5-Week Data Analysis Workshop for Economics & Social Sciences

Zahid Asghar

2024-02-01

Workshop Plan

1 Workshop Overview

This workshop is designed for participants in economics and social sciences interested in developing hands-on skills in data analysis using R. Sessions will run for 2 hours weekly across two days (Friday and Saturday, 6-8 PM). The workshop will focus on practical activities with guided facilitation for online learning.

1.1 Workshop Objectives

- Develop skills in data handling, merging, and cleaning.
- Build competence in descriptive statistics, visualizations, and balance tables.
- Apply regression techniques like OLS, Logit/Probit, IV, and Difference-in-Differences.
- Encourage peer collaboration and hands-on data analysis.

---

2 Week 1: Research Question & Data Handling

Friday & Saturday (2-hour slots)

#	Duration	Activity	Facilitator Actions	Output
1	10 min	Welcome & Ice-breaker	Share workshop goals, personal intros using “two truths and a lie”	Participants connected

#	Duration	Activity	Facilitator Actions	Output
2	30 min	Identifying a Research Question	Guide brainstorming using a mind map	List of potential research questions
3	20 min	Sources of Data	Present key sources for economic and social science datasets	Awareness of data sources
4	40 min	Hands-on: Loading & Exploring Datasets	Guide participants through loading CSV datasets using R	Loaded datasets
5	20 min	Group Discussion: Variables & Structures	Facilitate data structure analysis	Identified key variables
6	20 min	Merging Datasets	Live demonstration and practice	Successfully merged datasets

---

3 Week 2: Cleaning Data & Descriptive Statistics

Friday & Saturday (2-hour slots)

#	Duration	Activity	Facilitator Actions	Output
1	10 min	Energizer	Quick quiz on previous topics	Active recall
2	30 min	Identifying Missing Data	Guide practical exercise using real data	Cleaned dataset
3	30 min	Outliers & Data Transformation	Explain handling outliers & transformations	Cleaned and transformed data
4	30 min	Descriptive Statistics	Hands-on calculations and interpretation	Descriptive statistics
5	20 min	Descriptive Charts	Group activity to create histograms and boxplots	Completed visualizations

---

4 Week 3: Balancing Tables & Ordinary Least Squares

Friday & Saturday (2-hour slots)

#	Duration	Activity	Facilitator Actions	Output
1	10 min	Icebreaker	Group puzzle challenge	Engaged participants
2	40 min	Balancing Tables	Hands-on creation of balancing tables	Created balancing tables
3	50 min	OLS Regression Explanation	Visualizing OLS through scatter plots and coding in R	Clear understanding of OLS
4	20 min	Hands-on: Running OLS	Guide participants through coding an OLS model	Implemented OLS
5	20 min	Group Reflection	Discuss results and common errors	Clarified common mistakes

---

5 Week 4: Logit, Probit & Instrumental Variables

Friday & Saturday (2-hour slots)

#	Duration	Activity	Facilitator Actions	Output
1	10 min	Icebreaker	Quick check-in	Participants feel connected
2	40 min	Logit & Probit Models	Explain theory with real-world examples	Conceptual clarity
3	50 min	Hands-on: Logit & Probit	Walkthrough coding and interpretation	Participants ran models
4	40 min	Instrumental Variables	Explain concepts and coding practice	Implemented IV analysis

---

6 Week 5: Difference-in-Differences & Wrap-Up

Friday & Saturday (2-hour slots)

#	Duration	Activity	Facilitator Actions	Output
1	10 min	Icebreaker	Reflection: “One thing I learned…”	Reflective participants
2	45 min	Difference-in-Differences Theory	Explain concept with historical examples	Theoretical understanding
3	50 min	Hands-on: Difference-in-Differences	Code-along for DiD estimation	Ran DiD models
4	15 min	Wrap-Up & Next Steps	Feedback collection and next steps	Participant feedback

---

7 Technical Requirements

- **Software**: R, RStudio, Quarto
- **Packages Needed**: `tidyverse`, `haven`, `fixest`, `dplyr`, `ggplot2`
- **File Sharing**: All materials and datasets will be shared via a cloud link before the session.

8 Next Steps & Feedback

- Workshop slides and datasets will be provided in a shared folder.
- Feedback will be collected at the end of the last session using a Google Form.

8.1 License

This workshop material is shared under the Creative Commons Attribution 4.0 International License.