SAS and Data Management (EPID 700)
Syllabus, Fall 2017

Description:
EPID 700 will introduce you to the basic concepts and skills needed to use SAS for epidemiological research. Emphasis is on data management, but you will also be well versed in performing descriptive statistical analysis and you will get a preview of the statistical analyses taught in subsequent EPID methods and BIOS classes. My aim is to teach you "how SAS thinks" as well as SAS syntax and coding structure. With these tools you will not need to memorize specific code, though some memorization of commonly used code will occur naturally. You will also leave this class with a plethora of resources to help you with the SAS coding you will encounter in future courses, as well as in your research. In addition, a brief introduction to R will be included to teach you the underworkings of R as well as basic programming in R. Four lecture hours per week (3 credits).

Instructor: Paula Strassle, MSPH
e-mail: pstrass@live.unc.edu
Department of Epidemiology
Gillings School of Global Public Health
McGavran-Greenberg Hall
Office hours: EPID student room Wed 12:00 – 1:00 PM or Fri 11:00 AM – 12:00 PM

Faculty: Stephen Cole, Ph.D.
e-mail: cole@unc.edu
Department of Epidemiology
Gillings School of Global Public Health
2104 McGavran-Greenberg Hall

Time and Place:			
	Wednesday	1:25 – 3:15 PM	McGavran-Greenberg 2306
	Friday	12:20 – 2:10 PM	McGavran-Greenberg 2306

Prerequisites: Some familiarity with Windows and statistics would be useful, but not necessary. This course assumes that you have little to no prior knowledge of SAS or R.

Course Materials: **Required:** BIOS 511 course notes & supplemental readings– Available on Sakai. I will point you to selected sections from these notes.

A device to transfer files you create in class to your personal computer.
(H: drive, memory stick (recommended), online file back-up software, etc.)

Optional: These books are helpful, but you are not required to purchase them.

The Little SAS Book: A Primer (5th Edition) Lora D. Delwiche and Susan J. Slaughter (2012)
Note: the 4th edition is available electronically and on reserve via UNC libraries but does not cover new additions from the SAS 9.3 release. For the purposes of this class, you can use either the 4th or 5th editions.

Learning SAS by Example: A Programmer's Guide Ron Cody (2007)
This text has been placed on reserve at the UNC Health Sciences Library.

Applied Statistics and the SAS Programming Language Ronald P. Cody and Jeffrey K. Smith (2005)
This text has been placed on reserve at the UNC Health Sciences Library.

SAS and R: Data management, statistical analysis, and graphics Ken Kleinman and Nicholas J. Horton (2014)
This text available electronically via the UNC libraries.

Learning Objectives:
At the conclusion of this course you will be able to:

1. Apply appropriate data management techniques to data used in a research setting
2. Use SAS and R to perform epidemiologic summaries using basic tabular and statistical procedures.

Format:
Each class session will incorporate a topic-specific lecture, interactive class discussion, hands-on coding practice and self-paced exercises. There will also be ample opportunity for you to ask questions and discuss any technical difficulties you are experiencing.

Grading:

Homework (11 total)	50% (5.6% each for 9 best, drop 2 lowest grades)
Labs	10%
Project 1	10%
Project 2	10%
Project 3	15%
Forums Posting	5%

Final grades are based on the standard graduate school scale (H, P, L, F) with the following breakdown:

H: ≥95%; P: 70-<95%; L: 65 –<70%; F: <65%

Due dates: Please refer to the course schedule, posted on Sakai. Except for labs, which must be turned in on time for any credit, all homework and projects will be penalized 10% per day late. This policy may be relaxed in the event that we have come to a prior agreement about a deadline extension due to unplanned events.

Honor Code:
The UNC honor code and detailed information pertaining to it can be found at http://honor.unc.edu/

Accessing course materials:
The course website is accessed through Sakai (https://sakai.unc.edu) and is where lecture notes, assignments, and SAS resources are posted. Log on to Sakai using your ONYEN and password (the same as your email). From the set of tabs near the top, click on EPID 700 to access the class website. If EPID 700 is not included in your list of courses please contact the instructor.

Coursework-specific information

Submitting course materials:
All homework and project assignments must be submitted via the Sakai course site by 11:59 PM on the date they are due unless you receive other instructions.

