

Epidemiologic Analysis Using SAS

COURSE DURATION

This is an on-line, distance learning course and material will be available from:
June 1 – 30, 2019

INSTRUCTOR

Ryan Demmer, PhD MPH
Associate Professor
Demm0009@umn.edu

COURSE DESCRIPTION

The course will teach students to apply basic analytical epidemiologic methods using SAS statistical software. Heavy emphasis will be placed on basic data manipulation and operationalization skills in preparation for applied epidemiologic analysis that builds on concepts introduced in epidemiology and biostatistics core courses. Each session is a combination of didactic lecture and hands-on practice. Students will conduct epidemiological analyses on actual data sets, and learn the importance of data preparation and cleaning, descriptive analyses, as well as how to conduct basic statistical analyses in SAS.

PREREQUISITES

No experience with SAS is assumed, but students should be familiar with computing in the Windows environment. A basic understanding of statistical and epidemiological principles is strongly encouraged but brief introductions to key concepts will be offered as necessary although focus will naturally be on the application and interpretation of statistical and epidemiological concepts in the context of SAS.

TECHNICAL REQUIREMENTS

Students will need access to a computer with high-speed internet access and SAS software version 9.0 or higher installed. At minimum, Base SAS and SAS/STAT should be installed. Please read “Obtaining and installing SAS” below for helpful tips on acquiring and installing SAS software.

Obtaining SAS Software

Please find information below regarding various approaches to installing SAS. Please note that these are general guidelines that have worked well for students in previous years. However, these instructions might not work for all students and we cannot offer technical assistance for SAS installations. It is the responsibility of registrants to obtain access to SAS software prior to the beginning of this course.

SAS for FREE on a Mac or PC

See below for details regarding the use of SAS University Edition, a free SAS product for teaching and research. Be aware that the SAS University Edition can be slow compared with a full SAS installation. If at all possible, it is strongly recommended that registrants gain access to a full SAS installation (more details below).

http://www.sas.com/en_us/software/university-edition/download-software.html

Purchasing and Installing SAS

For non-Columbia registrants, SAS can be ordered by contacting SAS customer support:

<http://www.sas.com/technologies/analytics/statistics/stat/index.html#section=1>

It is also recommended that non-Columbia registrants check to be sure that free or discounted versions of SAS are not available for purchase through their respective institutions (e.g., many universities have negotiated reduced licensing fees for SAS).

Columbia students, staff and faculty can obtain licenses and installation disks through CUIT.

Please see the following link for details:

http://www.cumc.columbia.edu/it/getting_help/sas.html

Installing SAS Software

Please find below, a guide for purchasing and installing SAS on a personal computer. Note that this information is meant as a guide and should not replace a careful reading of the installation instructions provided by SAS. Note that we are unable to offer any technical support for acquiring or installing SAS software.

PC USERS

1. Both 32-bit and 64-bit versions of SAS 9.3 and 9.2 are available for Windows 7, and Vista Operating Systems. Be sure you purchase the version of SAS that is compatible with your operating system.
 - a. For Windows OS 7, you can determine if you have a 32-bit or 64-bit OS by going to the Windows start menu and selecting "Control Panel", then "System and Security", then "System". A summary of your Windows OS type will be displayed under "System".

2. To the best of the instructor's knowledge, SAS is not compatible with Windows Home versions; you must be running Windows Professional, Enterprise or Ultimate.
 - a. For Windows OS 7, you can find this information by going to the Windows start menu and selecting "Control Panel", then "System and Security", then "System". A summary of your Windows OS edition will be displayed under "Windows edition".
 - b. Columbia students, faculty and staff can download a free Windows OS 7 Enterprise edition from CUMC-IT at the following link:

<http://cuit.columbia.edu/cuit/software-downloads/operating-system-software/windows-7-enterprise>

MAC USERS

Current versions of SAS are not available for Mac computers. Nevertheless, it is quite possible to use SAS on a Mac. Several faculty in the department of Epidemiology at Columbia University run SAS on Mac, including the course instructor. There are various approaches to getting SAS on a Mac but the following is recommended:

1. Install a virtualization software on your Mac. A virtualization software allows the user to install a nonMac operating system on their Mac computer that can run simultaneously with their Mac OS. This does not change the look or function of your Mac computer. Two options are listed below. The course instructor has had good success with Parallels although VM Fusion seems to work well also.

