

SPPH 553 DL (SAS)

Foundations of Public Health Computing
UBC School of Population and Public Health (SPPH 143)
In-class Sessions: Sept 19, Oct 17, Nov 14 (Thursdays)

Course Description

The work of public health professionals is increasingly becoming dependent on large and complex datasets. In Canada, particularly British Columbia, typical sources of these data include Vital Statistics, Medical Services Plan (MSP) billings, the Canadian Institute for Health Information (CIHI) hospital Discharge Abstract Database (DAD), and PharmaNet. Because these data are not collected primarily for research purposes, they are not in a format that is ready for statistical analysis. They need to be imported into a software package, checked, cleaned, and linked to other datasets. Handling and working with these types of massive datasets (e.g. the MSP data alone covers up to 4.8 million of British Columbians with multiple visits to their physicians each year) requires a disciplined approach to data preparation and management.

In this course, students will gain practical, hands-on experience in cleaning, preparing, linking, and managing large health data. The specific focus is on ensuring correctness, efficiency, security, and reproducibility when managing and preparing data for analysis.

There are many tools that can be used to work with large data sets. **SAS is the software package that will be used in this version of the course.** Students who are already proficient in SAS or who are interested in learning another software package should take the traditional version offered via the 'R' programming language.

Note: This course was adapted from the SPPH 553 course originally developed by Sarah Henderson using the R software package.

Learning objectives

By the end of this course, students should be able to:

1. Identify common mistakes in entering and organizing quantitative data
2. Import/load and explore large health datasets in SAS
3. Use common SAS functions
4. Clean, organize, and create data fields or columns, such as dates, age, diagnostic codes and postal codes
5. Subset, collapse/aggregate, append, and merge datasets
6. Write SQL statements
7. Run descriptive statistics
8. Generate graphs
9. Write loops and macros to automate repetitive tasks
10. Create analytic cohorts using linked health administrative data

Prerequisites

Currently, the course has no prerequisite. However, many of the students who did well in the course last year were those who either have previous experience with other statistical software packages (i.e. R, Stata, SPSS) or have completed SPPH 400.

Instructor

Joseph H. Puyat, PhD, MSc, MA (Psychology)

Assistant Professor (Partner), UBC School of Population & Public Health

Scientist & Research Methodologist, CHEOS, Providence Health Care

Epidemiologist, BC Centre for Disease Control

588 - 1081 Burrard St (Room 583A), St. Paul's Hospital, Vancouver, BC V6Z 1Y6

604.682.2344 x 63578 | jpuyat@cheos.ubc.ca (preferred)

Office hours by appointment

Teaching Assistant

TBA

Equipment and Software

1. A laptop computer that meets or exceeds the [System Requirements](#)¹ for running SAS.
2. [SAS University Edition](#),² which can be downloaded and used for free in academic settings (see schedule of activities for week 1 for installation procedures). The commercial version of the software, SAS 9.4, may also be used but installation and troubleshooting support for this version will not be provided in this course.

Learning Resources

Recommended Text:

Delwiche LD, Slaughter SJ. *The Little SAS Book: A Primer, Fifth Edition*. SAS Institute, 2012.

Students are also encouraged to use other online resources, such as the SAS [tutorials](#)³ and reference manuals ([Base SAS 9.3 Procedures Guide](#),⁴ [SAS 9.3 SQL Procedure](#)⁵).

Course Format

There will be 3 in-class (~7 hours each) and 10 web-based (~ 1.5 – 2 hours each) sessions for this course. The in-class sessions will give students an overview of the topics that will be covered and reinforced in the web-based sessions. There will be lectures and group-

¹ https://www.sas.com/en_us/software/university-edition.html#m=system-requirements

² https://www.sas.com/en_us/software/university-edition/download-software.html

³ <https://support.sas.com/en/software/university-edition-support.html#tutorials>

⁴ [Base SAS 9.3 Procedures Guide](#)

⁵ [SAS 9.3 SQL Procedure](#)

based activities in the in-class sessions while the web-based sessions will consist of videos, readings, and written instructions. The majority of time will be devoted to data management, preparation and to some extent, statistical analysis.

Grading/Assessment

Assignments	35%
Examinations (3)	60%
Participation – in class	2%
Participation – online	3%

Assignments are due at the end of the week a specific module is scheduled to be completed. The three exams are timed (3 hours maximum) and are available within a specific period of time.

Academic Integrity

Academic dishonesty will not be tolerated and students should know what constitutes academic dishonesty. For further information and guidance, please consult the [UBC Academic Calendar](#), specifically the section on [Academic Honesty and Standards](#) and [Academic Misconduct](#).

Course Schedule

Date	Topics/Activities
<i>Module 1</i>	<i>Installing and Configuring SAS</i>
Sep 2 – 8	<ol style="list-style-type: none"> 1. Download the SAS University Edition. Download a PDF version of the instructions or watch a video to get an overview of the process. 2. Set-up a work folder. 3. Run SAS using your browser. If necessary, activate virtualization on your computer and update your version of SAS. 4. Explore the SAS Studio Environment. <p>Introduction to SAS</p> <ul style="list-style-type: none"> • SAS statements: DATA step and PROC step • Subsetting by observation and by variables • SAS operators: Comparison and Logical Operators • Creating Variables
<i>Module 2</i>	Introduction to the BC Vital Statistics data
Sep 9 -15	Renaming variables, running basic SAS functions, working with DATE/TIME variables, creating and using SAS MACRO variables, and generating bar plots.
<i>In-Class Session I</i>	Review of Modules 1 and 2

Sep 19	Working with CCHS data <ul style="list-style-type: none"> • Generating frequency counts and cross tabulations • Calculating means and standard deviations Importing CSV files
<i>Module 3</i> Sep 23 – 29	Merging data through DATA step; collapsing data across observations; generating high-resolution, scalable scatterplots via PROC SGPLOT (SAS Statistical Graphics); running linear regression analyses
<i>Module 4</i> Sep 30 – Oct 6	Automating repetitive tasks using loops and MACROs
Oct 7 – 13	<i>First Exam</i>
<i>In-Class Session II</i>	Review of Modules 3 and 4
Oct 17	Introduction to the Structured Query Language
<i>Module 5</i> Oct 21 – 27	Using SQL to create data sets, define variables, and merge data sets Generating basic maps in SAS
<i>Module 6</i> Oct 28 – Nov 3	Preparing hospital data (Discharge Abstract Database) for analysis; handling compressed (gzipped) flat files
November 4 – 10	<i>Second Exam</i>
<i>In-Class Session III</i> Nov 15	Review of Modules 5 and 6 Examining hospital length of stay and readmission rates Exploring Health Canada's Drug Product database
<i>Module 7</i> Nov 18 – 24	Working with prescription drug use data (PharmaNet data) Linking PharmaNet data with Health Canada Drug Product Database
<i>Module 8</i> Nov 25 – Dec 1	Creating analytic cohorts based on administrative data I
Dec 2 – 8	<i>Third Exam</i>