

# STATISTICAL COMPUTATION

**W**elcome to Math 475! My name is Professor Edward Spitznagel. This is a course that introduces you to the theory and practice of statistical computation using SAS software.

## Textbook

**T**he text is Marasinghe and Kennedy's *SAS for Data Analysis*. You can download it *for free* in pdf format through Olin Library at:

<http://catalog.wustl.edu:80/record=b4314747~S2>

When you download the text, you will also be given the opportunity to buy a printed copy for \$24.99 and have it mailed to you. I have asked the bookstore to stock a few copies of the book, in case you find buying it from them to be more convenient. The bookstore price is \$79.95 plus sales tax. Even though this is more than triple the direct-from-Springer-Verlag price, it's still a lot less than most other course books.

When you get a chance to review the book's table of contents, you will find that three main topics are covered: programming, graphics, and statistical modeling. One software package, SAS, can do all of these (and a lot more besides).

## SAS Software

**T**here are two major statistics packages, SAS and R. Each one has its unique strengths. Despite R being free and SAS being very expensive, SAS remains the skill set that companies look for in prospective statistics employees.

Why? Well, it has to do with history. Once upon a time, there were three major computing languages, FORTRAN, COBOL, and ALGOL, and one major computer company, IBM. This was in 1964, before most or all of

you were born and maybe even before your parents were born. IBM decided that there ought to be a fourth, super, language, which they called PL/I. ("One Ring to rule them all.") So they built PL/I as a gigantic language to encompass *everything* in the "big three," and even more.

If IBM had maintained its dominance, we might all be programming in that common language. Two years later, in 1966, several statisticians at North Carolina State University decided to build a statistics package, and they patterned its syntax after that of PL/I. Besides giving it the power of PL/I syntax, they added features whereby entire sets of data could be referred to as objects and manipulated with single commands such as:

```
PROC SORT DATA=A; BY ID; RUN;
```

If you've only programmed in a lower level language like VISUAL BASIC or JAVA, you'll appreciate what a timesaver that can be. As computing evolved, SAS evolved with it, so that today SAS continues to contain every last capability you might ever want. This does make SAS rather "bloated" and perhaps inelegant, but most of us revel in that "inelegance" because it enables us to get stuff done quickly. Here's a cool example, which I'll treat you to some day when we have a little class time to spare: *Prove that the 13<sup>th</sup> of the month is more likely to fall on a Friday than on any other day of the week.* It takes less than 100 characters of SAS code.

## Course Schedule

**O**ur course meets Tuesdays and Thursdays, 4-5:30 in Cupples I Room 115. **Before you come to class, please preview the section or sections of the book to be covered that day.** Our schedule is as follows:

Day	Sections
Tuesday, August 26	1.1 TO 1.4
Thursday, August 28	1.5 TO 1.6
Tuesday, September 02	1.7
Thursday, September 04	1.8
Tuesday, September 09	2.1
Thursday, September 11	2.1
Tuesday, September 16	2.2
Thursday, September 18	2.3
Tuesday, September 23	3.1 TO 3.2
Thursday, September 25	3.3 TO 3.5
Tuesday, September 30	3.6 TO 3.7
Thursday, October 02	3.8 TO 3.9
Tuesday, October 07	4.1
Thursday, October 09	4.2
Tuesday, October 14	4.3
Thursday, October 16	MIDTERM EXAM
Tuesday, October 21	4.4
Thursday, October 23	4.5
Tuesday, October 28	5.1
Thursday, October 30	5.2
Tuesday, November 04	5.3
Thursday, November 06	5.4
Tuesday, November 11	5.5 TO 5.6
Thursday, November 13	5.7
Tuesday, November 18	6.1 TO 6.2
Thursday, November 20	6.3
Tuesday, November 25	6.4
Tuesday, December 02	6.5
Thursday, December 04	6.6
Thursday, December 17	FINAL EXAM

## Homework Schedule

With the exception of the first day of class (August 26<sup>th</sup>) and the class after the midterm exam (October 21<sup>st</sup>) a small amount of homework is due at the beginning of each class. The schedule is:

Day	Exercises
Tuesday, August 26	Nothing due
Thursday, August 28	1.1,1.4,1.6
Tuesday, September 02	1.7,1.9,1.10
Thursday, September 04	1.5,1.8,1.11
Tuesday, September 09	1.2,1.12,1.13
Thursday, September 11	2.1bcd,2.3,2.5
Tuesday, September 16	2.7,2.8,2.9
Thursday, September 18	2.10,2.11
Tuesday, September 23	2.14,2.15
Thursday, September 25	3.2,3.3
Tuesday, September 30	3.4,3.6
Thursday, October 02	3.9,3.10
Tuesday, October 07	3.8,3.13
Thursday, October 09	4.1,4.2
Tuesday, October 14	4.5,4.7
Thursday, October 16	4.4
Tuesday, October 21	Nothing due
Thursday, October 23	4.14
Tuesday, October 28	4.8
Thursday, October 30	5.2
Tuesday, November 04	5.1,5.3
Thursday, November 06	5.9
Tuesday, November 11	5.15
Thursday, November 13	5.18
Tuesday, November 18	5.20
Thursday, November 20	6.1
Tuesday, November 25	6.2
Tuesday, December 02	6.5
Thursday, December 04	6.12

For the most part, one, two, or three homework problems are required per class. They are due at the beginning of class. Up to three of you will then be randomly selected to present solutions to those problems on the blackboard. (To do so, you temporarily “borrow back” the work you just handed in.)

While this kind of classroom participation may be unusual in mathematics courses, it is very common in other departments. It encourages you to stay current in your studies.

My *official* office hours are from 5:30 to 6:30 on Tuesday and Thursday. That’s right after class. My office is Room 118 in Cupples I, just a few feet east of our classroom. As far as unofficial hours are concerned, you are *welcome* to knock anytime you see the light on. However, if you are elsewhere on campus or off-campus, I recommend calling in advance to see if I’m in. My telephone number is 935-6745.

## Calculators

The Texas Instruments calculators TI-83, TI-84, and TI-89 (and the new TI-Nspire series) can perform some but not many statistical calculations. You may find them helpful in checking your work in some of the exercises. They will also be useful on your midterm and final exams. However, there will be nothing on your exams that requires their special capabilities. A scientific “four-banger” like a TI-30 should suffice. That is, you will need a calculator that can add, subtract, multiply, divide and (occasionally) take a square root, logarithm, or exponential. If you have a more powerful calculator made by TI, HP, Sharp, and Casio, etc., you are more than welcome to use it on the exams. In fact, feel free to bring a bunch of calculators if you wish. Just make sure you have new (or spare) batteries.

## Accessing SAS

If you were a visitor from another planet deciding on whether to apply to WU, you would be mystified by the apparent non-

availability of SAS on WU public computers. In fact, the only publicly accessible computers that have SAS on their desktops are all in one room, Seigle L012. (To use them, just click on the inverted pyramid on the “desktop”.) There are a total of 14 such computers, PC’s and Macs combined. However, SAS can be accessed from *any* PC or Mac on campus through Microsoft’s “Remote Desktop Application.” It just takes some elbow grease, as described below.

## Remote Desktop

**T**he following fifteen steps will enable you to access SAS from any computer connected to the campus network. I have written them in terms of using one of the computers in Olin Library, but with modifications, it should work with any computer (PC or Mac) as long as it is connected to the campus network. That is, you can access SAS from your dorm room or dorm public area, but you won’t be able to do it from your parents’ home.