Parallels: <http://www.parallels.com>

VM Fusion: <http://www.vmware.com/products/fusion/overview.html>

2. Install Windows 7 Enterprise OS as a virtual machine via either Parallels or VM Fusion (see part 1 above).
 - a. Columbia students, staff and faculty can download a free Windows OS 7 Enterprise from CUMC-IT at the following link:
<http://cuit.columbia.edu/cuit/software-downloads/operating-system-software/windows-7-enterprise>
3. Start your Windows 7 OS and install SAS on your Windows 7 OS.

The following link also provides a helpful overview of getting SAS installed on a Mac:
http://www.sascommunity.org/wiki/Running_SAS_on_Macintosh_OS_X_10.4_and_Above

COURSE LEARNING OBJECTIVES

The primary objective of this course is to provide students with the tools to use the SAS statistical software package in the practical conduct of basic epidemiological research analyses. To reach that objective, students will gain facility in the use of SAS to (1) understand the SAS working environment; (2) generate SAS data sets from nonSAS data files; (3) read, clean, manipulate and operationalize data; and (4) conduct basic statistical analyses.

By the end of the course, students will be able to:

- Read raw data into SAS.
- Manipulate and clean data sets in SAS through printing, sorting, merging and the use of conditional expressions.
- Apply simple statistical and graphical procedures for the descriptive analysis of normally distributed data.

- Conduct TTESTS and ANOVA analyses in SAS and interpret SAS output.
- Conduct basic linear regression analysis in SAS.
- Interpret SAS output for these analyses.

If time allows, basic logistic regression analysis and multivariable regression techniques will be introduced.

COURSE READINGS

There are no required readings but the following textbook is highly recommended:

"The Little SAS Book: A Primer" by Lora D. Delwiche and Susan J. Slaughter, 3rd Edition, Copyright 2003, SAS Institute. (ISBN 1-59047-333-7). As necessary, readings from additional sources will be recommended.

"SAS for Epidemiologists. Applications and Methods". By Charles DiMaggio. 1st Edition. Copyright 2012. Springer. ISBN: 978-1-4614-4853-2.

Additional material for the course is drawn from the following sources:

- Applied Statistics and the SAS Programming Language, 4th addition by Ronald P. Cody & Jeffrey K. Smith, Copyright 1997, Prentice-Hall Inc.
- Statistics I: Introduction to ANOVA, Regression and Logistic Regression (developed by Melinda Thielbar, Mike Patetta and Paul Marovich), Copyright 2005, SAS Institute, Cary NC
- Categorical Data Analysis Using Logistic Regression (developed by Mike Patetta), Copyright 2005, SAS Institute, Cary NC.
- Psychiatric Epidemiology, by Ezra Susser, Sharon Schwartz, Alfredo Morabia and Evelyn J. Bromet. Copyright 2006, Oxford University Press
- Applied Regression Analysis and Multivariable Methods by David Kleinbaum, Lawrence Kupper, Keith Muller and Azhar Nizam, Copyright 1998, Duxbury Press

COURSE STRUCTURE

The on-line version of this course is meant to be a highly self-directed learning style that enables greater flexibility for course participants to complete the course objectives at the times and pace most conducive to the respective schedules of participants.

This course utilizes the learning management software, Canvas:

<https://canvas.instructure.com/login>

To get started all registrants will receive an e-mail inviting them to join on the first day of the course offering. Upon receiving the e-mail, participants should follow the instructions to get signed up for a Canvas account.

