Applied Statistics for Economists

Margaret Lewis

Applied Stat
Economists

Margaret Lewis

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2012 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2012 Margaret Lewis

The right of Margaret Lewis to be identified as author of this work has been asserted by her in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Lewis, Margaret, 1956–
Applied statistics for economists/Margaret Lewis.
p. cm.
1. Economic–Statistical models. 2. Statistics. I. Title.
HB137.L49 2011
519.5–dc22

ISBN: 978-0-415-77798-8 (hbk)
ISBN: 978-0-415-55468-8 (pbk)
ISBN: 978-0-203-80845-0 (ebk)

Typeset in Times New Roman
by Sunrise Setting Ltd, Torquay, UK

Preface

Economists have used empirical information to understand economic phenomena for many years. While economic data, and the statistical methods for evaluating it, are increasingly sophisticated, the basic reason for using such evidence remains the same – learning about an economy by systematically measuring economic activity.

The following text presents quantitative methods frequently used by economists to gain such understanding. The methods examined are divided into two categories – descriptive statistics used to measure individual economic variables, and statistical inference used to measure relationships between economic variables and to test hypotheses about them. In addition to demonstrating how to generate a numerical result with each method, the current text applies each method to specific and contemporary economic questions. The text also provides detailed economic interpretations of the statistical results, a feature missing from virtually every current text, and it explains the strengths and limitations associated with each method. The goal is for students to learn to "do" empirical economics as they work through the text.

The text was initially developed as class notes for a course in quantitative methods in economics and grew out of a dissatisfaction with the scope and emphasis in existing textbooks. Many currently available texts, designed for business and economics courses in statistics, focus primarily on business examples, rarely using actual data for economic variables and seldom discussing questions of interest to economists. A second group of texts focuses on econometric analysis, an important statistical tool in economics, but certainly not the only quantitative method useful to economists.

Most contemporary texts also emphasize the mechanics of generating statistical measures, but provide little on the important questions of economic interpretation and economic application of the methods. As a result, students may gain technical knowledge from such an emphasis, but, when it comes to conducting their own economic research, they may not fully understand how the methods can be applied to their own research, nor can they always accurately interpret the statistical results generated. The current text attempts to bridge this gap.

Textbooks for statistical methods in economics have not always been so focused. Several years ago, I found a used copy of Harold W. Guthrie's *Statistical Methods in Economics* (Richard D. Irwin, Homewood, IL, 1966). Guthrie's book, like contemporary texts, provided instruction on how to use a wide range of quantitative methods, but it did more than that: it used actual economic data in most examples and, most importantly, it provided economic interpretations of the empirical results generated. The text also emphasized the appropriate, and inappropriate, use of statistics and quantitative methods in economics, another feature sometimes given short shrift in contemporary textbooks.

Guthrie's book inspired the class notes that evolved into the current text. While substantially different, it is my hope that the current text retains the clarity, accessibility, and applicability of Guthrie's text. I also want today's students to better understand, as those who used Guthrie's text surely did, the empirical dimensions of contemporary economic problems and how quantitative methods can increase our understanding of them. Indeed, to put a twist on an old economic adage, theory without measurement can lead us to conclusions that have little basis in reality.

Margaret Lewis
College of Saint Benedict/Saint John's University
St. Joseph/Collegeville, Minnesota

December 2010

试读结束：需要全本请在线购买： www.ertongbook.com

A note on the textbook's focus and scope

This textbook is designed for undergraduate students of economics who have completed an introductory course in statistics. While we review key findings from probability theory, sampling theory, and probability distribution functions in Part III, the purpose is to apply these concepts to hypothesis testing of sample means, correlation analysis, and regression analysis. Students without the prerequisite statistics course may find this discussion too brief and require additional study in basic statistics.

The economic questions raised throughout the text are based on the assumption that students have completed courses in micro- and macroeconomic theory, and that they have acquired some sophistication in applying economic theory to economic issues. Since it is the goal for students to learn these methods so they can conduct empirical analysis, this background is crucial. The text is thus pitched to students in the second half of their undergraduate education in economics.

The text also contains detailed instructions at the end of each chapter for using Microsoft Excel 2007/2010 to apply the quantitative methods described. While those familiar with dedicated statistical software may find Excel's tools less efficient, the widespread use and availability of Excel recommends its use here.

As economists know, economic statistics are governed by detailed definitions that can vary significantly across nations. Thus the text uses data primarily from the United States (reflecting the author's background), although the economic variables measured here are available for most economies. For those unfamiliar with a data series, the text will include information about who collects the data and how the data is constructed. Those wanting additional information can access the agency responsible, where detailed handbooks of methods and other information can be found.

A note on the "Where can I find. . .?" boxes

These boxes, found throughout the text, provide information about several popular locations for online economic data. Online accessibility has greatly facilitated empirical economic analysis, much as the development of computing power and the availability of statistical software has transformed such work.

