



Munich Personal RePEc Archive

Random Variables, Their Properties, and Deviational Ellipses: In Map Point and Excel, v 5.0

Goodwin, Roger L

1 January 2015

Online at <https://mpra.ub.uni-muenchen.de/65329/>
MPRA Paper No. 65329, posted 29 Jun 2015 13:04 UTC

RANDOM VARIABLES, THEIR PROPERTIES, AND DEVIATIONAL ELLIPSES

RANDOM VARIABLES, THEIR PROPERTIES, AND DEVIATIONAL ELLIPSES

**In Map Point and Excel, v 5.0
(A Work in Progress)**

Roger L. Goodwin

Copyright ©2015 by Roger L. Goodwin, All rights reserved.

Published by Roger L Goodwin, West Virginia.

*To my parents and in
memory of Winston*

CONTENTS IN BRIEF

1	Introduction to VBA	1
2	Introduction to Map Point	13
3	Mathematics Review	23
4	Geo-Coding	41
5	Standard Deviatonal Ellipse	59
6	The Exponential Ellipse	83
7	The Weibull Ellipse	109
8	Spherical Statistics	139

CONTENTS

List of Figures	xiii
List of Tables	xix
Foreword	xxiii
Preface	xxv
Acknowledgments	xxix
Acronyms	xxxix
1 Introduction to VBA	1
1.1 The Development Environment	1
1.2 Variables	3
1.3 Arrays and Records	4
1.4 Branching	5
1.5 Loops	6
1.6 Setting Properties	7
1.7 The Excel Examples	8
2 Introduction to Map Point	13
	ix

x CONTENTS

2.1	User Data	14
2.2	Plotting the Mean Center	14
2.3	Visualizing the Survey Data	15
2.4	Drawing Axes	19
2.5	Drawing Ellipse Boundaries	19
2.6	Programmer Notes	21
3	Mathematics Review	23
3.1	Notation	23
3.2	Derivatives	24
3.3	Integrals	25
3.4	Quadratic Equation	26
3.5	Ellipse Equation	26
3.6	Trigonometry Functions and Conversion Factors	26
3.7	Modulo Arithmetic	27
3.8	Weighted Data	28
3.9	Area	30
3.10	Eccentricity	30
3.11	Axes Length	31
3.12	Rotation	32
3.13	Base 10 Logarithm	33
3.14	Stratification	33
	3.14.1 Latitude Estimates	34
	3.14.2 Longitude Estimates	35
3.15	Coordinate Systems	36
3.16	Testing for Randomness	37
3.17	Sampling without Replacement	38
4	Geo-Coding	41
4.1	Kentucky June Area Survey	43
4.2	US Crime Statistics	49
4.3	Airline Data	53
	4.3.1 Air France Flight 447 Data	53
	4.3.2 Malaysia Flight MH 370 Data	53
4.4	OECD Countries	54
5	Standard Deviation Ellipse	59
5.1	A Weighted Regression Model	60

5.2	Mean Latitude and Mean Longitude	61
5.3	Standard Deviational Ellipse	61
	5.3.1 Weighted Mean Center	62
5.4	Ellipse Properties	63
5.5	Testing for Randomness	65
5.6	Kentucky Example	66
5.7	Crime Example	67
5.8	Airline Examples	68
5.9	GDP Example	69
	5.9.1 Analyses A	69
	5.9.2 Analyses B	71
5.10	Exercises	74
5.11	VBA Code	75
6	The Exponential Ellipse	83
6.1	Mean Latitude	83
6.2	Mean Longitude	86
6.3	Joint Distribution	87
6.4	Exponential Ellipse	89
	6.4.1 Mean Center	89
	6.4.2 Ellipse	89
6.5	Axis Rotation	90
6.6	Ellipse Properties	90
6.7	Combining Strata Measures	91
6.8	Testing for Randomness	92
6.9	Kentucky Example	94
6.10	Crime Example	94
6.11	Airline Examples	96
6.12	GDP Example	97
6.13	Comparison to SDE	98
6.14	Exercises	101
6.15	VBA Code	104
7	The Weibull Ellipse	109
7.1	Latitude Mean Center	109
7.2	Longitude Mean Center	112
7.3	Joint Distribution	115
7.4	Weibull Ellipse	118

7.4.1	Ellipse Angle	118
7.5	Ellipse Properties	119
7.6	Combining Strata Measures	120
7.7	Kentucky Example	121
7.8	Crime Example	123
7.9	Airline Examples	124
7.10	GDP Example	125
7.11	Axes Length Comparison	126
7.12	Sample Distribution Fitting	126
7.13	Exercises	128
7.14	VBA Code	130
8	Spherical Statistics	139
8.1	Introduction	139
8.2	Concepts	141
8.3	<i>X</i> -Axis Mean Center and Resultant Length	142
8.4	<i>Y</i> and <i>Z</i> -Axis Mean Centers	143
8.4.1	Spherical Variance	143
8.5	Test and Characterizations	144
8.5.1	Finding the Eigenvalues	147
8.5.2	Finding the Eigenvectors	147
8.6	Kentucky Example	148
8.7	Crime Example	150
8.8	Airline Example	152
8.9	GDP Example	154
8.10	Exercises	157
8.11	VBA Code for Spherical Statistics	159
8.12	VBA Code for Numerical Techniques	162
A	Standard Deviatonal Ellipse VBA Driver	167
B	Exponential VBA Driver	173
C	Weibull VBA Driver	179
D	Newton and Gaussian Elimination Driver	187
E	References	193
	Index	199

LIST OF FIGURES

1.1	<i>This figure shows Excel without the Developer menu.</i>	2
1.2	<i>This figure shows Excel Options dialog box. Excel uses the dialog box for adding the Developer menu for VBA programming.</i>	2
1.3	<i>This figure shows Excel with the Developer menu at the top of the screen.</i>	3
1.4	<i>This figure shows the Excel workbook. A data worksheet at the bottom named KY 2003 has been circled. The worksheet named STATS has been circled. This application always saves the results to the STATS worksheet.</i>	10
1.5	<i>This figure shows the Excel spreadsheet with data. For this application to run properly, the first row can contain any column names the reader wishes. Column 1 must contain the latitude observations. Column 2 must contain the longitude observations. Column 3 must contain the random variable of interest. It is not advisable to put any observations in the remaining columns as this application may over-write them.</i>	11
2.1	<i>This figure shows the Map Point dialog box for finding a point on the Earth using the latitude and longitude.</i>	14

2.2	<i>This figure shows the Map Point dialog box for browsing Excel files, which have the coordinates and the random variable.</i>	15
2.3	<i>This figure shows the Map Point dialog box that lists the Excel spreadsheets in the Excel workbook.</i>	15
2.4	<i>This figure shows the Map Point dialog box containing the spreadsheet column names and the data.</i>	16
2.5	<i>This figure shows the Map Point Data mapping wizard.</i>	17
2.6	<i>This figure shows the Map Point choices for the pushpin colors.</i>	17
2.7	<i>This figure shows the Map Point choice for the column chart.</i>	17
2.8	<i>This figure shows the Map Point choices for scaling the data and change the label text.</i>	18
2.9	<i>This figure shows the Microsoft Word shapes for drawing an ellipse.</i>	19
2.10	<i>This figure shows the Microsoft Office dialog box for formatting a shape.</i>	20
3.1	<i>This figure shows the Options dialog box in Map Point. We use it to convert the latitude and longitude coordinates to decimal degrees.</i>	28
3.2	<i>This figure shows the Options dialog box in Google Earth. We use it to convert the latitude and longitude coordinates to decimal degrees.</i>	28
3.3	<i>This figure shows the Earth using Map Point. We identify points on the Earth using latitude and longitude.</i>	36
4.1	<i>This figure shows Google Earth. It shows how to geo-code a county with-in a state.</i>	42
4.2	<i>This figure shows the Excel spreadsheet of the Kentucky data. The spreadsheet has not been geo-coded. It still contains the regional data.</i>	44
4.3	<i>This figure shows the Excel spreadsheet of the crime data. The spreadsheet has not been geo-coded and still contains the regional data, two years of data in the same spreadsheet, merged columns, and percentages.</i>	49
4.4	<i>This figure shows the Excel spreadsheet of the Gross Domestic Product data. The spreadsheet has not been geo-coded.</i>	57

- 5.1 *This figure shows the Kentucky data from 2003 and the standard deviational ellipse. It shows a sketch of the standard deviational ellipse with Ohio County at the center. The semi-major axis is approximately 184 miles long and the semi-minor axis is approximately 52 miles long. We rotated the major axis roughly 70° from the Y-axis in an imaginary Cartesian coordinate system.* 66
- 5.2 *This figure shows the standard deviational ellipse of violent crime in the US for the year 2007. It also shows the values of the violent crime using a bar graph. The ellipse is rather large compared to the area in the survey.* 69
- 5.3 *This figure shows the plot of the mean center for the Air France Flight 447 data.* 70
- 5.4 *This figure shows the plot of the mean center for the Malaysia Flight MH 370 data.* 71
- 5.5 *This figure shows the plot of the mean centers of global GDP on a map. As the years progress, the mean center measurements move closer to the Chinese coastline.* 72
- 5.6 *This figure shows the plot of the mean centers of global GDP on a map. As the years progress, the mean center measurements move closer to the Greece coastline.* 73
- 6.1 *This figure shows the probability plot of the weighted latitude transformation. The exponential distribution fits the weighted latitude data since it shows a straight line.* 84
- 6.2 *This figure shows the probability plot of the weighted longitude transformation. The exponential distribution fits the absolute value of the weighted longitude data since it shows a straight line. It was necessary to take absolute values since longitudinal data is negative.* 84
- 6.3 *This figure shows sketches of both the standard deviational ellipse and the exponential deviational ellipse from the Kentucky data for the year 2003. The exponential ellipse is the smaller of the two ellipses. It covers one-third the area of the standard deviational ellipse, yet one-half of the same points.* 93
- 6.4 *This figure shows a raised bar graph of the weights in their respective counties for the Kentucky data for the year 2003. The majority of the weighted data comes from the lower-left corner of the state of Kentucky. The majority of the un-weighted data comes from the center of the state and the right-hand side.* 93

6.5	<i>This figure shows both the exponential deviational ellipse and the standard deviational ellipse for the violent crime data in the US for 2007. Both have the same center of gravity. The exponential ellipse is much smaller than the standard deviational ellipse.</i>	95
7.1	<i>This figure shows both the Weibull deviational ellipse for the Kentucky data for the year 2003. The zoom level is 200 miles above the earth. The standard deviational ellipse and the exponential deviational ellipse were too large to fit on the map. It has the same center of gravity as the other ellipses. The Weibull ellipse is much smaller than the other two.</i>	123
7.2	<i>This figure shows both the Weibull deviational ellipse for the Crime data for the year 2007. The zoom level is 550 miles above the earth. The standard deviational ellipse was too large to fit on the map. It has the same center of gravity as the other ellipses. The Weibull ellipse is much smaller than the other two.</i>	124
8.1	<i>This figure shows a map of the U.S. relative to the Equator (represented by the black line).</i>	141
8.2	<i>This figure shows the diagram of a triangle with the quantities for the longitude θ, n_0, and the resultant length R.</i>	146
8.3	<i>This figure shows the histogram for the Kentucky 2003 data.</i>	148
8.4	<i>This figure shows the Kentucky 2003 converted data.</i>	148
8.5	<i>This figure shows the histogram for the Violent Crime 2008 data. Three possible modes appear at California, Texas, and Florida.</i>	151
8.6	<i>This figure shows the 2008 US Crime converted data.</i>	151
8.7	<i>This figure shows the plotted, unweighted data for Malaysia Flight MH 370.</i>	153
8.8	<i>This figure shows the histogram for the GDP 2008 data. The U.S. accounts for approximately 25% of the total GDP and China accounts for approximately 14%.</i>	155
8.9	<i>This figure shows the 2008 Gross Domestic Product converted data.</i>	155
A.1	<i>In Excel, double click on the module named Ch4StandardDeviationalEllipse on the left-hand pane circled in red to display the VBA code for the Standard Deviational Ellipse. The subroutine called Standard_Dev_Ellipse() runs the programs to calculate the statistics in Chapter 5.</i>	169

A.2	<i>This figure shows the VBA environment for the module named Ch4StandardDeviationalEllipse. It also highlights the driver subroutine Standard_Dev_Ellipse().</i>	169
A.3	<i>This figure shows the STATS worksheet and the results from the Standard_Dev_Ellipse() subroutine.</i>	170
B.1	<i>This figure shows the VBA environment for the module Ch5ExponentialDeviationalEllipse. It also highlights the driver subroutine Exponential_Dev_Ellipse().</i>	174
B.2	<i>This figure shows the STATS worksheet and the results from the Exponential_Dev_Ellipse() subroutine.</i>	175
C.1	<i>This figure shows the VBA environment for the module Ch6Weibull Deviatonal Ellipse. It also highlights the driver subroutine Weibull_Dev_Ellipse().</i>	180
C.2	<i>This figure shows the STATS worksheet and the results from the Weibull() subroutine.</i>	182
D.1	<i>This figure shows the VBA environment for the module Ch7NewtonGaussRoutines. It also highlights the driver subroutine NewtonGauss().</i>	188

LIST OF TABLES

3.1	Values of the Eccentricity e	31
3.2	Stratification Definitions	34
3.3	Example of Estimation with Stratification on Latitude	35
3.4	Example of Estimation with Stratification on Longitude	35
4.1	Acreage Planted Data	46
4.2	Acreage Planted Data (con't)	47
4.3	Acreage Planted Data (con't)	48
4.4	Violent Crime Data	51
4.5	Violent Crime Data (con't)	52
4.6	Air France Flight 447 Data	53
4.7	Malaysia MH 370 Data	54
4.8	Gross Domestic Product for OECD Countries	56
5.1	2003 Results on Corn acreage in Kentucky	67

5.2	2004 Results on Corn acreage in Kentucky	67
5.3	2007 Results on Violent Crime in the U.S.	68
5.4	2008 Results on Violent Crime in the U.S.	68
5.5	Results on Air France Flight 447	69
5.6	Results on Malaysia Flight 370	70
5.7	2008 Results on Gross Domestic Product of OECD Countries	71
5.8	2009 Results on Gross Domestic Product of OECD Countries	72
5.9	2008 Results on Gross Domestic Product of OECD Countries	73
5.10	2009 Results on Gross Domestic Product of OECD Countries	74
6.1	2003 Results for the Exponential Model for Kentucky	94
6.2	2004 Results for the Exponential Model for Kentucky	94
6.3	2007 Results for the Exponential Model on Violent Crime in the U.S.	95
6.4	2008 Results for the Exponential Model on Violent Crime in the U.S.	96
6.5	Results for the Exponential Model on Air France Flight 447	96
6.6	Results for the Exponential Model on Malaysia Flight MH 370	97
6.7	2008 Results for the Exponential Model on Gross Domestic Product	97
6.8	2009 Results for the Exponential Model on Gross Domestic Product	97
6.9	Relationship of angles of rotation for large x'_i s.	101
6.10	Relationship of angles of rotation for small x'_i s.	101
7.1	2003 Results for the Weibull Model for Kentucky	122
7.2	2004 Results for the Weibull Model for Kentucky	122
7.3	2007 Results for the Weibull Model on Violent Crime in the U.S.	123
7.4	2008 Results for the Weibull Model on Violent Crime in the U.S.	124
7.5	Results for the Weibull Model on Air France Flight 447	125
7.6	Results for the Weibull Model on Malaysia Flight MH 370	125
7.7	2008 Results for the Weibull Model on Gross Domestic Product	126

7.8	2009 Results for the Weibull Model on Gross Domestic Product	126
8.1	Interpretation of the Eigenvalues	146
8.2	Eigenvalues and Eigenvectors for the 2003 Kentucky Data	149
8.3	2003 Results for the Spherical Model for Kentucky	150
8.4	Eigenvalues and Eigenvectors for the 2008 Vilonent Crime Data	152
8.5	2008 Results for the Spherical Model for 2008 Crime	153
8.6	Eigenvalues and Eigenvectors for Malaysia Flight MH 370	154
8.7	Results for the Spherical Model for Malaysia Flight MH 370	154
8.8	Eigenvalues and Eigenvectors for the 2008 Gross Domestic Product	156
8.9	2008 Results for the Spherical Model for GPD	156

FOREWORD

This book is a practical reference guide accompanied with an Excel Workbook. This book gives an elementary introduction of the weighted standard deviational ellipse. This book also presents the computational aspects of the weighted exponential distributions as well. For the examples given, calculations are performed using VBA for Excel. This book makes comparisons (and shows the computations via VBA for Excel) using the likelihood functions with spatial data of the weighted ellipses. Lastly, the book covers spherical statistics. Throughout the text, the reader can see how to perform these difficult calculations and learn to adapt the code for his research.

This book is a work in progress. Meaning,

- Revising existing text.
- Modularizing source code.
- Adding new text to existing text (e.g. new examples, new source code).

PREFACE

Scientific articles exist since the 1960's for the use of an area frames for collecting and interpreting survey data in the United States, and in the 1990's in Europe and Africa. Cartographers divide units of land from satellite imagery into enumerable segments. Statisticians assign the segments to defined strata. Typically, crop estimation is performed using imagery analyzes, field surveys, and mail surveys. The following list of articles includes both practical results and theoretical results typically found in the literature on area frame surveys.

[Pratt, Bird, Taylor, and Carter (53)] wrote a paper that covers the topics of planning and implementing an area frame survey in Nigeria and choosing the estimation procedures. They were careful to choose their satellite imagery and coordinate it with the fieldwork. In Europe and Africa, Statisticians use regression to classify the satellite images from field data collection and from image classification. Image classification without field data collection does tend to lead to failure due to the following reasons: 1) intercropping practices, 2) fallow/cultivated continuum and 3) small, irregular field structure. The authors more generally state those reasons as "spectral confusion." The authors state that the best time to perform their fieldwork is shortly after harvest because it would be easiest to distinguish harvested rice fields from wet, green swamp grasslands. [page 70] calculates the sample size in terms of segments according to several criteria: 1) the size of the study size area, 2) the need to maximize the area covered, 3) the speed to cover the area during enumeration, 4)

reduce locational error within the segment during field mapping. The authors obtain a 4.8% sampling fraction by choosing forty-nine segments (500×500 meters with a sampling fraction of 5%) out of 75. Three estimators of the data collected are presented: 1) direct expansion of the survey data, 2) pixel count, and 3) regression estimator. The authors recommend the regression estimator, which take into account both the irrigation mapped by enumerators on the ground in each segment and the satellite image classification. This is because it was highly dependent on the classifier and gave a narrower range of possible classifications. They used a ratio of the direct expansion variance and the regression model to prove this mathematically.

