
Vikram Dayal

Quantitative Economics with R

A Data Science Approach

Contents

Part I Introduction to the Book and the Data Software

1	Introduction	3
1.1	A Data Science Approach	3
1.2	Quick Tour of the Book	4
1.2.1	Part 1: Introduction to the Book and the Data Software	4
1.2.2	Part 2: Managing and Graphing Data	4
1.2.3	Part 3: Mathematical Preliminaries for Data Analysis	4
1.2.4	Part 4: Inference from Data	5
1.2.5	Part 5: Accessing, Analysing and Interpreting Growth Data	5
1.2.6	Part 6: Basic Time Series Data	5
1.2.7	Part 7: Introduction to Statistical/Machine Learning from Data	5
1.3	How to Use the Book	6
1.4	Help	6
1.5	R Code and Output	6
1.6	An Overview of Typical R Code	6
1.7	Resources	8
	References	8
2	RStudio and R	9
2.1	Introduction	9
2.2	R and RStudio	9
2.3	Use Projects	10
2.4	Use a Script	10
2.5	Different Objects in R	11
2.5.1	Vectors	11
2.5.2	Matrices	13
2.5.3	Data Frames	14
2.5.4	Lists	14

2.6	Toy Example: Net Present Value	15
2.7	The Tidyverse Approach	16
2.7.1	Data Analysis Workflow	16
2.7.2	The Tidyverse Package	17
2.7.3	Input and Wrangle Synthetic Data	17
2.7.4	Five Data Verbs	20
2.7.5	Graphs	23
2.7.6	Linear Model	24
2.8	Resources	26
	References	27

Part II Managing and Graphing Data

3	Getting Data into R	31
3.1	Introduction	31
3.2	Data in R or a Package	31
3.3	Data in a csv File	33
3.4	Data in a Stata File	34
3.5	Data from the World Development Indicators	35
3.6	Resources	36
	References	36
4	Wrangling and Graphing Data	37
4.1	Introduction	37
4.2	Example: Anscombe's Synthetic Data	37
4.3	Example: Carbon and Livelihoods Data	40
4.4	Example: WDI Data on CO2 and Per Capita Income	46
4.4.1	Getting the Data	46
4.4.2	Graphing the Data	48
4.4.3	Mapping the Data	51
4.5	Resources	54
	References	59
5	Networks	61
5.1	Introduction	61
5.2	Simple Example with Synthetic Data	62
5.3	Example: Medici Network	66
5.4	Example: Bali Terrorist Network	68
5.5	Simulating Network Formation	70
5.6	Example: Electrical Automotive Goods Production Network	73
5.7	Resources	78
	References	80