The growing and widespread availability of economic data necessarily means much more information is available than can be recorded in this book. While each data source is current as of its writing, it is possible that some details, or even the web address itself, may have changed. (Readers are directed to use a search engine if forwarding links are not provided.)

Information is also increasingly available through the agencies that generate the data. It is recommended that students first visit the agency responsible for a particular data series to see if it is available and easily downloadable from that source.

Contents

List of figures xii
List of tables xiv
List of charts xviii
List of boxes xxi
Preface xxiii
A note on the textbook's focus and scope xxv
A note on the "Where can I find...?" boxes xxvi

1 The role of statistics in economics 1

1.1 Understanding the economy using empirical evidence 1
1.2 Measuring economic welfare 2
1.3 Distribution of household income 2
1.4 Assessing full employment 4
1.5 Calculating economic growth 6
1.6 Measuring inflation 6
1.7 Theoretical relationships between unemployment and economic growth 9
1.8 The connection between economic theory and statistical evidence 11
Summary 11
Concepts introduced 11
Exercises 20

2 Visual presentations of economic data 22

2.1 Economic graphs and charts 22
2.2 Time-series data and charts 23
2.3 Cross-section data and charts 27
2.4 Panel data and charts 38
2.5 Creating effective charts 38
2.6 Constructing charts using Excel 40
Summary 41

Concepts introduced 42
Exercises 55

PART I
Descriptive statistics of an economic variable **57**

3 Observations and frequency distributions 59

3.1 The design of observations: an introduction 59
3.2 Attributes and measured variables 60
3.3 Organizing data: absolute frequency distributions 63
3.4 Organizing data: relative frequency distributions 64
3.5 Visual presentations of frequency distribution: histograms 65
3.6 Classes in a frequency distribution 65
3.7 Constructing a frequency distribution 68
3.8 Frequency polygons 72
3.9 Cumulative frequency distributions and ogives 75
Summary 80
Concepts introduced 80
Exercises 85

4 Measures of central tendency 89

4.1 Desirable properties for descriptive statistics 89
4.2 Three measures of central tendency: mean, median, and mode 90
4.3 Measures of central tendency for frequency distributions 94
4.4 Weighted arithmetic means 97
4.5 Geometric means 100
4.6 Positional measures for ungrouped data 104
4.7 Positional measures for a frequency distribution 106
Summary 108
Concepts introduced 109
Exercises 111

5 Measures of dispersion 115

5.1 The concept of dispersion 115
5.2 Populations and samples 115
5.3 Range 117
5.4 Interquartile range 117
5.5 Average deviation 118
5.6 The concept of the standard deviation 119
5.7 Calculating the standard deviation for ungrouped data 121
5.8 Calculating measures of dispersion for frequency distributions 122
5.9 Locating extreme values 123
5.10 The shape of frequency distributions 125
5.11 Choosing the appropriate descriptive statistics 126
5.12 Assessing relative dispersion: coefficient of variation 128

5.13 Assessing relative dispersion: index of dispersion 132
5.14 Depicting relative dispersion: the Lorenz curve 133
5.15 Assessing relative dispersion: Gini coefficient of inequality 141
Summary 144
Concepts introduced 146
Exercises 152

PART II
Temporal descriptive statistics **155**

6 Measuring changes in price and quantity 157

6.1 Important index numbers in empirical economics 157
6.2 Why economists use index numbers 157
6.3 Constructing a simple price index 159
6.4 Constructing a weighted price index 161
6.5 Selecting appropriate weights for an index number 163
6.6 Chained price indices 166
6.7 Price index applications 169
6.8 Shifting an index's reference period 172
6.9 Quantity indices 175
6.10 Composite indices 176
Summary 179
Concepts introduced 179
Exercises 189

7 Descriptions of stability: short-run changes 192

7.1 Measuring economic change over time 192
7.2 Calculating percentage growth 192
7.3 Compound growth 196
7.4 Annualized growth rates from sub-annual rates 198
7.5 Annualized growth rates from supra-annual rates 201
7.6 Continuous compound growth 202
7.7 Continuously compounded annual growth and logarithms 205
Summary 208
Concepts introduced 208
Exercises 214

8 Patterns of long-term change 217

8.1 Economic growth over time 217
8.2 Constant long-run rates of growth 217
8.3 Growth by constant amounts 222
8.4 Change over time by constant rates or by constant amounts? 223
8.5 A complete model for describing change 228
8.6 Seasonal effects 228
8.7 Cyclical effects 230