[Kelly (33), Chhikara and Deng (7), Faulkenberry and Garoui, (16)] discuss area-frame data collection and estimation in the United States. [Kelly (33)] collects and summarizes survey data in a production environment. [Chhikara and Deng (7)] consider the problem of using the area frame and rotating segments amongst years. On page 926, they develop an ANOVA model to capture the stratum means per year and the segment mean effects for a particular random variable. The author also rotates and overlaps segments amongst years via a simulation study that shows the optimal rotation of segments should be 40% to 60%. [Faulkenberry and Garoui, (16)] discuss the topic of estimating population totals and variances from an area survey frame. The authors point out that because of the association of a farm with more than one segment does not lead to cluster sampling. Based on four classifications, the authors show that the Horvitz-Thompson estimator for totals is probably the best choice of estimators. It is a good estimator when the probability of selection π_k is proportional with the random variables y'_k 's. Depending on the number of farms or the number of known segments, four classifications arise. The author assigns an estimator to each class that will work well, and then takes expectations and variances.

Although most of the authors so far do present maps with their articles and their statistical estimation methods, they lack discussing the latitude and longitude of their random variables. The remainder of this section will discuss the history of associating random variables with the latitude and longitude (i.e. the placement on a map).

[Ebdon, (14), Chapter 7, 1985] discusses spatial statistics and several measures. The author gives interpretations to the concepts of the mean center and standard deviational ellipse. For instance, the mean center can be thought of the "center of gravity" of the distribution of the given points on a map. He defines the standard deviational ellipse as the "spread of points" about the mean center. Without using modern software, the author has an interesting way to draw the ellipse. It involves plotting the deviation for each point parallel to the rotated axes and fitting the ellipse [Ebdon, (14), p. 137].

The author Lefever first used the standard deviational ellipse in 1926 to study the geographic position of social problems [Lefever, (41)]. The author argues that a uniform distribution will not fit well for location problems of any kind because ecological and sociological problems occur more frequently on a select area of the map than other areas. The author shows the correspondence between calculating the standard deviation between the mean center of the points and each location; and then showing its geometric relationship on a map. An example is given. In the conclusion, the author states four characteristics of a unit of locations:

1. The center of the system determined by the distances from the central point (versus the extreme unit locations).
2. The direction or trend of the system given by the angle θ_m of the axis of maximum standard deviational variation value. This is the line of best fit for the entire system of unit locations.
3. The concentration of the system (or dispersion) shown in terms of a standard deviational ellipse.
4. The relative concentration given by the ratio of observations within the ellipse compared to the entire population expressed in area units.

The author gives no references in the paper. To summarize, the article presents the mean and standard deviation of a set of numbers — longitude and latitude. So, who came up with the idea of weighting the longitude and latitude data?

The weighted standard deviational ellipse appears in the literature in 1971 in [Yuill, (68)]. The author begins with a discussion of the work by Lefever and the comments by Furtey in 1927. The paper proceeds to defend using the ellipse for geographic applications. The author derives the formulas on pages 30-31 for the weighted standard deviational ellipse. It is on page 32 that the weighted mean center is introduced (a new notion in the literature). The author weights the latitude and longitudinal observations using the random variable. The author gives the following computations for comparing ellipses.

1. The enclosed area within the ellipse.
2. The number of points enclosed within the ellipse (more is better).
3. The shape of the ellipse measured by its eccentricity .

The shape of the ellipse determines the distribution of the points . Points concentrated at the pole of the ellipse (or circle) should have a non-uniform distribution while those scattered from the pole will have a uniform distribution . Finally, the author applies the concepts to several sets of data. References appear in the footnotes throughout the paper.

So, where did the weighted mean center originate? Yuill's paper stated the mean center, but did not derive it. Books and research articles have methods on the weighted least squares dating back to the 1970's. The weight assigned to each longitude and latitude pair is between zero and one; additionally, the weights sum to one. This is due to the formula, not due to the nature of the application (e.g. ecological versus sociological problem). Other authors [Magee (43), and Woolson and Clarke (66)] assign least squares weights according to the sample design. Magee assigns weights based on sample selection probabilities. Woolson and Clarke assign weights to compensate for missing data. [Rubin, (56)] models the leverage values as weights in the regression model. This in turn shows which data points in the data set exert more influence over the others. [Lee and Wong, (37), pages 38-39] assigns weights

according to the distribution of the phenomenon. If a city's population is the random variable of interest, then the city's population is the weight for the latitude and longitude. We apply weights in the context of [Lee and Wong, (37)] in this book.

Finally, to extend the concepts of the weighted mean center and the standard deviational ellipse, at times, we prove that the data has a particular distribution. By transforming the data and plotting the points $[t, F(t)]$ where $F(t)$ is a known cumulative distribution function, one should obtain a straight regression line [Lee, (36)] as long as the data has that distribution. The author covers data transformations for distributions with a location parameter, a scale parameter and a shape parameter. Knowing the distribution of the data is advantageous. Using maximum likelihood estimators, we can compare likelihood functions, [Lee (36), page 227]. Generally, we choose the likelihood function with the largest value.

ACKNOWLEDGMENTS

Map Point, Excel, and MS Word are products of Microsoft Corporation

Google Earth is a product of Google, Inc.

R. L. G.

ACRONYMS

GDP	Gross Domestic Product
OECD	Organization for Economic Cooperation and Development
SDE	Standard Deviation Ellipse
VBA	Visual Basic for Applications

CHAPTER 1

INTRODUCTION TO VBA

1.1 The Development Environment

The calculations in this book are difficult at times. Having a programming environment becomes advantageous. A programming environment provides the flexibility to calculate statistics and likelihood functions. This textbook uses the Visual Basic Application (VBA) for Excel. It is a programming environment based on the Visual Basic programming language. The VBA Development environment does not automatically appear as a menu option in Excel. See Figure 1.1. The user must make this option visible by following these steps:

1. Click on `File | Options | Customize Ribbon`. The dialog box in Figure 1.2 will appear.
2. From the `Choose Commands` drop-down list, select `Main Tabs`.
3. Highlight `Developer`.
4. Click on the `Add>>` button in the middle of the screen.

2 INTRODUCTION TO VBA

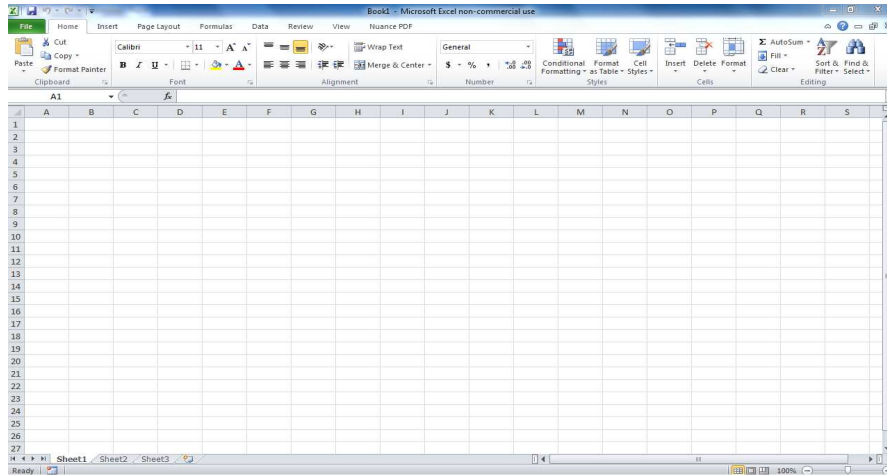


Figure 1.1 This figure shows Excel without the Developer menu.

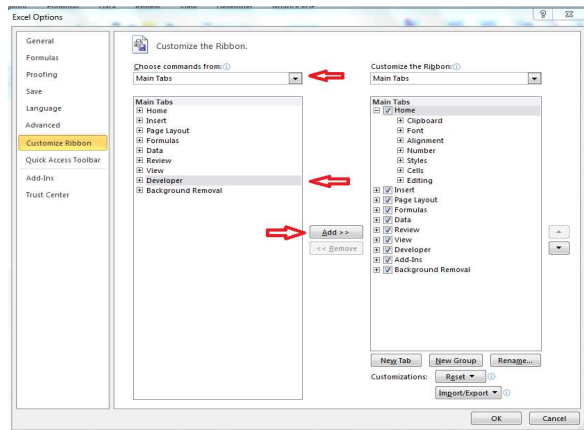


Figure 1.2 This figure shows Excel Options dialog box. Excel uses the dialog box for adding the Developer menu for VBA programming.

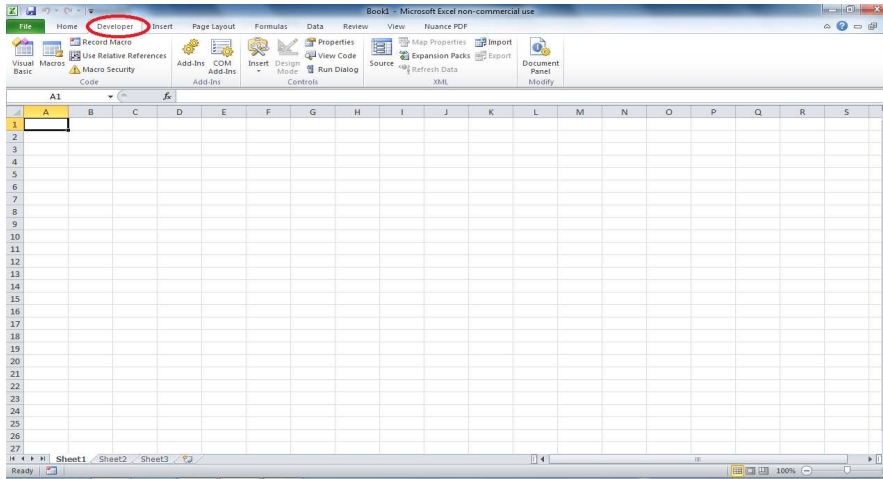


Figure 1.3 This figure shows Excel with the Developer menu at the top of the screen.

The `Developer` menu option will appear somewhere along the top of the screen. See Figure 1.3.

1.2 Variables

To declare variables explicitly in VBA for Excel, use the `Dim` statement. Excel supports the following basic data types:

- Integer numbers — **Dim X As Integer** The **Integer** data type holds integer variables stored as 2-byte whole numbers in the range of -32,768 to 32,767.
- Real numbers — **Dim Y As Double**. The **Real** data type holds double-precision floating-point numbers as 64-bit numbers in the range of -1.79769313486231E308 to -4.94065645841247E-324 for negative values and 4.94065645841247E-324 to 1.79769313486232E308 for positive values.
- String characters — **Dim S As String**. The **String** data can include letters, numbers, spaces, and punctuation. The **String** data type can store fixed-length strings ranging in length from 0 to approximately 63,000 characters.
- Date — **Dim BirthDay As Date**. Excel allows the **Date** data type to store dates and times as a real number. **Date** variables are stored as 64-bit (8-byte) numbers. The value to the left of the decimal represents a date, and the value to the right of the decimal represents a time. Excel considers the date expression to be any expression such as a date, including date literals, numbers that look like dates, strings that look like dates, and dates returned from functions. A date expression is limited to numbers or strings, in any combination, that can represent a date from January 1, 100 to December 31, 9999.

Dates are stored as part of a real number. Values to the left of the decimal represent the date; values to the right of the decimal represent the time. Negative numbers represent dates prior to December 30, 1899.

A date can be any sequence of characters with a valid format surrounded by number signs (#). Valid formats include the date format specified by the locale settings for your code or the universal date format. For example, use #12/31/92# in the VBA editor when explicitly referring to a date.

- **Currency — Dim Q As Currency.** The **Currency** data type has a range of -922,337,203,685,477.5808 to 922,337,203,685,477.5807. Use this data type for calculations involving money and for fixed-point calculations where accuracy is particularly important. In the VBA Editor, use the "@" sign when referring to currency.
- **Long integer — Dim R As Long.** The **Long** data type is a 4-byte integer ranging in value from -2,147,483,648 to 2,147,483,647. In the VBA Editor, use the "&" symbol when referring to long integers.
- **Logical — Dim L As Boolean.** The **Boolean** data type has only two possible values, True (-1) or False (0).

Variables defined inside a subroutine are visible only inside that subroutine. Variables defined at the module level are visible to the subroutines defined in that module.

1.3 Arrays and Records

We can build on the basic data types listed in Section 1.2. Consider arrays and records.

- **One dimensional array — Dim Name_List(1 To 10) As String.** This array defines a set of sequentially indexed elements having the data type **String**. Each element of the array has a unique identifying index number. Changes made to one element of an array do not affect the other elements.
- **Two dimensional array — Dim AList(1 To 5, 1 To 10) As Double.** This array defines a matrix of indexed elements having the numeric data type **Double**.
- **A user defined record —** Excel allows the definition of the user-defined record using the **TYPE-END-TYPE** statement. To define a list of customer address records, use the following VBA code:

```

1 3 [Type Customers
    [First_Name, Last_Name As String
    Address As String
    City As String
    State As String
    Zip_Code As Integer
    Phone As Integer
    End Type

```

```

2 [Dim List As Customers

```

Do not use the **Dim** inside the **TYPE-END-TYPE** statement. Use the dot "." notation to reference the items in the record. For example in the VBA Editor, to reference the customer phone number, use the code:

```

1 [List.Phone = 3223223

```

1.4 Branching

VBA has an **IF-THEN-ELSE** statement for conditionally executing code. The syntax is as follow:

```

1 2 [If <condition> Then
    [VBA statements
    Else
    3 [VBA statements
    End If

```

The **IF-THEN-ELSE** construct must appear in a subroutine. It cannot appear as open code in a module. Aside from variable definitions and subroutine declarations, this is true for the majority of VBA programming. There can only be one **ELSE** statement in any **IF-END-IF** construct. An alternative to the above **IF-THEN-ELSE** construct is the **IF-ELSE-IF** construct.

```

1  [
2  | If <condition> Then
3  | VBA statements
4  | ElseIf <condition> Then
5  | VBA statements
6  | ElseIf <condition> Then
7  | VBA statements
8  | Else
9  | VBA statements
10 | End If

```

If the user has a short, single VBA statement for any of the **ELSEIF** statements, it is not advisable to put it on the same line after the **THEN**. It usually creates a run-time error even though the syntax looks correct.

1.5 Loops

The two loops covered in this section are the **FOR-NEXT** loop and the **WHILE-WEND** loop. To execute a set of VBA statements a given number of times in a given sequence, the **FOR-NEXT** loop has the following syntax:

```

1  [
2  | For <counter>= start To end
3  | VBA statements
4  | Next <counter>

```

As an example, we can generate two digits in a phone number.

```

1  [
2  | multiplier = 1
3  | List.phone = 0
4  | [
5  | | For i = 1 To 7
6  | | List.phone = List.phone + i*multiplier
7  | | multiplier = multiplier * 10
8  | | Next i

```

The value of `List.phone` is 28 because $\frac{n(n+1)}{2} = \frac{7(8)}{2} = 28$. The only purpose of the example is to show the syntax of the **FOR-NEXT** loop.

The **WHILE-WEND** loop has the following syntax.

```

1  [
2  | While <condition>
3  | VBA statements
4  | Wend

```

It is more likely to program infinite loops with the **WHILE-WEND** loop than with the **FOR-NEXT** loop. It is important to initialize the condition variable(s) before entering the loop and to update the condition variable(s) inside the loop.

1.6 Setting Properties

Both the Excel spreadsheet and the VBA environment allow the user to change the cell contents properties to bold, italic, underline, strikethrough, and the font size. Except for the font size, most of these properties are Boolean valued.

- Object.Bold — Boolean
- Object.Italic — Boolean
- Object.Size — Integer
- Object.StrikeThrough — Boolean
- Object.Underline — Boolean

In our case, the VBA object is a cell in a spreadsheet. We use the following code to reference the properties of the cell.

```

1  ActiveSheet.Cells(1, 3).Font.Bold = True
   ActiveSheet.Cells(1, 3).Font.Italic = True
   ActiveSheet.Cells(1, 3).Font.Size = 25
   ActiveSheet.Cells(1, 3).Font.Strikethrough = True
   ActiveSheet.Cells(1, 3).Font.Underline = True

```

`Cells(1, 3)` references row 1, column C in the last spreadsheet viewed before entering the VBA environment. Alternatively, we could have substituted the `ActiveSheet` with a specific spreadsheet name such as `Worksheets("Sheet1")`. We have hard-coded `Sheet1`. `Sheet1` will always be the spreadsheet referenced.

When updating a set of cells in a spreadsheet, it is convenient to use the **WITH-END-WITH** statement. It can save some typing and make the code easier to read. For example, we can re-write the VBA code for the font updates as follow:

```

1  With ActiveSheet
   .Cells(1, 3).Font.Bold = True
   .Cells(1, 3).Font.Italic = True
   .Cells(1, 3).Font.Size = 25
   .Cells(1, 3).Font.Strikethrough = True
   .Cells(1, 3).Font.Underline = True
End With

```

The **WITH-END-WITH** statement is convenient to use in complicated calculations as well.

1.7 The Excel Examples

The Excel spreadsheet and the Excel VBA environment contain many of the same functions. Some common syntax and conventions for the worksheet follows.

1. Columns in an Excel worksheet always begin with a letter.
2. Rows in an Excel worksheet always begin with a number.
3. The three default names of the worksheets in an Excel workbook are `Sheet1`, `Sheet2`, and `Sheet3`. Microsoft Corporation capitalized the "S" in the word "Sheet." In the VBA for Excel environment, upper and lower case counts.
4. Character and numeric data can ordinarily be copy and pasted into the worksheet cells.
5. Precede numeric data with leading zeros with the single quote " ' "to retain those leading zeros. Entering the single quote in front of the data is a manual operation. Changing the column to the TEXT format usually causes other problems later on.
6. Formulas begin with an equal sign. Some useful formulas used in this application include:
 - `mod(number, divisor)` — where the argument `number` can either be reference to a cell or a hard coded number such as 180 or 360; and the argument `divisor` can either be a cell reference or a hard coded number such as 180 or 360.
 - `average(range)` — where the argument `range` is a range of cells such as `A2..A91`.
 - `sum(range)` — where the argument `range` is a range of cells such as `C2..C39`.

