Using Excel
For Principles of Econometrics, Fourth Edition
Genevieve Briand dedicates this work to Tom Trulove

Carter Hill dedicates this work to Todd and Peter

To order books or for customer service call 1-800-CALL-WILEY (225-5945).

Copyright © 2011 John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA 01923, (978)750-8400, fax (978)646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201)748-6011, fax (201)748-6008.

ISBN 978-0-470-58581-8
Preface

This book is a supplement to Principles of Econometrics, 4th Edition by R. Carter Hill, William E. Griffiths and Guay C. Lim (Wiley, 2011). This book is not a substitute for the textbook, nor is it a stand alone computer manual. It is a companion to the textbook, showing how to perform the examples in the textbook using Excel 2007. This book will be useful to students taking econometrics, as well as their instructors, and others who wish to use Excel for econometric analysis.

In addition to this computer manual for Excel, there are similar manuals and support for the software packages EViews, Gretl, Shazam, and Stata. In addition, all the data for Principles of Econometrics, 4th in various formats, including Excel, are available at http://www.wiley.com/college/hill. Individual data files, as well as errata for this manual and the textbook, can also be found at http://principlesofeconometrics.com.

The chapters in this book parallel the chapters in Principles of Econometrics, 4th. Thus, if you seek help for the examples in Chapter 11 of the textbook, check Chapter 11 in this book. However within a Chapter the sections numbers in Principles of Econometrics, 4th do not necessarily correspond to the Excel manual sections.

This work is a revision of Using Excel 2007 for Principles of Econometrics, 3rd Edition by Genevieve Briand and R. Carter Hill (Wiley, 2010). Genevieve Briand is the corresponding author.

We welcome comments on this book, and suggestions for improvement.*

Genevieve Briand
School of Economic Sciences
Washington State University
Pullman, WA 99164
gbriand@wsu.edu

R. Carter Hill
Economics Department
Louisiana State University
Baton Rouge, LA 70803
eohill@lsu.edu

* Microsoft product screen shot(s) reprinted with permission from Microsoft Corporation. Our use does not directly or indirectly imply Microsoft sponsorship, affiliation, or endorsement.
BRIEF CONTENTS

1. Introduction to Excel 1
2. The Simple Linear Regression Model 19
3. Interval Estimation and Hypothesis Testing 67
4. Prediction, Goodness-of-Fit and Modeling Issues 95
5. The Multiple Linear Regression 143
6. Further Inference in the Multiple Regression Model 154
7. Using Indicator Variables 180
8. Heteroskedasticity 204
9. Regression with Time Series Data: Stationary Variables 228
10. Random Regressors and Moment-Based Estimation 262
11. Simultaneous Equations Models 278
12. Nonstationary Time-Series Data and Cointegration 294
13. Vector Error Correction and Vector Autoregressive Models 310
14. Time-Varying Volatility and ARCH Models 328
15. Panel Data Models 355
16. Qualitative and Limited Dependent Variable Models 391
   A. Mathematical Tools 402
   B. Review of Probability Concepts 416
   C. Review of Statistical Inference 431
Index 466
CONTENTS

CHAPTER 1 Introduction to Excel  1
1.1 Starting Excel  1
1.2 Entering Data  3
1.3 Using Excel for Calculations  3
  1.3.1 Arithmetic Operations  3
  1.3.2 Mathematical Functions  4
1.4 Editing your Data  6
1.5 Saving and Printing your Data  8
1.6 Importing Data into Excel  10
  1.6.1 Resources for Economists on the Internet  10
  1.6.2 Data Files for Principles of Econometrics  13
  1.6.2a John Wiley & Sons Website  13
  1.6.2b Principles of Econometrics Website  14
  1.6.3 Importing ASCII Files  14

CHAPTER 2 The Simple Linear Regression Model  19
2.1 Plotting the Food Expenditure Data  19
  2.1.1 Using Chart Tools  21
  2.1.2 Editing the Graph  23
    2.1.2a Editing the Vertical Axis  23
    2.1.2b Axis Titles  24
    2.1.2c Gridlines and Markers  25
    2.1.2d Moving the Chart  26
2.2 Estimating a Simple Regression  27
  2.2.1 Using Least Squares Estimators’ Formulas  27
  2.2.2 Using Excel Regression Analysis Routine  31
2.3 Plotting a Simple Regression  34
  2.3.1 Using Two Points  34
  2.3.2 Using Excel Built-in Feature  38
  2.3.3 Using a Regression Option  38
  2.3.4 Editing the Chart  40
2.4 Expected Values of $b_1$ and $b_2$  44

2.4.1 Model Assumptions  45
2.4.2 Random Number Generation  47
2.4.3 The LINEST Function  49
2.4.4 Repeated Sampling  50
2.5 Variance and Covariance of $b_1$ and $b_2$  52

CHAPTER 3 Interval Estimation and Hypothesis Testing  67
3.1 Interval Estimation  68
  3.1.1 The t-Distribution  68
    3.1.1a The t-Distribution versus Normal Distribution  68
    3.1.1b t-Critical Values and Interval Estimates  69
  3.1.1c Percentile Values  69
    3.1.1d TINV Function  69
    3.1.1e Appendix E: Table 2 in POE  71
3.1.2 Obtaining Interval Estimates  71
3.1.3 An Illustration  71

2.6 Nonlinear Relationships  53
  2.6.1 A Quadratic Model  53
    2.6.1a Estimating the Model  53
    2.6.1b Scatter Plot of Data with Fitted Quadratic Relationship  55
  2.6.2 A Log-Linear Model  57
    2.6.2a Histograms of $PRICE$ and ln($PRICE$)  57
    2.6.2b Estimating the Model  61
    2.6.2c Scatter Plot of Data with Fitted Log-Linear Relationship  62

2.7 Regression with Indicator Variables  63
  2.7.1 Histograms of House Prices  63
  2.7.2 Estimating the Model  65
3.1.3a Using the Interval Estimator Formula 71
3.1.3b Excel Regression Default Output 73
3.1.3c Excel Regression Confidence Level Option 74
3.1.4 The Repeated Sampling Context (Advanced Material) 75
3.1.4a Model Assumptions 75
3.1.4b Repeated Random Sampling 75
3.1.4c The LINES Function Revisited 77
3.1.4d The Simulation Template 78
3.1.4e The IF Function 79
3.1.4f The OR Function 79
3.1.4g The COUNTIF Function 80
3.2 Hypothesis Tests 81
3.2.1 One-Tail Tests with Alternative “Greater Than” (>) 81
3.2.2 One-Tail Tests with Alternative “Less Than” (<) 82
3.2.3 Two-Tail Tests with Alternative “Not Equal To” (≠) 82
3.3 Examples of Hypothesis Tests 82
3.3.1 Right-Tail Tests 83
3.3.2 Left-Tail Tests 84
3.3.3 Two-Tail Tests 86
3.3.4 The p-Value 88
3.4.1 The p-Value Rule 88
3.4.1a Definition of p-value 88
3.4.1b Justification for the p-Value Rule 89
3.4.2 The TDIST Function 91
3.4.3 Examples of Hypothesis Tests Revisited 92
3.4.3a Right-Tail Test from Section 3.3.1b 92
3.4.3b Left-Tail Test from Section 3.3.2 92
3.4.3c Two-Tail Test from Section 3.3.3a 93
3.4.3d Two-Tail Test from Section 3.3.3b 93

CHAPTER 4 Prediction, Goodness-of-Fit and Modeling Issues 95
4.1 Least Squares Prediction 96
4.2 Measuring Goodness-of-Fit 98
4.2.1 Coefficient of Determination or $R^2$ 98
4.2.2 Correlation Analysis and $R^2$ 98
4.2.3 The Food Expenditure Example and the CORREL Function 99
4.3 The Effects of Scaling the Data 100
4.3.1 Changing the Scale of x 100
4.3.2 Changing the Scale of y 101
4.3.3 Changing the Scale of x and y 102
4.4 A Linear-Log Food Expenditure Model 104
4.4.1 Estimating the Model 104
4.4.2 Scatter Plot of Data with Fitted Linear-Log Relationship 105
4.5 Using Diagnostic Residual Plots 108
4.5.1 Random Residual Pattern 108
4.5.2 Heteroskedastic Residual Pattern 111
4.5.3 Detecting Model Specification Errors 112
4.6 Are the Regression Errors Normally Distributed? 115
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6.1</td>
<td>Histogram of the Residuals</td>
<td>115</td>
</tr>
<tr>
<td>4.6.2</td>
<td>The Jarque-Bera Test for Normality using the CHINV and CHIDIST Functions</td>
<td>118</td>
</tr>
<tr>
<td>4.6.3</td>
<td>The Jarque-Bera Test for Normality for the Linear-Log Food Expenditure Model</td>
<td>121</td>
</tr>
<tr>
<td>4.7</td>
<td>Polynomial Models: An Empirical Example</td>
<td>122</td>
</tr>
<tr>
<td>4.7.1</td>
<td>Scatter Plot of Wheat Yield over Time</td>
<td>123</td>
</tr>
<tr>
<td>4.7.2</td>
<td>The Linear Equation Model</td>
<td>125</td>
</tr>
<tr>
<td>4.7.2a</td>
<td>Estimating the Model</td>
<td>125</td>
</tr>
<tr>
<td>4.7.2b</td>
<td>Residuals Plot</td>
<td>126</td>
</tr>
<tr>
<td>4.7.3</td>
<td>The Cubic Equation Model</td>
<td>126</td>
</tr>
<tr>
<td>4.7.3a</td>
<td>Estimating the Model</td>
<td>126</td>
</tr>
<tr>
<td>4.7.3b</td>
<td>Residuals Plot</td>
<td>128</td>
</tr>
<tr>
<td>4.8</td>
<td>Log-Linear Models</td>
<td>129</td>
</tr>
<tr>
<td>4.8.1</td>
<td>A Growth Model</td>
<td>129</td>
</tr>
<tr>
<td>4.8.2</td>
<td>A Wage Equation</td>
<td>130</td>
</tr>
<tr>
<td>4.8.3</td>
<td>Prediction</td>
<td>132</td>
</tr>
<tr>
<td>4.8.4</td>
<td>A Generalized $R^2$ Measure</td>
<td>135</td>
</tr>
<tr>
<td>4.6.5</td>
<td>Prediction Intervals</td>
<td>136</td>
</tr>
<tr>
<td>4.9</td>
<td>A Log-Log Model: Poultry Demand Equation</td>
<td>139</td>
</tr>
<tr>
<td>4.9.1</td>
<td>Estimating the Model</td>
<td>139</td>
</tr>
<tr>
<td>4.9.2</td>
<td>A Generalized $R^2$ Measure</td>
<td>140</td>
</tr>
<tr>
<td>4.9.3</td>
<td>Scatter Plot of Data with Fitted Log-Log Relationship</td>
<td>140</td>
</tr>
</tbody>
</table>