Part III Mathematical Preliminaries for Data Analysis

6 Functions 83

6.1 Introduction 83

6.2 Making Your Own Functions in R 83

6.3 Plotting Functions with the Curve Function 84

6.4 Statistical Loss Functions 85

6.5 Supply and Demand 86

6.6 Cobb–Douglas Production Function 87

6.7 Resources 91

References 92

7 Difference Equations 93

7.1 Introduction 93

7.2 Simple Toy Example 93

7.3 Example: Global Carbon Stocks 94

7.4 Fish 96

7.4.1 Numerical Simulation 96

7.4.2 Example: North Sea Herring 98

7.5 Example: Conrad’s Model of a Stock Pollutant 101

7.5.1 Commodity Residual Transformation Function 101

7.5.2 Stock Pollutant 102

7.5.3 Firm’s Choice of Commodity Q Given a Tax
on Waste S 103

7.5.4 What Is the Optimal Tax? 104

7.6 Resources 105

References 108

8 Matrices 109

8.1 Introduction 109

8.2 Simple Statistics with Vectors 109

8.3 Matrix Operations 111

8.4 Example: Poverty Rate and Relative Income 112

8.5 Resources 115

References 115

Part IV Inference from Data

9 Statistical Inference 119

9.1 Introduction 119

9.2 Box Models 119

9.2.1 Sample 119

9.2.2 Binomial Distribution 121

9.2.3 Function for Binomial Distribution 124

9.3	Sampling Distribution	125
9.3.1	Six-Sided Dice Simulation	126
9.3.2	Function for Sampling Distribution	129
9.3.3	Sampling Distribution for the T-Statistic	130
9.3.4	Inference from One Sample	132
9.3.5	Confidence Intervals	133
9.4	Bootstrap	136
9.4.1	Function to Understand Bootstrap	139
9.5	Permutation Tests	141
9.6	Example: Verizon	145
9.6.1	Permutation Test	147
9.6.2	Bootstrapping Confidence Intervals	148
9.7	Cautionary Example with Synthetic Data	150
9.8	Resources	151
	References	151
10	Causal Inference	153
10.1	Introduction	153
10.2	Causal Graphs and Potential Outcomes	154
10.2.1	Simple Example with Synthetic Data	154
10.2.2	Randomized Assignment of Treatment (Causal Graphs)	156
10.2.3	Randomized Assignment of Treatment (Potential Outcomes)	157
10.2.4	Covariate Adjustment	160
10.2.5	Selecting Regressors by Statistical Significance	163
10.3	Experiments	165
10.3.1	Example: Anchoring	165
10.3.2	Example: Women as Policymakers	169
10.3.3	Example: Educational Programme	172
10.3.4	Example: Star	175
10.4	Matching	178
10.4.1	Simple Example with Synthetic Data	179
10.4.2	Example: Labour Training Programme	183
10.4.3	Sensitivity Analysis	187
10.4.4	Example: Lead Exposure	190
10.4.5	Example: Compensation for Injury	192
10.5	Regression Discontinuity	197
10.5.1	Simple Example with Synthetic Data	197
10.5.2	Example: Minimum Legal Drinking Age (MLDA)	198

10.6	Difference-in-Difference	202
10.6.1	Example: Scrap Rate and Training	202
10.6.2	Simulation	203
10.6.3	Example: Banks in Business	206
10.7	Example:Manski Bounds for Crime and Laws	208
10.7.1	Bounds with Maryland as Counterfactual	210
10.7.2	Bounds Based on Difference-in-Difference	212
10.8	Instrumental Variables	215
10.8.1	Simulation	215
10.8.2	Example: Demand for Cigarettes	218
10.9	Resources	222
	References	222

Part V Accessing, Analysing and Interpreting Growth Data

11	Growth Data and Models	227
11.1	Introduction	227
11.2	Example: Growth	227
11.3	Example: Production Model and Crosscountry Data	230
11.4	Solow Model Simulation	232
11.5	Romer Model Simulation	235
11.6	Example: Growth in Recent Decades	237
11.7	Resources	243
	References	244
12	Growth Causes	245
12.1	Introduction (Institutions and Growth Example)	245
12.2	Geography and Growth	249
12.3	Exclusion Restriction Simulation	252
12.4	Other Support for AJR	254
12.5	Resources	254
	References	254

Part VI Time Series Data

13	Graphs for Time Series	259
13.1	Introduction	259
13.2	Simple Example with Synthetic Data	259
13.3	Example: Air Passengers	260
13.4	Example: Stock Market Volatility	262
13.5	Example: Inflation and Unemployment	263
13.6	Example: Historical Unemployment Data	266
13.7	Resources	269
	References	271

14	Basic Time Series Models	273
14.1	Introduction	273
14.2	Simulations	273
14.2.1	White Noise	273
14.2.2	Autoregressive Model	277
14.2.3	Random Walk	280
14.2.4	Moving Average	281
14.2.5	Autoregressive Moving Average	282
14.3	Example: Forecasting Inflation	285
14.4	Cointegration	292
14.4.1	Simulating Spurious Regression	292
14.4.2	Simulating Cointegration	293
14.4.3	Example: Federal Funds and Bond Rate	296
14.5	Example: Dynamic Causal Effects of Weather	298
14.6	Resources	299
	References	301
Part VII Introduction to Statistical/Machine Learning from Data		
15	Smoothers and Generalized Additive Models	305
15.1	Introduction	305
15.2	Simple Example with Synthetic Data	305
15.3	Example: GAMS with Wages Data	307
15.4	Example: Housing in Texas	310
15.5	Resources	313
	References	313
16	From Trees to Random Forests	315
16.1	Introduction	315
16.2	Simple Tree Example with Synthetic Data	316
16.3	Example: Arsenic in Wells in Bangladesh	318
16.4	Example: Home Mortgage Disclosure Act	322
16.5	Resources	326
	References	326