

OER for Teaching Advanced Epidemiological Methods

M. Ehsan Karim <ehsan.karim@ubc.ca> and Epi-OER team: Hossain MB, Frank HA, Yusuf FL, Ahmed SS, Asamoah-Boaheng M, Zheng C

2024 OER Virtual Showcase and Poster Session, March 7, 2024

Introduction

In response to the growing need for advanced methodological training among health researchers and trainees, our project introduces a comprehensive OER designed to bridge the gap between theoretical knowledge and practical application in epidemiology.

Project Summary

The SPPH 604 course, enhanced with a computational lab component, is at the forefront of our initiative, focusing on the implementation of advanced epidemiological methods using open-source data and software.

Goals and Objectives

1. Develop an Interactive Open Textbook: A core resource that combines epidemiological theory with hands-on statistical practice.
2. Incorporate Real-World Data Analysis: Using government repositories such as Statistics Canada and the US CDC to provide authentic learning experiences.
3. Facilitate Self-Paced Learning: Through formative quizzes and interactive modules, catering to diverse learning needs.

Methodology

Leveraging the R software, we guide learners from novice to proficient levels, enabling them to conduct reproducible research and communicate findings effectively. The curriculum spans ten core modules, covering topics from confounding adjustment to machine learning in causal inference.

Open-access Online Textbook [CC-BY 4.0]

Advanced Epidemiological Methods

The Project

- Data wrangling
- R basics
- Data types
- Automating tasks
- Importing dataset
- Data manipulation
- Import external data
- Summary tables
- R Markdown
- R Functions (W)
- Quiz (W)
- App (W)
- Exercise (W)

Advanced Epidemiological Methods

AUTHORS
M. Ehsan. Karim

AFFILIATIONS
School of Population and Public Health, The University of British Columbia

Epi-OER team includes co-applicants of the UBC OER Fund Implementation Grant - Hossain MB, Frank HA, Yusuf FL, Ahmed SS, Asamoah-Boaheng M, Zheng C (all affiliated with the University of British Columbia)

PUBLISHED
January 29, 2024

The Project

Welcome to a place crafted to bridge a unique gap in the health research world. This website offers valuable resources for those who are taking their first steps into health research and advanced statistics. Even if you are familiar with health research, but advanced statistical methods seem daunting, you are in the right

Core + Bonus Modules

Module	Topics.Indicators
1	R for Data Wrangling (W)
2	Accessing (A) Survey Data Resources
3	Crafting Analytic Data for Research Questions (Q)
4	Causal Roles (R)
5	Predictive (P) issues
6	Complex Survey Data (D) Analysis
7	Missing (M) Data Analysis
8	Propensity Score (S) Analysis
9	Machine Learning (L)
10	Integrating Machine Learners in Causal (C) Inference
11	Non-binary Outcomes (N)
12	Longitudinal Analysis (T)
13	Mediation Analysis (I)
14	Scientific Writing Tools (G)

Impact & Accessibility

This OER initiative is set to revolutionize health research education by:

- Enhancing accessibility of methodological instructional materials.
- Promoting analytical expertise around various complex topics.
- Our resources will be openly shared through the UBC OER Catalogue, GitHub, and other digital platforms, ensuring widespread accessibility.
- The transition to OER significantly reduces reliance on traditional resources (e.g., printed textbooks), leading to substantial cost savings and enhanced sustainability.

Evaluation Strategy

Feedback mechanisms are incorporated at every stage, from pilot testing with SPPH 604 students to broader community input.

Reference

- Karim M. E., Epi-OER team (2024). Advanced Epidemiological Methods. Retrieved from <https://ehsanx.github.io/EpiMethods/> on January 07, 2024.

Acknowledgement

We extend our deepest gratitude to

- [UBC OER Fund Implementation Grant](#)
- support from UBC Work Learn and
- GTA allocation from SPPH.

Organized Materials

Overview of tutorials

Confounding
The first tutorial provides a thorough particular focus on its impact on treatment effect estimates, emphasizing the importance of proper adjustment for accurate estimates.

Mediator
This tutorial focuses on the role of mediators in causal inference. It assesses how adjusting for mediators affects the estimated treatment effect, exploring both scenarios where the mediator is either non-null or null.

Collider
This tutorial serves as a practical guide to understanding and adjusting for collider bias in observational studies.

Video Contents [Concepts + Computing]

Video Lessons

Target parameters
Average Treatment Effect (ATE) vs. Average Treatment effect on the Treated (ATT)

Average Treatment Effect (ATE) vs. Average Treatment effect on the Treated (ATT)

John takes Rosuvastatin (A=1) and his total cholesterol level is $Y(A=1) = 195$ mg/dL (milligrams per deciliter) after 3 months

John does not take Rosuvastatin (A=0) and his total cholesterol level is $Y(A=0) = 245$ mg/dL after 3 months

Effect of Rosuvastatin on John is $Y(A=1) - Y(A=0) = 195 - 245 = -50$

Exercises with Hints

Question III: [20% grade]

3(a) Subset
Subset the dataset excluding 'Very good or excellent'

3(b) Recode
Recode self-rated mental health variable and make it category labels only). Convert that variable to a factor

3(c) Regression
Run a logistic regression model for finding the relationship between self-rated mental health (Reference: Poor) and disorders. Adjust the model for three confounders: se

Annotated Software Codes

Step 4

```
1 fit.psm.s1 <- glm(Death ~ RHC.use,
2 data = matched.data.s1,
3 family = binomial)
```

Summary of fit.psm.s1:

Variable	Odds_Ratio	CI_Lower	CI_Upper
RHC.use	1.26	1.09	1.45

The interpretation is the same as before. In the propensity score model, the odds of death was 26% higher among those participants without RHC use.

Interactive Web Apps

Subset Data Info | Visually Exploring | Numerically Exploring | Table 1

DataExplorer package is used.

Select Plot Type:
Missing Data

Introduction
Missing Data
Histogram
Density Plot
Bar Plot

Features

Feature	Percentage
INDMPEL	5.16%
ENDMPEL	35.23%
RDEXPNG	57.44%

Missing Rows

Self-assessing Interactive Quizzes

Q1. Which is the correct way to access the NHANES 2017-18 data cycle in your RStudio environment?

A. Directly download from the CDC website and open it

B. Use R package 'nhanesA' to download

C. Both A and B are correct

D. Neither A nor B is correct

Incorrect. Try again.

Hints: The continuous NHANES files are stored on the NHANES website as SAS transport file formats (.xpt). You can import these data files in any statistical software that supports this file format. Some packages are also available to open those data files without downloading them.

