semester goes on, the answer might well be "What do you think?".
* Alas, I will not be able review preliminary drafts of your projects, as the class size makes this arrangement unworkable.
* You should feel free to get help from classmates or people outside the class on computational issues, such as how to do something in Minitab or how to pull data from the Web.
* You should not get outside help on the substantive statistical issues.

Some additional notes:

(1) Laptop use in class is not permitted. If however you have a special need for note taking, please let me know.

(2) The final grade for the course will be based on the grades on the assigned homeworks. There will be no makeup or extra credit work. An incomplete grade for the course will not be considered simply to make up assignments that were not done, except for extraordinary circumstances. Assignments for which you receive no credit will have a strong detrimental effect on your grade. Missing two assignments could result in a failing grade in the course. The actual grades for the course will depend on the performance of the class.

(3) This course will be time-consuming. It is critical to plan carefully. Locating data suitable for your homework should get high priority.

(4) We take very seriously the Stern Honor Code. The Stern Honor Code is as follows: "I will not lie, cheat or steal to gain an academic advantage, or tolerate those who do." The Honor Code, as applied to this class, means that analyses must be done independently.

(5) If you have a qualified disability that affects your learning style, please contact the Moses Center for Students with Disabilities at 998-4980. Please let me know accommodations that they recommend. New York University is serious about enabling its students to work around their disabilities.

The prerequisite for this course is an introductory statistics core course. The Stern courses C22.0103 and B01.1305 are adequate prerequisites.

It is assumed that you already have a basic understanding of the simple regression model. You should review this material from your introductory statistics course before the first class session. You should download, print out, and read the document INTROnotes2010.pdf. This is on the course website www.stern.nyu.edu/~gsimon.

There is a small regression exercise in the file Startup.pdf (and also in Startup.docx). This is not assigned homework, but it will be covered in the first class. It will be helpful if you work through this before class. This exercise uses the data file NewsCirc; it's provided in formats NewsCirc.mtp (Minitab protable) and NewsCirc.xls (Excel).

You are responsible for all of the material in those handouts, although we will briefly discuss them in class. You should also download Homework 1 and answer all of the questions. I will give out the answers to these questions on the first day of class.