Please follow the detailed instructions below:

1. Every file you submit must be named using the course name, the assignment, and your first initial and last name, i.e. "epid700_hw5_ pstrassle", "epid700_project1_pstrassle" and "epid700_lab2_ pstrassle".
2. Please include your name and email address as part of each document you submit.
3. Follow the instructions specific to each assignment or lab to know which files you are required to submit. Please do not submit any files not specified in the instructions. Submit multiple files with the same file name but the appropriate file extension, i.e. "epid700_hw5_ pstrassle.docx" and "epid700_hw5_ pstrassle.sas".
4. Follow the instructions specific to each assignment to know what format each file needs to be in for submission. For example, I will sometimes require you to submit a SAS log file (.log) as a text file (.txt). Below are the file extensions for each type of file you will encounter in this course:

File type	Extension
Portable Document Format	.pdf
Microsoft Word	doc or .docx
Microsoft Excel	.xls or .xlsx
Text	.txt
SAS editor	.sas
SAS log	.log
SAS output	.mht
SAS data set	.sas7bdat
Access database	.mdb

Homework:
Homework assignments will comprise short problems allowing you to practice material covered during class sessions. Repetition is key to familiarizing yourself with SAS syntax, and a part of the homework may feel like busy-work (unavoidable!). The remainder of the homework questions will lead you through more complex and thought-provoking coding.

Collaboration in homework is allowed, but use good judgment to prevent your collaboration from becoming detrimental to your learning. Each student is required to submit his/her own homework assignment. Submitted assignments should be your individual effort, even if you consult with other students about your strategy for obtaining these solutions. To facilitate your learning it is always best to first try each problem on your own before collaborating with others.

Homework grading:

- Half of the homework credit will be given for an honest attempt at completing all of the problems in a given assignment. If you are unable to complete a problem, you must do the following three things:

    1) Complete as much of the problem as you can.
    2) State which part of the problem you are stuck on and why.
    3) Specify the resources you have consulted for assistance.

- If you are not able to finish all of the problems in a given assignment, turn in what you have by the completion due date. Partial credit will be given based on the proportion of the homework problems completed. Your two lowest grades will be dropped from the final grade.

- Please code your answers to homework assignment problems using the homework completion template provided on Sakai. In order to encourage professional coding practices, points worth up to 10% of the total homework grade will be added for code that is easy to read (details on style will be posted to Sakai and discussed in class). Sloppy code may result in a deduction of up to 10% for the homework assignment.

- Following the completion due date/time for each assignment, a solution key will be provided via Sakai.

Homework assignments should be submitted according to the homework guidelines document found under **Sakai > Resources > Homework.**

Labs
There will be a number of labs that will typically be held on Wednesdays. Typically, labs will be more challenging than homework assignments and will require you to apply the skills you have learned in novel ways. Since the labs are more difficult, you will score 100% for completing the lab. Each lab will be held immediately following lecture. I will provide materials for each lab on Sakai. If you are unwilling/unable to attend the lab section, you will still be graded on lab participation.

You may complete your lab assignments independently or in groups (maximum group size of 4), but each lab must be submitted **via Sakai before the end of the day (11:59 PM)** on the day we have the lab (students who work in groups may submit one set of lab assignments provided that all group member names are included on the documents). No late labs will be accepted. Completed labs will be submitted in the same fashion as homework.

Lab assignments should also be submitted according to the homework guidelines document found under **Sakai > Resources > Homeworks.**

Projects
In lieu of midterm and final exams, you will complete three projects as a way of synthesizing the skills you learn over the semester. The projects will work with data to develop publication-quality tables, graphs and apply advanced SAS programming skills. Each project will be cumulative, with a focus on the most recent skills developed at the time the projects are assigned.

You may consult any print or electronic resources (excluding prior EPID 700 exams, SAS competency exams or EPID 715/716 materials from past years) while completing the projects; however, the honor code is in effect. Questions and points of clarification should be directed to the course instructor. Each project is worth 100 points. Points will be deducted for sloppy code.

Project 1: Project 1 will assess whether you are mastering the most basic and fundamental aspects of SAS programming. The project is placed early in the semester in order to allow for extra review and/or changes to the pacing of material delivery before we move on to more advanced topics. Project 1 must be completed independently.