This course is anchored around ~15 recorded lectures organized into five "modules". The recorded lectures are audio and screen recordings that will enable the instructor to teach by moving between a PowerPoint lecture setting and the SAS software environment with relative ease so that conceptual points can be made while also embedding applied examples of course

concepts using SAS software. In general, each recorded lesson will have a PowerPoint file as well as a SAS editor file that can be used by participants and kept for future reference.

Data: There will also be several data sets provided to participants so they can replicate the examples shown in the lessons and also experiment on their own. The data sets can be found on-line at the Canvas course site. A few of the data sets used for “in-class” examples are pseudo data sets that are created for teaching purposes. Several other data sets are publically available and arise from The First National Health and Nutrition Examination Survey (NHANES):

<http://www.cdc.gov/nchs/nhanes/nhanesi.htm>

To learn SAS properly, it is absolutely essential that you use it. Therefore, all participants should have SAS installed on the computer from which they are watching streamed lectures. This will allow participants to frequently pause the lecture and apply concepts in SAS.

COURSE SCHEDULE

Module 1 – Introduction to SAS	
June 1st – June 6th	<p><u>Learning Objectives:</u> Understand: the SAS working environment; difference between SAS editor, log and output windows; differences between SAS editor files and SAS data sets. Introduction to the following procedures: PROC PRINT, PROC CONTENTS, PROC MEANS, PROC FREQ; OBS= and FIRSTOBS= options; WHERE statements;</p> <p><u>Lessons:</u> #1 – #2</p> <p><u>Exercises:</u> #1</p> <p><u>Recommended Reading:</u> Delwiche and Slaughter, Chapters 1 & 2</p>
Module 2 – Creating SAS Data Sets	
June 7th – June 12th	<p><u>Learning Objectives:</u> Understand: data set naming conventions; permanent vs. temporary data sets; LIBNAME statement; how to merge and/or modify SAS data sets using MERGE or SET statements; how to use PROC SORT; how to use the SAS import wizard; how to read in data using list and column formats; reading in data from internal and external sources; handling dates; INFORMATS; OPTIONS statement;</p> <p><u>Lessons:</u> #3 - #6</p> <p><u>Exercise:</u> #2</p> <p><u>Recommended Reading:</u> Delwiche and Slaughter, Chapters 1 & 2</p>

Module 3 – Refining, Operationalizing and Describing Data

June 13th – June 18th – Learning Objectives: Understand: different variables types (continuous, ordinal, nominal, binary); how to create new variables with arithmetic operations; how to use SAS functions; arrays, do loops, IF/THEN statements; PROC UNIVARIATE, PROC MEANS, PROC FREQ, PROC RANK. How to create and apply FORMATS in SAS procedures and data steps.

Lessons: #7 - #9

Exercises: #3 - #5

Recommended Reading: Delwiche and Slaughter, Chapters 3, 4 & 6

Module 4 – Operationalizing Data Cont., Chi-Square, TTESTs, ANOVA

June 19th – June 24th – Learning Objectives: Understand: how to use SAS functions; arrays, do loops, IF/THEN statements; necessary data structure for conducting Chi-Square tests and obtaining odds ratios and risk ratios using PROC FREQ; how to conduct TTESTS using PROC TTEST.

Lessons: #10 - #11

Exercises: #6 & #7

Recommended Reading: Cody & Smith chapters 3 & 7

Module 5 – ANOVA & Linear Regression

June 25th – June 30th – Learning Objectives: Understand: basic principles of ANOVA and linear regression; when these methods should/can be used; underlying assumptions of linear regression analysis; how to use PROC ANOVA, PROC REG, PROC GLM, PROC PLOT, PROC BOXPLOT.

Lessons: #12 - #13.2

Exercises: #8 - #10

Recommended Reading: Cody & Smith chapters 5 & 9