

PubH 6325 – Sections 001-003
Data Processing with PC-SAS
Course Syllabus
Spring 2009

Credits:	1
Meeting Days:	Monday 1/12/2009 – Friday 1/16/2009
Meeting Time:	Lecture: Sec. 001 – 9:00-11:00 am Labs: Sec. 002 – 12:00-2:00 pm Sec. 003 – 2:00-4:00 pm (Location TBA)
Meeting Place:	Lecture: Mayo 3-100; Labs: Mayo C381
Instructor:	J. Michael Oakes, PhD
Office Address:	431 West Bank Office Building (WBOB)
Office Phone:	612-624-6855
Fax:	612-624-0315
E-mail:	oakes007@umn.edu
Office Hours:	by appointment

I. 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 an PC with a MS Windows operating system. The intended audience is public health masters and doctoral students.

Please note that although different in emphasis, software, and timing, there is overlap between this course and Biostat 6420 (Introduction to SAS Programming), typically taught by Mr. Greg Grandits. In short, 6420 has relied on Unix SAS, meets over the entire fall semester, and includes discussion about statistical tests, such as t-tests and regression.

Because of the overlap between this course and PubH 6420, if Division of Epidemiology students takes 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.

II. Course Prerequisites

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

III. Course Goals and 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, including (1) reading into SAS at least moderately complex ASCII and EBCDIC raw data, (2) writing fixed and delimited ASCII files from SAS, and (3) exploiting commercially available data conversion software;
- 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.

IV. Methods of Instruction and Work 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 5 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 5 hours per each of five days of this class. The typical student should allow 1 hour of homework each night.

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.
- 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.

V. Course Text and 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.

VI. Course Outline/Week Schedule

DAY 1: 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)

Navigating SAS Help: Native help files, online resources, texts

Introduction to the Program Data Vector and Data Step Processing

Ethics of Secondary Data Analysis

DAY 2: DATA STEP AND PROGRAMMING BASICS; SIMPLE PROCEDURES;

Professional programming convention, 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, Proc Freq, Proc Means

DAY 3: DATA SET MANIPULATION; FUNCTIONS; FORMATS; DATES

Sorting; Appending; Merging; Collapsing

Random sampling/Subsetting

String and Numeric Functions

Formats / Informats

Dates, Date formats, and intervals

DAY 4: TRANSFERRING DATA BETWEEN PROGRAMS RETAIN; ARRAYS; MACROS

Reading and Writing ASCII data

Importing Excel data; Stat/Transfer & DBMS Copy

Retain

Arrays and do loops

Macro Basics

DAY 5: STUDENT CHOICE, ODS, GRAPHS, STATA

ODS

SAS Graph Basics

VII. Homework/Readings

The instructor will offer notes/data and other such materials for each session via email and websites.

DAY 1: BACKGROUND; DATA AND DATABASES; PC-SAS INTERFACE; HELP; ETHICS

Surf www.sas.com

Surf [SAS online help documents](#)

Explore UCLA site www.ats.ucla.edu/stat/sas

Study 'Concepts' (chaps: 1-3 basics)

Study 'Concepts' (chaps 18-19 Windows stuff)

Skim Oakes, J. Michael. 2002. "Risks and Wrongs in Social Science Research: An Evaluator's Guide to the IRB." *Evaluation Review* 26:443-478.

DAY 2: DATA STEP AND PROGRAMMING BASICS; SIMPLE PROCEDURES;

Study 'Concepts' (chaps: 7 Expressions; 20 Data step processing)

Read 'Advanced Data Step Topics' PDF

Study 'Procedures' (chaps: 3-2 Freq Procedure; 2-29 Means)

Study online help for [Proc Contents](#)

UCLA SAS Module for [Subsetting Data](#)

UCLA SAS Module for [Common system options](#)

UCLA SAS Module for [Creating Variables](#)

UCLA SAS Module for [Labeling data and variables](#)

DAY 3: DATA SET MANIPULATION; FUNCTIONS; FORMATS; DATES

Study 'Concepts' (chaps: 23 Modifying data sets)

Study 'Concepts' (chaps: 8 Dates)

Study 'Dictionary' (chaps: 3 Formats; 5 Informat)

UCLA SAS Module for [Functions](#)

DAY 4: TRANSFERRING DATA BETWEEN PROGRAMS RETAIN; ARRAYS; MACROS; ODS; GRAPHS

Review www.stattransfer.com

UCLA SAS Module for [inputting data](#)