- (1) Find the Olin computers. There are 36 (egad, I’ve died and gone to heaven) in an area on the main floor past the staff desk.
- (2) Sit down and sign on to one of them. (Skip the few that have small screens.) Click the Windows symbol at the lower left.
- (3) You should see a line called “Remote Desktop Connection”. Click it. (If you don’t see that line, then look for “Remote Desktop Connection” under “All Programs” followed by “Accessories”.)
- (4) It will ask you what you want to connect to. Tell it you want to connect to lab.vlab.wustl.edu .
- (5) It (lab.vlab.wustl.edu) will ask you to sign on. If you’ve gotten this far, that’s a good sign! Give it your WUSTLkey and password.
- (6) It may reject you, but try again and again. (I once got in on my fourth try.)
- (7) On the new desktop, first column, about half-way down you should see an orange cubical icon labeled SAS 9.2(32) (English). The label runs over

two lines, just so you’re not surprised. Click (or double-click) it, and SAS should start up.

- (8) Great! You’re now running SAS, just like in Seigle.
- (9) Well, almost like in Seigle. You’ll search around for a way to print your results, and you’ll almost be ready to tear your hair out.
- (10) The big secret: You have to have TWO desktops open at the same time, the one containing SAS, and the one you went through to get there.
- (11) You should see a bar at the very top of the desktop, with a map-tack symbol at its left end and two overlapping squares at its right end. (The whole bar is only about five inches long, so it doesn’t occupy the whole top of the screen.) Click the overlapping squares symbol.
- (12) Your SAS desktop screen should shrink somewhat, revealing the original desktop screen behind it. From that original screen, you can start up MSWord, to receive output from SAS.
- (13) You can resize your SAS screen to something that you’re comfy with, and start computing. You will be able to copy both text and graphics from SAS into MSWord.
- (14) Once things are in MSWord and arranged to your satisfaction, you can print your results.
- (15) To finish, exit from SAS and then close the remote desktop. Finally, sign off of the computer, pack up your things and go off to have some fun.

Note that Seigle L012 is open very generous hours:

Monday:	8:30am-8:30pm
Tuesday:	8:30am-8:30pm
Wednesday:	8:30am-8:30pm
Thursday:	8:30am-8:30pm
Friday:	8:30am-5:30pm
Saturday:	11:00am-5:00pm
Sunday:	11:00am-8:00pm

Therefore, you may never have to use Remote Desktop, but it’s good to have it available as backup. For my part, I am staggering the due dates for homework in my three courses, to avoid everybody showing up in Seigle at the same day/time.

However, if you get the feeling that just 14 public computers with SAS on their desktops are insufficient for more than 4000 A&S undergraduates, you might mention to your aca-

demic advisors that you expect a little more for the tuition you're paying.

## Examinations

**W**e will have a midterm exam on October 16<sup>th</sup> and a final exam on December 17<sup>th</sup>. The midterm is at the regular class time, 4:00-5:30, while the final exam is 6:00-8:00. The midterm is in our regular classroom, Cupples I 115. The final exam may be in a different room, to be announced at the end of the semester. Each examination will consist of ten long-answer problems, each worth up to two points. You may bring one 8.5×11 inch notesheet to each examination. You may write on both sides of the notesheet. In the past, some students have formed “cooperatives” and composed their notesheets using MS Word. If you wish to do that, it's fine with me.

## Course Grades

**T**he first six homework sets contain 3 problems each, worth 18 points total. Points are awarded in discrete fashion, 1 point if a solution is mostly correct and 0 if a solution is mostly incorrect. The remaining twenty-one homework sets contain one or two problems. If there are two problems in a set, each solution is scored 1 or 0. If there is a single problem in a set, the solution is scored 2 (mostly or completely correct), 1 (about half correct) or 0 (mostly or completely incorrect). These twenty-one sets are worth 42 points total, so the entire set of homework problems is worth  $18+42 = 60$  points. The two exams are worth a total of 40 points, so homeworks plus exams are worth 100 points. Correct presentations are worth up to two extra points each. Typically I will invite you to present three times during the semester, so ideally you can accumulate up to 106 points by the end of the course.

I anticipate that the course grades will follow the modern convention, in which the A range will be 90 to 100, the B range will be 80 to 90, the C range will be 70 to 80, and the D range

will be 60 to 70, with plus and minus grades at the tops and bottoms of each of these ranges. If you are registered pass/fail, you must achieve at least 70 points to pass, which is the lowest score for a C-.)