**CHAPTER 5  The Multiple Linear Regression**

4.1 Least Squares Estimates Using the Hamburger Chain Data 143
4.2 Interval Estimation 145
4.3 Hypothesis Tests for a Single Coefficient 145
4.3.1 Tests of Significance 145
4.3.2 One-Tail Tests 146

**CHAPTER 6  Further Inference in the Multiple Regression Model**

6.1 Testing the Effect of Advertising: the $F$-test 154
6.1.1 The Logic of the Test 154
6.1.2 The Unrestricted and Restricted Models 155
6.1.3 Test Template 158
6.2 Testing the Significance of the Model 159
6.2.1 Null and Alternative Hypotheses 159
6.2.2 Test Template 159
6.2.3 Excel Regression Output 160
6.3 The Relationship between $t$- and $F$-Tests 161
6.4 Testing Some Economic Hypotheses 163
6.4.1 The Optimal Level of Advertising 163
6.4.2 The Optimal Level of Advertising and Price 164
6.5 The Use of Nonsample Information 166
6.6 Model Specification 167
6.6.1 Omitted Variables 167
6.6.2 Irrelevant Variables 169
6.6.3 The RESET Test 172
6.7 Poor Data, Collinearity and Insignificance 176
6.7.1 Correlation Matrix 176
6.7.2 The Car Mileage Model Example 177
CHAPTER 7 Using Indicator Variables 180
7.1 Indicator Variables: The University Effect on House Prices Example 180
7.2 Applying Indicator Variables 182
7.2.1 Interactions Between Qualitative Factors 182
7.2.2 Qualitative Factors with Several Categories 185
7.2.3 Testing the Equivalence of Two Regressions 187
7.3 Log-Linear Models: a Wage Equation Example 191
7.4 The Linear Probability Model: A Marketing Example 192
7.5 The Difference Estimator: The Project STAR Example 193
7.6 The Differences-in-Differences Estimator: The Effect of Minimum Wage Change Example 198

CHAPTER 8 Heteroskedasticity 204
8.1 The Nature of Heteroskedasticity 204
8.2 Detecting Heteroskedasticity 206
8.2.1 Residual Plots 206
8.2.2 Lagrange Multiplier Tests 206
8.2.2a Using the Lagrange Multiplier or Breusch-Pagan Test 206
8.2.2b Using the White Test 209
8.2.3 The Goldfeld-Quandt Test 210
8.2.3a The Logic of the Test 210
8.2.3b Test Template 211
8.2.3c Wage Equation Example 212
8.2.3d Food Expenditure Example 216
8.3 Heteroskedasticity-Consistent Standard Errors or the White Standard Errors 219
8.4 Generalized Least Squares: Known Form of Variance 221
8.4.1 Variance Proportional to x: Food Expenditure Example 221
8.4.2 Grouped Data: Wage Equation Example 222
8.4.2a Separate Wage Equations for Metropolitan and Rural Areas 222
8.4.2b GLS Wage Equation 223

8.5 Generalized Least Squares: Unknown Form of Variance 224

CHAPTER 9 Regressions with Time Series Data: Stationary Variables 228
9.1 Finite Distributed Lags 228
9.1.1 US Economic Time Series 228
9.1.2 An Example: The Okun’s Law 230
9.2 Serial Correlation 232
9.2.1 Serial Correlation in Output Growth 232
9.2.1a Scatter Diagram for $G_t$ and $G_{t+1}$ 232
9.2.1b Correlogram for $G$ 233
9.2.2 Serially Correlated Errors 237
9.2.2a Australian Economic Time Series 237
9.2.2b A Phillips Curve 239
9.2.2c Correlogram for Residuals 240
9.3 Lagrange Multiplier Tests for Serially Correlated Errors 241
9.3.1 $t$-Test Version 241
9.3.2 $T \times R^2$ Version 243
9.4 Estimation with Serially Correlated Errors 245
9.4.1 Generalized Least Squares Estimation of an AR(1) Error Model 245
9.4.1a The Prais-Winsten Estimator 245
9.4.1b The Cochrane-Orcutt Estimator 248
9.4.2 Autoregressive Distributed Lag (ARDL) Model 252
9.5 Forecasting 254
   9.5.1 Using an Autoregressive (AR) Model 254
   9.5.2 Using an Exponential Smoothing Model 257
9.6 Multiplier Analysis 258

CHAPTER 10 Random Regressors and Moment-Based Estimation 262
10.1 OLS Estimation of a Wage Equation 262
10.2 Instrumental Variables Estimation of the Wage Equation 264
   10.2.1 With a Single Instrument 264
      10.2.1a First Stage Equation for EDUC 264
      10.2.1b Stage 2 Least Squares Estimates 265
   10.2.2 With a Surplus Instrument 268
      10.2.2a First Stage Equation for EDUC 268
      10.2.2b Stage 2 Least Squares Estimates 270
10.3 Specification Tests for the Wage Equation 273
   10.3.1 The Hausman Test 273
   10.3.2 Testing Surplus Moment Conditions 274

CHAPTER 11 Simultaneous Equations Models 278
11.1 Supply and Demand Model for Truffles 278
   11.1.1 The Reduced Form Equations 279
      11.1.1a Reduced Form Equation for Q 279
      11.1.1b Reduced Form Equation for P 280
   11.1.2 The Structural Equations or Stage 2 Least Squares Estimates 281

CHAPTER 12 Nonstationary Time-Series Data and Cointegration 294
12.1 Stationary and Nonstationary Variables 294
   12.1.1 US Economic Time Series 294
   12.1.2 Simulated Data 296
12.2 Spurious Regressions 299
12.3 Unit Root Tests for Stationarity 301
12.4 Cointegration 306

CHAPTER 13 Vector Error Correction and Vector Autoregressive Models 310
13.1 Estimating a VEC Model 310
   13.1.1 Test for Cointegration 312
   13.1.2 The VEC Model 315
13.2 Estimating a VAR Model 317
   13.2.1 Test for Cointegration 318
   13.2.2 The VAR Model 321
13.3 Impulse Responses Functions 323
   13.3.1 The Univariate Case 323
   13.3.2 The Bivariate Case 325
### CHAPTER 14 Time-Varying Volatility and ARCH Models 328

14.1 Time-Varying Volatility 328  
14.1.1 Returns Data 328  
14.1.2 Simulated Data 334  
14.2 Testing and Forecasting 341  
14.2.1 Testing for ARCH Effects 341  
14.2.1a Time Series and Histogram 342  
14.2.1b Lagrange Multiplier Test 344  
14.2.2 Forecasting Volatility 347  
14.3 Extensions 349  
14.3.1 The GARCH Model 349  
14.3.2 The T-GARCH Model 350  
14.3.3 The GARCH-In-Mean Model 352

### CHAPTER 15 Panel Data Models 355

15.1 Pooled Least Squares Estimates of Wage Equation 355  
15.2 The Fixed Effects Model 357  
15.2.1 Estimates of Wage Equation for Small $N$ 357  
15.2.1a The Least Squares Dummy Variable Estimator for Small $N$ 357  
15.2.1b The Fixed Effects Estimator: Estimates of Wage Equation for $N = 10$ 361  
15.2.2 Fixed Effects Estimates of Wage Equation from Complete Panel 365  
15.3 The Random Effects Model 371  
15.3.1 Testing for Random Effects 371  
15.3.2 Random Effects Estimation of the Wage Equation 373  
15.4 Sets of Regression Equations 381  
15.4.1 Estimation: Equal Coefficients, Equal Error Variances 381  
15.4.2 Estimation: Different Coefficients, Equal Error Variances 383  
15.4.3 Estimation: Different Coefficients, Different Error Variances 384  
15.4.4 Seemingly Unrelated Regressions: Testing for Contemporaneous Correlation 388

### CHAPTER 16 Qualitative and Limited Dependent Variable Models 391

16.1 Least Squares Fitted Linear Probability Model 391  
16.2 Limited Dependent Variables 393  
16.2.1 Censored Data 393  
16.2.2 Simulated Data 395

### APPENDIX A Mathematical Tools 402

A.1 Mathematical Operations 402  
A.1.1 Exponents 408  
A.1.2 Scientific Notation 409  
A.1.3 Logarithm and the Number $e$ 410  
A.2 Percentages 413

### APPENDIX B Review of Probability Concepts 416

B.1 Binomial Probabilities 416  
B.1.1 Computing Binomial Probabilities Directly 417  
B.1.2 Computing Binomial Probabilities Using BINOMDIST 419  
B.2 The Normal Distributions 422  
B.2.1 The STANDARDIZE Function 422  
B.2.2 The NORMSDIST Function 423  
B.2.3 The NORMSINV Function 423  
B.2.4 The NORMDIST Function 424  
B.2.5 The NORMINV Function 424  
B.2.6 A Template for Normal Distribution Probability Calculations 424
B.3 Distributions Related to the Normal
   B.3.1 The Chi-Square Distribution 426
   B.3.2 The t-Distribution 428
   B.3.3 The F-Distribution 429

APPENDIX C Review of Statistical Inference 431
C.1 Examining a Sample of Data 431
C.2 Estimating Population Parameters 436
   C.2.1 Creating Random Samples 436
   C.2.2 Estimating a Population Mean 438
   C.2.3 Estimating a Population Variance 438
   C.2.4 Standard Error of the Sample Mean 439
C.3 The Central Limit Theorem 439
C.4 Interval Estimation 444
   C.4.1 Interval Estimation with $\sigma^2$ unknown 446
   C.4.2 Interval Estimation with the Hip Data 447
C.5 Hypothesis Tests About a Population Mean 449
   C.5.1 An Example 450
   C.5.2 The $p$-value 450
   C.5.3 A Template for Hypothesis Tests 451
C.6 Other Useful Tests 454
   C.6.1 Simulating Data 454
   C.6.2 Testing a Population Variance 456
   C.6.3 Testing Two Population Means 459
   C.6.4 Testing Two Population Variances 461
C.7 Testing Population Normality 463
   C.7.1 A Histogram 463
   C.7.2 The Jacque-Bera Test 465

Index 467
CHAPTER 1

Introduction to Excel

CHAPTER OUTLINE
1.1 Starting Excel
1.2 Entering Data
1.3 Using Excel for Calculations
  1.3.1 Arithmetic Operations
  1.3.2 Mathematical Functions
1.4 Editing your Data
1.5 Saving and Printing your Data
1.6 Importing Data into Excel
  1.6.1 Resources for Economists on the Internet
  1.6.2 Data Files for Principles of Econometrics
    1.6.2a John Wiley & Sons Website
    1.6.2b Principles of Econometrics Website
  1.6.3 Importing ASCII Files

1.1 STARTING EXCEL

Find the Excel shortcut on your desktop. Double click on it to start Excel (left clicks).

Alternatively, left-click the Start menu at the bottom left corner of your computer screen.