In this application, we keep the data and results in the worksheets. We use VBA to perform the calculations. Some common conventions in VBA for Excel follow.

1. Declare global variables at the top of a module.
2. VBA program code must appear inside a subroutine. Some useful, recurring VBA statements used in this application include:
 - `Selection.Rows.Count` — Visual Basic applies the `selection` expression to a range of data in the active spreadsheet. It returns a count of the number of rows. Since we included the first row, the count *n* is larger by one. This application adjusts for this.

- `ActiveSheet.Cells(row, col).Value` — The `activesheet` object references the data in the spreadsheet highlighted before entering the VBA editor. This entire statement allows reading the data.
- `Worksheets("name").Cells(row, col).Value` — The `worksheets` object references the spreadsheet named NAME. In the application name = stats. This entire statement allows writing data to a spreadsheet other than the active spreadsheet.
- `For-Next` — This is the common looping structure used for calculating the sums, the areas, the probability distributions, eccentricities, and so on.
- `While-Wend` This application uses this looping in Module 3 in the Secant algorithms because the termination condition required more knowledge that the number of observations in the data set.
- VBA numeric calculations — The Visual Basic for Excel numeric calculations are similar in style to those of other programming language statements.

Some useful, recurring VBA functions used in this application include:

- `log(number)` — This function returns the natural logarithm of the given number. The user must write his own code for other base logarithms.
- `exp(number)` — This function returns the exponential function of the given number. e is approximately 2.718282.
- `worksheetfunction.pi()` — This function returns the value of π .

3. Modules are a collection of related subroutines and variables.
4. Declare local variables inside a subroutine.
5. Variable names are not case sensitive. For all declared variables, the VBA editor will change the upper and lower case spelling.

In this application, the results are stored in the spreadsheet called STATS. See Figure 1.4. The data are stored in the remaining spreadsheets. The first three columns of the data must be in a particular order. See Figure 1.5.

1. Latitude x_i .
2. Longitude y_i .
3. Random variable of interest w_i .

Further, select the data (and only the data) so that the subroutines can identify the beginning and ending rows. To get into the VBA editor, follow these steps.

1. Click on the `Developer` ribbon.

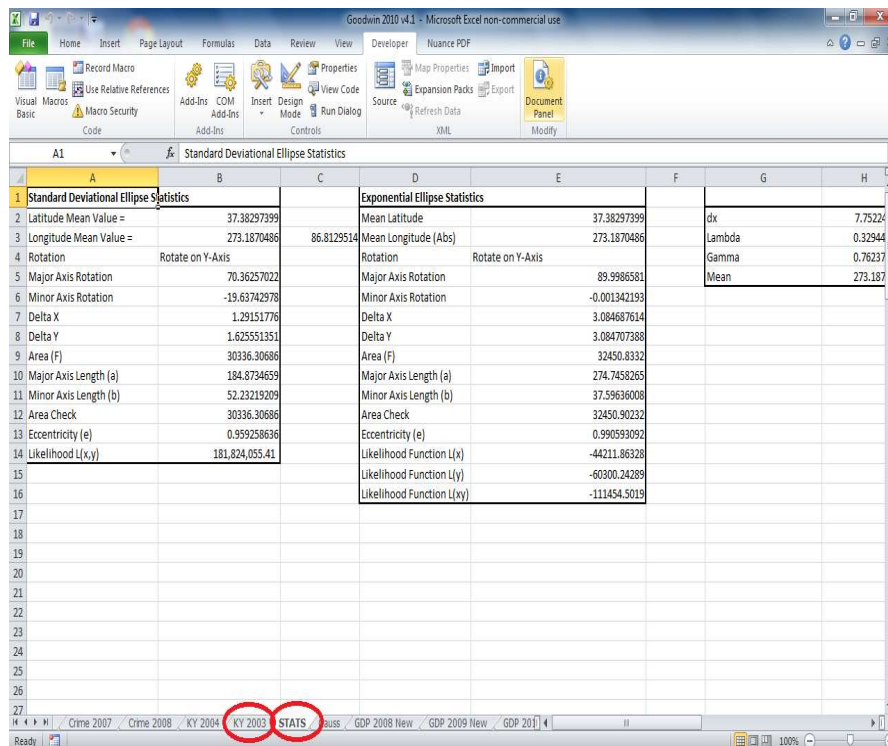


Figure 1.4 This figure shows the Excel workbook. A data worksheet at the bottom named KY 2003 has been circled. The worksheet named STATS has been circled. This application always saves the results to the STATS worksheet.

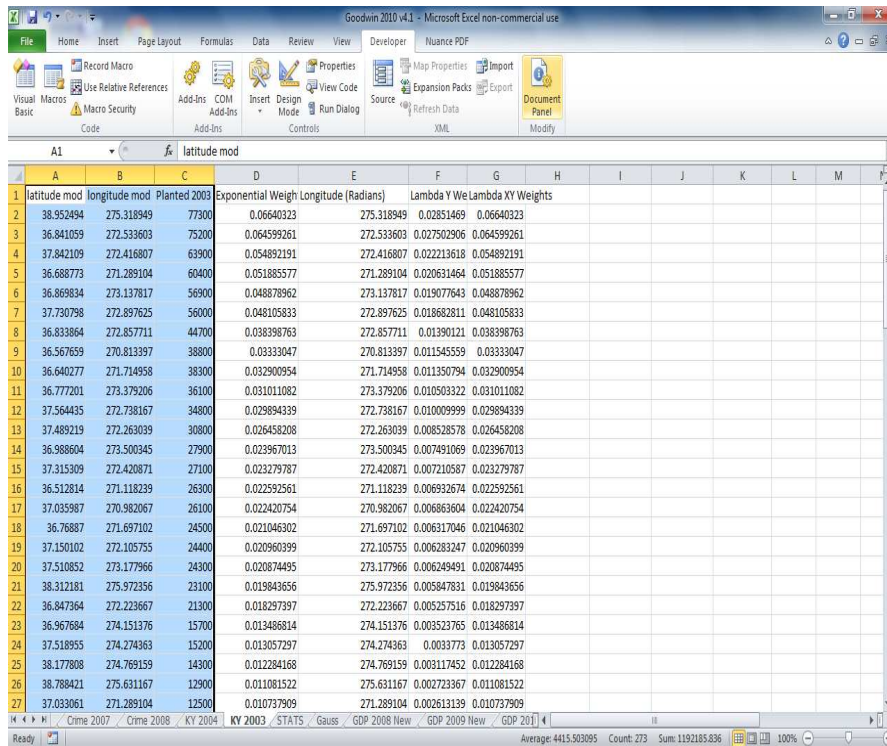


Figure 1.5 This figure shows the Excel spreadsheet with data. For this application to run properly, the first row can contain any column names the reader wishes. Column 1 must contain the latitude observations. Column 2 must contain the longitude observations. Column 3 must contain the random variable of interest. It is not advisable to put any observations in the remaining columns as this application may over-write them.

2. Click on the `Visual Basic` button on the left side.
3. For a new module, click on the `Insert` menu item and select `Module`.

A blank VBA Editor screen will appear. Just like any other programming language, either you know the syntax or you do not. VBA for Excel is similar in style to the programming language Basic. However, the objects tend to be centered around the Excel workbook — cells, spreadsheets, columns, rows, and so on.

CHAPTER 2

INTRODUCTION TO MAP POINT

Map Point is a software package that maps points on the Earth. The user provides the latitude and longitude, and Map Point plots the point. The points in this textbook are either geo-coded observations from Google Earth or calculations such as the mean center (also called the center of gravity) or the semi-major and semi-minor axis lengths of an ellipse. It is possible to enter, say a county name into Map Point. Map Point will identify the county on the Earth. However, it will not give the user the latitude and longitude coordinates. This is why we must use an alternate software product.

There are several useful concepts when using Map Point:

1. Plotting the mean center on the Earth.
2. Plotting and graphing the entire survey of points on the Earth.
3. Drawing lines with an exact length in miles (or kilometers).
4. Saving work.
5. Copying and pasting images to other software applications.

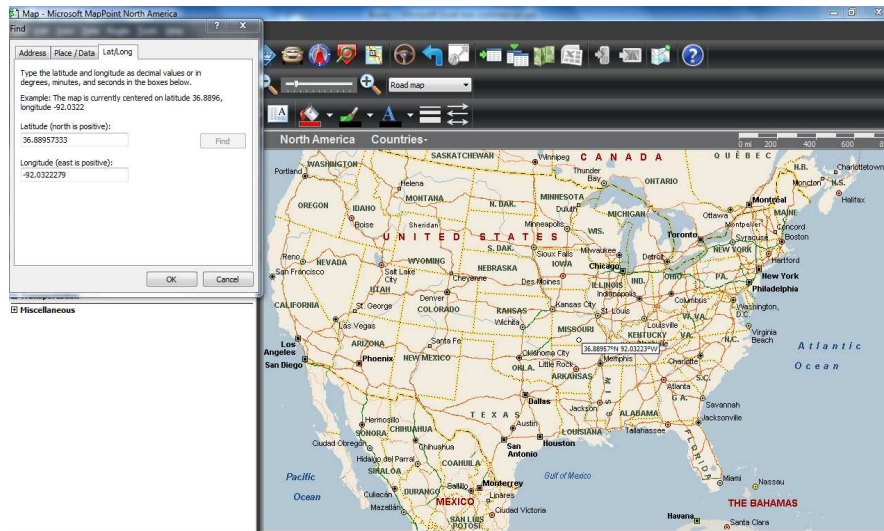


Figure 2.1 This figure shows the Map Point dialog box for finding a point on the Earth using the latitude and longitude.

Map Point does allow us to draw ellipses. However, it does not allow us to rotate them. Because of this, we must draw ellipses outside of Map Point. We use Map Point to draw the semi-major axis and the semi-minor axes.

2.1 User Data

Map Point has two ways to enter data:

1. Manually one data point at a time using `Tools | Find | Lat\Long`.
2. Import an Excel spreadsheet using `Data | Data import wizard`.

When the user enters data manually, one data point at a time, Map Point requests the latitude and the longitude. When the user imports multiple data points via an Excel spreadsheet, Map Point requires column names with "latitude" and "longitude." In addition, Map Point expects the user to plot a third variable called a "data" variable or an analysis variable along with the latitude and longitude. Naming the third variable is at the user's discretion. Notes on entering data manually appear in Section 2.2. More notes follow on importing data in Section 2.3.

2.2 Plotting the Mean Center

To plot the mean center (a single data point), select `Tools | Find`. The dialog box on the left side in Figure 2.1 will appear. Upon selecting the `Lat/Long` option

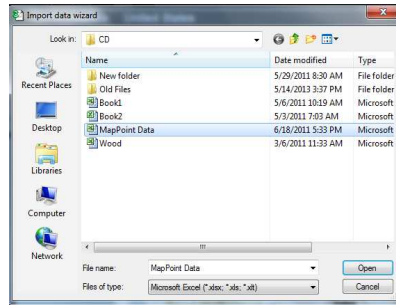


Figure 2.2 This figure shows the Map Point dialog box for browsing Excel files, which have the coordinates and the random variable.

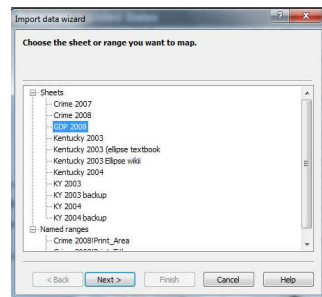


Figure 2.3 This figure shows the Map Point dialog box that lists the Excel spreadsheets in the Excel workbook.

at the top of the dialog box, the user types in the coordinates and clicks on the Find button. Map Point will locate that latitude and longitude on the Earth and identify those coordinates with a box.

2.3 Visualizing the Survey Data

Sometimes the user may want to visualize the survey data (i.e. many data points) and the ellipse together (i.e. drawing). Alternatively, maybe, the user may want to visualize the plotted points on the Earth with the random measurements (i.e. graphing). Map Point does provide the software tool to plot or graph an entire data set onto the Earth.

2.3.0.1 Plotting To plot the data, first prepare an Excel file with the latitude, longitude, and the random variable of interest. To plot the data set, select Data | Import data wizard.

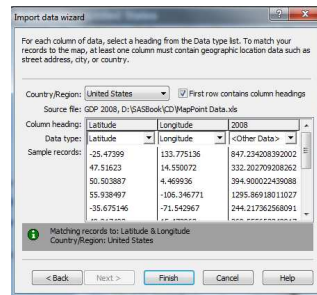


Figure 2.4 This figure shows the Map Point dialog box containing the spreadsheet column names and the data.

- The dialog box in Figure 2.2 allows the user to browse the hard drive and select the Excel file that contains the survey data.
- Map Point displays the dialog box in Figure 2.3 showing the spreadsheet names in the Excel workbook. Select one of the spreadsheets.
- Click the `Next` button in the Import Data Wizard. The screen in Figure 2.4 appears showing the columns and the data in the spreadsheet.
- Click the `Finish` button.
- The Data mapping wizard in Figure 2.5 appears. There are some choices here. Do we want to use shaded circles, sized circles or pushpins to mark the points on the Earth? The remaining options pertain to graphing the data. We will discuss graphing later. Let us choose pushpins since they stand out the best.
- Click on the `Pushpin` button.
- Click on the `Next` button at the bottom of the dialog box. The dialog box in Figure 2.6 appears. The user can change the color of the pushpins here, if desired. Map Point automatically recognized the column names `latitude` and `longitude`. This is important because the user would have had to identify these had this not been the case.
- Click on the `Finish` button at the bottom of the dialog box. Map Point plots the data in the spreadsheet on the Earth. Plotted data on the Earth are shown throughout the textbook.

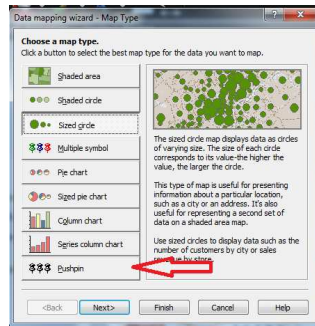


Figure 2.5 This figure shows the Map Point Data mapping wizard.

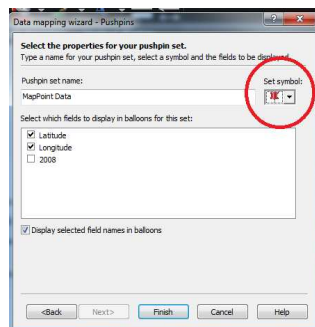


Figure 2.6 This figure shows the Map Point choices for the pushpin colors.



Figure 2.7 This figure shows the Map Point choice for the column chart.

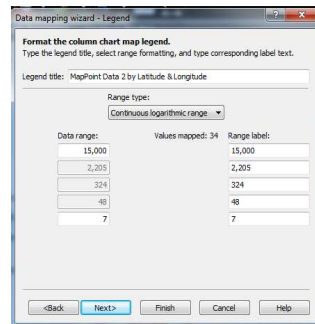


Figure 2.8 This figure shows the Map Point choices for scaling the data and change the label text.

2.3.0.2 Graphing To graph the data, first prepare an Excel file with the survey data. Save the Excel file to the hard drive. Select `Data | Import data wizard`.

- The dialog box in Figure 2.2 allows the user to browse the hard drive and select the Excel file that contains the data to analyze.
- Map Point displays the dialog box in Figure 2.3 showing the spreadsheet names in the Excel workbook. Select one of the spreadsheets.
- Click the `Next` button in the Import Data Wizard. The screen in Figure 2.4 appears which shows the column names and the data.
- Click the `Finish` button.
- The `Data mapping wizard` in Figure 2.5 appears. There are some choices here. Do we want our data displayed as a pie chart, sized pie chart, column chart, or as a series column chart? The remaining options pertain to plotting the data. We discussed plotting data in the previous section.
- Click on the `Column chart` button.
- Click on the `Next` button at the bottom of the dialog box. The dialog box in Figure 2.7 appears. Map Point automatically recognized the column names `latitude` and `longitude`. The user must confirm to Map Point to graph the column name `2008`.
- Click the check box next to the field `2008` to confirm graphing this field.
- Click the `Next` button. The dialog box in Figure 2.8 appears. This lets the user scale the data if needed. It also allows the user to change the label text.
- Click on the `Finish` button at the bottom of the dialog box. Map Point displays a graph on the Earth of the data. Short bars represent small values and tall

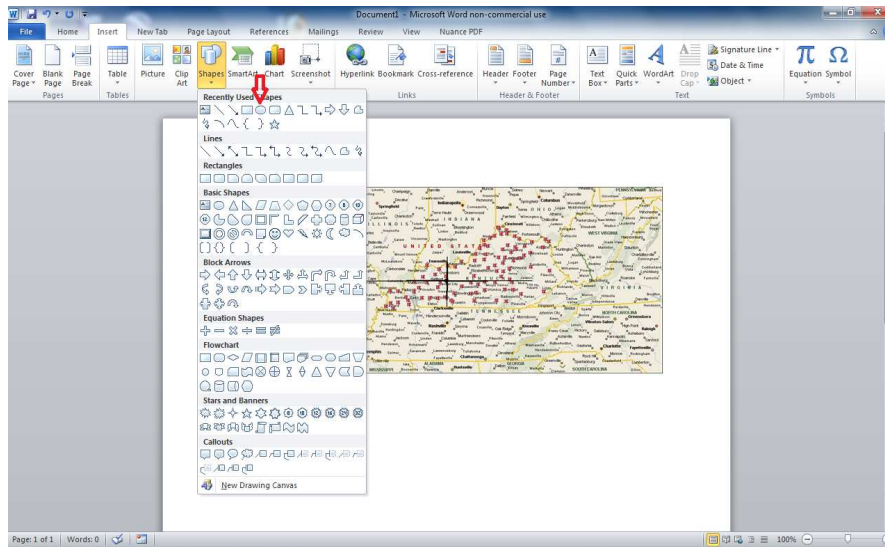


Figure 2.9 This figure shows the Microsoft Word shapes for drawing an ellipse.

bars represent large values. Map Point places the bars on the Earth using the latitude and longitude coordinates.

2.4 Drawing Axes

Before drawing an axis, it may be necessary to zoom into the survey area. To draw an axis for an ellipse which has an exact length measured in miles or kilometers, select **Tools | Measure distance**. To toggle between miles and kilometers, select **Tools | Options** and the dialog box in Figure 3.1 will appear. The section on **Units** changes the unit of measure.

