

An Introduction to Data Management and Cleaning for Analysis

Overview

This webinar series provides an overview of basic data management and data cleaning techniques using SAS software.

In taking this course, you will learn how to develop a systematic approach to managing and cleaning your data for statistical analyses. This approach involves understanding the big picture first, and then developing a strategy for translating the big picture into concrete problem-solving steps.

The workflow involved in these steps will be illustrated using a synthesized administrative data set and honed through a variety of applied exercises. During the course, you will be provided with access to a variety of practical tools that will help you to develop a sustainable and effective workflow for all of your future data analysis projects: SAS code, case studies, web resources and more. The overall goal of the course is to provide the conceptual and practical tools you need to handle your data preparation needs with confidence.

Homework activities will be provided for practice between sessions.

Webinar format

The interactive webinar software will provide remote access for students to view the instructor's screen, listen to the lecture in real time, and ask questions. The instructors will provide lecture slides (PowerPoint) for pre-reading prior to the start of the webinar. A training dataset with associated SAS code will be provided both for webinar demonstrations and homework practice activities.

Session 1: Tuesday April 4

Session 2: Thursday April 6

Session 3: Tuesday April 11

Session 4: Thursday April 13

Session 5: Tuesday April 18

Session 6: Thursday April 20

10:00am to 12:00pm PST

Prior required knowledge

Participants will be expected to have familiarity with the use of administrative data, basic knowledge of SAS functions (i.e.: descriptive statistics, merging and sorting) and an understanding of logistic regression.

Instructor: Brandon Wagar, PhD

Brandon is the Director of Clinical Analytics for Island Health, and an Adjunct Assistant Professor at the University of Victoria (UVic), School of Health Information Science. Previously, he was a Methodologist at the Canadian Institute for Health Information for eight years. He received his PhD in Behavioural Neuroscience from the University of Waterloo, and completed a post-doctoral fellowship in Cognition and Brain Sciences at the University of Victoria. Brandon developed and has taught "From Data to Meaningful Information: Tools and Techniques for working with Large Healthcare Datasets" within the Health Information Science Master's program offered by UVic.

WEBINAR SERIES: An Introduction to Data Management and Cleaning for Analysis

Webinar objectives

By the end of this webinar series, participants will be able to:

- Identify key types of data errors commonly found in the use of administrative data
- Address and correct data errors using a systematic process
- Subset, filter and aggregate data in preparation for statistical analyses
- Define the role of key variables for statistical analyses
- Recode qualitative variables as required
- Transform quantitative variables as required

Course content

Session 1: Understand the 'big picture' and know your research variables

- Course outline and homework assignments
- Identify and assess use of research variables

Session 2: Bring all your variables together under the same roof: Building a study data set

- Take stock of all the data sets involved in the research project
- Examine each data set briefly to understand what information it contains, why that information was collected and how it fits into the "big picture"
- Determine whether the data exhibit any natural grouping
- Extract the relevant portions from each data set and merge the appropriate portions together to bring all of the available variables under the same roof
- Perform simple linkage techniques

Session 3: Data cleaning/screening, diagnosing and editing

- Examine the distribution of each variable using both visual and numerical means
- Find "errors" in the data
- Rectify the "errors" found in the data
- Make a mental note of the distribution of the "error"-free data

Session 4: Date and time manipulation and sequencing

- Work with variables recorded at different time scales
- Perform various operations involving date and time variables

Session 5: Transformation, grouping, deriving

- Create new (numeric) variables by transforming available (numeric) variables using transformations such as log, square root, etc.
- In data sets where the data exhibits a natural grouping structure, create new variables by aggregating information across the groups present in the data
- Cluster variables that need to be modelled as random variables
- Create categorical variables from continuous variables

Session 6: Analysis and interpretation

- Making comparisons
- Understanding limitations
- Interpreting results

Course texts and pre-webinar reading

A recommended text for this course is:

Cody's Data Cleaning Techniques Using SAS by Ron Cody

Webinar fee

- Regular rate: \$295
- Student rate: \$185

For more information or to register, contact:
ann.greenwood@popdata.bc.ca