Statistics 6740 Data Management and Graphics for Statistical Analysis

Instructor: Mike Fligner – 408A, Cockins Hall - 292-0463

Office Hours: Monday and Wednesday 2:00 – 3:00, or by appointment

Text: SAS APPLICATIONS PROGRAMMING
by FRANK DI IORIO

Grading: Your grade will be based on the projects turned in throughout the semester.

Objectives:

This course is intended to familiarize you with the most common data management and presentation tasks that arise in Applied Statistics work. The data sets used are real and the tasks that need to be done have arisen in consulting work. The course emphasizes the use of the statistical software SAS, although the movement of data between software packages will be addressed. The course is composed of approximately 10 units, and for each unit there is a reading assignment and a project that needs to be turned in. The pacing will be such that a project is due around every 1 1/2 weeks. You can discuss project details with classmates or with me, but the final project you turn in should be your own work.

SAS software will run on a variety of platforms. SAS programs run similarly irrespective of which platform you are using in terms of windows, output appearance etc., although most students prefer to run SAS on the PC. A primary difference between platforms is the path names used to refer to files (programs and data sets) that are either saved or you are going to save. Statistics graduate students can run SAS in our department computer lab which has SAS version 9.2 installed. Graduate students outside the department can access SAS in their departments or at a public computing site.

Although the text is quite out of date in terms of the version of SAS, it covers the basic programing ideas nicely and makes a good reference. The topics covered are all essential to SAS programming and will form the basis for much of the course. The most recent information in terms of capabilities and options is contained in the HELP documentation and you will need to get some of your information to complete the projects from this source. Becoming familiar with the organization of the documentation in the HELP window will allow you to learn about additional SAS features after you have completed the course.

There are several useful features of SAS that have been added since the text was written, and there are also a few important topics that were not included in the text. I will discuss these and include them in the projects. The text has nice exercises every few pages, which students have found helpful. All the answers are in the back. My recommendation is that you do the exercises in the book as you are reading the required material. (You don’t need to write out the solutions formally. Just think about them for a few minutes and then check with the answers in the back to make sure you are picking up the major points. These exercises do not require actually running programs and are not all that time consuming). If you are serious about learning SAS, reading the text is important to obtain a good overview of the SAS programming language.

The emphasis in the course is on data management and presentation, not particular statistical applications, as the statistical backgrounds of students in the course vary widely. However we will assume basic knowledge of descriptive statistics including graphics. The regression analysis capabilities of SAS will be covered briefly.
Finding the Course Data, Handouts and Class Programs and Projects

The course data, handouts, class programs and projects will all be posted on Carmen (login to www.carmen.osu.edu using your name.number OSU account). The data sets are not large and are in a ZIP file which allows them to be downloaded easily. If you are having trouble downloading the data, e-mail me (maf@stat.osu.edu) and I can attach the ZIP file to my reply. The current handouts, projects and class programs will be made available throughout the quarter.

The projects are intended to get you to use SAS on your own, but if you are having trouble with a particular part of a project feel free to e-mail me or make an appointment to come in. Typically it is best if you try and solve the problems you encounter on your own, but don’t waste hours getting the same error message, particularly at the beginning of the course. When you e-mail me, make sure to include the code you are having trouble with as well as the error messages from the LOG that you are getting. Often your code is fine but it is referring to files or variables that have not been defined correctly earlier in the program.

The course data sets are saved in several formats including text files, excel files, transport files and .csv files. To begin we will be reading the data into SAS using programming statements, although we will use the menu driven import features later in the course using common formats. Regardless of the availability of menu driven import features, knowledge of the programming statements gives you a great deal of flexibility in the kinds of raw data sets that you can read into SAS and is also important when writing macros.

First Homework

Read Chapters 1 – 3 of the text. If you are new to SAS this will start to give you a feel for some of the basics of a SAS program. Don’t worry about understanding what each line of the program is doing – we will cover all of this as the course progresses. Do the pre-project handed out.

NOTE – DO NOT turn in electronic copies of the projects. ONLY paper copies will be accepted. If you have an important midterm on the due date of a project and are trying to decide whether to complete the SAS project or study for your midterm, STUDY for your midterm. Our schedule is fairly flexible.