Slide your mouse over All programs, Microsoft Office, and finally Microsoft Office Excel 2007. Left-click on this last one to start Excel—or better yet, if you would like to create a shortcut, right-click on it; slide your mouse over Send to, and then select (i.e. drag your mouse over and left-click on) Desktop (create shortcut). An Excel 2007 short-cut is created on your desktop. If you right-click on your shortcut and select Rename, you can also type in a shorter name like Excel.
Excel opens to a new file, titled Book1. You can find the name of the open file on the very top of the Excel window, on the **Title bar**. An Excel file like Book1 contains several sheets. By default, Excel opens to Sheet1 of Book1. You can figure out which sheet is open by looking at the **Sheet tabs** found in the lower left corner of your Excel window.

There are lots of little bits that you will become more familiar with as we go along. The **Active cell** is surrounded by a border and is in Column A and Row 1; its **Cell reference** is A1.

Below the title bar is a **Tab list**. The **Home tab** is the one Excel opens to. Under each tab you will find groups of commands. Under the home tab, the first one is the **Clipboard group of commands**, named after the tasks it relates to. The wide bar including the tab list and the groups of commands is referred to as the **Ribbon**. The content of the **Active cell** shows up in the **Formula bar** (right now, there is nothing in it). Perhaps the most important of all of this is to locate the **Help button** on the upper right corner of the Excel window. Finally, you can use the **Scroll bars** and the arrows around them to navigate up-down and right-left in your worksheet. And you have a long way to go: each worksheet in Microsoft Excel 2007 contains 1,048,576 rows and 16,384 columns!!!!

Note that your **Ribbon** might look slightly different than the one shown above. If your screen is bigger, Excel will automatically display more of its available options. For example, in the **Styles** group of command, instead of the **Cell styles** button, you might have a colorful display of cell styles.
1.2 ENTERING DATA

We will use Excel to analyze data. To enter labels and data into an Excel worksheet move the cursor to a cell and type. First type X in cell A1. Press the Enter key on your keyboard to get to cell A2 or navigate by moving the cursor with the mouse, or use the Arrow keys (to move right, left, up or down). Fill in the rest as shown below:

1.3 USING EXCEL FOR CALCULATIONS

What is Excel good for? Its primary usefulness is to carry out repeated calculations. We can add, subtract, multiply and divide; and we can apply mathematical and statistical functions to the data in our worksheet. To illustrate, we are going to compute the squares of the numbers we just entered and then add them up. There are two main ways to perform calculations in Excel. One is to write formulas using arithmetic operators; the other is to write formulas using mathematical functions.

1.3.1 Arithmetic Operations

Select the Excel Help button in the upper right corner of your screen. In the window of the Excel Help dialog box that pops up, type arithmetic operators and select Search. In the list of results, select Calculation operators and precedence.

Standard arithmetic operators are defined as shown below. To close the Excel help dialog box, select the X button found on its upper right corner.
4 Chapter 1

Place your cursor in cell B1, and type X-squared. In cells B2 through B6 below (henceforth referred to as B2:B6), we are going to compute the squares of the corresponding values from cells A2:A6. Let us emphasize that the trick to using Excel efficiently is NOT to re-type values already stored in the worksheet, but instead to use references of cells where the values are stored. So, to compute the square of 1, which is the value stored in cell A1, instead of using the formula =1*1, you should use the formula =A2*A2 or =A2^2. Place your cursor in cell B2 and type the formula.

Then press Enter. Note that: (1) a formula always starts with an equal sign; this is how Excel recognizes it is a formula, and (2) formulas are not case sensitive, so you could also have typed =a2^2 instead. Now, we want to copy this formula to cells B3:B6. To do that, place your cursor back into cell B2, and move it to the south-east corner of the cell, until the fat cross turns into a skinny one, as shown below:

Left-click, hold it, drag it down to the next four cells below, and release!

Excel has copied the formula you typed in cell B2 into the cells below. The way Excel understands the instructions you gave in cell B2 is “square the value found at the address A2”. Now, it is important to understand how Excel interprets “address A2”. To Excel “address A2” means “from where you are at, go left by one cell”—because this is where A2 is located vis-à-vis B2. In other words, an address gives directions: left-right, up-down, and distances: number of cells away—all in reference to the cell where the formula is entered. So, when we copied the formula we entered in cell B2, which instructed Excel to collect the value stored one-cell away from its left, and then square it—those exact same instructions were given in cells B3:B6. If you place your cursor back into B3, and look at the Formula bar, you can see that, in this cell, these same instructions translate into “=A3^2”.

1.3.2 Mathematical Functions

There are a large number of mathematical functions. Again, the list of functions available in Excel can be found by calling upon our good friend Help button and type Mathematical functions. If you try it, you will be able to see that the list is long. We will not copy it here.
We did compute the squares of the numbers we had. Now we will add them up—the numbers, and the squares of the numbers, separately. For that, we will be using the **SUM** function.

We first need to select or highlight all the numbers from our table. There are several ways to highlight cells. For this small area the easiest way is to place your cursor in A2, hold down the left mouse button and drag it across the area you wish to highlight—i.e. all the way to cell B6. Here is how your worksheet should look like:

Next, go to the **Editing** group of command, which is found in the extreme right of the **Home** tab, and select **Σ AutoSum**.

Excel sums the numbers from each column and places the sum in the bottom cell of each column. The result is:

Notice that if you select the arrow found to the right of **Σ AutoSum** you can find a list of additional calculations that Excel can automatically perform for you.

Alternatively, you could have placed your cursor in cell A7, typed `=SUM(A2:A6)`, and pressed the **Enter** key (and then copied this formula to cell B7).

Note that: (1) as soon as you type the first letter of your function, a list of all the other available functions that start with the same letter pops up. This can be very useful: if you left click on any of them, Excel gives you its definition; if you double left-click on any of them, it automatically finishes typing the function name for you, and (2) once the function name and the opening parenthesis are typed, Excel reminds you of what the needed **Arguments** are, i.e. what else you need to specify in your function to use it properly.
Now, you could also have used the **Insert function** button, which you can find on the left side of the **Formula bar**.

Once your cursor is placed in **A7**, select the **Insert function** button. An **Insert function** dialog box pops up. You can **Select a function** you need (highlight it, and select **OK**), or **Search for a function** first (follow the instructions given in that window).

In the **Function Arguments** dialog box that pops up, you need to specify the cell references of the values you want to add. If they are not already properly specified, you can type **A2:A6** in the **Number 1** window, or place your cursor in the window, delete whatever is in it, and then select **A2:A6**. Select **OK**. Now that you have the formula in **A7**, copy it into **B7**.

### 1.4 Editing your data

Before wrapping-up, you want to polish the presentation of your data. It actually has less to do with appearance than with organization and communication. You want to make sure that anyone can easily make sense of your table (like your instructor for example, or yourself for that matter—when you come back to it after you let it sit for a while).

We are going to add labels and color/shade to our table. Hold your cursor over cell **A** until it turns into an arrow-down; left-click to select the whole column; and select **Insert** in the **Cells** group of commands, found left to the **Editing** group of commands.

Excel adds a new column to the left of the one you selected. That’s where we are going to write our labels. In the new **A1** cell, type **Variables**; in cell **A2**, type **Values**; in cell **A7** type **Sum**.
Select column A again, make it **Bold** (Font group of commands, right to the Clipboard one), and align it **Left** (Alignment group of commands, right to the Font one).

Select cells B1 and C1, and make them **Bold**. Repeat with cells B7 and C7. Better, but not there yet. Select row 7, make it **Italic** (next to **Bold**). Select column B, hold your left-click and drag your mouse over cell C to select column C too; select **Center** alignment (next to Left). Next, select A2:A6; left-click the arrow next to **Merge & Center** (on the Alignment group of commands), and select **Merge cells**.

Immediately after, select **Middle Align**, which is found right above the **Center** alignment button.

Select A1:C7, left-click the arrow next to the **Bottom Border** button and select **All Borders**.

Select A7:C7 (A7:C7, not A1:C7 this time), left-click the arrow next to the **Fill Color** button, and select a grey color to fill in the cell with. Choose a different color for A1:C1.
Finally, put your cursor between cells C and D until it turns to a left and right arrow as shown here:

Hold it there and double left-click so that the width of column C gets resized to better accommodate the length of the label “X-squared”. The result is:

Next, drag your cursor over the Sheet1 tab, right-click, select Rename and type in a descriptive name for your worksheet like Excel for POE 1.2-1.4, for Using Excel for Principles of Econometrics, 4e—sections 1.2 through 1.4. Press the Enter key on your keyboard or left-click anywhere on your worksheet.

1.5 SAVING AND PRINTING YOUR DATA

All you need to do now is to save your Excel file. Select the Save button on the upper left corner of the Excel window.

A Save As dialog box pops up. Locate the folder you want to save your file in by using the arrow-down located at the extreme right of the Save in window or browsing through the list of folders displayed below it.
In the **File name** window, at the bottom of the **Save As** dialog box, the generic name Book1 should be outlined. Type the descriptive name you would like to give to your Excel file, like **POE Chapter 1**. Finally, select **Save**.

If you need to create a new folder, use the **Create New Folder** button found to the right of the **Save in** window.

A **New Folder** dialog box pops up; it is prompting you for the name you want to give to your new folder, **Excel for POE** for example. Type it in the **Name** window and select **OK**. Finally, select **Save**.

If you would like to print your table, select the **Office Button**, next to the **Save** button; go to **Print**, and select one of the print options.

For more print options, you might want to check out the **Page Layout** tab, on the upper left of your screen, as well as the **Page Layout** button on the bottom right of your screen.

To close your file, select the **X** button on the upper right corner of your screen.
In the next section, we show you how to import data into an Excel spreadsheet. Getting data for economic research is much easier today than it was years ago. Before the Internet, hours would be spent in libraries, looking for and copying data by hand. Now we have access to rich data sources which are a few clicks away.

First we will illustrate how convenient sites that make data available in Excel format can be. Then we illustrate how to import ASCII or, text files, into Excel.

1.6 IMPORTING DATA INTO EXCEL

1.6.1 Resources for Economists on the Internet

Suppose you are interested in analyzing the GDP of the United States. The website Resources for Economists contains a wide variety of data, and in particular the macro data we seek. Websites are continually updated and improved. We guide you through an example, but be prepared for differences from what we show here.

First, open up the website http://rfe.org/.

Select the Data link and then select U.S. Macro and Regional Data.
This will open up a range of sub-data categories. For the example discussed here, select the **Bureau of Economic Analysis (BEA)**.
Finally, select **Gross Domestic Product (GDP)**.

The result shows the point we are making. Many government and other web sites make data available in **Excel format**. Select **Current-dollar and “real” GDP**.

You have the option of saving the resulting Excel file to your computer or storage device, or opening it right away—which we proceed to do next.

What opens is a workbook with headers explaining the variables it contained. We see that there is a series of annual data and a quarterly series.
The opened file is “Read Only” so you must save it under another name to work with it, graph, run regressions and so on.

1.6.2 Data Files for Principles of Econometrics

The book *Principles of Econometrics, 4e*, uses many examples with data. These data files have been saved as workbooks and are available for you to download to your computer. There are about 150 such files. The data files and other supplementary materials can be downloaded from two web locations: the publisher website or the book website maintained by the authors.