- Single click on the label box for the mean center. This will anchor the ruler. Map Point displays the unit of measure and number of measurements at the right side of the line.
- Drag the mouse the length required.
- Single click. This will anchor the right end of the line.
- Hit the <Esc> key to stop drawing lines.

2.5 Drawing Ellipse Boundaries

Given that we drew the major axis and the minor axis, and we identified the mean center of the ellipse, it is a simple task to draw the ellipse using MS Word. The steps are as follow:

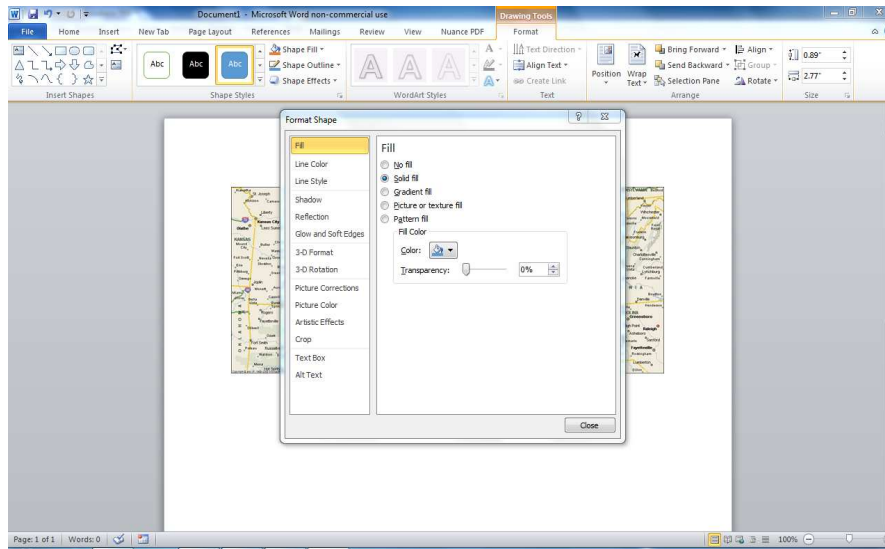


Figure 2.10 This figure shows the Microsoft Office dialog box for formatting a shape.

- Draw the axes described in Section 2.4 using an imaginary X and Y -axes laid on top of the Earth.
- Note the axis of rotation (e.g. X or Y) from the statistical computations.
- Copy and paste the image from Map Point into MS Word. The imaginary X and Y -axes form 90° angles. Half of 90° is 45° , and so on. If the rotation is from the Y -axis, then begin counting from there.
- In MS Word, select `Insert | Shapes`. The shapes in Figure 2.9 will appear.
- Select the fifth shape from the top, `oval`. Left click and right click around the axes lengths. The oval will be dark filled.
- Right click the oval. Left click the oval and select `Format shape`. The dialog box in Figure 2.10 will appear.
- Click on the `No fill` radio button.
- Click on `Close`.

We drew the ellipse following the steps above. We need to rotate the ellipse. Suppose we want to rotate the ellipse from the Y -axis 70° . Then, follow these steps:

- Click on one of the edges of the ellipse to highlight it.

- A green circle appears at the top of the ellipse. This is for rotating the ellipse. Single click on the green circle.
- Drag the mouse to obtain a 70° angle between the Y-axis and the semi-major axis. Certainly you know where 90° and 45° are at. This is not difficult.
- Select `File | Save As` to save the ellipse.

To avoid the appearance of *cross hairs* in any published graphs, the user can alternatively identify the ends of the major and minor axes using pushpins or some other symbol. When in Map Point, delete any lines drawn when using the ruler.

2.6 Programmer Notes

- Microsoft Corporation did not provide Map Point with a development environment as they did with Excel and MS Word. We need to perform repetitive research manually. The one advantage is that the registered version of Map Point does accept Excel files for plotting and graphing data.
- Not being able to rotate ellipses in Map Point is a draw back. This is just another reason to leave Map Point for another software product.
- Map Point does not geo-code data *and* return the coordinates to the user for computational purposes. This is just another reason to leave Map Point for another software product.

CHAPTER 3

MATHEMATICS REVIEW

3.1 Notation

We discuss some common notation used throughout this book here.

Notation	Meaning
x	Subscript denoting latitude.
y	Subscript denoting longitude.
t	A dummy variable used in integration. Usually denotes the latitude.
u	A dummy variable used in integration. Usually denotes the longitude.
w	Subscript denoting the weight variable.
γ_x	Shape parameter to the Weibull distribution for the latitude.
γ_y	Shape parameter to the Weibull distribution for the longitude.
λ_x	Scale parameter to the Exponential distribution and the Weibull distribution for the latitude. Generally, there is no implied equality.
λ_y	Scale parameter to the Exponential distribution and the Weibull distribution for the longitude. Generally, there is no implied equality.
k	The iteration number in the Secant algorithm.
n	Denotes the sample size .
ϕ_i	Represents the observed values of the latitude in spherical statistics.
θ_i	Represents the observed values of the longitude in spherical statistics.
R	Called the resultant length in spherical statistics.
$\bar{l}_0, \bar{m}_0, \bar{n}_0$	Called the mean direction of cosines in spherical statistics.
S^*	Denotes the spherical variance.
T	Denotes the matrix T . The diagonal elements sum to the sample size n .
B	Denotes the matrix B . The diagonal elements sum to $2n$.

The set $\{x_1, x_2, \dots, x_n\}$ represents the observed values of the latitude. The set $\{y_1, y_2, \dots, y_n\}$ represents the observed values of the longitude. The set $\{w_1, w_2, \dots, w_n\}$ represents the observed values of the random variable at those corresponding latitude and longitude. Using these three sets of observed numbers (w_i, x_i, y_i) , we estimate parameters from known distributions.

3.2 Derivatives

Some common derivatives used in this text are given next. Using the same notation as in most calculus books and the CRC Standard Math Tables, u represents a function of x , a and n represent fixed real numbers.

Example 1:

$$\frac{d(a)}{dx} = 0.$$

Example 2:

$$\frac{d \log_e n}{dx} = 0.$$

Example 3:

$$\frac{d \log_e u}{dx} = \frac{1}{u} \frac{du}{dx}.$$

Example 4:

$$\frac{d e^u}{dx} = e^u \frac{du}{dx}.$$

Take the derivative of $f(w, \lambda, x) = w \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$ with respect to x .

$$\frac{d w \frac{1}{\lambda} e^{-\frac{x}{\lambda}}}{dx} = -w \left(\frac{1}{\lambda} \right)^2 e^{-\frac{x}{\lambda}}$$

Example 5:

$$\frac{d x^a}{da} = x^a \log x.$$

We are not asking for the derivative with respect to x in this example.

3.3 Integrals

Some common integrals used in this text are given in this section.

Example 6:

$$\int_0^x a \, du = ax.$$

Example 7:

$$\int_0^a \ln n \, dx = x \ln n \Big|_0^a = a \ln n.$$

Example 8:

$$\int \frac{1}{x} \, dx = \log x$$

Example 9:

$$\int e^{ax} = e^{ax}/a$$

In this textbook, sometimes the symbol $\exp\{\}$ represents the exponential function instead of e .

3.4 Quadratic Equation

Most of the material in this section appears in CRC's Standard Math Tables. A quadratic equation has the form $ax^2 + bx + c = 0$. Then, the roots are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

If a , b , and c are real, then

1. If $b^2 - 4ac$ is positive, then the roots are real and unequal.
2. If $b^2 - 4ac$ is zero, then the roots are real and equal.
3. If $b^2 - 4ac$ is negative, then the roots are imaginary and unequal.

The roots to the quadratic equation give the equation for the standard deviational ellipse. Cases (1) and (2) have more practicable importance than (3). We will discuss the standard deviational ellipse and other ellipses in later Chapters.

3.5 Ellipse Equation

The elliptical formula appears in Equation (3.1).

$$\frac{(x - h)^2}{a^2} + \frac{(y - k)^2}{b^2} = 1, a > b > 0 \quad (3.1)$$

It may be necessary to complete the quadratic equations in the numerators to obtain an ellipse.

3.6 Trigonometry Functions and Conversion Factors

In Excel, the `Atn()` function is used to obtain the argument x from the tangent function $Tan(x)$. The function takes degrees as input, but returns radians as output. Given that, it is convenient to know the following conversion factors:

- To convert radians to degrees,

$$1 \text{ radian} = \frac{180}{\pi} = 57.2957795 \text{ degrees.}$$

- To convert degrees to radians,

$$1 \text{ degree} = \frac{\pi}{180} = 0.0174532925 \text{ radians.}$$

3.7 Modulo Arithmetic

Modulo arithmetic arises when we convert spherical data to directional data. The spherical data comes from the observed latitude and longitude for a random event. To apply some statistical techniques, we need to convert the longitude to a circular reference system, which results in directional data. To accomplish this, we use modulo arithmetic.

To convert the latitude coordinates x_i to the circular reference system x'_i we use the general rule:

$$x'_i = \text{mod}(x_i, 180). \quad (3.2)$$

To convert the summary statistic for the mean latitude \bar{x} back to the original coordinate system, we use the general rule:

$$\bar{x}' = \begin{cases} \bar{x}, & \text{if } \bar{x} \leq 90^\circ. \\ -\text{mod}(180, \bar{x}), & \text{if } \bar{x} > 90^\circ. \end{cases} \quad (3.3)$$

The reader should not worry about how to calculate \bar{x} in this Chapter. Realize that when samples with mixed signs for the latitude and longitude arise, we must have a mechanism to determine the sign of the summary statistics.

Example 10: The coordinates of Australia are $(x_i, y_i) = (25^\circ 16' 27.84 \text{ South}, 133^\circ 46' 30.49 \text{ East})$. The decimal degree coordinates are $(x_i, y_i) = (-25.47399, 133.775136)$.

To convert the coordinates to decimal degrees in Map Point or Google Earth:

- Select `Tools | Options`.
- If using Map Point, the dialog box in Figure 3.1 will appear. If using Google Earth, the dialog box in Figure 3.2 will appear.
- Click on the radio button `Decimal degrees`.

The latitude is negative. The longitude is positive. We convert the latitude using modulo arithmetic. The new latitude observation becomes $x'_i = \text{mod}(-25.47399, 180) = 154.52601$. For computation purposes, we use the coordinates $(x_i, y_i) = (154.52601, 133.775136)$.

To convert the longitude coordinates y_i to the circular reference system y'_i we use the general rule:

$$y'_i = \text{mod}(y_i, 360). \quad (3.4)$$

To convert the summary statistic for the mean longitude \bar{y} back to the original coordinate system, we use the general rule:

$$\bar{y}' = \begin{cases} \bar{y}, & \text{if } \bar{y} \leq 180^\circ. \\ -\text{mod}(360, \bar{y}), & \text{if } \bar{y} > 180^\circ. \end{cases} \quad (3.5)$$

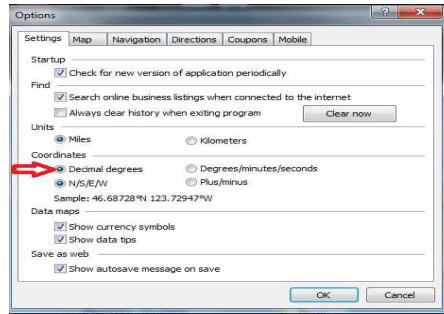


Figure 3.1 This figure shows the Options dialog box in Map Point. We use it to convert the latitude and longitude coordinates to decimal degrees.

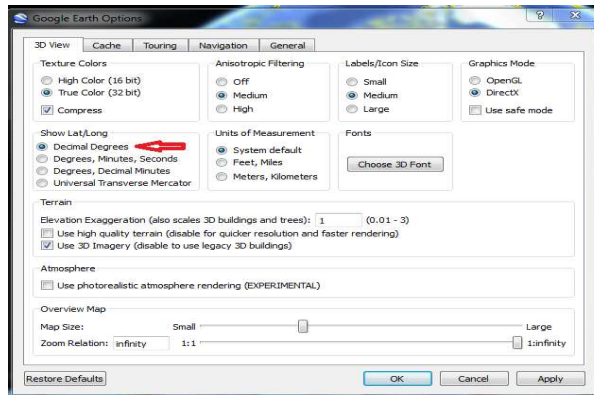


Figure 3.2 This figure shows the Options dialog box in Google Earth. We use it to convert the latitude and longitude coordinates to decimal degrees.

Example 11: The coordinates of Canada are $(x_i, y_i) = (55^\circ 56' 18.59 \text{ North}, 106^\circ 20' 48.37 \text{ West})$. The decimal degree coordinates are $(x_i, y_i) = (55.938497, -106.346771)$. The longitude is negative. The latitude is positive. We convert the longitude using modulo arithmetic. The new longitude becomes $y'_i = \text{mod}(-106.346771, 360) = 253.653229$. For computation purposes, we use the coordinates $(x_i, y_i) = (55.938497, 253.653229)$.

3.8 Weighted Data

Most of the chapters deal with weighted data. What is the motivation for weighting the latitude and longitude observations? Consider the case of un-weighted observations. For example, the Bureau of Justice Statistics (BJS) conducts a survey to measure violent crime in the United States. Hypothetically, let us say that the lowest

level of geography reported by the BJS in the survey is at the state level. To find the center of gravity of violent crime, we take the average of the latitude observations and the average of the longitude observations. Averages are un-biased estimators. To say that each state contributes equally to the center of gravity is misleading. Some states have a high number of incidents of crime while others have a low number of incidents of crime. We should reflect this fact in the estimator. The literature on weighted data typically multiplies the latitude and longitude by a weight. In this case, the weight is the number of incidents of violent crime divided by the sum of the number of incidents of violent crime for the entire U.S. This ensures the weights are between zero and one. It ensures those states with a high number of incidents of violent crime will have a correspondingly proportional effect on the center of gravity — those states will draw the mean center of gravity to themselves. *Vise-versa*, those states with a low number of incidents of violent crime will have the mean center of gravity pulled away from them.

Throughout the textbook, the examples use weighted data. Equation (3.6) shows the weighting scheme for observation i out of n observations.

$$w'_i = \frac{w_i}{\sum_{i=1}^n w_i} \quad (3.6)$$

```

1  Sub Set_Weights(sum_weight)
   'this procedure calculates the weights between 0 and 1 in column D
   'columns A, B, and C must be highlighted
   n = Selection.Rows.Count
   sum_weight = 0
   For i = 2 To n
     2  With ActiveSheet
         4  sum_weight = sum_weight + .Cells (i, 3).Value
         End With
     Next
   For i = 2 To n
     3  5  .Cells (i, 4).Value = .Cells (i, 3).Value / sum_weight
         End With
     Next
   End Sub

```

Comments in VBA for Excel begin with a single tick mark "'". Loop 2 shows the code for summing the weight column in an Excel spreadsheet. In this case, Column C (referred to as 3) and rows 2 thru n are being summed. The Excel provided function `Selection.Rows.Count` counts the number of rows highlighted in the active spreadsheet. It initializes the value of n . Loop 3 shows the code for the final calculation of the weights. It divides each observation w_i by the sum $\sum_{i=1}^n w_i$

and stores the results in Column D. We do not weight observation i equal to 1 (i.e. row 1 in the active spreadsheet) because it is a label.

Upon running this subroutine, column D will always sum to one. The use of the `WITH-END-WITH` statement is optional. It is a short-cut to not having to keep typing in the Excel object — in this case, the spreadsheet last referenced "Activsheet."

3.9 Area

Consider the simple formula for calculating the area of an ellipse (or a circle). Let a be the length of the semi-major axis. Let b be the length of the semi-minor axis. If $a > b$, then we have an ellipse. If $a = b$, then we have a circle. A function to calculate the area of an ellipse is given next. The parameters to the function are the semi-major axis length a and the semi-minor axis length b . Equation (3.7) gives the formula for the area of an ellipse.

$$f = \pi ab. \quad (3.7)$$

```

1  [Function Area(a, b)
    'a= semi-major axis length
    'b= semi-minor axis length
    2 [Area = WorksheetFunction.Pi * a * b
    End Function

```

Line 2 calculates the area of the ellipse using the formula in Equation (3.7). The value of the calculation is returned via the function name `Area`.

3.10 Eccentricity

Equation (3.8) gives the eccentricity of an ellipse (or circle) where a is the length of the semi-major axis and b is the length of the semi-minor axis, $a \geq b$.

$$e = \frac{\sqrt{a^2 - b^2}}{a} \quad (3.8)$$

```

1  [Function Eccentricity(a, b)
    'a= semi-major axis length
    'b= semi-minor axis length
    [If a >= b Then
    1  Eccentricity = Sqr(a ^ 2 - b ^ 2) / a 'eccentricity
    2  Else
    Eccentricity = Sqr(b ^ 2 - a ^ 2) / b 'eccentricity
    End If
    End Function

```

Some defensive programming has been incorporated into the `Eccentricity()` function. If a is the semi-major axis, then Equation (3.8) works fine. If the user forgets the variable definitions, then the function would return an error message if only one formula was programmed. If $a < b$, the roles of a and b are switched in Equation (3.8). The `IF-THEN-ELSE` statment (2) demonstrates the defensive programming concept.

Table 3.1 summarizes the values of the eccentricity e and the accompanying shapes.

Table 3.1 Values of the Eccentricity e

Eccentricity	Axes Lengths	Comment
$e = 0$	$a = b$	Circle
$0 < e < 1$	$a > b$	Ellipse
$e = 1$	$b = 0$	Straight line

3.11 Axes Length

Equation (3.9) gives the formula for the semi-major axis length a . Equation (3.10) gives the formula for the semi-minor axis length b .

Calculating the values x , y , and xy are complicated sums depending on the probability distribution. We will cover the details on calculating the sums in later Chapters. For now, assume that the variables x , y , and xy are given.