Project 2: Project 2 will involve a data analysis project using information about the demographics, prenatal care, and smoking habits of mothers who gave birth to a live infant in North Carolina in 2006, with an opportunity to practice creating figures. All students who plan to enroll in EPID715/716 are required to complete Project 2, either as the SAS competency exam or during the first week of EPID 716. You will use the final dataset you create for this project in EPID 716. Project 2 must be completed independently.

Project 3: Project 3 will stretch your SAS coding skills to create a macro, practice arrays and do loops and other advanced skills using longitudinal data (the data are hypothetical for the purposes of the class). This project is

designed to be challenging, so you may work independently or with a small group (groups will be randomly assigned).

Projects should be submitted according to guidelines found under **Sakai > Resources > Projects.**

Forum Posting
On the EPID700 Sakai site, there is a section called 'Forums'. This will be a forum for posting questions related to homework and labs, but not projects since those are independent assignments. This is not a replacement for office hours, rather a supplement to emailing me. And it allows other students to review common questions and more advanced students to post answers. You are required to post either questions or answers to questions in this forum as part of a participation grade. Your grade will be determined by the following formulae: $1 - 0.4^{\#\ of\ posts}$, so if you post 1 time you get a 60%, 2 times a 84%, …., and anything greater than 5 is 100%.

SAS Availability:
You can obtain a free copy of SAS for your personal computer using the UNC Software Acquisition Ordering system. Log in at https://software.unc.edu/portal/ using your onyen and password. Click *Continue to Student Ordering Portal*. Type 'SAS' into the search field and click *Go*. To your cart, add the SAS Complete Set version 9.4 TS1M1 for your computer's platform. Follow the instructions to complete your order. When your order is ready you will receive an email asking you to pick up your installation discs or USB key from the UNC Software Acquisition Office located in the basement of the Undergraduate Library.

Once you receive the installation discs or USB key, if you choose to attempt to install the software yourself, it is highly recommended that you use the installation wizard's preset selections when you install the software. These preset selections will provide you with components and options in excess of what you will need during your time as a student in the Epidemiology Department. If you want to install additional components or options, know what you are doing or seek assistance as installation is not always straightforward and errors will prevent your software from operating correctly. Instructions for installing SAS can be found at http://help.unc.edu/help/installing-sas-9-4-on-windows/. If you would prefer not to attempt to install SAS yourself, you have the option of either taking your computer and the installation disks to IT services in the basement of the Undergraduate Library (across the hall from the Software Acquisition Office) or to Spencer Gee in the Epidemiology Department (2nd floor of McGavran-Greenberg Hall) to have them install SAS for you. If you choose this option, be prepared to leave your computer with them for 1-3 days and be aware that there is often a flood of IT problems at the beginning of the semester and you should seek help early and be patient with their work load!

Note that as of July 2014 SAS 9.4 TS1M1 is the latest version of SAS available from the Software Acquisition Office. If you already have SAS 9.3 or earlier installed on your computer, you can remove it and install SAS 9.4. SAS 9.4 is NOT supported on Windows XP Home edition, Vista Home Basic and Vista Home Premium edition, but you can upgrade to Windows 7 Enterprise for free at the ITS Walk-in Service in the Undergraduate Library. For Mac users: Many Epidemiology students run SAS on their Macintosh machines using virtualization software such as Parallels or Bootcamp. More information can be found at http://sils.unc.edu/it-services/macs/virtualization.

If you experience difficulties while installing SAS you have three sources of help:

1) UNC ITS Response Center – see http://help.unc.edu/2083 for phone, online and in person contact information.
2) SAS Technical Support - call 919-677-8008. They will ask for your site number. It is 39466002.
3) Epidemiology Department ITS – submit a help ticket on http://sph.unc.edu/epid/epid-tech-support/

Finally, be aware that your SAS license will expire on December 31st each year. To update your license you will need to update the SAS Installation Data File. The following website directs you on how to do this: https://software.unc.edu/sas

As an alternative to installing SAS on your own computer (especially for Mac users), if you have a reliable internet connection you can use UNC's Virtual Lab or Virtual Computing Lab. See the help document in Sakai under **Readings and Resources > Options for Installing_Running SAS > Virtual Computing at UNC.pdf**