1.6.2a John Wiley and Sons Website

Using your web browser, enter the address [www.wiley.com/college/hill](http://www.wiley.com/college/hill). Find, among the authors named “Hill”, the book *Principles of Econometrics, 4e*.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Current Dollar and “Real” Gross Domestic Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Annual</td>
<td>Quarterly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>GDP in billions of current dollars</td>
<td>GDP in billions of chained 2005 dollars</td>
<td>(Seasonally adjusted annual rates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1929</td>
<td>103.6</td>
<td>977.0</td>
<td>1947q1</td>
<td>237.2</td>
<td>1,722.2</td>
</tr>
<tr>
<td>7</td>
<td>1930</td>
<td>91.2</td>
<td>892.8</td>
<td>1947q2</td>
<td>240.4</td>
<td>1,769.5</td>
</tr>
<tr>
<td>8</td>
<td>1931</td>
<td>76.5</td>
<td>834.9</td>
<td>1947q3</td>
<td>244.5</td>
<td>1,760.8</td>
</tr>
<tr>
<td>9</td>
<td>1932</td>
<td>58.7</td>
<td>725.8</td>
<td>1947q4</td>
<td>254.3</td>
<td>1,794.8</td>
</tr>
<tr>
<td>10</td>
<td>1933</td>
<td>56.4</td>
<td>716.4</td>
<td>1948q1</td>
<td>260.3</td>
<td>1,823.4</td>
</tr>
</tbody>
</table>

Follow the link to Resources for Students, and then Student Companion Site. There, you will find links to supplement materials, including a link to Data Files that will allow you to download all the data definition files and data files at once.
14 Chapter 1

1.6.2b Principles of Econometrics Website

The address for the book website is www.principlesofeconometrics.com. There, you will find links to the Data definitions files, Excel spreadsheets, as well as an Errata list. You can download the data definition files and the Excel files all at once or select individual files. The data definition files contain variable names, variable definitions, and summary statistics. The Excel spreadsheets contain data only; those files were created using Excel 2003.

1.6.3 Importing ASCII Files

Sometimes data that you want to use may be provided but in ASCII or text format. To illustrate go to http://principlesofeconometrics.com. There you will find that one of the formats in which we provide data is ASCII or text files. These are used because they contain no formatting and can be used by almost every software once imported.

Select ASCII files and then go to the food data.
ASCII data files (*.dat) are text files containing only data.

Download all the *.dat files in (a) ZIP format or (b) a self-extracting EXE file (download and double-click)

Select individual *.dat files from the table below.

<table>
<thead>
<tr>
<th>airline</th>
<th>coke</th>
<th>gold</th>
<th>meat</th>
<th>profit</th>
<th>tax</th>
</tr>
</thead>
<tbody>
<tr>
<td>alcohol</td>
<td>coke2</td>
<td>golf</td>
<td>medical</td>
<td>pub</td>
<td>tax2</td>
</tr>
<tr>
<td>andy</td>
<td>commute</td>
<td>growth</td>
<td>metrics</td>
<td>pubexp</td>
<td>term</td>
</tr>
<tr>
<td>asparagus</td>
<td>computer</td>
<td>gunfeld</td>
<td>mexico</td>
<td>pm</td>
<td>texas</td>
</tr>
<tr>
<td>bangla</td>
<td>consumption</td>
<td>gunfeld2</td>
<td>mining</td>
<td>quizzes</td>
<td>theories</td>
</tr>
<tr>
<td>beer</td>
<td>cps</td>
<td>gunfeld3</td>
<td>money</td>
<td>returns</td>
<td>tobit</td>
</tr>
<tr>
<td>bond</td>
<td>cps_small</td>
<td>hitsurvey</td>
<td>manop</td>
<td>rice</td>
<td>tobitnc</td>
</tr>
<tr>
<td>br</td>
<td>cps1</td>
<td>hip</td>
<td>mroz</td>
<td>robbery</td>
<td>tooday</td>
</tr>
<tr>
<td>br2</td>
<td>cps2</td>
<td>house_state</td>
<td>music</td>
<td>salary</td>
<td>transport</td>
</tr>
<tr>
<td>broker</td>
<td>crime</td>
<td>housing</td>
<td>nals</td>
<td>sales</td>
<td>truffle</td>
</tr>
<tr>
<td>brun</td>
<td>csi</td>
<td>hwave</td>
<td>nals_small</td>
<td>savings</td>
<td>tuna</td>
</tr>
<tr>
<td>byd</td>
<td>demand</td>
<td>indpro</td>
<td>newbroiler</td>
<td>share</td>
<td>uk</td>
</tr>
<tr>
<td>canada</td>
<td>demo</td>
<td>inflation</td>
<td>nls</td>
<td>sheep</td>
<td>unit</td>
</tr>
<tr>
<td>cap2</td>
<td>edj_inc</td>
<td>ineur</td>
<td>nls_panel</td>
<td>ninings</td>
<td>usa</td>
</tr>
<tr>
<td>cars</td>
<td>euro</td>
<td>inreg1</td>
<td>nls_panel2</td>
<td>sp</td>
<td>utown</td>
</tr>
<tr>
<td>castle</td>
<td>exrace</td>
<td>inreg2</td>
<td>oil</td>
<td>spurious</td>
<td>vacan</td>
</tr>
<tr>
<td>ces</td>
<td>fair</td>
<td>jobs</td>
<td>olympics</td>
<td>sterling</td>
<td>vacation</td>
</tr>
<tr>
<td>cpass</td>
<td>figurer</td>
<td>kings</td>
<td>orang</td>
<td>stockton</td>
<td>var</td>
</tr>
<tr>
<td>ch19</td>
<td>flomba</td>
<td>learn</td>
<td>scar</td>
<td>stockton2</td>
<td>vec</td>
</tr>
<tr>
<td>chard</td>
<td>food</td>
<td>liquor</td>
<td>oz</td>
<td>stockton96</td>
<td>vote</td>
</tr>
<tr>
<td>cloth</td>
<td>fullwin</td>
<td>kent</td>
<td>phillips</td>
<td>surplus</td>
<td>vote2</td>
</tr>
</tbody>
</table>

Right-click on the file name. Select Save Target As. A Save As dialog box pops up. Locate the folder you want to save your file in by using the arrow-down located at the extreme right of the Save in window or browsing through the list of folders displayed below it. Finally, select Save.

Once the download of the file is completed, a Download complete window pops up. Choose Close.

Start Excel. Select the Office Button on the upper left corner of the Excel window, then Open.
Navigate to the location of the data file. Make sure you have selected **All Files** in the **Files of Type** window. Select your **food.dat** file and then select **Open**.

What begins is a Windows “Wizard” that will take you through 3 steps to import the data into Excel. Our ASCII data files are neatly lined up in columns with no commas or anything else separating the columns. Select **Fixed width**, and then **Next**.

In the next step the data are previewed. By clicking on the vertical black line you could adjust the column width, but there is no need most of the time. For neatly arrayed data like ours, Excel can determine where the columns end and begin. Select **Next** again.
In the third and final step Excel permits you to format each column, or in fact to skip a column. In our case you can simply select Finish.

This step concludes the process and now the data is in a worksheet named food.
Next, you need to save your food data in an Excel File format. To do that, select the Office Button, Save As, and finally Excel Workbook.

A Save As dialog box pops up. Locate the folder you want to save your file in by using the arrow-down located at the extreme right of the Save in window or browsing through the list of folders displayed below it.

Excel has automatically given a File name, food.xlsx, and specify the file format in the Save as type window, Excel Workbook (*.xlsx). All you need to do is select Save.

From this point you are ready to analyze the data.

This completes our introductory Chapter. The rest of this manual is designed to supplement your readings of Principles of Econometrics, 4e. We will walk you through the analysis of examples found in the text, using Excel 2007. We would like to be able to replicate most of the plots of data and tables of results found in your text.
In this chapter we estimate a simple linear regression model of weekly food expenditure. We also illustrate the concept of unbiased estimation. In the first section, we start by plotting the food expenditure data.

2.1 PLOTTING THE FOOD EXPENDITURE DATA

Open the Excel file food. Save it as POE Chapter 2.

Compare the values you have in your worksheet to the ones found in Table 2.1, p. 49 of Principles of Econometrics, 4e. The second part of Table 2.1 shows summary statistics. You can
compute and check on those by using Excel mathematical functions introduced in Chapter 1, if you would like.

Select the **Insert** tab located next to the **Home** tab. Select A2:B41. In the **Charts** groups of commands select **Scatter**, and then **Scatter with only Markers**.

The result is:

![Scatter chart example]

Each point on this **Scatter chart** illustrates one household for which we have recorded a pair of values: weekly food expenditure *and* weekly income. This is very important. We *chose Scatter chart* because we wanted to keep track of those pairs of values. For example, the point highlighted below illustrates the pair of values (187.05, 12.47) found in row 6 of your table.

When we select two columns of values to plot on a **Scatter chart**, Excel, by default, represents values from the *first column on the horizontal axis* and values from the *second column on the vertical axis*. So, in this case, the expenditure values are illustrated on the horizontal axis and income values on the vertical axis. Indeed, you can see that the scale of the values on the
horizontal axis corresponds to the one of the food expenditure values in column A, and the scale of the values on the vertical axis corresponds to the one of the income values in column B.

We actually would like to illustrate the food expenditure values on the vertical axis and the income values on the horizontal axis—opposite of what it is now. By convention, across disciplines, the variable we monitor the level of (the dependent variable) is illustrated on the vertical axis (Y-variable). And by convention, across disciplines, the variable that we think might explain the level of the dependent variable is illustrated on the horizontal axis (X-variable).

In our case, we think that the variation of levels of income across households might explain the variation of levels of food expenditure across those same households. That is why we would like to illustrate the food expenditure values on the vertical axis and the income values on the horizontal axis.

2.1.1 Using Chart Tools

If you look up on your screen, to the right end of your tab list, you should notice that Chart Tools are now displayed, adding the Design, Layout, and Format tabs to the list. The Design tab is open. (If, at any time, the Chart Tools and its tabs seem to disappear, all you need to do is to put your cursor anywhere in your Chart area, left-click, and they will be made available again.)

Go to the Data group of commands, to the left, and select the Select Data button.
A Select Data Source dialog box pops up. Select Edit.

In the Edit Series dialog box, highlight the text from the Series X values window. Press the Delete key on your keyboard. Select B2:B41. Highlight and delete the text from the Series Y values window. Select A2:A41. Select OK.

The Select Data Source dialog box reappears. Select OK again. You have just told Excel that income are the X-values, and food expenditure are the Y-values—not the other way around.

The result is:
2.1.2 Editing the Graph

Now, we would like to do some editing. We do not need a Legend, since we have only one data series. Our expenditure values do not go over 600, so we can restrict our vertical axis scale to that. We definitely would like to label our axes. We might want to get rid of our Gridlines, and change the Format of our data series. Finally, we would like to move our chart to a new worksheet.

Select the Layout tab. On the Labels group of commands, select Legend and None to delete the legend.