Given the quantities x , y , xy , the subroutine `Axes.Length()` returns the length of the semi-major axis a and the length of the semi-minor axis b .

$$a = \sqrt{\frac{y}{n} + \frac{2(xy)^2}{n \left[-(x-y) + \sqrt{(x-y)^2 + 4(xy)^2} \right]}} \quad (3.9)$$

$$b = \sqrt{\frac{x}{n} + \frac{2(xy)^2}{n \left[-(x-y) + \sqrt{(x-y)^2 + 4(xy)^2} \right]}} \quad (3.10)$$


```

Sub Axes.Length(m, x, y, xy, a, b)
'm= number of observations
'x= sum of diffs from mean center squared x direction (input)
'y= sum of diffs from mean center squared y direction (input)
'xy= diff sums from mean centers x and y direction (input)
'a= semi-major axis (output)
'b= semi-minor axis (output)
1  If x >= y Then
    a = Sqr(y / m + 2 * (xy) ^ 2 / (m * (-1 * (x - y) + Sqr((x - y) ^ 2 + 4 * xy ^ 2))))
    b = Sqr(x / m - 2 * (xy) ^ 2 / (m * (-1 * (x - y) + Sqr((x - y) ^ 2 + 4 * xy ^ 2))))
2  Else
    a = Sqr(x / m + 2 * (xy) ^ 2 / (m * (-1 * (y - x) + Sqr((y - x) ^ 2 + 4 * xy ^ 2))))
    b = Sqr(y / m - 2 * (xy) ^ 2 / (m * (-1 * (y - x) + Sqr((y - x) ^ 2 + 4 * xy ^ 2))))
    End If
End Sub

```

3.12 Rotation

The subroutine `Angle_of_Rotation()` calculates the angle of rotation θ in Equation (5.10). It also gives the axis of rotation. The first four VBA statements calculate the angle of rotation with the plus and minus sign accounted for. The angle of rotation θ is stored in the variables `theta1` and `theta2`. The IF-THEN-ELSE statement (2) determines the axis of rotation using the values `theta1` and `theta2`. If `theta1 > theta2` then the axis of rotation is the Y-axis because the square root term is positive. Otherwise, the square root term is negative (or equal) and the axis of rotation is the X-axis.

```

Sub Angle_of_Rotation(x, y, xy, note, atheta, itheta)
'calculate the rotation from the axes
'x= sum of diffs from mean center squared x direction (input)
'y= sum of diffs from mean center squared y direction (input)
'xy= diff sums from mean centers x and y direction (input)
'note= states direction of rotation (output)
'atheta= angle of rotation in the x direction (output)
'itheta= angle of rotation in the y direction (output)

tan_theta1 = -1 * (x - y) / (2 * xy) + Sqr((x - y) ^ 2 + 4 * xy ^ 2) / (2 * xy)
tan_theta2 = -1 * (x - y) / (2 * xy) - Sqr((x - y) ^ 2 + 4 * xy ^ 2) / (2 * xy)
1 theta1 = Atn(tan_theta1) * 180 / WorksheetFunction.Pi()
theta2 = Atn(tan_theta2) * 180 / WorksheetFunction.Pi()
    If theta1 > theta2 Then
        note = "Rotate on Y-Axis"
        atheta = theta1
        itheta = theta2
    2 Else
        note = "Rotate on X-Axis"
        atheta = theta2
        itheta = theta1
    End If
End Sub

```

3.13 Base 10 Logarithm

The VBA logarithm function has a natural exponential base. VBA for Excel does not provide a means to change the base of the logarithm. If the programmer wishes to change the base, then the programmer must write a function for that particular logarithm with that particular base. The static function $\text{Log}_{10}(X)$ returns the logarithm of the number X with the base 10.

```

1 Static Function Log10(X)
    Log10 = Log(X) / Log(10#)
End Function

```

3.14 Stratification

The regression model does not have restrictions on the data. The latitude and longitude can be negative and positive in the same model. For the exponential model and

the Weibull model, the latitude and the longitude must be positive. The OECD example in Section 4.4 is the first example to contain the full range of data. The other data sets contained mixed data where either the latitude had a full range $-90 \leq x_i \leq 90$ or the longitude had a full range $-180 \leq y_i \leq 180$, but never in the same data set.

For this reason, we must define strata for the exponential model and the Weibull model to return the same results as the regression model. An estimator will combine the strata estimates into a single estimate. Table 3.2 gives the strata definitions for the OECD data. The definitions work for both the latitude and longitude calculations. The reader will have to separate the data, as the software does not give strata estimates directly.

Table 3.2 Stratification Definitions

Latitude	Longitude	Stratum
+	+	1
-	+	2
+	-	3
-	-	4

3.14.1 Latitude Estimates

The procedure for creating a latitude estimate for the exponential model and the Weibull model is as follow:

1. Stratify the data based on sign. This creates four strata. See Table 3.2.
2. Calculate stratum averages.
3. Combine stratum averages for latitude by:
 - (a) Stratum 1 and 3 latitude average gives $\bar{x}_{1,3}$.
 - (b) Stratum 2 and 4 abs(latitude) average gives $\bar{x}_{2,4}$.
 - (c) Multiply the averages by the number of observations n_i in each strata where $\sum_{i=1}^n n_i = n$.
 - (d) Average $\frac{n_i \bar{x}_{1,3} + n'_i \bar{x}_{2,4}}{n}$. Note the average of the difference (the sign is negative for strata 2,4).

Example 12: Consider the example in Table 3.3. It shows how to obtain a consistent estimate when the signs change for the latitude. This example also accounts for unequal sample sizes in each stratum. $\bar{x} = -2.65$ is incorrect because it assumes equal sample sizes in the strata. To obtain $\bar{x} = -0.34$,

$$\bar{x} = \frac{9 \times 8.9 + 6 \times -14.2}{15} = -0.34.$$

Table 3.3 Example of Estimation with Stratification on Latitude

Estimate	n_i	Regression	Exp	Weibull	Comment
$\bar{x}_{1,3}$	9	8.9	8.9	8.9	Latitude strata 1,3 always positive
$\bar{x}_{2,4}$	6	-14.2	14.2	14.2	Latitude strata 2,4 always negative
\bar{x}	15	-0.34	-0.34	-0.34	

3.14.2 Longitude Estimates

The procedure for creating a longitude estimate for the exponential model and the Weibull model is as follow:

1. Stratify the data based on sign. This creates four strata. See Table 3.2.
2. Calculate stratum averages.
3. Combine stratum averages for longitude by:
 - (a) Stratum 1 and 2 longitude average gives $\bar{y}_{1,2}$.
 - (b) Stratum 3 and 4 abs(longitude) average gives $\bar{y}_{3,4}$.
 - (c) Multiply the averages by the number of observations n_i in each strata where $\sum_{i=1}^n n_i = n$.
 - (d) Average $\frac{n_i \bar{y}_{1,2} + n_i' \bar{y}_{3,4}}{n}$. Note the average of the difference (the sign is negative for strata 3,4).

Example 13: Consider the example in Table 3.4. It shows how to obtain a consistent estimate when the signs change for the longitude. The answer $\bar{y} = 40.775$ is incorrect. That answer assumes equal stratum sizes. The correct longitude estimate is

$$\bar{y} = \frac{9 \times 102.3 + 6 \times -20.75}{15} = 53.08$$

Table 3.4 Example of Estimation with Stratification on Longitude

Estimate	n_i	Regression	Exp	Weibull	Comment
$\bar{y}_{1,2}$	9	102.3	102.3	102.3	Longitude strata 1,2 always positive
$\bar{y}_{3,4}$	6	-20.75	20.75	20.75	Longitude strata 3,4 always negative
\bar{y}	15	53.08	53.08	53.08	

3.15 Coordinate Systems

The Cartesian coordinate system is the classical coordinate system taught in most textbooks. In two dimensions, we have an X -axis and a Y -axis. This is a single pole coordinate system with $(0, 0)$ at the center. In three dimensions, we have the additional Z -axis — still a single pole coordinate system.

We can alternatively define the function $\mathbf{r} = \mathbf{i}x + \mathbf{j}y$, where \mathbf{i} and \mathbf{j} are vectors defined as:

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$\mathbf{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



Figure 3.3 This figure shows the Earth using Map Point. We identify points on the Earth using latitude and longitude.

In the spherical statistics literature, a one-pole coordinate system appears most often because these coordinate systems have the most practical applications and interpretations. Suppose we are measuring the hands on a clock, a roulette wheel, an experiment where a person in a room must identify the source of a sound, and so on. All of these examples have one common theme in that they have one pole. The center of the clock is the pole. The center of the roulette wheel is the pole. The person's head is the pole. [Leong and Carlile (39)] describe the *hoop* coordinate system, which has two dimensions and one pole. They mention that data is restricted to observations lying on the sphere as opposed to measuring distance (i.e. the sound experiment).

Exact formulas exist for converting from the polar coordinate system to the Cartesian coordinate system (and vice-versa). For most problems of practicable importance from these two coordinate systems, we restrict the space to one-dimension or two-dimensions so that it becomes a *circular coordinate system*. Otherwise, interpreting the results becomes difficult.

Another set of practicable problems use latitude and longitude coordinates. We identify points on the Earth using latitude and longitude. See Figure 3.3. We look-up the latitude and longitude coordinates for our data. This is called *geo-coding*. The latitude ranges from $(-90^\circ, 90^\circ)$. The longitude ranges from $(-180^\circ, 180^\circ)$. We can always convert geographic data to circular data. The mean results can be converted back to the original coordinate system while preserving the sign. We must interpret the results such as the mean center, which is also called the *center of gravity*. Using the calculations from the spherical statistics literature, these calculations strive to project the data onto an X , Y and Z -axis as opposed to the Earth. Using the calculations from the geography and statistics literature, the center of gravity will always be in the survey area. Some careful forethought goes into what you expect from spherical statistics.

Many times the data in this textbook comes from several sources. The random variable w_i augments with a pair of coordinates called latitude x_i and longitude y_i for each observation i . We omit the altitude from the analysis. Using special software, we plot the random variable on a map using the pair (x_i, y_i) . To make calculations simple, we measure the latitude and longitude pairs (x_i, y_i) in *degrees*, then *decimal minutes and seconds* as opposed to degrees, minutes, and seconds. Having knowledge of the Cartesian coordinate system for plotting points does help to understand the subtle differences between mapping and plotting.

When drawing an ellipse on a map, knowledge of the polar coordinate system is most useful. We derive several pieces of information:

- The center of the ellipse.
- The semi-major and the semi-minor axis lengths.
- The angle of rotation from an imaginary X or Y -axis.

We have enough information to draw an ellipse and to rotate the ellipse from the imaginary axis. The imaginary axis (or axes) is simply the X and Y -axes from the Cartesian coordinate system super-imposed onto a map using specialty software. Chapter 2 discusses some of the specialty software available.

- The software comes with a ruler to draw lines. The software can measure lines in miles or kilometers. We use the ruler to draw the semi-major and semi-minor axes.
- The user must use another software product to draw and rotate the ellipse.

3.16 Testing for Randomness

This section presents a simple test for randomness for a given probability distribution. The statistical test statistic bases the test on the chi-square distribution. Given a random sample from a known distribution, we test the latitude with the following test statistic in Equation 3.14.

$$V = \sum_{i=1}^n \frac{[x_i - E(x_i)]^2}{Var(x_i)} \sim \chi_{0.05}^2(n-1) \quad (3.11)$$

where $E(x_i)$ is the expected value of the sample data and $Var(x_i)$ is the variance of the sample data. Both $E(x_i)$ and $Var(x_i)$ may depend on an assumed, underlying distribution. The test for the longitude is similar to Equation (3.14). Simply replace x_i with y_i .

Example 14: Consider the regression model. The weighted expected value $E(x_i) = \bar{x}$ and the weighted error $Var(x_i) = e_x^2$. Then, the test for randomness v becomes

$$V = \frac{1}{n-1} \sum_{i=1}^n e_x^2 = \frac{1}{n-1} \sum_{i=1}^n w_i [x_i - \bar{x}]^2 \sim \chi_{0.05}^2(n-1) \quad (3.12)$$

for the latitude. The weighted estimator for \bar{x} will be given in Chapter 5. Similar results hold true for the longitude.

Example 15: Consider the exponential model. The weighted expected value $E(x_i) = \bar{x}$ and the weighted variance $Var(x_i) = \bar{x}^2$. Then, the test for randomness v becomes

$$V = \sum_{i=1}^n \frac{[x_i - \bar{x}]^2}{\bar{x}^2} \sim \chi_{0.05}^2(n-1) \quad (3.13)$$

for the latitude. The weighted estimator for \bar{x} will be given in Chapter 6. Similar results hold true for the longitude.

Example 16: Consider the Weibull model. The weighted expected value $E(x_i) = \bar{x}$. The weighted variance $Var(x_i) = \left(\frac{1}{\bar{x}}\right)^2 \left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma\left(1 + \frac{1}{\gamma}\right)^2 \right]$. Then, the test for randomness v becomes

$$V = \sum_{i=1}^n \frac{[x_i - \bar{x}]^2}{\left(\frac{1}{\bar{x}}\right)^2 \left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma\left(1 + \frac{1}{\gamma}\right)^2 \right]} \sim \chi_{0.05}^2(n-1) \quad (3.14)$$

for the latitude. The weighted estimator for \bar{x} will be given in Chapter 7.

3.17 Sampling without Replacement

Sampling without replacement is a finite sampling scheme. Select an observation and assign a probability from the sample; do not return the observation to the sample. Given this, the total number of observations changes as removing an observation reduces the sample size. This in turn, changes the probability of selection for each sample element. This is different from sampling with replacement where the number of observations, called the sample size n , does not change.

Under the sampling without replacement scheme and using the airline data in Chapter 4, we assign weights in the Excel spreadsheet as follow:

1. Assign the departure location (or closest observation) with the largest weight of n .
2. Assign the next closest location to the departure location a weight of $n - 1$.
3. Assign the first closest location to the arrival location a weight of 2.
4. Assign the arrival location (or closest observation) with the smallest weight of 1.

Upon running the VBA code, the program will calculate the following weights for the regression model and the exponential model:

$$\frac{w_i}{\sum_{i=1}^n w_i} = \frac{j}{\sum_{i=1}^n i} = \frac{j}{n(n+1)/2}, j = n, n-1, \dots, 1.$$

The weights for the Weibull model have a shape parameter. Chapter 7 discusses the Weibull model.

CHAPTER 4

GEO-CODING

The data used in this book comes from many sources. The majority of the data resides on government web sites today. It would have been nice if the data were prepared before downloading it. The data provided (free) did have the random variables of interest. However, the data did not have the latitude and longitude observations. Thus, the data must be *geo-coded* before performing any analyzes in this textbook.

Existing textbook authors use either outdated software or expensive software to calculate their statistics. Outdated software such as FORTRAN 77 and 10 Statement Fortran has been around since at least the 1960's. Arc GIS and Map Soft are modern Geographic Software Information systems that can geo-code data; however, products such as those are expensive. We can always geo-code the data free on Google Earth — but this becomes time consuming. Academic software products do exist such as *CrimeStat*. It performs many statistical calculations including weighted calculations. That one comes with an extensive software manual, too. Map Point, a Microsoft product, is fairly inexpensive. The user can save work to the hard drive. The user can copy and paste maps into other Microsoft products to draw the ellipse and to rotate it. [Hammond (27)] has a thirty-one page printed "Student Notebook Atlas" of the world. The atlas has the land masses enclosed in a box with the latitude and longitude references on the sides of the box. The Hammond Atlas is an inexpensive

product, too. It may be difficult to get the same precise coordinates that the software products give.

It is a trade-off among data precision, time, cost, and scope of the research as to what software one chooses. Data is data — of course, obtain it in the most pristine condition possible before performing the research.

Do we need the software or data provider to geo-code the data and calculate the statistics or just geo-code the data? In this case, we just want the data geo-coded. We will write the code for calculating the statistics for the various ellipses in Visual Basic for Applications (VBA) in Excel 2010. This application is further accessible on Android tablets and smart phones using *Quick Office*.

Most of the geo-coding in this textbook was done using Google Earth. Begin by downloading the software at <http://www.google.com/earth/index.html>. Google Earth periodically updates their images of the Earth. This causes slight differences in latitude and longitude readings. As long as the data is geo-coded under one or another satellite images, it does not alter the survey data or the statistical techniques presented in this textbook. Figure 4.1 shows Google Earth. As an example, for the June Area Survey in Kentucky, the user types-in the county and state in the upper-left corner of Google Earth. Google Earth returns the decimal degrees of the county and state in the lower-right corner.

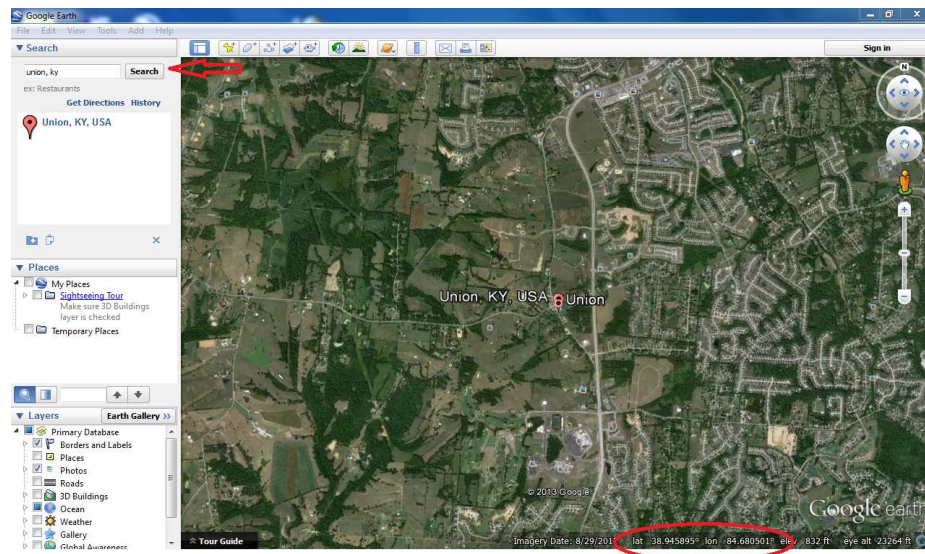


Figure 4.1 This figure shows Google Earth. It shows how to geo-code a county within a state.

Similarly, the user can type in a state name to obtain the geo-coding or the user can type in a country name such as Canada to obtain the geo-coding. It is important not to zoom into or out of the Earth when collecting these coordinates. Otherwise,

the altitude (which is not being collected) will change. This in turn will change the latitude and longitude observations.

4.1 Kentucky June Area Survey

This is a survey conducted by the US Department of Agriculture (USDA). Within the USDA, the National Agricultural Statistics Service (NASS) supervises and oversees the data collection and reporting responsibilities for this survey. Their website is <http://www.nass.usda.gov/index.asp>. Plenty of data is available and downloadable that is ready to use with little or no formatting issues.

