

PUBH 6325 SECTION 001-002

Data processing with PC-SAS
Spring/2019

COURSE & CONTACT INFORMATION

Credits: 1

Meeting Day(s): Four Saturdays (January 26, February 2, 9, 16)

Meeting Time: Lecture: Sec 001-Online
Labs: Sec. 002 – 9:00am – 11:00 am

Meeting Place: Lab: Mayo C381

Instructor: J. Michael Oakes

Email: Oakes007@umn.edu

Office Phone: 612-624-6855

Fax: 612-624-0315

Office Hours: Contact Libby Fortner at fortn002@umn.edu for an appointment

Office Location: 433 West Bank Office Building (WBOB)

COURSE DESCRIPTION

To introduce School of Public Health (SPH) students to methods for transferring and processing existing data sources. This time-intensive short course emphasizes hands-on and otherwise practical approaches to *pre-statistical data processing* and analysis with PC-SAS statistical software on a PC with a MS Windows operating system. The intended audience is public health masters and doctoral students.

Because of the overlap between this course and PubH 6420, if Division of Epidemiology students take both courses they can only use one towards their degree requirements, including electives. Students in other SPH programs will have to check with their Chair/DGS to see if they can count both or only one course toward a degree requirement.

COURSE PREREQUISITES

Prospective students must be matriculated graduate or professional students. Students should have passed a (bio)statistics course. Prior experience with SAS and/or STATA is helpful but not required. The Instructor may waive prerequisites for special circumstances.

COURSE GOALS & OBJECTIVES

Upon completion of this course the student should be able to:

- Understand basic and moderately advanced data structures, including rudiments of the binary number system, flat-files, relational and hierarchical data resources;
- Discuss the benefits and limitations of the PC-SAS interface;
- Deal with raw and non-SAS data;
- Know how to find, understand, and exploit SAS help-files, online resources, and texts;
- Write moderately complex “data steps” for manipulating data sets of any size, including the ability to (1) write professional SAS programs and interpret related conventia; (2) sort, merge, append, and collapse data sets; (2) generate new variables, especially with functions and formats; (3) label variables and data sets; (4) produce simple graphs; (5) write code for cleaning data and detecting missing values; (6) write simple macros for efficiency, be familiar with SAS’s ODS or graphics subroutines.
- Understand the essential *syntax* and commands for SAS analysis procedures (proc steps), such as tabulate, means, and regression. Interpretation of such methods is beyond the scope of this course.

METHODS OF INSTRUCTION AND WORK EXPECTATIONS

Course Workload Expectations

Class sessions will be held in both a lecture hall and a computer-lab. Lectures will introduce ideas and issues to students. Labs will permit hands-on training, group discussions, and interactive computer exercises. Homework will be readings and related internet access tasks (e.g., tutorials). The course is a time-intensive short course that meets approximately 4 hours per day for 4 days.

This is a time-intensive 1-credit short course. Since 1 credit is roughly equal to 3 hours of effort per week, and there are 15 weeks in a typical semester, you should expect to devote about 45 hours of effort to this short course. Subtracting both in-class and in-lab time, there are 25 hours for homework, or a maximum of 6.5 hours per each of four days of this class. The typical student should allow 1-2 hours of homework each week.

Students must complete all of the following to earn a passing grade in the course:

- A. Satisfactory completion of all assigned readings: Readings are critical to useful class discussions. (20% of total grade)
- B. Satisfactory completion of computer in-lab homework assignments: Such assignments will include essential programming and data manipulation techniques. Assignments will be given and must be completed in lab time (second hour of lab). Late homework will not be accepted. Homework must be presented in a neat and careful manner, as one would present a professional report. Detailed directions will be offered. (20% of total grade)
- C. Completion of an in-lab final exam on the last day of class. (60% of total grade)

COURSE TEXT & READINGS

(a) Two texts are recommended but not required, others are recommended for HPM students or those interested in such data. You may purchase these from SAS.com or other places such as Amazon.com. The instructor has not ordered them through the UMN Bookstore.