2.1.2a Editing the Vertical Axis

Select the Axes button on the Axes group of commands. Go to Primary Vertical Axis, and select More Primary Vertical Axis Options.

A Format Axis dialog box pops up. Change the Maximum value illustrated on the axis from Auto to Fixed, and specify 600.

Next select Alignment, and use the arrow-down in the Text direction window to select Rotate all text 270°.
Place your cursor on the upper blue border of your **Format Axis** dialog box.

Left-click, hold it, and drag the box over so you can see your chart; release. Look at the vertical axis of your chart.

The numbers are now displayed vertically instead of horizontally, but less of them are displayed as well:

We want to change that back.

Select **Axis Options** again. Change **Major unit** from **Auto** to **Fixed**, and specify **100**. Select **Close**.

### 2.1.2b Axis Titles

Back to the **Labels** group of commands; select **Axis Titles**, go to **Primary Horizontal Axis Title**, and select **Title Below Axis**.
Select the generic Axis Title in the bottom of your chart and type in $x = \text{weekly income in }$ 100.

Go back to Axis Titles, then to Primary Vertical Axis Title this time. Select Rotated Title.

Select the generic Axis Title on the left of your chart and press Delete, or put your cursor on top of the Axis Title box, left-click, and press the Backspace key to delete the generic Axis Title. Type in $y = \text{weekly food expenditure in }$.

2.1.2c Gridlines and Markers

Back to the Axes group of commands now. Select Gridlines. Go to Primary Horizontal Gridlines, and select None.

Change the Current Selection (group of commands to the far left) to Series 1 (use the arrow down button to the right of the window to make that selection). Select Format Selection.
A **Format Data Series** dialog box pops up. Select **Marker Options**. Change the **Marker Type** from **Automatic** to **Built-in**. Change the **Type** and the **Size** as shown below:

Next, select **Marker Fill**. Change it from **Automatic** to **Solid fill**. **Color** options pop up. Change the **Color** to black. Select **Marker Line Color**, and change it from **Automatic** to **No line**. Select **Close**.

The result is a replica of Figure 2.6 p. 50 in *Principles of Econometrics, 4e*: (if it looks like some of your dots are little flowers, left-click your cursor anywhere on your screen first)

### 2.1.2d Moving the Chart

Go back to the **Design** tab. (Remember if you don’t see your **Chart Tools** tabs, what you need to do is place your cursor in your chart area and left-click). Select the **Move Chart** button on the **Location** group of commands to the far right of your screen.
A Move Chart dialog box pops up. Select New sheet and give it a name like Figure 2.6. Select OK.

 Rename Sheet 1 Data (if needed, see Section 1.4 of this manual on how to do that).

We have plotted our data, and edited our chart. Next, we want to estimate the regression line that best fit the data, and add this line to the chart.

2.2 ESTIMATING A SIMPLE REGRESSION

In this section, we are going to use two different methods to obtain the least squares estimates of the intercept and slope parameters $\beta_1$ and $\beta_2$. Method 1 consists of plugging in values into the $b_1$ and $b_2$ least squares estimators’ formulas. Method 2 consists of making use of Excel built-in regression analysis routine.

2.2.1 Using Least Squares Estimators’ Formulas

The least squares estimators are:

$$b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (2.1)$$

$$b_1 = \bar{y} - b_2 \bar{x} \quad (2.2)$$

These formulas are telling us two things: (1) which values we need, and (2) how we need to combine them to compute $b_1$ and $b_2$.

(1) Which values do we need?

We need the $(x_i, y_i)$ pairs of values—they do appear explicitly in equation (2.1). We also need $\bar{x}$ and $\bar{y}$, which are the sample means, or simple arithmetic averages of the $x_i$ values and $y_i$ values—those averages appear both in equation (2.1) and equation (2.2). Note that the subscript $i$ in $x_i$ and $y_i$ keeps count of the $x$ and $y$ values. In other words, $i$ denotes the $i$th value or $i$th pair of values. Also, $\bar{x}$ and $\bar{y}$, are referred to as “x-bar” and “y-bar”.


(2) How do we combine those values?

Equation (2.1): \[ b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \]

The numerator is the sum of products; \( \Sigma \) is the Greek capital letter “sigma” which denotes sum. The first term of each product is the deviation of an \( x \) value from its mean \( (x_i - \bar{x}) \). The second term of each product is the deviation of the corresponding \( y \) value from its mean \( (y_i - \bar{y}) \). The products are computed for each \( (x_i, y_i) \) pair of values before they are added together.

The denominator is the sum of the squared deviations from the mean, for the \( x \) values only. In other words, each \( x \) value deviation from its mean is first squared, and then all those squared deviations values are summed.

Equation (2.2): \[ b_1 = \bar{y} - b_2 \bar{x} \]

This equation tells us to multiply \( b_2 \) by \( \bar{x} \), and then subtract this product from \( \bar{y} \). Note that \( b_2 \) must be computed first—before \( b_1 \) can be computed.

There is actually no magic to this. We use the food expenditure and income values we have collected from our random sample of 40 households, and perform simple arithmetic operations to compute the estimates the intercept and slope coefficient of our regression line.

As for the computation of \( b_1 \) and \( b_2 \) itself, there is only one trick. We need to make sure we know which values are the \( x \)’s and which ones are the \( y \)’s. So, we are going to start by adding labels to our columns of data.

You should be in your Data worksheet. If not, you can go back to it by selecting its tab on the bottom of your screen.

Select row 2 and insert a new row (see Section 1.4 of this manual if you need help on that). In the new cell A2, type \( y \); and in the new cell B2, type \( x \). Right-align A1:B2.

Next, we need to lay out the frame of the table where we are going to store our intermediate and final computations. Type \( x_{\text{bar}} \) in cell D2, \( y_{\text{bar}} \) in cell D3, \( b_2 \) in cell D6, and \( b_1 \) in cell D7. In cell G2:J2, type \( \text{x_deviation} \), \( \text{y_deviation} \), \( (\text{x_dev})(\text{y_dev}) \), and \( (\text{x_deviation})^2 \), respectively. (Note that you can use your Tab key, instead of moving your cursor or using the Arrow key, to move to the next cell to your right).
Below \texttt{x\_deviation} we are going to compute and store the deviations of the \(x\) values from their mean. Below \texttt{y\_deviation}, we are going to compute and store the deviations of the \(y\) values from their mean. Below \((x\_dev)(y\_dev)\), we are going to compute and store the products of the \(x\) deviation and the \(y\) deviation for each pair of values. Finally, below \((x\_deviation)^2\) we are going to compute and store the \(x\) deviations squared.

To show the 2 of \((x\_deviation)^2\) as a square, place your cursor in \texttt{J2}, if it is not already in it. Move to the \textbf{Formula bar} to select the 2, and select the arrow to the right corner of the \textbf{Font} group of commands.

A \textbf{Format cells} dialog box pops up. Select \textbf{Superscript} and then \textbf{OK}.

In cells \texttt{D6} and \texttt{D7} proceed to format the 2 and 1 of \(b_2\) and \(b_1\) as \textbf{Subscripts} instead. \textbf{Bold} all the labels you just typed, and \textbf{Align Right} the ones from \texttt{G2:J2}. Finally, resize the width of columns \texttt{G:J} to accommodate the width of its labels (see Section 1.4 of this manual if you need help on that).
Now, your worksheet should look like this one:

We have computed averages before. The formula you should have in cell E2 is `=AVERAGE(B3:B42)`, and the one in cell E3 is `=AVERAGE(A3:A42)`. Compare the averages you get to the sample means of Table 2.1 in *Principles of Econometrics, 4e* (p. 49); they should be the same.

Next, we want to compute the deviations. Think about what you are trying to compute. And then type the needed formulas in G3:J3.

You should type `=B3 – E2` in cell G3, `=A3 – E3` in cell H3, `=G3*H3` in cell I3, and `G23^2` in cell J3. Here are the values you should get:

Now, in cells G3 and H3, we gave cell references E2 and E3, where the averages are stored. Note that we will need to use those averages again, and get those averages from these same exact locations, to compute the deviations of the next 39 observations.

So, what we actually need to do is to transform these Relative cell references (E2 and E3) into Absolute cell references ($E2$ and $E3$). This will allow us to copy the formula from G3:H3 down below without losing track of the fact that the values for the averages are stored in cells E2 and E3.

A Relative cell reference is made into an Absolute cell reference by preceding both the row and column references by a dollar sign. Place your cursor back in cell G3 (i.e. move your mouse over and left-click); in the Formula bar, place your cursor before the E and insert a dollar sign (press the Shift-key and the $ key at the same time); move your cursor before the 2 and insert another dollar sign; place your cursor at the end of the formula and press Enter.
Go to cell H3, and add the needed dollar signs there too. Now, you can select G3:J3. Select Copy on the Clipboard group of command. Select G4:J42, and select Paste (next to Copy). You have just copied the formulas to compute the needed deviations for the rest of the \((x_i, y_i)\) pairs.

Your worksheet should look like this:

![Worksheet Image]

We have everything we need to finalize the computation of \(b_1\) and \(b_2\).

Place your cursor in cell E6, and again think about what you need to compute \(b_2\). Recall that the least squares estimators are:

\[
b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \tag{2.1}
\]

\[
b_1 = \bar{y} - b_2\bar{x} \tag{2.2}
\]

If you refer back to equation (2.1), you can see that \(=\text{SUM(I3:I42)/SUM(J3:J42)}\) is the formula you need in cell E6. The one you need in cell E7 is \(=E3 - E6*E2\) for equation (2.2).

Your worksheet should look like this:

![Worksheet Image]

In the table above we obtain the same exact least squares estimates as those reported on p. 53 of Principles of Econometrics, 4e.

That was Method 1 of obtaining the least squares estimates of the intercept and slope parameters \(\beta_1\) and \(\beta_2\). For Method 2, we are going to use the Excel built-in regression analysis routine.

### 2.2.2 Using Excel Regression Analysis Routine

Select the Data tab, in the middle of your tab list. On the Analysis group of commands to the far right of the ribbon, select Data Analysis.
Chapter 2

If the Data Analysis tool does not appear on the ribbon, you need to load it first.

Select the Office Button in the upper left corner of your screen, Excel Options on the bottom of the Office Button tasks panel, Add-Ins in the Excel Options dialog box, Excel Add-ins in the Manage window at the bottom of the Excel Options dialog box, and then Go.

In the Add-Ins dialog box, check the box in front of Analysis ToolPak. Select OK.

Now Data Analysis should be available on the Analysis group of commands. Select it.

A Data Analysis dialog box pops up. In it, select Regression (you might need to use the scroll up and down bar to the right of the Analysis Tools window to find it), then select OK.

