

Statistical Computing and Research Data Management Programs (i.e., What programming package should I use?)

Background

Conduct of modern epidemiologic analyses requires use of sophisticated statistical software. The most flexible of these applications require knowledge of a high-level programming language (e.g., SAS, Stata, R). A high-level language has greater abstraction from the machine, as opposed to low-level programming languages (e.g. C, FORTRAN).

Currently, the core Epidemiology methods courses (i.e., EPID 715/6, 718, 722) support SAS only. Other courses in Epidemiology may require use of SAS, Stata, R or another language, at the preference of the Instructor. The core Biostatistics service courses require SAS (e.g. BIOS 545, BIOS 663, etc.).

The choice of your primary high-level language depends largely on your environment, as most the above listed high-level languages perform typical epidemiologic analyses equally well. For example, in the clinical setting there is often a preference for Stata due to its low cost, rapid learning curve and integrated point-and-click functionality, while in the conduct of large-scale studies there is a preference for SAS (e.g., ARIC, MACS, Nurses' Health Study). R is preferred by some epidemiologists and biostatisticians for its flexibility, quality graphics and specialized packages (it's also free). Stata and R can be run natively on Macintosh, Linux, and Windows based computers, while SAS does not run natively on a Macintosh (virtual operating system solutions are available, however). Many epidemiologists will eventually learn more than one statistical programming language.

Regardless of your choice, all UNC Epidemiology degree programs, except for the MSCR, require proficiency in SAS. Students must achieve a working knowledge of their statistical software package of choice prior to beginning EPID 715/6 in the spring of their first year. Learning opportunities are listed below for each program.

SAS Learning Opportunities

Courses:

1. EPID 700 -- SAS/Data Management (3 credits). Basic introduction to data management with an in-depth introduction to reporting and statistical analysis. SAS JMP, Stata and R will also be introduced. Compared to BIOS 511, EPID 700 has a slightly smaller breadth and only a basic introduction to the underlying workings of SAS; all the necessary information to prepare for EPID 715/6 is contained in this course.
 2. BIOS 511 -- Introduction to Statistical Computing and Research Data Management (3 credits). In-depth introduction to data management with a basic introduction to reporting and statistical analysis. Compared to EPID 700, BIOS 511 provides a more general introduction to SAS and further explores its underlying workings.
 3. UNC Odum Institute, SAS Statistical Computing – a 4-day short course, offered twice during fall and spring semester and once during summer semester. For more information: <http://www.odum.unc.edu/>, click on Short Courses. [This provides a good orientation to SAS, but is *not sufficient to prepare you for EPID 715/6.*]
-

Websites:

1. SAS Online Documentation: (for 9.2, but easy to navigate and everything you need should be here): (<http://support.sas.com/documentation/onlinedoc/bookshelf/92/>)
2. SAS learning modules at UCLA: (<http://www.ats.ucla.edu/stat/sas/>)

Recommended books:

1. *The Little SAS Book* by Lora D. Delwiche and Susan J. Slaughter. SAS Institute, Inc.: 2003.
2. *Applied Statistics and the SAS Programming Language* by Ronald P. Cody and Jeffrey K. Smith. Prentice-Hall, Inc: 2006.
3. *Learning SAS by Example: A Programmer's Guide* by Ron Cody. SAS Institute, Inc: 2007.

Note: To get SAS on your home PC for **free**, go to <https://software.unc.edu> . You will find links for "SAS 9.4" under "Science and Statistics." Select "SAS 9.4;" this is the most recent version of SAS available to students. Enter your onyen and password and follow the instructions to place your order. You will receive an email telling you when your order is ready to be picked up in the basement of the undergraduate library. SAS 9.4 is NOT supported on Windows Home edition, Basic edition and Vista Home Premium edition. Instructions for installing SAS can be found at <http://help.unc.edu/help/installing-sas-9-4-on-windows/>. Alternatively, the ITRC Walk-in Service will install SAS for you. This usually requires leaving your computer for 1-3 days.

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/macsvirtualization>.

Virtual Computing Lab (VCL): As a UNC student, you can access a suite of computer programs online through the Virtual Computing Lab (<https://its.unc.edu/research-computing/virtual-computing-lab-vcl/>).

Stata & R Learning Opportunities

(Optional – The Department does not currently support Stata or R, but does encourage learning multiple statistical software packages.)

STATA:

Courses:

UNC Odum Institute, Stata Statistical Computing. 4-day short course, offered twice during fall and spring semester and once during summer semester. For more information: <http://www.odum.unc.edu/>, click on Upcoming Short Courses under Education.

Websites:

1. Stata Learning Modules at UCLA: (<http://www.ats.ucla.edu/stat/stata/modules>)
2. Stata Tutorial from Princeton: (<http://data.princeton.edu/stata/>)
3. Stata Corp Website: (<http://www.stata.com>)

Note: To find newly posted introductions on the web, you may want to "google" search, type "introduction to stata."

Recommended books:

The tutorial book packaged with new orders of Stata provides a nice introduction to basics of the language.

Note: To purchase Stata, go to <https://software.unc.edu/available.php>. You will find links for “Stata GradPlan” under “Science and Statistics.” Select “Stats GradPlan.” Enter your onyen and password and follow the instructions to place your order. You will receive an email telling you when your order is ready to be picked up in the basement of the Undergraduate Library. Alternatively, you can order directly from the company at www.stata.com.

R:

Courses:

UNC Odum Institute, R Statistical Computing. 2-day short course, offered during fall and spring semesters. For more information: <http://www.odum.unc.edu> and click on Short Courses.

EPID 701-001. R for Epidemiologists, offered in spring semester.

Websites:

1. Quick R: <http://www.statmethods.net/>
2. R Studio tutorials: <http://www.rstudio.com/training/>
3. Applied Epidemiology using R: <http://medepi.com/courses/applied-epi-using-r/>

Recommended Books:

R in Action: Data analysis and graphics in R by Robert Kabacoff.
<http://www.manning.com/kabacoff/>

Note: R is freely available at <http://www.r-project.org>. R Studio <http://www.rstudio.com> is a recommended user-interface.