SAS Programming by Example (Recommended for all)

By: Ron Cody and Ray Pass

List price: \$42.95

360 pages

SAS Order No: 55126

ISBN: 1-55544-681-7

Year: 1995

Professional SAS Programmer's Pocket Reference, 5th ed. (Recommended for all)

By: Rick Aster

List price: \$17.95

239 pages

SAS Order No: 58128

ISBN: 1-891957-12-4 0

Year: 2000

Health Care Data and the SAS System (Recommended for HPM students)

By: Marge Scerbo, Craig Dickstein, and Alan Wilson

List price: \$37.95

300 pages

SAS Order No: 57638

ISBN: 1-58025-865-4

Year: 2001

Longitudinal Data and SAS: A Programmer's Guide (Recommended for the interested)

By: Ron Cody

List Price: \$29.95

208 pages

SAS Order No: 58176

ISBN: 1-58025-924-3

Year: 2001

(b) Standard UMN Internet access to online SAS help files, and the internet more generally. All students should automatically have this access.

COURSE OUTLINE/WEEKLY SCHEDULE

Week	Topic	Readings	Activities/Assignments
Week 1 January 26 – February 1	<ul style="list-style-type: none"> • Background; Data and Databases; PC-SAS Interface and Help • The computer (binary) number system; Character representation; Numeric and String representations; ASCII and EBCDIC • Databases; Flat files (fixed, delimited, freeform); Binary files; Hierarchical Data; Relational Data; Append, Collapse and Merge (conceptually) • The PC-SAS Interface, and its relation to the Windows OS (libname, system options; external files) • Introduction to the Program Data Vector and Data Step Processing • Ethics of Secondary Data Analysis 	<ul style="list-style-type: none"> • Explore UCLA Site • Surf SAS online help documents • Skim- Oakes, J. Michael. 2002. "Risks and Wrongs in Social Science Research: An Evaluator's Guide to the IRB." <i>Evaluation Review</i> 26:443-478 	<ul style="list-style-type: none"> • Study 'Concepts' (Chaps: 1-3 basics) • Study 'Concepts' (chaps 18-19 Windows stuff)
Week 2 February 2 – February 8	<ul style="list-style-type: none"> • Data Step and Programming Basics; Simple Procedures • Professional programming conventia, including comments and system options • Reading simple SAS data • Subsetting – keep, drop; obs, nobs • Introduction to Boolean Logic and SAS Operators (if, where, by, in, or, and, not) • Simple manipulation: variable assignment, rename, label, title • Basic Procedures: Proc Contents (simple), Proc Print, Prov Freq, Proc Means 	<ul style="list-style-type: none"> • Read 'Advanced Data Step Topics' PDF • UCLA SAS Module Subsetting Data • UCLA SAS Module for Common system options • UCLA SAS Module for Creating Variables • UCLA SAS Module for Labeling data and variables 	<ul style="list-style-type: none"> • Study 'Concepts' (Chaps: 7 Expressions; 20 Data step processing) • Study 'Procedures' (chaps: 3-2 Freq procedure; 2-29 Means) • Study online help for Proc Contents
Week 3 Date February 9 – February 15	<ul style="list-style-type: none"> • Data Set Manipulation; Functions; Formats; Dates • Sorting; Appending; Merging; Collapsing • Random sampling/Subsetting • String and Numeric Functions • Formats/Informats • Dates, Date formats, and intervals 	<ul style="list-style-type: none"> • UCLA SAS Module for Functions 	<ul style="list-style-type: none"> • Study 'Concepts' (chaps: 23 Modifying data sets) • Study 'Concepts' (chaps 8 Dates) • Study 'Dictionary' (chaps:3 Formats; 5 Informats)
Week 4 Date-Date February 16 – February 22	<ul style="list-style-type: none"> • Transferring Data Between Programs Retain; Arrays; Macros; ODS Graphs • Reading and Writing ASCII data • Importing excel data; Stat/Transfer & DBMS Copy • Retain • Arrays and do loops • Macro Basics • ODS • SAS Graph Basics 	<ul style="list-style-type: none"> • UCLA SAS Module for inputting data 	<ul style="list-style-type: none"> • Review www.stattransfercom

SPH AND UNIVERSITY POLICIES & RESOURCES

The School of Public Health maintains up-to-date information about resources available to students, as well as formal course policies, on our website at www.sph.umn.edu/student-policies/. Students are expected to read and understand all policy information available at this link and are encouraged to make use of the resources available.