The Regression dialog box that pops up next is very similar to the Edit Series box we encountered before (see Section 2.1.1). Place your cursor in the Input Y Range window, and select A3:A42 to specify the y-values you are working with. Similarly, place your cursor in the Input X Range window, and select B3:B42 to specify the x-values you are working with. Next, place your cursor in the New Worksheet Ply window and type Regression—this is going to be the name of the new worksheet where Excel regression analysis results are going to be stored. Select OK.
The Summary Output that Excel just generated should be highlighted as shown below:

![Excel Output](image)

Select the Home tab. In the Cells group of commands, select Format, and AutoFit Column Width; this is an alternative to adjust the width of the selected columns to fit their contents.
Your worksheet should now look like this:

The least squares estimates are given under the **Coefficients** column in the last table of the **Summary Output**. The estimate for the **Intercept** coefficient or \( b_1 \) is the first one; followed by the estimate of the slope coefficient (\( X \) variable 1 coefficient) or \( b_2 \). The summary output contains many other items that we will learn about shortly. For now, notice that the number of **observations** or pairs of values, 40, is given in cell B8.

A convenient way to report the values for \( b_1 \) and \( b_2 \) is to write out the equation of the estimated regression line:

\[
\hat{y}_i = 83.42 + 10.21x_i
\]  

(2.3)

Now that we have the equation of our straight line, we would like to graph it. This is what we are doing in the next section.

### 2.3 PLOTTING A SIMPLE REGRESSION

There are different ways to draw a regression line. One way is to plot two points and draw the line that passes through those two points—this is the method we are going to use first. Another way is plot many points, and then draw the line that passes through all those points—this is the method that Excel uses in its built-in features we are going to look at next.

#### 2.3.1 Using Two Points

When we draw a line by hand, on a piece of paper, using a pen and a ruler, we can use *any* two points. We can extend our line between the points, as well as beyond the points, up and down, or right and left. Excel does not use a ruler. Instead, it uses the coordinates of two points to draw a line, and it draws the line *only* between them. So, to have Excel draw a line that spans over the whole range of data we have, we need to choose those two points a little bit more strategically than usual.
If you look back at your scatter chart (Figure 2.6 worksheet) or back in your table (Data worksheet), you can see that our $x$ values range from about 0 to 35 (from 3.69 to 33.4 exactly). So, we choose our first point to have an $x$ value equal to 0, and our second point an $x$ value of 35.

The point with an $x$ value of zero is our $y$ intercept. It is the point where the line crosses the vertical axis. Its coordinates are $x = 0$ and $y = b_1$ or $(0, 83.42)$. This is our first point.

For our second point, we let $x = 35$; plug this $x$ value in equation (2.3), and compute its corresponding or predicted $y$ value. We obtain:

$$\hat{y} = 83.42 + 10.21(35) = 440.77$$

(2.4)

This is our second point, with coordinates (35, 440.77).

Go back to your Data worksheet (if you are not already there). In cell L1, type Points to graph regression line. In columns L and M we are going to record the coordinates of the two points we are using to draw our regression line. In cell L2, type $y$; in cell M2, type $x$. In cell M3, type 0; in cell M4, type 35. In cell L3, we actually want to record the value for our $y$ intercept or $b_1$, which we already have in cell E7. So, we are going to get it from there: in cell L3, type = E7, and press Enter. In cell L4, we want to have the computed predicted $y$ value from (2.4). So we type =E7+E6*M4, and press Enter. Note that instead of typing all those cell references, you can just move your cursor to the cells of interest as if you were actually getting the needed values—this is a very good way to avoid typing errors. So, you would type the equal sign, move your cursor to E7 and left-click to select it, type the plus sign, move your cursor to cell E6 and left-click to select it, type the asterisk, move your cursor to cell M4 and left-click to select it, and finally press Enter. Once you have done all of that, your worksheet should look like this:

![Worksheet Image]

Note that the predicted $y$ value we obtain in the worksheet for $x = 35$ is slightly different than the one we just computed in equation (2.4) due to rounding number differences.

Now, go back to your Figure 2.6 worksheet. The data we have plotted on the chart represent one set or series of data. The two new pairs of values we want to add to this chart represent a second set or series of data.

Select the Design tab, then the Select data button from the Data group of commands.
In the Legend Entries (Series) window of the Select data source dialog box, select the Add button.

Place your cursor in the Series X values window of the Edit series dialog box, and select M3:M4 in the Data worksheet. Place your cursor in the Series Y values window (delete whatever is in there), and select L3:L4 in the Data worksheet. Select OK.

The Select data source dialog box reappears. A second data series, Series2, was created from the selection you just specified. Select OK.

The two points from your new series are plotted on your chart (squares below):
Now, we need to draw a line across those two points. Go to the **Layout** tab. Change the **Current selection** (group of command to the far left) to **Series 2** (use the arrow down button to the right of the window to make that selection). Select **Format selection**.

A **Format data series** dialog box pops up. Select **Line color** and change its selection from **No line** to **Solid line**. Select **Close**.

The result is:

Note that while you *need* only two points to be able to draw a straight line, you *can use* more than two points. So we could have computed a predicted level of food expenditure for every level of income we have in our original data set, and use the $40 \,(x_i, \hat{y}_i)$ pairs of values as our data Series 2. This is actually what Excel does when it adds a **Linear Trend Line** to a **Scatter** chart or a **Line** of best **Fit** to **Plots** of data as part of the **Regression Analysis** routine.

We are going to delete the line and two points we just added to our graph and successively look at these other two ways to plot our regression line.
2.3.2 Using Excel Built-in Feature

In the Design tab, go back to the Data group of commands, and select the Select Data button. In the Select Data Source dialog box, select Series2 and Remove. Finally select OK.

To add a Linear Trend Line, select the Layout tab. Go to the Analysis group of commands, select Trendline, and then Linear Trendline.

Your chart should look like this (see also Figure 2.8 p. 54 in Principles of Econometrics, 4e):

2.3.3 Using a Regression Option

You can also have Excel add the Line that best Fit your data by choosing that option on the Regression dialog box.

Go back to your Data worksheet (bottom left corner of your screen).
Select the Data tab, located in the middle of your tab list. Select Data Analysis on the Analysis group of commands to the far right of the ribbon. Select Regression in the Data Analysis dialog box, and then OK.

In the Regression dialog box, proceed as you did before, except this time, name your worksheet Regression and Line, and check the box in front of Line Fit Plots. Select OK.

In addition to the Summary Output you now have a Residual Output table and a Chart in your new worksheet. The Residual Output table is only partially shown below, and shown after AutoFitting the Column Width (see Section 2.2.2 for more details on that).

The Predicted Y or \( \hat{y}_i \) values have been computed for all the original observed \( x_i \) values, similarly to the way we computed \( \hat{y} \) for \( x = 35 \) (see Section 2.3.1).

The least squares Residuals are defined as

\[
\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i
\]  

You can compare the Predicted Y and Residuals values reported in the Excel Residual Output to the ones reported in Table 2.3 of Principles of Econometrics, 4e (p. 66). They should be the same.
2.3.4 Editing the Chart

Now, the chart needs a little bit of editing. For one it looks like it is a Column chart as opposed to a Scatter one. The scales could be changed. Finally, Chart and Axis titles are not currently very helpful.

Place your cursor anywhere in the Chart area, and left-click, so that Chart Tools are made available to you again. Select the Design tab. Go to the far left group of commands, Type, and select Change Chart Type. In the Change Chart Type dialog box, select X Y (Scatter) chart, and then Scatters with only Markers. Finally, select OK.

The result is:

Now that we have the correct chart type, we would like to draw a line through all the Predicted Y points. Actually, since we are using those points to draw our regression line, what we want to show is only the line. So, we will use the points to draw the line, and then get rid of those big square points. This way our chart won’t be as busy.

On your chart, select the Predicted Y points with your cursor. Your cursor should turn into a fat cross as shown below:
Right-click and select **Format Data Series**. A **Format Data Series** dialog box pops up. Select **Line Color** and **Solid line**. Change the line color to something different from the Y points. Select **Marker Options**, and change the **Marker Type** from **Automatic** to **None**. Select **Close**.

The result is:

On your chart, select the **Legend** with your cursor, right-click and select **Delete**.

Change the **Chart** and **Axis titles** as you see fit. Below, we show you how you can change the **Chart title**. You can follow a similar process to change the **Axis titles**.

Place your cursor in the title area and left click.
Select the generic title.

![Image of chart](image)

Type in your new title.

You can select any of the titles and change the **Font** size by going back to the **Home** tab. Select what you need on the **Font** group of commands.

![Font size options](image)

You can reformat the y-axis (and/or the x-axis) by selecting it with your cursor, right-clicking and selecting **Format Axis**.

![Format Axis options](image)

If you proceed as you did before to edit your vertical axis (see Section 2.1.2a), you should obtain the following:

![Vertical axis formatting](image)

To resize the whole **Chart area**, put your cursor over its lower border until it turns into a double cross arrow as shown below.

![Chart area resize](image)
Left click, and it should turn into a skinny cross.

Hold it, and drag it down until you are satisfied with the way your chart looks.

You can delete the Gridlines by first selecting them, right-clicking and then selecting Delete.

You can also reformat the Data Series Y by selecting the points, right-clicking and selecting Format Data Series. Then proceed as you did before to change your markers’ options (see Section 2.1.2c).
Your result might be (see also Figure 2.8 p. 54 in *Principles of Econometrics, 4e*):

In this next section we illustrate the concept of unbiased estimators.

### 2.4 Expected Values of $b_1$ and $b_2$

To show that under the assumptions of the simple linear regression model, $E(b_1) = \beta_1$ and $E(b_2) = \beta_2$, we first put ourselves in a situation where we know our population and regression parameters (i.e. we know the truth). We then use the least squares regression technique to unveil the truth (which we already know). This allows us to check on the validity of the least squares regression technique, and specifically to check on the unbiasedness of the least squares estimators.
2.4.1 Model Assumptions

First, let us restate the assumptions of the simple linear regression model (see p. 45 of *Principles of Econometrics, 4e*):

- The mean value of $y$, for each value of $x$, is given by the linear regression function:
  \[ E(y|x) = \beta_1 + \beta_2 x \]  
  (2.6)

- For each value of $x$, the values of $y$ are distributed about their mean value, following probability distributions that all have the same variance:
  \[ \text{var}(y|x) = \sigma^2 \]  
  (2.7)

- The sample values of $y$ are all uncorrelated and have zero covariance, implying that there is no linear association among them:
  \[ \text{cov}(y_i, y_j) = 0 \]  
  (2.8)

- The variable $x$ is not random and must take at least two different values.

- *(optional)* The values of $y$ are normally distributed about their mean for each value of $x$:
  \[ y \sim N(\mu y|x, \sigma^2) \]  
  (2.9)

In the specific and simplified case we are considering in this section, half of our hypothetical population of three person households has a weekly income of $1000 (x = 10)$, and half of it has a weekly income of $2000 (x = 20)$. Because we are all mighty, we know the values of our population parameters, and consequently the values of our regression parameters. Let $\mu_{y|x=10} = 200$, $\mu_{y|x=20} = 300$, and $\text{var}(y|x = 10) = \text{var}(y|x = 20) = \sigma^2 = 2500$. This implies $\beta_1 = 100$ and $\beta_2 = 10$.