VIII. Evaluation and 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."

Grade Mapping			
Class Points	Letter Grade	Grade Points	Interpretation (relative to requirements)
>94	A	4.00	Outstanding
90-94	A-	3.67	
87-89	B+	3.33	
83-86	B	3.00	Significantly Above
80-82	B-	2.67	
77-79	C+	2.33	
73-76	C	2.00	Satisfactory
70-72	C-	1.67	Less than Satisfactory
70+	S	-	At Least Satisfactory
-	N/F	-	Not Satisfactory, No Credit
-	I	-	Incomplete

The maximum number of points a student can earn is 100. **Class participation** demonstrating comprehension of assigned readings and notes will account for **50 points (50%)** of the final grade. Up to fifteen points per day will be given for active participation, such as asking probing questions and being prepared to answer questions from the Instructor. This means that class attendance (both lecture and lab) is necessary but not sufficient to earn the maximum allowable points. **Four in-lab assignments** demonstrating knowledge and ability to perform required tasks will account for **50 points (50%)** of the final grade. Each assignment will be worth equal amounts. Numeric grades will map to letter grades as per the table above.

Course Evaluation

Beginning in fall 2008 the SPH will collect student course evaluations electronically using a software system called CoursEval. The system will send email notifications to students when they can access and complete their course evaluations. Students who complete their course evaluations promptly will be able to access their final grades just as soon as the faculty member renders the grade. All students will have access to their final grades two weeks after the last day of the semester regardless of whether they completed their course evaluation or not. Student feedback on course content and faculty teaching skills are important means for improving our work. Please take the time to complete a course evaluation for each of the courses for which you are registered.

Incomplete Grade

An incomplete grade is permitted only in cases of exceptional circumstances and following consultation with the instructor. In such cases an “I” grade will require a specific written agreement between the instructor and the student specifying the time and manner in which the student will complete the course requirements. Extension for completion of the work will not exceed one year.

University of Minnesota Uniform Grading and Transcript Policy

A link to the policy can be found at onestop.umn.edu.

IX. Other Course Information and Policies

Grade Option Change

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.

Course Withdrawal

Students should refer to the Refund and Drop/Add Deadlines for the particular term at onestop.umn.edu for information and deadlines for withdrawing from a course. As a courtesy, students should notify their instructor and, if applicable, advisor of their intent to withdraw.

Students wishing to withdraw from a course after the noted final deadline for a particular term must contact the School of Public Health Student Services Center at sph-ssc@umn.edu for further information.

Student Conduct, Scholastic Dishonesty and Sexual Harassment Policies

Students are responsible for knowing the University of Minnesota, Board of Regents' policy on Student Conduct and Sexual Harassment found at www.umn.edu/regents/polindex.html.

Students are responsible for maintaining scholastic honesty in their work at all times. Students engaged in scholastic dishonesty will be penalized, and offenses will be reported to the Office of Student Academic Integrity (OSAI, www.osai.umn.edu).

The University's Student Conduct Code defines scholastic dishonesty as “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; or altering, forging, or misusing a University academic record; or fabricating or falsifying of data, research procedures, or data analysis.”

Plagiarism is an important element of this policy. It is defined as the presentation of another's writing or ideas as your own. Serious, intentional plagiarism will result in a grade of "F" or "N" for the entire course. For more information on this policy and for a helpful discussion of preventing plagiarism, please consult University policies and procedures regarding academic integrity: <http://writing.umn.edu/tww/plagiarism/>.

Students are urged to be careful that they properly attribute and cite others' work in their own writing. For guidelines for correctly citing sources, go to <http://tutorial.lib.umn.edu> and click on “Citing Sources”.

In addition, original work is expected in this course. It is unacceptable to hand in assignments for this course for which you receive credit in another course unless by prior agreement with the instructor. Building on a line of work begun in another course or leading to a thesis, dissertation, or final project is acceptable.

If you have any questions, consult the instructor.

Disability Statement

It is University policy to provide, on a flexible and individualized basis, reasonable accommodations to students who have a documented disability (e.g., physical, learning, psychiatric, vision, hearing, or systemic) that may affect their ability to participate in course activities or to meet course requirements. Students with disabilities are encouraged to contact Disability Services to have a confidential discussion of their individual needs for accommodations. Disability Services is located in Suite 180 McNamara Alumni Center, 200 Oak Street. Staff can be reached by calling 612/626-1333 (voice or TTY).