Experience says the Area Frame Section at USDA / NASS uses a stratified survey design to obtain population totals, standard errors, and CVs. The June Area Survey is conducted every year to measure, among other things, corn acreage, cotton acreage, soybean acreage, three seasons of wheat acreage, NOL cattle, and the number of farms in the United States. Each state has a set of land use strata specific to it. Within each strata, replicates are assigned to sub-strata. A sample of land called a segment is finally sampled. Since a minimum of two replicates is assigned to the sub-strata, the minimum number of segments per strata is two (assuming one sub-strata). Segment sizes range from 1/10th of a square mile to 1 square mile (640 acres per mile). Often, multiple sub-strata are assigned to each stratum. A complete, comprehensive set of designs for the United States can be found in the Area Frame Design Books.

[Kelly (33)] describes the June Area Survey conducted by the USDA / National Agricultural Statistics Service. That paper describes techniques intended to identify differences in variation, not to visualize them. Given that the spatial data is available and the computations described so far, it is possible to visualize the variation in the data. The author draws the sample from a sample of land called a segment. Segment sizes range from 1/10th of a square mile to 1 square mile (640 acres per mile). Enumerators completely cover each segment during the survey. Cartographers and survey researchers know every segments center point coordinates. Instead of using the simplistic weighted estimators for the total crop acreage where the weights are simply the total number of segments divided by the number of sampled segments, the alternative weights will be defined using spatial data. We will define alternative weights using spatial data, instead of using the simplistic weighted estimators for the total crop acreage where the weights are simply the total number of segments divided by the number of sampled segments. From these alternative weights, we can derive what are called what are called *weighted mean centers* and *weighted standard deviational ellipses*.

We obtain the data in this book by augmenting the NASS data available to the public with Google Earth¹ coordinates for each county. Data consumers can easily download the crop data. Figure 4.2 shows the data downloaded from the NASS web

¹The web site is www.earthgoogle.com. Decimal coordinates are preferred since we can more easily integrate them than degrees, minutes, and seconds.

site. Regional level observations have arrows identifying them. The yellow arrows point to the following formatting issues for us.

- A. The arrow points to non-geographical data. For analytical purposes, we delete non-geographical observations.
- B. The arrow points to the sum of a region. For analytical purposes, we delete regional observations.

Obtaining the latitude and longitude for the counties does take time. Given that, our sample will be the county-level data in Kentucky for corn for grain. We are not so much interested in the quantity of acreage of corn grown in June for the state. *We are interested in the location that corresponds to that single quantity for the state.* Hence, we augment the data with latitude and longitude coordinates.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Commodity	Practice	Year	State	cofips	StFips	District	county	CommCode	PracCode	Planted All Purposes	Harvested	Yield	Production
2	Corn For Grain	Total For Crop	2003	Kentucky	Ballard	21	10	7	11199199	9	26100	25500	128	326400
3	Corn For Grain	Total For Crop	2003	Kentucky	Calloway	21	10	35	11199199	9	38300	37600	112	421120
4	Corn For Grain	Total For Crop	2003	Kentucky	Carlisle	21	10	39	11199199	9	23100	22700	141	320070
5	Corn For Grain	Total For Crop	2003	Kentucky	Fulton	21	10	75	11199199	9	26300	24200	144	348480
6	Corn For Grain	Total For Crop	2003	Kentucky	Graves	21	10	83	11199199	9	60400	59400	131	778140
7	Corn For Grain	Total For Crop	2003	Kentucky	Hickman	21	10	105	11199199	9	38800	37000	138	510600
8	Corn For Grain	Total For Crop	2003	Kentucky	Livingston	21	10	139	11199199	9	7000	6600	114	75240
9	Corn For Grain	Total For Crop	2003	Kentucky	Lyon	21	10	143	11199199	9	6000	5700	132	75240
10	Corn For Grain	Total For Crop	2003	Kentucky	McCracken	21	10	145	11199199	9	12500	12000	117	140400
11	Corn For Grain	Total For Crop	2003	Kentucky	Marshall	21	10	157	11199199	9	9300	8700	100	87000
12	Corn For Grain	Total For Crop	2003	Kentucky	Trigg	21	10	221	11199199	9	21300	21100	126	265860
13	Corn For Grain	Total For Crop	2003	Kentucky	D10 Purchase	21	10	999	11199199	9	269100	260500	128.5	3348550
14	Corn For Grain	Total For Crop	2003	Kentucky	Caldwell	21	20	33	11199199	9	24400	23700	120	284400
15	Corn For Grain	Total For Crop	2003	Kentucky	Christian	21	20	47	11199199	9	75200	73700	144	1061280
16	Corn For Grain	Total For Crop	2003	Kentucky	Crittenden	21	20	55	11199199	9	9000	8400	124	104160
17	Corn For Grain	Total For Crop	2003	Kentucky	Daviess	21	20	59	11199199	9	56000	54700	139	760330
18	Corn For Grain	Total For Crop	2003	Kentucky	Hancock	21	20	91	11199199	9	6500	6500	118	76700
19	Corn For Grain	Total For Crop	2003	Kentucky	Henderson	21	20	101	11199199	9	63900	61800	140	865200
20	Corn For Grain	Total For Crop	2003	Kentucky	Hopkins	21	20	107	11199199	9	27100	26900	135	363150
21	Corn For Grain	Total For Crop	2003	Kentucky	Logan	21	20	141	11199199	9	56900	55000	168	924000
22	Corn For Grain	Total For Crop	2003	Kentucky	McLean	21	20	149	11199199	9	34800	34500	153	527850
23	Corn For Grain	Total For Crop	2003	Kentucky	Muhlenberg	21	20	177	11199199	9	12100	11900	123	146370
24	Corn For Grain	Total For Crop	2003	Kentucky	Ohio	21	20	183	11199199	9	24300	24000	141	338400
25	Corn For Grain	Total For Crop	2003	Kentucky	Simpson	21	20	213	11199199	9	36100	35100	163	572130
26	Corn For Grain	Total For Crop	2003	Kentucky	Todd	21	20	219	11199199	9	44700	43200	151	652320
27	Corn For Grain	Total For Crop	2003	Kentucky	Union	21	20	225	11199199	9	77300	75300	139	1046670
28	Corn For Grain	Total For Crop	2003	Kentucky	Webster	21	20	233	11199199	9	30800	30400	147	446880
29	Corn For Grain	Total For Crop	2003	Kentucky	D20 Midwestern	21	20	999	11199199	9	579100	565100	144.6	8169840
30	Corn For Grain	Total For Crop	2003	Kentucky	Adair	21	30	1	11199199	9	7800	4500	83	37350
31	Corn For Grain	Total For Crop	2003	Kentucky	Allen	21	30	3	11199199	9	3000	2700	147	39690
32	Corn For Grain	Total For Crop	2003	Kentucky	Barren	21	30	9	11199199	9	15700	9500	135	128250

Figure 4.2 This figure shows the Excel spreadsheet of the Kentucky data. The spreadsheet has not been geo-coded. It still contains the regional data.

The data lacks latitude and longitude observations. These observations are obtainable from Google Earth. Since, the geography attached to the observations from USDA are at the county level, the most precision we can hope for from Google Earth is the latitude and longitude of the counties in the reported data for the given counties.

Table 4.1 shows the data for 2003. It contains the county names, the geo-coding, the stratum, and the acreage of crops planted for 2003. The 2004 data set is similar

and contains a different number of observations. For analysis purposes, drop the county name and FIPS code from the data set.

Table 4.1 Acreage Planted Data

County	County FIPS	Transformed Latitude	Transformed Latitude	Transformed Longitude	Transformed Longitude	Stratum	2003 Planted
Adair	1	37.08	37.08	-85.31	274.69	3	7,800
Allen	3	37.61	37.61	-82.73	277.27	3	3,000
Ballard	7	37.04	37.04	-89.02	270.98	3	26,100
Barren	9	36.97	36.97	-85.85	274.15	3	15,700
Bath	11	38.11	38.11	-83.72	276.28	3	2,600
Boone	15	38.99	38.99	-84.73	275.27	3	2,200
Bourbon	17	38.22	38.22	-84.23	275.77	3	3,800
Boyle	21	37.65	37.65	-84.82	275.18	3	2,300
Bracken	23	38.71	38.71	-84.06	275.94	3	1,500
Breckinridge	27	37.80	37.80	-86.46	273.54	3	11,700
Bullitt	29	37.98	37.98	-85.68	274.32	3	2,200
Butler	31	38.79	38.79	-84.37	275.63	3	12,900
Caldwell	33	37.15	37.15	-87.89	272.11	3	24,400
Calloway	35	36.64	36.64	-88.29	271.71	3	38,300
Campbell	37	38.90	38.90	-84.40	275.60	3	600
Carlisle	39	38.31	38.31	-84.03	275.97	3	23,100
Carroll	41	38.67	38.67	-85.06	274.94	3	800
Casey	45	37.32	37.32	-84.90	275.10	3	4,100
Christian	47	36.84	36.84	-87.47	272.53	3	75,200
Clark	49	37.96	37.96	-84.14	275.86	3	1,500
Clinton	53	36.67	36.67	-88.99	271.01	3	1,200
Crittenden	55	38.78	38.78	-84.61	275.39	3	9,000
Cumberland	57	36.98	36.98	-82.99	277.01	3	1,600
Daviess	59	37.73	37.73	-87.10	272.90	3	56,000
Edmonson	61	37.19	37.19	-86.22	273.78	3	2,600
Fayette	67	38.06	38.06	-84.48	275.52	3	2,500
Fleming	69	38.40	38.40	-83.68	276.32	3	4,300
Franklin	73	36.73	36.73	-86.58	273.42	3	1,200
Fulton	75	36.51	36.51	-88.88	271.12	3	26,300
Gallatin	77	38.73	38.73	-84.88	275.12	3	700
Garrard	79	37.64	37.64	-84.56	275.44	3	1,600
Graves	83	36.69	36.69	-88.71	271.29	3	60,400
Grayson	85	38.33	38.33	-82.95	277.05	3	7,200
Green	87	37.26	37.26	-85.56	274.44	3	4,800
Greenup	89	38.57	38.57	-82.83	277.17	3	1,600
Hancock	91	37.83	37.83	-86.76	273.24	3	6,500
Hardin	93	36.77	36.77	-88.30	271.70	3	24,500
Harrison	97	38.43	38.43	-84.35	275.65	3	3,300
Hart	99	37.31	37.31	-85.85	274.15	3	4,400

Table 4.2 Acreage Planted Data (con't)

County	County FIPS	Latitude	Transformed Latitude	Longitude	Transformed Longitude	Stratum	2003 Planted
Henderson	101	37.84	37.84	-87.58	272.42	3	63,900
Henry	103	38.43	38.43	-85.15	274.85	3	3,500
Hickman	105	36.57	36.57	-89.19	270.81	3	38,800
Hopkins	107	37.32	37.32	-87.58	272.42	3	27,100
Jefferson	111	38.19	38.19	-85.64	274.36	3	1,100
Jessamine	113	37.90	37.90	-84.56	275.44	3	900
Knox	121	36.93	36.93	-83.89	276.11	3	800
Larue	123	37.52	37.52	-85.73	274.27	3	15,200
Laurel	125	37.07	37.07	-84.19	275.81	3	1,400
Lewis	135	38.51	38.51	-83.38	276.62	3	1,600
Lincoln	137	37.48	37.48	-84.65	275.35	3	8,600
Livingston	139	37.30	37.30	-84.21	275.79	3	7,000
Logan	141	36.87	36.87	-86.86	273.14	3	56,900
Lyon	143	37.02	37.02	-88.09	271.91	3	6,000
McCracken	145	37.03	37.03	-88.71	271.29	3	12,500
McLean	149	37.56	37.56	-87.26	272.74	3	34,800
Madison	151	37.71	37.71	-84.31	275.69	3	2,500
Marion	155	37.33	37.33	-88.08	271.92	3	11,100
Marshall	157	36.86	36.86	-88.40	271.60	3	9,300
Mason	161	38.62	38.62	-83.89	276.11	3	5,600
Meade	163	37.96	37.96	-86.22	273.78	3	10,700
Mercer	167	37.83	37.83	-84.90	275.10	3	2,600
Metcalfe	169	37.00	37.00	-85.64	274.36	3	4,900
Monroe	171	36.75	36.75	-85.73	274.27	3	5,000
Montgomery	173	38.03	38.03	-83.89	276.11	3	1,600
Muhlenberg	177	37.18	37.18	-87.14	272.86	3	12,100
Nelson	179	37.33	37.33	-87.05	272.95	3	12,500
Nicholas	181	38.38	38.38	-84.06	275.94	3	1,200
Ohio	183	37.51	37.51	-86.82	273.18	3	24,300
Oldham	185	38.36	38.36	-85.48	274.52	3	4,100
Pendleton	191	38.46	38.46	-85.30	274.70	3	1,400
Pulaski	199	37.09	37.09	-84.52	275.48	3	8,800
Rockcastle	203	37.37	37.37	-84.31	275.69	3	1,300
Rowan	205	38.18	38.18	-83.46	276.54	3	1,100
Russell	207	38.51	38.51	-82.70	277.30	3	4,800
Scott	209	38.32	38.32	-84.56	275.44	3	2,200
Shelby	211	38.18	38.18	-85.23	274.77	3	14,300
Simpson	213	36.78	36.78	-86.62	273.38	3	36,100

Table 4.3 Acreage Planted Data (con't)

County	County FIPS	Transformed Latitude	Transformed Longitude	Stratum	2003 Planted
Spencer	215	38.01	-85.31	3	3,200
Taylor	217	37.33	-85.31	3	8,200
Todd	219	36.83	-87.14	3	44,700
Trigg	221	36.85	-87.78	3	21,300
Trimble	223	38.60	-85.31	3	1,100
Union	225	38.95	-84.68	3	77,300
Warren	227	36.99	-86.50	3	27,900
Washington	229	37.75	-85.15	3	3,900
Wayne	231	36.76	-84.86	3	5,500
Webster	233	37.49	-87.74	3	30,800
Whitley	235	36.72	-84.47	3	800
Wolfe	237	37.76	-83.46	3	600
Woodford	239	38.07	-84.73	3	1,700

4.2 US Crime Statistics

The Department of Justice established the Bureau of Justice Statistics (BJS) on December 27, 1979, under the Justice Systems Improvement Act of 1979 as an amendment to the Omnibus Crime Control and Safe Streets Act of 1968. According to the Bureau of Justice Statistic’s website <http://www.bjs.gov/>, their mission is to collect, analyze, publish, and disseminate information on crime, criminal offenders, victims of crime, and the operation of justice systems at all levels of government.

This data appears on the Fed Stats web site. Figure 4.3 shows the downloaded data from the Department of Justice. Examining the left column (Column A), some of the rows are summarized to the regional level . The data provider merged the headings across the top. The data include Puerto Rico and Hawaii. These states and islands are disconnected from the US. We are interested in the main-land part of the United States.

Area	Year	Population ¹	Violent crime		Murder and nonnegligent manslaughter		Forcible rape		Robbery		Aggravated assault	
			Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000
United States Total ^{2,3,4}	2007	301,621,157	1,408,337	466.9	16,929	5.6	90,427	30.0	445,125	147.6	855,856	283.8
	2008	304,059,724	1,382,012	454.5	16,272	5.4	89,000	29.3	441,855	145.3	834,885	274.6
	Percent change			-1.9	-2.7	-3.9	-4.7	-1.6	-2.4	-0.7	-1.5	-2.5
Northeast	2007	54,680,626	203,632	372.4	2,261	4.1	10,821	19.8	75,326	137.8	115,224	210.7
	2008	54,924,779	203,654	370.8	2,293	4.2	10,981	20.0	76,048	138.5	114,332	208.2
	Percent change			*	-0.4	+1.4	+1.0	+1.5	+1.0	+0.5	-0.8	-1.2
New England	2007	14,264,185	43,334	303.8	357	2.5	3,395	23.8	12,225	85.7	27,357	191.8
	2008	14,303,542	46,682	326.4	380	2.7	3,580	25.0	12,696	88.8	30,026	209.9
	Percent change			+7.7	+7.4	+6.4	+6.1	+5.4	+5.2	+3.9	+3.6	+9.8
Connecticut	2007	3,502,309	8,965	256.0	106	3.0	658	18.8	3,607	103.0	4,594	131.2
	2008	3,501,252	10,427	297.8	123	3.5	674	19.3	3,907	111.6	5,723	163.5
	Percent change			+16.3	+16.3	+16.0	+16.1	+2.4	+2.5	+8.3	+8.3	+24.6
Maine	2007	1,317,207	1,554	118.0	21	1.6	391	29.7	349	26.5	793	60.2
	2008	1,316,456	1,547	117.5	31	2.4	375	28.5	333	25.3	808	61.4
	Percent change			-0.5	-0.4	+47.6	+47.7	-4.1	-4.0	-4.6	-4.5	+1.9
Massachusetts	2007	6,449,755	27,832	431.5	184	2.9	1,634	25.3	7,006	108.6	19,008	294.7
	2008	6,497,967	29,174	449.0	167	2.6	1,736	26.7	7,069	108.8	20,202	310.9

Figure 4.3 This figure shows the Excel spreadsheet of the crime data. The spreadsheet has not been geo-coded and still contains the regional data, two years of data in the same spreadsheet, merged columns, and percentages.

To address cleaning the data from this spreadsheet will take more work than that for the Kentucky spreadsheet. This data contains merged cells, regional data and two years of data in the same spreadsheet. In addition, we are not interested in the percentages. We must contend with some of the problems indicated by the yellow arrows. Obviously, the data provider created this spreadsheet for presentation purposes, not analytical. The yellow arrows point to the following formatting issues for us.

- A. The arrow points to the sum of the U.S. totals for the U.S. This is the first clue that other summary data exists in the spreadsheet. We need to delete summary data.
- B. The arrow points to multiple years of data in the same spreadsheet. We need to separate this into two different spreadsheets for the algorithms in the VBA code to work properly.
- C. The arrow points to a percentage change. Although useful for a presentation, we do not need the percentage change for any analyzes in this book. We need to delete percentage change data.

Table 4.4 gives the data for violent crime. As the reader can see, all of the data appears in a single stratum. This makes the data analyses easier.