The probability distribution functions of weekly food expenditure, $y$, given an income level $x = 10$ and an income level $x = 20$, are assumed to be Normal. They look like this:
Chapter 2

The linear relationship between weekly food expenditure and weekly income looks like the following:

\[ E(y|x) = \beta_0 + \beta_1 x \]

Let us emphasize the difference between this section and Chapter 2 in *Principles of Econometrics, 4e*. In this section, we do *know* the truth. In other words, we have information regarding weekly food expenditure and weekly food income on *all* three person households that constitute our population. In Chapter 2 of *Principles of Econometrics, 4e*, like it is the case in real-life, you do not have that population information. You must thus rely *solely* on your random sample information to make inferences about your population.

Now, as an exercise, and as a way to prove the unbiasedness of the least squares estimators, we are going to use the least square regression technique to unveil the truth.

Insert a new worksheet in your workbook by selecting the **Insert Worksheet** tab at the bottom of your screen (or Press the **Shift** and **F11** keys). Name it **Simulation**.

We are going to draw a random sample of 40 households from our population. Half of the sample is drawn from the first type of households, with weekly income \( x = 10 \); and half of the sample is drawn from the second type of households, with weekly income \( x = 20 \).

Let us keep records of the level of weekly income for our 40 households in column **A** of our **Simulation** worksheet: in cell **A1**, type \( x \) and **Right-Align** it; in cells **A2:A21**, record the value 10; in cells **A22:A41**, record the value 20.
2.4.2 Random Number Generation

We use the Random Number Generation analysis tool to draw our random sample of households. We keep record of their weekly food expenditure in column B of our Simulation worksheet: type \( y \) in B1, and Right-Align it.

Select the Data tab, in the middle of your tab list. On the Analysis group of commands to the far right of the ribbon, select Data Analysis.

The Data Analysis dialog box pops up. In it, select Random Number Generation (you might need to use the scroll up and down bar to the right of the Analysis Tools window to find it), then select OK.

A Random Number Generation dialog box pops up. Since we are drawing one random sample, we specify 1 in the Number of Variables window. We first draw a random samples of 20 from
households with weekly income of \( x = 10 \), so we specify the **Number of Random Numbers** to be 20. For simplicity we assumed that our population of households has weekly food expenditure that is normally distributed, so this is the distribution we choose. Once you have selected **Normal** in the **Distribution** window, you will be able to specify its **Parameters**: for \( x = 10 \), its **Mean** is \( \mu_{y|x=10} = 200 \) and its **Standard deviation** is \( \sqrt{\text{var}(y|x = 10)} = \sigma = 50 \). Select the **Output Range** in the **Output options** section, and specify it to be B2:B21 in your **Simulation** worksheet. Finally, select **OK**.

Repeat to draw a random sample of 20 from households with weekly income of \( x = 20 \). Change the **Mean** to \( \mu_{y|x=10} = 300 \) and the **Output Range** to B22:B41.

Here is the random sample that we obtained. NOTE: you will obtain a *different* random sample, due to the nature of random sampling.
2.4.3 The LINEST Function

Next, we use the **LINEST** function to obtain the least squares estimates for the intercept and slope parameters, based on the random sample we just drew. The **LINEST** function is an alternative to using the Least Squares Estimators’ Formulas (see Section 2.2.1) or the Excel Regression Analysis Routine (see Section 2.2.2). It allows us to quickly get the least squares estimates for the intercept and slope parameters. For this purpose, the general syntax of the **LINEST** function is as follows:

\[
\text{= LINEST}(y's, x's)
\]

The first argument of the **LINEST** function specifies the \( y \) values, and the second argument specifies the \( x \) values, the least squares estimates are based on. In our case, we thus need to specify:

\[
\text{= LINEST(B2:B41,A2:A41)}
\]

The **LINEST** function creates a table where it stores the least squares estimates in Excel memory. It first reports the slope coefficient estimate, and then the intercept coefficient estimate. So, if we were to look into Excel memory, the estimates would be reported as shown below:

<table>
<thead>
<tr>
<th>column 1</th>
<th>column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_2 )</td>
<td>( b_1 )</td>
</tr>
</tbody>
</table>

We nest the **LINEST** function in the **INDEX** function to get the estimated coefficients, *one at a time*. The **INDEX** function returns values from within a table. In the case of a table with only one row, the **INDEX** function general syntax is as follows:

\[
\text{= INDEX(table of results, column_num)}
\]
The first argument of the INDEX function specifies which table to get the results from. In our case, this is the table of results generated by the LINEST function above. So, we replace “table of results” by “LINEST(B2:B41,A2:A41)”. The second argument indicates from which column of the table to retrieve the result of interest to us. So, if we want to retrieve the estimate of the intercept coefficient, \( b_1 \), from the table above, we would indicate that it can be found in column 2 by replacing “column_num” by “2”.

We are going to report our estimated coefficients at the bottom of our table. In cell A43, type \( b_1 = \); in cell A44, type \( b_2 = \). Bold those labels. In cell B43 and B44, type the following equations, respectively:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>( b_1 = ) INDEX(LINEST(B2:B41,A2:A41),2)</td>
</tr>
<tr>
<td>44</td>
<td>( b_2 = ) INDEX(LINEST(B2:B41,A2:A41),1)</td>
</tr>
</tbody>
</table>

Here are the estimates that we get:

The estimates of the intercept and slope coefficients are based on one random sample. Our random sample is different than yours, and each random sample yields different estimates, which may or may not be close to the true parameter values. The property of unbiasedness is about the average values of \( b_1 \) and \( b_2 \) if many samples of the same size are drawn from the same population. In the next section, we are thus going to repeat our sampling and least squares estimation exercise.

### 2.4.4 Repeated Sampling

Note that in Chapter 2 of Principles of Econometrics, 4e, the repeated samples given to you were randomly collected from a population with unknown parameters. In this section, we draw our samples from a population with known parameters.

Go back to the Random Number Generation dialog box. We would like to draw 9 additional random samples, so we specify 9 in the Number of Variables window. Again, we first draw random samples of 20 from households with weekly income of \( x = 10 \), so we specify the Number of Random Numbers to be 20. We also select Normal in the Distribution window, and specify its Parameters. For \( x = 10 \), its Mean is \( \mu_{y|x=10} = 200 \) and its Standard Deviation is \( \sqrt{\text{var}(y|x=10)} = \sigma = 50 \). Specify the Output Range to be C2:K21. Finally, select OK.
Repeat to draw a random sample of 20 from households with weekly income of $x = 20$. Change the Mean to $\mu_{y|x=10} = 300$ and the Output Range to C22:K41.

Next, before we copy the formula to get our coefficient estimates, we need to transform their Relative cell references A2:A41 into Absolute cell references $A$2:$A$41, since we will be using the same $x$-values for our next 9 rounds of least squares estimates.

Copy the formulas from B43:B44 into C43:K44. In cells L43:L44 compute the AVERAGEs of your estimates from your 10 samples. In cell L43, you should have =$AVERAGE(B43:K43)$; in cell L44, you should have =$AVERAGE(B44:K44)$. The estimates and average values that we get for our 10 samples are:

If we took the averages of estimates from many samples, these averages would approach the true parameter values $\beta_1$ and $\beta_2$. To show you that this is the case, we repeated the exercise again. Here are the average values of $b_1$ and $b_2$ that we did get as we increased the number of samples from 10, to 100, and finally to 1000:

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>10</th>
<th>100</th>
<th>1000</th>
<th>Parameter Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average value of $b_1$</td>
<td>89.14425</td>
<td>98.44593</td>
<td>99.48067</td>
<td>100</td>
</tr>
<tr>
<td>Average value of $b_2$</td>
<td>10.48296</td>
<td>10.08958</td>
<td>10.04135</td>
<td>10</td>
</tr>
</tbody>
</table>
The next section of this chapter is very short. It points out how you can compute an estimate of the variances and covariance of the least squares estimators $b_1$ and $b_2$ using Excel. It also outlines other numbers you can recognize in the Excel summary output. Note that for this section we are getting back to our food expenditure and income data of Sections 2.1-2.3, i.e. data from one sample of 40 households that was drawn from a population with unknown parameters.

### 2.5 VARIANCES AND COVARIANCE OF $b_1$ AND $b_2$

You can compute an estimate of the variances and covariance of the least squares estimators $b_1$ and $b_2$, the same way you computed $b_1$ and $b_2$. Consider their algebraic expressions (see below or p. 65 of *Principles of Econometrics, 4e*), and perform the simple arithmetic operations needed. You might want to do that as an exercise; you will be able to check on your work by comparing your estimates to the one reported on pp. 66-67 of *Principles of Econometrics, 4e*.

Estimates of the variances and covariance of the least squares estimators $b_1$ and $b_2$ are given by:

\[
\text{var}(b_1) = \hat{\sigma}^2 \left[ \frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} \right] \tag{2.10}
\]

\[
\text{var}(b_2) = \hat{\sigma}^2 \left[ \frac{1}{\sum (x_i - \bar{x})^2} \right] \tag{2.11}
\]

\[
\text{cov}(b_1, b_2) = \hat{\sigma}^2 \left[ \frac{-\bar{x}}{\sum (x_i - \bar{x})^2} \right] \tag{2.12}
\]

where: $N$ is the total number of pairs of values,

and 
\[
\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N-K} \text{ is an estimate of the error variance,} \tag{2.13}
\]

where: $K$ is the number of regression parameters, $K = 2$,

and 
\[
\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i \text{ are the least squares residuals.}
\]

The square roots of the estimated variances are the standard errors of $b_1$ and $b_2$. They are denoted as $se(b_1)$ and $se(b_2)$.

\[
se(b_1) = \sqrt{\text{var}(b_1)} \quad \text{and} \quad se(b_2) = \sqrt{\text{var}(b_2)} \tag{2.14}
\]

Excel regression routine does not automatically generate estimates of the variances and covariance of the least squares estimators $b_1$ and $b_2$, but it does compute the standard errors of $b_1$ and $b_2$, as well as other intermediary results.
Specifically, the following estimates can be found in the Excel **Summary Output** you generated earlier:

- \( \sum \hat{e}_i^2 \): Sum of Squared Residuals (SS Residual) in C13
- \( \hat{\sigma}^2 \): Mean Square Residual (MS Residual) in D13
- \( \hat{\sigma} \): Standard Error of the Regression in B7
- \( se(b_1) \) and \( se(b_2) \): Standard Errors of Intercept and X Variable 1 in C17:C18

Note that \( \sum \hat{e}_i^2 \), the **Sum of Squared Residuals (SS Residual)**, is also referred to as the **Sum of Squared Errors**—hence the abbreviation **SSE** used in p. 51 of *Principles of Econometrics, 4e*.

### 2.6 NONLINEAR RELATIONSHIPS

#### 2.6.1 A Quadratic Model

##### 2.6.1a Estimating the Model

Open the Excel file *br*. Excel opens the data set in Sheet 1 of a new Excel file. Since we would like to save all our work from Chapter 2 in one file, create a new worksheet in your POE Chapter 2 Excel file, name it **pr data**, and in it, copy the data set you just opened.

