SPPH 553 001 2021W (SAS)

Foundations of Public Health Computing
UBC School of Population and Public Health
September 7 – December 7, 2021 (Thursdays, 2-5PM, B151)

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 typical 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 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) require 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 course**.

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

1. Identify survey and health administrative data that can be used for research
2. Import/load and explore large health datasets using 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 and execute 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
SPPH 400 and SPPH 502

Instructor

Joseph H. Puyat, PhD, MSc, MA (Psychology)
Assistant Professor (Partner), UBC School of Population & Public Health
Scientist & Research Methodologist, CHEOS, Providence Health Care
588 - 1081 Burrard St (Room 583A), St. Paul's Hospital, Vancouver, BC V6Z 1Y6
jpuyat@cheos.ubc.ca Office hours by appointment

Teaching Assistant

Fidela Mushashi
PhD Student
Fidela3000@gmail.com

Office hours by appointment

Equipment and Software

1. A relatively recent laptop/desktop computer with at least 4Gb of RAM and 50Gb of free hard disk space.
2. Robust and Hi-speed Internet connection, preferably wired.
3. Free SAS OnDemand account (https://welcome.oda.sas.com/login)

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[1] and reference manuals (Base SAS 9.3 Procedures Guide,[2] SAS 9.3 SQL Procedure[3]).

Course Format

The course consists of 6 in-person classes (~3 hours each) and 5 modules (~ 2 – 3 hours each) that can be completed offline. There will be lectures and group-based activities in the in-person sessions while the web-based sessions will consist of videos, readings, and written instructions. Optional tutorials (structured) may also be scheduled. The majority of time will be devoted to data management and preparation, and to some extent, statistical analysis.

Grading/Assessment

Assignments (5)	30%
Examinations (3)	60%
Participation (discussion board)	10%

Assignments are due at the end of the week a specific module is scheduled to be completed, or at pre-specified deadlines. The three exams are timed and are available within a week. The timed assignments/exams must be taken at a time when they can be completed without interruption as

[1] https://support.sas.com/en/software/university-edition-support.html#tutorials
[2] Base SAS 9.3 Procedures Guide
[3] SAS 9.3 SQL Procedure

the system starts counting down the moment the first question/item is shown. All work (completed or not) will be submitted automatically at the end of the time limit. Starting an assignment/exam, just to view the questions/items, without working on or completing it could result in receiving the lowest possible score for that assignment/exam.

Participation in online discussions will be marked and assessed following this rubric:

0	No posts or contributions
1	Questions or requests for help or assistance
2	Short responses to questions or requests for assistance
3	Very helpful responses to questions or requests for assistance; sharing alternate code that facilitates communication and reproducibility

There will be a total of 11 online discussions, one for each module. Students are expected to participate or post in 8 of the 11 discussions. Those who participated in all 11 discussions will have only their top 8 contributions included in the calculation of the final marks for this course. Note that discussions will be open for 2 weeks only and will not be available for further posting once closed.

Academic Integrity

Academic dishonesty will not be tolerated and students are expected to know what constitutes academic dishonesty. For further information and guidance, please consult the UBC Academic Calendar, particularly the section on Academic Honesty and Standards and Academic Misconduct.

Course Schedule

Date	Topics/Activities
Module 1 Sep 7 – 12	Introduction to SAS • SAS statements: DATA step and PROC step • Subsetting by observation and by variables • SAS operators: Comparison and Logical Operators • Creating Variables **In-person class (1/6): September 9**
Module 2 Sep 13 - 19	Working with CCHS data • Generating frequency counts and cross tabulations • Calculating means and standard deviations • Importing CSV files **In-person class (2/6): September 16**

Module 3 Sep 20 – 26	Introduction to the BC Vital Statistics data Renaming variables, running basic SAS functions, working with DATE/TIME variables, creating and using SAS MACRO variables, and generating bar plots.
Module 4 Sep 27 – Oct 3	Automating repetitive tasks using loops and MACROs
Module 5 Oct 4 – 10	Merging data through DATA step; collapsing data across observations; generating high-resolution, scalable scatterplots via PROC SGPLOT (SAS Statistical Graphics); running linear regression analyses **In-person class (3/6): October 7**
Oct 11 – 17	***First Exam (3 hours)***
Module 6 Oct 18 – 24	Introduction to the Structured Query Language **In-person class (4/6): October 21**
Module 7 Oct 25 - Oct 31	Using SQL to create data sets, define variables, and merge data sets Generating basic maps in SAS
Module 8 Nov 1 – 7	Preparing hospital data (Discharge Abstract Database) for analysis; handling compressed (gzipped) flat files
Nov 8 – 14	***Second Exam (3 hours)***
Module 9 Nov 15 – 21	Examining hospital length of stay and readmission rates Exploring Health Canada's Drug Product database **In-person class (5/6): November 18**
Module 10 Nov 22 – 28	Working with prescription drug use data (PharmaNet data) Linking PharmaNet data with Health Canada Drug Product Database
Module 11 Nov 29 – Dec 5	Creating analytic cohorts based on administrative data **In-person class (6/6): December 2**
Dec 6 – 12	***Third Exam (4 hours)***