8.8 Irregular effects 235
Summary 235
Concepts introduced 237
Exercises 240

PART III
Statistical inferences about a single variable **243**

9 Basic concepts in statistical inference 245

9.1 Populations and samples revisited 245
9.2 Sampling procedures 246
9.3 Concepts of probability 247
9.4 Probability distributions 248
9.5 Continuous probability distributions: the normal distribution 249
9.6 Continuous probability distributions: standard normal distribution 251
9.7 Identifying a normal distribution 255
9.8 The concept of the sampling distribution of means 256
9.9 Sampling distribution of means and the Central Limit Theorem 261
9.10 Sampling distribution of the Z-statistic 261
Summary 262
Concepts introduced 262
Exercises 264

10 Statistical estimation 266

10.1 Sample surveys as a source of data 266
10.2 Interval estimates of the population mean when the variance of the population is known 267
10.3 Confidence levels and the precision of an interval estimate 270
10.4 The t-distribution 272
10.5 Confidence intervals for the population mean when the variance of the population is not known 274
10.6 Confidence intervals for proportions, percentages, and rates 276
10.7 Confidence intervals for differences between means and proportions 277
Summary 279
Concepts introduced 280
Exercises 285

11 Statistical hypothesis testing of a mean 287

11.1 Testing hypotheses in economics: an analogy to criminal trials 287
11.2 An overview of hypothesis testing in economics: evaluating truth in advertising 288
11.3 Economic hypothesis testing: stating the hypotheses 291
11.4 Economic hypothesis testing: selecting the level of significance 293
11.5 Economic hypothesis testing: establishing the decision rule 294

试读结束：需要全本请在线购买： www.ertongbook.com

11.6 Economic hypothesis testing: constructing the test statistic and making a decision about the null hypothesis 294
11.7 Testing hypotheses versus estimating confidence intervals 295
11.8 Evaluating a statistical rule in terms of a Type I error 296
11.9 Evaluating a statistical rule in terms of a Type II error 297
11.10 The p-value and hypothesis testing 298
Summary 301
Concepts introduced 301
Exercises 302

PART IV
Relationships between two variables **305**

12 Correlation analysis 307

12.1 Statistical relationships between two variables 307
12.2 Correlation analysis: descriptive statistics 308
12.3 Testing the significance of the correlation coefficient 310
12.4 Testing the sign on the correlation coefficient 314
Summary 315
Concepts introduced 315
Exercises 320

13 Simple linear regression analysis: descriptive measures 326

13.1 Introduction to simple linear regression analysis 326
13.2 The algebra of linear relationships for regression analysis 327
13.3 Simple linear regression analysis: education and GDP 328
13.4 The algebra of variations in linear regression relationships 330
13.5 The coefficient of determination in regression analysis 333
13.6 Simple linear regression analysis: infant mortality rates and skilled health personnel at birth 334
Summary 337
Concepts introduced 337
Exercises 338

14 Simple regression analysis: statistical inference 339

14.1 The need for statistical inference in regression analysis 339
14.2 Testing hypotheses about the GDP–education regression model's slope coefficient 339
14.3 Sampling distributions of the linear regression's slope and intercept coefficients 342
14.4 Establishing the null and alternative hypotheses for the slope coefficient 344
14.5 Levels of significance and decision rules 347
14.6 The test statistic and p-value for the slope coefficient 348

14.7 An example of a hypothesis about the slope coefficient: infant mortality rates 348
14.8 What hypothesis testing does and does not prove 350
Summary 350
Concepts introduced 350
Exercises 354

15 Simple regression analysis: variable scales and functional forms 355

15.1 Rescaling variables and interpreting the regression coefficients 355
15.2 Specifying the regression equation: functional forms 358
15.3 Semi-log functional forms: the log–lin model 359
15.4 Semi-log functional forms: the lin–log model 361
15.5 Double-log functional form 362
15.6 Other functional forms 363
15.7 Selecting the appropriate functional form 364
Summary 365
Concepts introduced 366
Exercises 366

PART V

Relationships between multiple variables **369**

16 Multiple regression analysis: estimation and interpretation 371

16.1 Simple to multiple regression analysis: an introduction 371
16.2 The multiple-variable linear regression model 371
16.3 Specifying the independent variables and functional form for the multiple regression model 372
16.4 Specifying a multiple regression model for infant mortality rates 373
16.5 Estimating a multiple regression model for infant mortality rates 375
Summary 377
Concepts introduced 377
Exercises 380

17 Multiple regression analysis: hypothesis tests for partial regression coefficients and overall goodness of fit 382

17.1 General procedures for testing the significance of the partial regression coefficients (b_i) 382
17.2 Testing the significance of partial regression coefficients (b_i): IMR model 384
17.3 Evaluating the overall goodness of fit 387
17.4 Interpreting the infant mortality model's overall goodness of fit 388