This data set contains data on 1080 houses sold in Baton Rouge, LA during mid-2005, which we are using to estimate the following quadratic model for house prices:

\[
PRICE = \alpha_1 + \alpha_2 SQFT^2 + e
\]  
(2.15)
In your br data worksheet, insert a column to the right of the sqft column B (see Section 1.4 for more details on how to do that). In your new cells C1:C2, enter the following column label and formula.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sqft$^2$</td>
</tr>
<tr>
<td>2</td>
<td>=B2^2</td>
</tr>
</tbody>
</table>

Copy the content of cells C2 to cells C3:C1081. Here is how your table should look (only the first five values are shown below):

In the Regression dialog box, the Input Y Range should be A2:A1081, and the Input X Range should be C2:C1081. Select New Worksheet Ply and name it Quadratic Model. Finally select OK.

The result is (matching the one reported on p. 70 in Principles of Econometrics, 4e):
2.6.1b Scatter of Data and Fitted Quadratic Relationship

Go back to your br data worksheet and select A2:B1081. Select the Insert tab located next to the Home tab. In the Charts group of commands select Scatter, and then Scatter with only Markers.

The result is:

You can see that our house price values are on the horizontal axis and square footage values are on the vertical axis; we would like to change that around and edit our chart as we did in Section 2.1 with our plot of food expenditure data. The result is (see also Figure 2.14 on p. 70 in Principles of Econometrics, 4e):

Finally, we add the fitted quadratic relationship to our scatter plot. In cells N1:N2 and O1:O3 of your br data worksheet, enter the following column label and formula.
Select cells O2:O3, move your cursor to the lower right corner of your selection until it turns into a skinny cross as shown below; left-click, hold it and drag it down to cell O22: Excel recognizes the series and automatically completes it for you. Next, copy the content of cell N2 to cells N3:N22. Here is how your table should look (only the first five values are shown below):

<table>
<thead>
<tr>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>quadratic price-hat sqft</td>
</tr>
<tr>
<td>2</td>
<td>=Quadratic Model'!$B$17+Quadratic Model'!$B$18*br data'!O2</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
</tr>
</tbody>
</table>

Go back to your scatter plot and right-click in the middle of your chart area. Select Select Data. In the Legend Entries (Series) window of the Select Data Source dialog box, select the Add button. In the Series name window, type Fitted Quadratic Relationship. Select O2:O22 for the Series X values and select N2:N22 for the Series Y values. Finally, select OK. The Fitted Quadratic Relationship series has been added to your graph.

Before you close the Select Data Source dialog box, select Series1 and Edit. Type the name Actual in the Series name window. Select OK. In the Select Data Source window that re-appears, select OK again.

Make sure you chart is selected so that the Chart Tools are visible. In the Layout tab, go to the Labels group of commands. Select the Legend button and choose either one of the Overlay
Legend options. Grab your legend with your cursor and move it to the upper left corner of your chart area.

Finally, we want to reformat our Fitted Quadratic Relationship values series. Select the plotted series in your chart area, right-click and select Format Data Series. A Format Data Series dialog box pops up. Select Line Color and Solid line. Change the line color to something different from the Actual series points. Select Marker Options, and change the Marker Type from Automatic to None. Select Close.

The result is (see also Figure 2.14 on p. 70 in Principles of Econometrics, 4e):

![Graph showing fitted quadratic relationship]

### 2.6.2 A Log-Linear Model

#### 2.6.2a Histograms of PRICE and ln(PRICE)

In your br data worksheet, insert a column to the right of the sqft\(^2\) column C (see Section 1.4 for more details on how to do that). In your new cells D1:D2, enter the following column label and formula.
Chapter 2

<table>
<thead>
<tr>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

Copy the content of cells D2 to cells D3:D1081. Here is how your table should look (only the first five values are shown below):

Next, we specify BIN values. These values will determine the range of PRICE and ln(PRICE) values for each column of the histogram. The bin values have to be given in ascending order. Starting with the lowest bin value, a PRICE or ln(PRICE) value will be counted in a particular bin if it is equal to or less than the bin value.

In cells S1:T3 of your br data worksheet, enter the following column labels and data.

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>price bin</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>50000</td>
</tr>
<tr>
<td></td>
<td>ln(price bin)</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>9.2</td>
</tr>
</tbody>
</table>

Select cells S2:S3, move your cursor to the lower right corner of your selection until it turns into a skinny cross as shown below; left-click, hold it and drag it down to cell S34: Excel recognizes the series and automatically completes it for you. Similarly, select cells T2:T3, move your cursor to the lower right corner of your selection until it turns into a skinny cross; left-click, hold it and drag it down to cell T29. Here is how your table should look (only the first five values are shown below):

Select the Data tab, in the middle of your tab list. On the Analysis group of commands to the far right of the ribbon, select Data Analysis.

The Data Analysis dialog box pops up. In it, select Histogram (you might need to use the scroll up and down bar to the right of the Analysis Tools window to find it), then select OK.
An **Histogram** dialog box pops up. For the **Input Range**, specify A2:A1081; for the **Bin Range**, specify S2:S34. The **Input Range** indicates the data set Excel will look at to determine how many values are counted in each bin of the **Bin Range**. Check the **New Worksheet Ply** option and name it **Price Histogram**; check the box next to **Chart Output**. Finally, select **OK**.

Select the columns in your chart area, right-click and select **Format Data Series**. The **Series Options** tab of the **Format Data Series** dialog box should be open. Select the **Gap Width** button and move it to the far left, towards **No Gap**.

Go to the **Border Color** tab and select **Solid line**, choose a different **Color** if you would like. Select **Close**.
Chapter 2

After editing our chart as we did in Section 2.1 with our plot of food expenditure data, the result is (see also Figure 2.16(a) on p. 72 in *Principles of Econometrics, 4e*):

Note that the frequencies given in the graph above are absolute ones, while the frequencies given in Figure 2.16(a) of *Principles of Econometrics, 4e* are relative ones.

Go back to your br data worksheet. In the **Histogram** dialog box, specify D2:D1081 for the **Input Range** and T2:T29 for the **Bin Range**. Check the **New Worksheet Ply** option and name it **lnPrice Histogram**; check the box next to **Chart Output**. Finally, select **OK**.

The final result is (see also Figure 2.16(b) on p. 72 in *Principles of Econometrics, 4e*):
Again, note that the frequencies given in the graph above are absolute ones, while the frequencies given in Figure 2.16(b) of *Principles of Econometrics, 4e* are relative ones.

### 2.6.2b Estimating the Model

We estimate the following log-linear model for house prices:

\[
\ln(PRICE) = \gamma_1 + \gamma_2 SQFT + e
\]  

(2.16)

In the Regression dialog box, the **Input Y Range** should be D2:D1081, and the **Input X Range** should be B2:B1081. Select **New Worksheet Ply** and name it **Log-Linear Model**. Finally select **OK**.

The result is (matching the one reported on p. 72 in *Principles of Econometrics, 4e*):
In cells Q1:Q2 of your br data worksheet, enter the following column label and formula.

```
Q
1 log-linear price-hat
2 =EXP('Log-Linear Model'!$B$17+'Log-Linear Model'!$B$18*'br data'!P2)
```

Next, copy the content of cells Q2 to cells Q3:Q22. Here is how your table should look (only the first five values are shown below):

Select your scatter plot of actual data points and fitted quadratic relationship and make a copy of it. Right-click in the middle of the copy of your chart. Select Select Data. In the Legend Entries (Series) window of the Select Data Source dialog box, select the Fitted Quadratic Relationship series, and then the Edit button. In the Series name window, replace the old name by Fitted Log-Linear Relationship. Select P2:P22 for the Series X values and select Q2:Q22 for the Series Y values. Finally, select OK, twice. The Fitted Log-Linear Relationship series has been added to your graph.
The result is (see also Figure 2.17 on p. 73 in *Principles of Econometrics, 4e*):

![Graph depicting house prices vs. total square feet](image)

2.7 REGRESSION WITH INDICATOR VARIABLES

2.7.1 Histograms of House Prices

Open the Excel file *utown*. Excel opens the data set in Sheet 1 of a new Excel file. Since we would like to save all our work from Chapter 2 in one file, create a new worksheet in your POE Chapter 2 Excel file, name it *utown data*, and in it, copy the data set you just opened.

This data file contains a sample of 1000 observations on house prices in two neighborhoods. One neighborhood is near a major university and called University Town. Another similar neighborhood, called Golden Oaks, is a few miles away from the university.

In cells H1:H3 of your *utown data* worksheet, enter the following column label and data.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>bin</td>
</tr>
<tr>
<td>1</td>
<td>125</td>
</tr>
<tr>
<td>2</td>
<td>137.5</td>
</tr>
</tbody>
</table>

Select cells H2:H3, move your cursor to the lower right corner of your selection until it turns into a skinny cross as shown below; left-click, hold it and drag it down to cell H20. Here is how your table should look (only the first five values are shown below):
In the **Histogram** dialog box, specify A2:A482 for the **Input Range** and H2:H20 for the **Bin Range**. Check the **New Worksheet Ply** option and name it **Golden Oaks Prices Histogram**; check the box next to **Chart Output**. Finally, select **OK**.

The final result is (see also Figure 2.18 on p. 74 in *Principles of Econometrics, 4e*):

Note that the frequencies given in the graph above are absolute ones, while the frequencies given in Figure 2.18 of *Principles of Econometrics, 4e* are relative ones.

Go back to your **utown data** worksheet. In the **Histogram** dialog box, specify A483:A1001 for the **Input Range** and H2:H20 for the **Bin Range**. Check the **New Worksheet Ply** option and name it **U Town Prices Histogram**; check the box next to **Chart Output**. Finally, select **OK**.
The final result is (see also Figure 2.18 on p. 74 in *Principles of Econometrics, 4e*):

![Histogram](image)

### 2.7.2 Estimating the Model

We estimate the following regression model for house prices

\[
PRICE = \beta_1 + \beta_2 U_{TOWN} + e
\]  

(2.17)

The indicator variable is

\[
U_{TOWN} = \begin{cases} 
0 & \text{house is in University Town} \\
1 & \text{house is in Golden Oaks} 
\end{cases}
\]  

(2.18)

Go back to your *utown data* worksheet.

In the Regression dialog box, the Input Y Range should be A2:A1001, and the Input X Range should be D2:D1001. Select New Worksheet Ply and name it Indicator Variable Model. Finally select OK.
The result is (matching the one reported on p. 75 in Principles of Econometrics, 4e):

![Regression Output](image)

This ends Chapter 2 of this manual. You might want to save your work before you close shop.
We hope you’ve enjoyed this complimentary preview of *Using Excel for Principles of Econometrics* by Genevieve Briand. To continue reading, please purchase the full manual:

Format: **Paperback**  
ISBN: 978-1-118-03210-7


Additional eText formats are available at [www.amazon.com](http://www.amazon.com).