The University of Minnesota has official policies, including but not limited to the following:

- Grade definitions
- Scholastic dishonesty
- Makeup work for legitimate absences
- Student conduct code
- Sexual harassment, sexual assault, stalking and relationship violence
- Equity, diversity, equal employment opportunity, and affirmative action
- Disability services
- Academic freedom and responsibility

Resources available for students include:

- Confidential mental health services
- Disability accommodations
- Housing and financial instability resources
- Technology help
- Academic support

EVALUATION & GRADING

Letter grades and associated points are awarded in this course as follows below. These will appear in the student's official transcript.

Pass/Fail Grading: An alternative to traditional A-F scale grades is the S/N grading scale. The "S" grade does not carry points but credit will count toward completion of student's degree if permitted by college or program. An "N" is given for student's exercising the S/N grading option but who fail to meet minimum course requirements. Students may change grading options without written permission as specified by the University and without penalty during the initial registration period or during the first two days of the class (since the course is a short-course). The grading option may not be changed after the second class.

Grades will not be normed within the class; it is possible for all students to receive an "A."

The maximum number of points a student can earn is 100. **Class participation, including lab attendance**, demonstrating comprehension of assigned readings and notes will account for **20 points (20%)** of the final grade. **Three in-lab assignments** demonstrating knowledge and ability to perform required tasks will account for **30 points (30%)** of the final grade. Each assignment will be worth equal amounts. **A final in-lab programming exam will account for 50 point (50%)** of your final grade. Numeric grades will map to letter grades as per the table above.

Grading Scale

The University uses plus and minus grading on a 4.000 cumulative grade point scale in accordance with the following, and you can expect the grade lines to be drawn as follows:

% In Class	Grade	GPA
93 - 100%	A	4.000
90 - 92%	A-	3.667
87 - 89%	B+	3.333
83 - 86%	B	3.000
80 - 82%	B-	2.667
77 - 79%	C+	2.333
73 - 76%	C	2.000
70 - 72%	C-	1.667
67 - 69%	D+	1.333
63 - 66%	D	1.000
< 62%	F	

- A = achievement that is outstanding relative to the level necessary to meet course requirements.
- B = achievement that is significantly above the level necessary to meet course requirements.
- C = achievement that meets the course requirements in every respect.
- D = achievement that is worthy of credit even though it fails to meet fully the course requirements.
- F = failure because work was either (1) completed but at a level of achievement that is not worthy of credit or (2) was not completed and there was no agreement between the instructor and the student that the student would be awarded an I (Incomplete).
- S = achievement that is satisfactory, which is equivalent to a C- or better
- N = achievement that is not satisfactory and signifies that the work was either 1) completed but at a level that is not worthy of credit, or 2) not completed and there was no agreement between the instructor and student that the student would receive an I (Incomplete).

For additional information, please refer to: <http://policy.umn.edu/Policies/Education/Education/GRADINGTRANSCRIPTS.html>

Evaluation/Grading Policy	Evaluation/Grading Policy Description
<p>Scholastic Dishonesty, Plagiarism, Cheating, etc.</p>	<p>You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis (As defined in the Student Conduct Code). For additional information, please see https://z.umn.edu/dishonesty</p> <p>The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: https://z.umn.edu/integrity.</p> <p>If you have additional questions, please clarify with your instructor. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class-e.g., whether collaboration on assignments is permitted, requirements and methods for citing sources, if electronic aids are permitted or prohibited during an exam.</p> <p>Indiana University offers a clear description of plagiarism and an online quiz to check your understanding (http://z.umn.edu/iuplagiarism).</p>
<p>Late Assignments</p>	<p>Not accepted</p>
<p>Attendance Requirements</p>	<p>No</p>
<p>Extra Credit</p>	<p>Yes</p>