Table 4.4 Violent Crime Data

State	Transformed		Transformed		Stratum	2007	2008
	Latitude	Latitude	Longitude	Longitude		Crime	Crime
Connecticut	41.60	41.60	-73.09	286.91	3	8,965	10,427
Maine	45.25	45.25	-69.45	290.55	3	1,554	1,547
Massachusetts	42.41	42.41	-71.38	288.62	3	27,832	29,174
New Hampshire	43.19	43.19	-71.57	288.43	3	1,807	2,069
Rhode Island	41.58	41.58	-71.48	288.52	3	2,404	2,621
Vermont	44.56	44.56	-72.58	287.42	3	772	844
New Jersey	40.06	40.06	-74.41	285.59	3	28,601	28,351
New York	40.71	40.71	-74.01	285.99	3	79,915	77,585
Pennsylvania	41.20	41.20	-77.19	282.81	3	51,782	51,036
Illinois	40.63	40.63	-89.40	270.60	3	68,528	67,780
Indiana	40.55	40.55	-85.60	274.40	3	21,165	21,283
Michigan	44.31	44.31	-85.60	274.40	3	53,988	50,166
Ohio	40.42	40.42	-82.91	277.09	3	39,360	39,997
Wisconsin	43.78	43.78	-88.79	271.21	3	16,296	15,421
Iowa	41.88	41.88	-93.10	266.90	3	8,805	8,520
Kansas	39.01	39.01	-98.48	261.52	3	12,566	11,505
Minnesota	46.73	46.73	-94.69	265.31	3	15,003	13,717
Missouri	37.96	37.96	-91.83	268.17	3	29,682	29,819
Nebraska	41.49	41.49	-99.90	260.10	3	5,367	5,416
North Dakota	47.55	47.55	-101.00	259.00	3	911	1,068
South Dakota	43.97	43.97	-99.90	260.10	3	1,347	1,620
Delaware	38.91	38.91	-75.53	284.47	3	5,960	6,141
District of Columbia	38.91	38.91	-77.03	282.97	3	8,320	8,509
Florida	27.66	27.66	-81.52	278.48	3	131,880	126,265
Georgia	32.16	32.16	-82.91	277.09	3	47,075	46,384
Maryland	39.05	39.05	-76.64	283.36	3	36,062	35,393
North Carolina	35.76	35.76	-79.02	280.98	3	42,262	43,099
South Carolina	33.84	33.84	-81.16	278.84	3	34,746	32,691
Virginia	37.43	37.43	-78.66	281.34	3	20,798	19,882
West Virginia	38.60	38.60	-80.45	279.55	3	4,987	4,968
Alabama	32.32	32.32	-86.90	273.10	3	20,732	21,111
Kentucky	37.84	37.84	-84.27	275.73	3	12,513	12,646
Mississippi	32.35	32.35	-89.40	270.60	3	8,502	8,373

Table 4.5 Violent Crime Data (con't)

State	Transformed		Transformed		Stratum	2007	2008
	Latitude	Latitude	Longitude	Longitude		Crime	Crime
Tennessee	35.52	35.52	-86.58	273.42	3	46,380	44,897
Arkansas	35.20	35.20	-91.83	268.17	3	15,007	14,374
Louisiana	31.24	31.24	-92.15	267.85	3	31,317	28,944
Oklahoma	35.01	35.01	-97.09	262.91	3	18,072	19,184
Texas	31.97	31.97	-99.90	260.10	3	122,054	123,564
Arizona	34.05	34.05	-111.09	248.91	3	30,600	29,059
Colorado	39.55	39.55	-105.78	254.22	3	16,906	16,946
Idaho	44.07	44.07	-114.74	245.26	3	3,589	3,483
Montana	46.88	46.88	-110.36	249.64	3	2,754	2,497
Nevada	38.80	38.80	-116.42	243.58	3	19,257	18,837
New Mexico	34.97	34.97	-105.03	254.97	3	13,085	12,896
Utah	39.32	39.32	-111.09	248.91	3	6,210	6,070
Wyoming	43.08	43.08	-107.29	252.71	3	1,251	1,236
Alaska	63.59	63.59	-154.49	205.51	3	4,519	4,474
California	36.78	36.78	-119.42	240.58	3	191,025	185,173
Oregon	43.80	43.80	-120.55	239.45	3	10,777	9,747
Washington	47.75	47.75	-120.74	239.26	3	21,546	21,691

4.3 Airline Data

Airline data does not have an obvious variable to weight with. There are several options:

1. Assign equal weights to all observations.
2. Assign larger weights to the observations closest to the departure location and smaller weights closest to the arrival location. See Section 3.17. This is a pessimistic weighting scheme.

4.3.1 Air France Flight 447 Data

This airline was a scheduled passenger flight from Rio de Janeiro, Brazil to Paris, France, which crashed on June 1, 2009. This airplane was eventually found. The website <http://www.weathergraphics.com/tim/af447/> has coordinates.

Table 4.6 shows the data for Air France Flight 447. This data set has two strata, which makes it more difficult to model than the acreage planted data and the violent crime data.

Table 4.6 Air France Flight 447 Data

	Transformed		Transformed	
Latitude	Latitude	Longitude	Longitude	Stratum
-1.65	178.35	-32.98	327.02	4
-0.49	179.51	-32.37	327.63	4
0.64	0.64	-31.76	328.24	3
1.682	1.682	-31.209	328.791	3
1.8	1.8	-31.15	328.85	3
2.726	2.726	-30.676	329.324	3
2.98	2.98	-30.59	329.41	3
3.103	3.103	-30.539	329.461	3

4.3.2 Malaysia Flight MH 370 Data

The data for this example appeared in the article [Ashton, Bruce, Colledge, and Dickinson (1)]. MH 370 refers to the airplane that took-off from a Malaysian airport to China. The airplane never arrived to its intended destination. As of June 12, 2015, the airplane is still missing. [Ashton, et al (1)] constructed a data set of 13 observations using Inmarsat satellite signals.

Table 4.7 shows the data for Flight MH 370. It has two strata, which makes it more difficult to model. This data has a latitude observation of zero. Knowing that

the exponential distribution and the Weibull distribution will not accept such a value, we omit the latitude of zero from the data analyses.

Table 4.7 Malaysia MH 370 Data

Latitude	Transformed Latitude	Longitude	Transformed Longitude	Stratum
2.7	2.7	101.7	101.7	1
2.8	2.8	101.7	101.7	1
4	4	102.2	102.2	1
5.3	5.3	102.8	102.8	1
6.9	6.9	95.8	95.8	1
7.5	7.5	94.4	94.4	1
0	0	93.7	93.7	1
-7.5	172.5	92.9	92.9	2
-15	165	92.7	92.7	2
-22.5	157.5	92.9	92.9	2
-26.6	153.4	93	93	2
-33.7	146.3	93	93	2
-34.7	145.3	93	93	2

4.4 OECD Countries

The Organization for Economic Cooperation and Development (OECD) has its headquarters in Paris, France; the organization produces 250 publications per year. Europe created OECD as a successor to the Marshall Plan for the reconstruction of Europe after World War II. OECD countries are committed to democracy and the market economy. This data comes from the *Organization for Cooperation and Economic Development* website at www.oecd.org. Obviously, some very large economies are missing from the data. Non-OPEC countries such as India are missing. Most OPEC countries (Saudi Arabia, Egypt, Iran, Libya, etc.) are missing. This data does provide a global aspect for the computations in this textbook. We must perform additional pre-computations on the latitude and longitude before performing any proposed calculations.

We need knowledge of the bounds on the latitude and longitude before modeling this data set. In prior data sets, the longitude was consistently negative. We can simply change the sign to positive, and then carry out certain computations. This is not the case with the OECD data. It has mixed signs for both the latitude and the longitude. See Figure 4.4. An understanding of the Prime Meridian and the Equator measurements helps. Very little formatting is necessary compared to the US Crime data to use it "as-is."

Table 4.8 shows the OECD data. It contains at least one observation from all four stratum defined in Section 3.14. This makes it more difficult to model and calculate the statistics and measures associated with the weighted ellipses. Chapter 8 shows that this data is bi-modal.

Table 4.8 Gross Domestic Product for OECD Countries

Country	Transformed		Transformed		Stratum	2007	2008	2009
	Latitude	Latitude	Longitude	Longitude		GDP	GDP	GDP
Australia	-25.47	154.53	133.78	133.78	2	831,024	848,668	884,706
Austria	47.52	47.52	14.55	14.55	1	316,109	331,804	325,192
Belgium	50.50	50.50	4.47	4.47	1	378,707	396,187	393,108
Canada	55.94	55.94	-106.35	253.65	3	1,263,042	1,298,956	1,277,080
Chile	-35.68	144.32	-71.54	288.46	4	230,663	244,217	242,429
Czech Republic	49.82	49.82	15.47	15.47	1	262,841	281,204	276,566
Denmark	56.26	56.26	9.50	9.50	1	206,010	216,902	208,068
Estonia	58.60	58.60	25.01	25.01	1	28,979	29,593	26,601
Finland	61.92	61.92	25.75	25.75	1	191,319	201,911	190,363
France	46.23	46.23	2.21	2.21	1	2,114,425	2,178,484	2,152,421
Germany	51.17	51.17	10.45	10.45	1	2,925,790	3,043,354	2,946,618
Greece	39.07	39.07	21.82	21.82	1	310,212	332,274	325,875
Hungary	47.16	47.16	19.50	19.50	1	190,431	206,172	199,844
Iceland	64.96	64.96	-19.02	340.98	3	11,577	12,498	11,718
Ireland	53.41	53.41	-8.24	351.76	3	198,309	189,464	177,818
Israel	31.05	31.05	34.85	34.85	1	189,867	201,662	205,254
Italy	41.87	41.87	12.57	12.57	1	1,893,938	1,990,548	1,951,037
Japan	36.20	36.20	138.25	138.25	1	4,290,184	4,316,608	4,082,561
Korea	35.91	35.91	127.77	127.77	1	1,269,102	1,306,387	1,322,666
Luxembourg	49.82	49.82	6.13	6.13	1	40,580	43,574	41,441
Mexico	23.63	23.63	-102.55	257.45	3	1,540,937	1,629,595	1,545,856
Netherlands	52.13	52.13	5.29	5.29	1	667,311	703,006	673,383
New Zealand	-40.90	139.10	174.89	174.89	2	121,154	124,478	126,272
Norway	60.47	60.47	8.47	8.47	1	259,029	288,428	263,400
Poland	51.92	51.92	19.15	19.15	1	638,908	688,458	721,873
Portugal	39.40	39.40	-8.22	351.78	3	256,787	265,100	266,395
Slovak Republic	48.67	48.67	19.70	19.70	1	112,684	125,498	123,578
Slovenia	46.15	46.15	15.00	15.00	1	54,961	59,084	56,089
Spain	40.46	40.46	-3.75	356.25	3	1,447,256	1,512,485	1,481,082
Sweden	60.13	60.13	18.64	18.64	1	352,073	363,958	346,025
Switzerland	46.82	46.82	8.23	8.23	1	325,737	351,515	349,794
Turkey	38.96	38.96	35.24	35.24	1	976,374	1,063,519	1,022,261
United Kingdom	55.38	55.38	-3.44	356.56	3	2,178,371	2,260,520	2,172,047
United States	37.09	37.09	-95.71	264.29	3	13,961,800	14,219,300	13,863,600
China	35.86	35.86	104.20	104.20	1	7,338,772	8,218,933	9,046,410
India	20.59	20.59	78.96	78.96	1	3,264,222	3,446,276	3,884,404
Indonesia	179.21	179.21	113.92	113.92	1	840,971	911,027	961,445
Russian Federation	60.60	60.60	105.32	105.32	1	2,387,519	2,878,201	2,677,803
South Africa	149.44	149.44	22.94	22.94	1	480,841	508,870	504,946

	A	B	F	G	H	I	J
1	Economics: Key tables from OECD - ISSN 2074-384x - © OECD 2011						
2	Gross domestic product in US dollars						
3	At current prices and current PPPs						
4	Billions of US dollars						
5			2006	2007	2008	2009	2010
6	Australia		774.3	828.2	847.2	876.5 e	901.5 e
7	Austria		299.9	313.8	332.2	324.7	334.5
8	Belgium		360.1	378.1	394.9	391.8	405.9
9	Canada		1 200.5	1 263.0	1 295.9	1 275.6	1 331.3 e
10	Chile		214.2	230.7	244.2	239.2	253.8
11	Czech Republic		229.5	253.7	269.6	268.2	274.1 e
12	Denmark		195.9	206.0	216.9	208.1	215.5
13	Estonia		25.7	28.5	29.2	26.6	28.3
14	Finland		174.3	191.2	200.8	188.1	196.7
15	France		2 000.2	2 123.9	2 195.7	2 173.3	2 232.7
16	Germany		2 778.8	2 930.5	3 052.5	2 975.3	3 108.4
17	Greece		302.1	316.2	338.0	330.7	316.7
18	Hungary		184.6	192.9	207.8	203.3	210.4
19	Iceland		10.9	11.6	12.5	11.7	11.5
20	Ireland		180.1	197.7	189.5	176.8	175.1
21	Israel (1)		176.1	190.9	202.3	205.8	218.0 e
22	Italy		1 781.5	1 893.9	1 990.5	1 951.5	1 995.2
23	Japan		4 071.4	4 290.2	4 322.9	4 135.2 e	4 330.3 e
24	Korea		1 173.0	1 269.1	1 306.4	1 321.0	1 423.7 e
25	Luxembourg		37.1	40.6	43.8	42.2	45.9
26	Mexico		1 432.2	1 540.9	1 629.6	1 540.4	1 647.2 e

Figure 4.4 This figure shows the Excel spreadsheet of the Gross Domestic Product data. The spreadsheet has not been geo-coded.

CHAPTER 5

STANDARD DEVIATIONAL ELLIPSE

Textbooks on geo-statistics cover the weighted mean center and the dispersion of the data, [Davis, (13), p. 25]. The author defines the weighted mean center used in Section 5. The author also defines the accompanying standard distance based on Pythagoras' theorem for a set of coordinates (x_i, y_i) as

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}. \quad (5.1)$$

The author objectively compares the standard distance of the dispersion of different point patterns. to be compared objectively. He gives as an example the number of clothing shops versus butchers or grocers. The standard distance is a single number and always results in a circle around the mean center to show the dispersion. [Gregory (23), and Lee (36)] discuss data transformations and regression estimators. To derive the estimators for the weighted mean centers, we can use a general linear means model. [Johnston, (32), p. 130-34] discusses the relationship between the standard deviational ellipse and the correlation coefficient in linear regression modeling. The author provides details to graphical representations in lieu of proofs.

We will introduce an extension to the standard deviational ellipse and use the above three criteria given in [Yuill, (68)] to compare the results. This book will consider these two spatial estimators:

1. The weighted mean center — shows the spatial mean of a random variable.
2. The weighted standard deviational ellipse — shows the spatial spread of a set of point locations for a random variable.

The *mean center* is simply the usual sample mean found in the statistics literature. For the *weighted mean center* found in the geo-statistics literature, the formulas and their distributional properties are neither too obvious nor intuitive. We will discuss those concepts next.

5.1 A Weighted Regression Model

Consider the weighted regression model $y_i = \alpha + \beta x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \frac{1}{\sqrt{w_i}}\sigma)$ where w_i represents the weighted observations. [Lee (36), page 236] gives the following estimators.

The parameter estimator for α appears in Equation (5.2).

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \quad (5.2)$$

The parameter estimator for β appears in Equation (5.3).

$$\hat{\beta} = \frac{\sum_{i=1}^n w_i(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n w_i(x_i - \bar{x})^2} - \frac{\sum_{i=1}^n w_i x_i y_i - \frac{(\sum_{i=1}^n w_i x_i)(\sum_{i=1}^n w_i y_i)}{\sum_{i=1}^n w_i}}{\sum_{i=1}^n w_i x_i^2 - \frac{(\sum_{i=1}^n w_i x_i)^2}{\sum_{i=1}^n w_i}} \quad (5.3)$$

where the averages are estimated using Equations (5.4) and (5.5).

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (5.4)$$

$$\bar{y} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \quad (5.5)$$

Under the normal distribution, the weighted residual sums of squares for σ^2 appears in Equation (5.6).

$$\sum_{i=1}^n \hat{\epsilon}_i^2 = \sum_{i=1}^n w_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2. \quad (5.6)$$

Equation (5.7) shows the weighted likelihood function for the regression model. It has a chi-square distribution.

$$L = \prod_{i=1}^n \left(\frac{1}{2\pi\hat{\epsilon}^2} \right)^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \frac{w_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{2\hat{\epsilon}^2} \right\} \quad (5.7)$$

The natural logarithm of the weighted likelihood function appears in Equation (5.8).

$$\ln L = \ln \left[\prod_{i=1}^n \left(\frac{1}{2\pi\hat{\epsilon}^2} \right)^{\frac{1}{2}} \right] - \frac{1}{2} \sum_{i=1}^n \frac{w_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{\hat{\epsilon}^2}. \quad (5.8)$$

Equation (5.8) is based on the Akaike information criterion. The entire equation is retained for comparison purposes in Section 7.12.

5.2 Mean Latitude and Mean Longitude

Let $(w_i x_i, w_i y_i)$ equal to a weighted, paired observation. Let w_i equal to the weight for the sample total of a random variable on observation i . Let x_i equal to the latitude of the i^{th} observation such that $0 \leq x_i \leq 90$. Let y_i equal to the longitude of the i^{th} observation such that $-180 \leq y_i \leq 180$.

$$(\bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \right). \quad (5.9)$$

Equation (5.9) is called the *center of gravity*.