17.5 Testing joint hypotheses about the regression model's overall significance: basic concepts and procedures 389
17.6 Testing joint hypothesis about the IMR multiple regression model's overall significance 393
17.7 Is each independent variable statistically related to the dependent variable? 393
17.8 The "best" statistical multiple regression model 394
17.9 The "better" statistical infant mortality model? 396
17.10 Closing observations about the IMR multiple regression results 400
Summary 400
Concepts introduced 401
Exercises 401

18 Multiple regression analysis: dummy variables and statistical problems 404

18.1 Independent dummy variables: basic concepts and considerations 404
18.2 Independent dummy variables: infant mortality rates and sub-Saharan Africa 408
18.3 Multicollinearity 410
18.4 Model misspecification and omitted variable bias 413
18.5 "Misbehaved" regression residuals: heteroskedasticity and serial correlation 416
Summary 420
Concepts introduced 421
Exercises 421

Notes 425
Bibliography 433
Index 439

List of figures

2.1 Theoretical graph of the U.S. market for coffee 22
3.1 Scales of measurement for economic variables 62
4.1 Desirable properties for descriptive statistics 89
4.2 Three measures of central tendency by desirable statistical properties 93
4.3 Positional locators of variables for ungrouped data 105
5.1 Algebraic definitions and symbolic notation for variations, variances, and standard deviations 120
5.2 Three measures of dispersion by desirable statistical properties 121
5.3 Shapes of frequency distributions 126
5.4 Steps for constructing a Lorenz curve 141
5.5 Lorenz curve and the Gini coefficient 142
5.6 Steps for calculating the Gini coefficient 143
6.1 Formulas for quantity indices 175
11.1 Types of null and alternative hypothesis tests, where μ_0 is the hypothesized value of the population parameter 292
11.2 Consequences in hypothesis testing 298
12.1 Range and strength of the correlation coefficient 309
12.2 Establishing hypotheses about the correlation coefficient 311
12.3 Setting the significance level and establishing the decision rule 311
12.4 The correlation coefficient test statistic and decisions regarding the null hypothesis 313
12.5 Determining the correlation coefficient's p-value 314
14.1 Desirable statistical properties of the OLS population parameters 344
14.2 Possible hypothesis tests for the simple regression slope coefficient 345
14.3 Decision rules for testing hypotheses about the simple linear regression model's slope coefficient 347
15.1 Effects of rescaling the dependent and independent variables by a constant (c) 358
15.2 Interpreting the slope coefficient for four functional forms of a simple regression model 364
15.3 ESS statistics for the linear, semi-log, and double-log functional forms for a simple regression model 365
17.1 Possible hypothesis tests and corresponding decision rules for partial regression coefficients 383
17.2 Possible outcomes of hypothesis tests for partial regression coefficients 392

17.3 Possible outcomes of hypothesis tests for the overall significance of the multiple regression model 395
17.4 Statistical tests and possible outcomes for the multiple linear regression model 395
18.1 Earnings model with intercept dummy variable for gender 405
18.2 Independent binary intercept dummy variables and independent quantitative variables 407

List of tables

1.1 Percentage distribution of total money income across U.S. households, 2008 3
2.1 Distribution of U.S. education services employees, 2009 37
2.2 Excel chart types and economic charts 41
3.1 Absolute frequency distribution of U.S. owner-occupied housing units, 2000 64
3.2 Relative frequency distribution of U.S. owner-occupied housing units, 2000 65
3.3 Relative frequency distribution of 2004 monthly housing costs from AHS Metropolitan Sample 70
3.4 Relative frequency distribution of monthly housing costs from 2004 AHS Metropolitan Sample with class intervals defined by the U.S. Census Bureau 72
3.5 Relative and relative cumulative frequency distributions of persons over 25 years of age by educational attainment and sex in June 2009 75
3.6 U.S. males 25 years and over by last year of school completed and race, June 2009 76
3.7 Relative and cumulative relative frequency distributions for U.S. owner-occupied housing units, 2000 78
4.1 Duration of unemployment for ten hypothetical unemployed persons 91
4.2 Measures of central tendency for monthly housing costs, 2004 Metropolitan Sample, American Housing Survey 94
4.3 Size of U.S. families, 2008 95
4.4 Percentage distribution of total money income across U.S. households, 2008 96
4.5 Calculation of mean consumer expenditures on necessities by size of consumer unit, 2007 98
4.6 Expenditures on household operations in four Midwestern cities, 2006–2007 99
4.7 Annual percentage change of real gross domestic product (billions of chained (2005) dollars) 101
4.8 Annual interest rate schedule for hypothetical ARM 103
4.9 Percentage distribution of income across U.S. households, 2008 107
4.10 Quintile calculations for the percentage distribution of U.S. household units by income classes, 2008 108
5.1 Annual incomes in two hypothetical occupations 116

试读结束：需要全本请在线购买： www.ertongbook.com