Learning R

Course Objectives

By the end of this workshop, participants will be able to design, implement, and critically evaluate quasi-experimental research designs for causal inference using R.

Knowledge Objectives

After completing the workshop, participants will be able to:

* Design causal research studies using quasi-experimental methods
* Implement key methods (RCT, DiD, RDD, Matching) in R with real data
* Diagnose and validate assumptions of each approach
* Interpret and communicate results effectively

Learning Approach

* Each concept introduced through data
* Theory explained as needed during implementation
* Active coding throughout
* Real-world examples and challenges
* Continuous hands-on practice

Pre-requisites

Please watch the video “Installing R and RStudio” at this link: <https://alexd106.github.io/intro2R/howto.html> and download and install R and R Studio.

For downloading R: <https://cran.r-project.org/mirrors.html>

For downloading R Studio: <https://posit.co/products/open-source/rstudio/>

Learning materials

<https://intro2r.com/> - Free online book that introduces to R

<https://alexd106.github.io/intro2R/index.html> - Companion to the previous course with examples and code.

<https://www.econometrics-with-r.org/> - Introduction to econometrics in R (check chapters 10 to 16).

Virtual Sessions

# Session 1: Foundations of R Programming (3 hours)

## Part A (60 min)

- Introduction to R and RStudio interface

- Creating projects

- Understanding the console, script editor, and environment pane

- Basic syntax and data types (numeric, character, logical)

- Variables and assignment operators

## Part B (60 min)

- Vector operations and creation

- Basic mathematical operations

- Functions in R: built-in functions

- Getting help in R (help(), ?, examples)

- Install packages

## Part C (60 min)

- Introduction to data frames

- Reading data into R (read.csv, read.table)

- Basic data inspection (head, tail, str, summary)

- Simple data manipulation

# Session 2: Data Manipulation (3 hours)

## Part A (90 min)

- Introduction to tidyverse

- Data filtering and selection using dplyr

- Sorting and arranging data

- Creating new variables with mutate()

## Part B (90 min)

- Grouping data with group\_by()

- Summarizing data with summarize()

- Joining datasets

- Handling missing values

# Session 3: Data Visualization and Basic Statistics (3 hours)

## Part A (60 min)

- Basic statistical concepts in R

- Descriptive statistics

- Basic probability distributions

- Simple hypothesis testing

## Part B (60 min)

- Introduction to ggplot2

- Grammar of graphics concept

- Basic plot types (scatter, line, bar)

- Customizing plot aesthetics

## Part C (60 min)

- Advanced plotting techniques

- Faceting and multiple plots

- Adding statistical layers

- Themes and color schemes

Econometrics with R: 4-Day Intensive Course

Duration: 5 days, 4 hours per day (20 hours total)

Format: Combined theory and practical sessions with R

## Course Description

This intensive course provides a comprehensive introduction to causal inference methods in econometrics, combining theoretical foundations with practical implementation in R. Students will learn fundamental econometric concepts and immediately apply them using R, culminating in the ability to conduct and interpret their own empirical analyses.

## Session 1: Regression & Randomized Experiments (2h)

### 1. Understanding Causality through Data (30min)

- Load dataset

- Explore treatment and control groups

- Calculate naive differences

- Introduce potential outcomes while visualizing data

- Implement first simple regression

### 2. Building Robust Analysis (30min)

- Add control variables based on balance analysis

- Implement and interpret clustered standard errors

- Visualize treatment effects across subgroups

## Session 2: Difference-in-Differences & Synthetic Control (2h)

### 1. Building Intuition for DiD (60min)

- Create basic two-period plots

- Implement simple DiD regression

- Visualize parallel trends

- Test parallel trends assumptions

- Extend to multiple periods

### 2. Synthetic Control Extension (60min)

- Transform data for synthetic control

- Build and visualize control units

- Compare DiD vs synthetic results

## Session 3: Regression Discontinuity & Time Series Breaks (2h)

### 1. Visualizing Discontinuities (40min)

- Create RD plots

- Discuss bandwidth through visualization

- Implement basic RD estimation

### 2. Refining RD Analysis (40min)

- Test manipulation with density plots

- Try different bandwidths

- Add covariates

- Validate results

### 3. Time Series Applications (40min)

- Create time series plot

- Implement ITS analysis

## Session 4: Matching & Selection (2h)

### 1. Understanding Selection (40min)

- Visualize selection bias

- Create propensity scores

- Match observations

### 2. Matching in Practice (80min)

- Try different matching approaches

- Check balance visually

- Estimate treatment effects

- Assess sensitivity