5.3 Standard Deviation Ellipse

To calculate the standard deviation ellipse, one must substitute the latitude and longitude coordinate pairs as follow.

$$x'_i = x_i - \bar{x}, \quad y'_i = y_i - \bar{y}.$$

[Yuill (68)] calculates the angle of rotation for the *weighted standard deviation ellipse* as

$$\tan \theta = - \frac{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)}{2 \sum_{i=1}^n x_i' y_i' w_i} \pm \frac{\sqrt{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}}{2 \sum_{i=1}^n x_i' y_i' w_i} \quad (5.10)$$

Observe that Equation (5.10) gives the roots to the quadratic equation where the constants

$$a = \sum_{i=1}^n x_i' y_i' w_i, \quad b = \sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i, \quad \text{and} \quad c = 1.$$

The $\text{Arctan}()$ function is applied to the right-hand-side of Equation (5.10) to obtain the angle of rotation θ . The value for θ is substituted into Equations (5.11) and (5.12) to calculate the standard error terms along the axes. The standard error terms for the semi-major axis and semi-minor axis are given by Equations (5.11) and (5.12).

$$\delta_x = \sqrt{\frac{\sum_{i=1}^n (y'_i \sin \theta - x'_i \cos \theta)^2 w_i}{\sum_{i=1}^n w_i}} \quad (5.11)$$

$$\delta_y = \sqrt{\frac{\sum_{i=1}^n (y'_i \cos \theta - x'_i \sin \theta)^2 w_i}{\sum_{i=1}^n w_i}} \quad (5.12)$$

The advantages of these two estimators are that they allow trends in the data to be visualized. The mean center is pulled closer together with counties with larger weights w_i using Equation (5.9). Section 5 will show how Equation 5.9 is derived using a general linear means model and a linear regression model. Section 5.6 will show how Equation (5.10) can be extended to the exponential model.

5.3.1 Weighted Mean Center

Although [Yuill, (68)] did provide a derivation for the weighted ellipse, the author did not provide the derivation for the mean centers. We will cover the derivation next. Let the variable W represent the random variable of interest. We wish to give those latitude observations with the largest values of W the most weight, and those latitude values with the smaller values of W the lesser weights. Consider the general linear means model

$$\text{model } x_i \quad y_i = \mu_x + \mu_y + \frac{\epsilon_x}{\sqrt{w_i}} + \frac{\epsilon_y}{\sqrt{w_i}}$$

where w_i equals to the weight for the sample total of a random variable on observation i ; x_i equals to the latitude of the i^{th} observation such that $0 \leq x_i \leq 90$; y_i equals to the longitude of the i^{th} observation such that $-180 \leq y_i \leq 180$; μ_x represents the weighted mean response of the latitude responses; and μ_y represents the weighted mean response of the longitude responses. The general linear means model weights the error terms by w_i .

Consider the first model (latitude). We wish to minimize the weighted sum of squares. Our error terms are the difference between the known random values and the predicted values (i.e. the epsilons).

$$e_x = \sqrt{w_i}(x_i - \mu_x).$$

We square e_x and sum to obtain the weighted sum of squares.

$$e_x^2 = \sum_{i=1}^n w_i(x_i - \mu_x)^2.$$

The w_i 's and x_i 's are known observations. We estimate μ_x using the w_i 's and x_i 's observations. We take the derivative of e_x with respect to μ_x ; set it to zero; and solve for μ_x .

$$\begin{aligned} \frac{d e_x}{d \mu_x} &= -2 \sum_{i=1}^n w_i (x_i - \mu_x) = 0 \Rightarrow \\ \sum_{i=1}^n w_i x_i - \mu_x \sum_{i=1}^n w_i &= 0 \Rightarrow \\ \sum_{i=1}^n w_i x_i &= \mu_x \sum_{i=1}^n w_i \Rightarrow \\ \hat{\mu}_x &= \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \bar{x}. \end{aligned} \quad (5.13)$$

A similar derivation holds for the longitude e_y by substituting y_i for x_i and μ_y for μ_x above.

$$\hat{\mu}_y = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} = \bar{y}. \quad (5.14)$$

The quantities (\bar{x}, \bar{y}) are the weighted mean center for the random variable W .

5.4 Ellipse Properties

The quantity a_1 in Equation (5.15) gives the length of the semi-major axis under the standard deviational ellipse.

$$\begin{aligned} a_1^2 - \frac{1}{n} \sum_{i=1}^n y_i'^2 w_i &= \\ + \frac{1}{n} \frac{2 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}{- \left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}} \end{aligned} \quad (5.15)$$

where $x_i' = x_i - \bar{x}$, $0 \leq x_i \leq 90$ and $y_i' = y_i - \bar{y}$, $-180 \leq y_i \leq 180$, $w_i \geq 0$. Both \bar{x} and \bar{y} are weighted means. Equation (5.15) is subject to the constraint $\sum_{i=1}^n x_i'^2 w_i > \sum_{i=1}^n y_i'^2 w_i$.

The quantity b_1 in Equation (5.16) gives the formula for the semi-minor axis length for the weighted standard deviational ellipse.

$$b_1^2 - \frac{1}{n} \sum_{i=1}^n x_i'^2 w_i = \quad (5.16)$$

$$-\frac{1}{n} \frac{2 \left(\sum_{i=1}^n x'_i y'_i w_i \right)^2}{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x'_i y'_i w_i \right)^2}}$$

There are two formulas for the area. We use one formula to check the other.

$$F = \tag{5.17}$$

$$\frac{\pi}{n} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 w_i \sum_{i=1}^n (y_i - \bar{y})^2 w_i - \left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) w_i \right)^2}$$

The second formula for the area uses the axes lengths. $F = \pi a_1 b_1$ gives the area where a_1 is the semi-major axis length and b_1 is the semi-minor axis length.

Most calculus books give the *eccentricity* for an ellipse (or circle). It is based on the axes lengths with the major axis a_1 being the longest.

$$e = \frac{\sqrt{a_1^2 - b_1^2}}{a_1} \tag{5.18}$$

$0 \leq e \leq 1$ binds the eccentricity e . It denotes the degree from circularity. When e is close to 1, it means that the shape of the ellipse is very elliptical and very close to a straight line. When the eccentricity e is close to 0, it means that the shape of the ellipse is close to a circle.

The *concentration* is a count of the number of data points that fall within (or on) the ellipse. We calculate this quantity as a percentage relative to the total number data points.

If $\sum_{i=1}^n x_i'^2 w_i \geq \sum_{i=1}^n y_i'^2 w_i$, then the lengths must be calculated using the two equations:

$$a_1^2 - \frac{1}{n} \sum_{i=1}^n y_i'^2 w_i =$$

$$+\frac{1}{n} \frac{2 \left(\sum_{i=1}^n x'_i y'_i w_i \right)^2}{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x'_i y'_i w_i \right)^2}}$$

and

$$b_1^2 - \frac{1}{n} \sum_{i=1}^n x_i'^2 w_i =$$

$$-\frac{1}{n} \frac{2 \left(\sum_{i=1}^n x'_i y'_i w_i \right)^2}{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x'_i y'_i w_i \right)^2}}$$

If $\sum_{i=1}^n x_i'^2 w_i < \sum_{i=1}^n y_i'^2 w_i$, then the lengths must be calculated using the two equations:

$$a_1^2 - \frac{1}{n} \sum_{i=1}^n x_i'^2 w_i =$$

$$+ \frac{1}{n} \frac{2 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}{\left(\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}}$$

and

$$b_1^2 - \frac{1}{n} \sum_{i=1}^n y_i'^2 w_i =$$

$$- \frac{1}{n} \frac{2 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}{\left(\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}}$$

This ensures we do not take the square root of a negative number in the denominator.

Avoiding taking the square root of a negative number, if $a_1 > b_1$, then a_1 is the semi-major axis and b_1 is the semi-minor axis. Otherwise, if $a_1 < b_1$, then a_1 is the semi-minor axis and b_1 is the semi-major axis.

5.5 Testing for Randomness

Consider the Malaysia MH 370 full data set with 13 observations. Using the regression estimators, calculate the test statistic v . Recall that all of the weights equal to the same constant, $w_i = c$ and sum to 1. For the latitude, $\bar{x} = -8.523076923$.

$$\frac{1}{n-1} \sum_{i=1}^n e_x^2 = \frac{1}{n-1} \sum_{i=1}^n w_i [x_i - \bar{x}]^2 =$$

$$\frac{1}{12} \sum_{i=1}^{13} [2.7 + 8.523076923]^2 + [2.8 + 8.523076923]^2 + \dots +$$

$$[-34.7 + 8.523076923]^2 = 254.4635897$$

The cut-off is $\chi^2(12) = 21.0261$. Since $v = 254.46 > 21.0261$, reject H_0 . The latitude observations did not come from a normally distributed sample. To repeat the analysis for the longitude, for the longitude $\bar{y} = 96.13846154$.

$$\frac{1}{n-1} \sum_{i=1}^n e_y^2 = \frac{1}{n-1} \sum_{i=1}^n w_i [y_i - \bar{y}]^2 =$$

$$\frac{1}{12} \sum_{i=1}^{13} [101.7 - 96.13846154]^2 + [101.7 - 96.13846154]^2 + \dots + [93 - 96.13846154]^2 = 17.8675641$$

Since $v = 17.87 < 21.0261$, accept H_0 . The longitude observations come from a normally distributed sample.

5.6 Kentucky Example

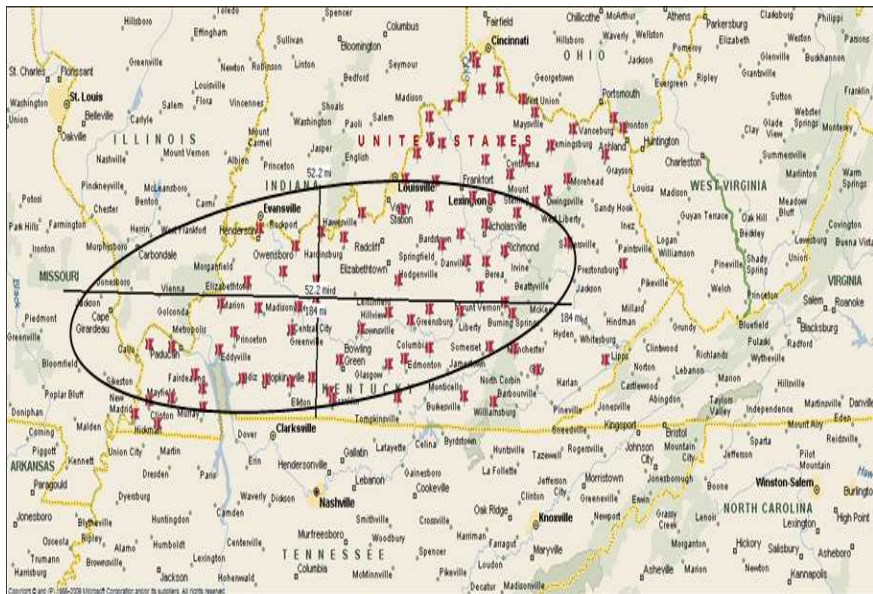


Figure 5.1 This figure shows the Kentucky data from 2003 and the standard deviational ellipse. It shows a sketch of the standard deviational ellipse with Ohio County at the center. The semi-major axis is approximately 184 miles long and the semi-minor axis is approximately 52 miles long. We rotated the major axis roughly 70° from the Y-axis in an imaginary Cartesian coordinate system.

Tables 5.1 and 5.2 summarize the results from the weighted standard deviational ellipse. Using Equation (5.10), we calculate the angle of rotation as $\theta = 70.4^\circ$, the semi-major axis length as $a = 184$ miles and the semi-minor axis length as $b = 52$ miles (using 2003 data as an example). With the semi-major and semi-minor axis lengths, θ , and the weighted mean centers from Section 5, we can draw the standard deviational ellipse in Figure 5.1 using the procedure outlined in [Lee and Wong (37)]. It shows Ohio County as the center.

Given the summary tables, there is sufficient information to draw the ellipse. The information needed includes (\bar{x}, \bar{y}) , the angle of rotation from the given axis, and

Table 5.1 2003 Results on Corn acreage in Kentucky

Description	Measurement
Center	(37.38297399, -86.8129514)
Axes	$a = 184.8734659$, $b = 52.23219209$
Area	30,336.30686 sq. mi.
Standard Deviations	$\delta_x = 1.29151776$, $\delta_y = 1.625551351$
Rotation	Y-Axis
Orientation	$\theta_x = 70.36257022^\circ$, $\theta_y = -19.63742978^\circ$
Eccentricity	0.959258636
Concentration	55/90

Table 5.2 2004 Results on Corn acreage in Kentucky

Description	Measurement
Center	(37.40588805, -86.8066601)
Axes	$a = 184.632769$, $b = 53.02565716$
Area	30,757.05259 sq. mi.
Standard Deviations	$\delta_x = 1.28529402$, $\delta_y = 1.613981198$
Rotation	Y-Axis
Orientation	$\theta_x = 70.42188382^\circ$, $\theta_y = -19.57811618^\circ$
Eccentricity	0.957872058

the lengths of the major and minor axes. The program does not draw the ellipse or calculate the percentage of concentration.

5.7 Crime Example

The example in this section comes from violent crime in the U.S. for 2007 and 2008. Tables 5.3 and 5.4 summarize the results for the weighted standard deviational ellipse. The data includes the continental U.S.

Figure 5.2 shows the standard deviational ellipse for the violent crime in the US for 2007. The ellipse covers most of the United States. We omitted Puerto Rico and Hawaii from the analysis. A histogram of the data shows the three highest crime states as California (13.6%), Texas (8.7%), and Florida (9.4%). These three states stretch from the west coast to the east coast and account for almost 32% (one-third) of the violent crime. This could explain why the weighted, standard deviational ellipse covers so much area.

Table 5.3 2007 Results on Violent Crime in the U.S.

Description	Measurement
Center	(36.8766, -92.1131260)
Axes	$a = 2, 604.517043, b = 821.0805658$
Area	6,718,353.465 sq. mi.
Standard Deviations	$\delta_x = 4.857083084, \delta_y = 4.857028283$
Rotation	X-Axis
Orientation	$\theta_x = 0.045206827^\circ, \theta_y = -89.95479317^\circ$
Eccentricity	0.949007823

Table 5.4 2008 Results on Violent Crime in the U.S.

Description	Measurement
Center	(36.88957333, -92.0322279)
Axes	$a = 2, 577.512836, b = 809.2827681$
Area	6,553,163.484 sq. mi.
Standard Deviations	$\delta_x = 15.39212752, \delta_y = 15.39222494$
Rotation	Y-Axis
Orientation	$\theta_x = 89.89264601^\circ, \theta_y = -0.107353986^\circ$
Eccentricity	0.949430202

5.8 Airline Examples

The data in these two examples came from Section 4.3. Table 5.5 summarizes the statistics for Air France Flight 447. The first observation in Table 5.5 is that the area is extremely small. This, of course, is because the axis lengths are not that long (when compared to other examples). The orientation is far away from the imaginary X-Y axes.

Figure 5.3 shows a plot of the mean center. The solid red line shows the departure location Rio de Janeiro, Brazil. It also shows the destination Paris, France. The mean center for the standard deviational ellipse falls on (or close to) the intuitive path of the airline.

Table 5.6 summarizes the statistics for Malaysia Flight 370. The area of the weighted standard deviational ellipse (square miles) is larger than that for the France airline data, but significantly smaller than the other examples. The rotation of the ellipse is greater than the other examples, but less than that of the France airline data.

Figure 5.4 shows a plot of the mean center. The solid red line shows the departure location of Kuala Lumpur International Airport. It also shows the destination Beijing, China. Unlike the Air France data set, the mean center is in the opposite direction of the intuitive path of the airline. For the given data set, $(-9.2333, 96.3417)$ is the correct and reasonable unweighted mean center.

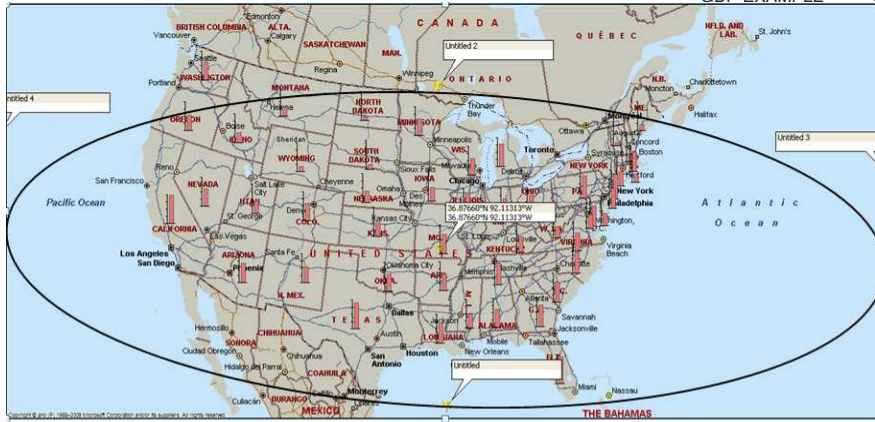


Figure 5.2 This figure shows the standard deviational ellipse of violent crime in the US for the year 2007. It also shows the values of the violent crime using a bar graph. The ellipse is rather large compared to the area in the survey.

Table 5.5 Results on Air France Flight 447

Description	Measurement
Center	(1.3489, -31.4093)
Axes	$a = 1.82, b = 0.02$
Area	0.12 sq. mi.
Standard Deviations	$\delta_x = 1.0487, \delta_y = 1.8169$
Rotation	Y-Axis
Orientation	$\theta_x = 27.3760^\circ, \theta_y = -62.6240^\circ$
Eccentricity	0.9999

5.9 GDP Example

Sections 5.9.1 and 5.9.1 will present two data analyses for the global gross domestic product data in Section 4.4. This is because, in Chapter 8, it will be shown that this data is bi-modal.

5.9.1 Analyses A

This section shows the weighted standard deviational ellipse and the statistics for the Gross Domestic Product of OECD Countries data set. Since this is a global data set, the quantities tend to be enormous compared to the other two examples. Tables 5.9 and 5.10 summarize the standard deviational ellipse for the GDP data for 2008 and 2009. We see that the ellipses are highly eccentric. The total area contracted from 2008 to 2009. This is the period where the Great Recession took place. The mean center changed an entire degree in both the latitude and longitude directions.



Figure 5.3 This figure shows the plot of the mean center for the Air France Flight 447 data.

Table 5.6 Results on Malaysia Flight 370

Description	Measurement
Center	$(-9.2333, 96.3417)$
Axes	$a = 16.02, b = 2.96$
Area	148.69 sq. mi.
Standard Deviations	$\delta_x = 14.9434, \delta_y = 16.0157$
Rotation	Y-Axis
Orientation	$\theta_x = 10.7359^\circ, \theta_y = -79.2641^\circ$
Eccentricity	0.9828

The GDP data for 2010 gives the center of gravity as $(45.94186507, -60.0743298)$. The GDP data for 2006 gives the center of gravity as $(46.02643851, -48.8800190)$ and for 2007 as $(45.94290023, -52.4580413)$. As we can see, the trend shows that the global GDP is moving away from the U.S. and toward the China coastline before and after the Great Recession.

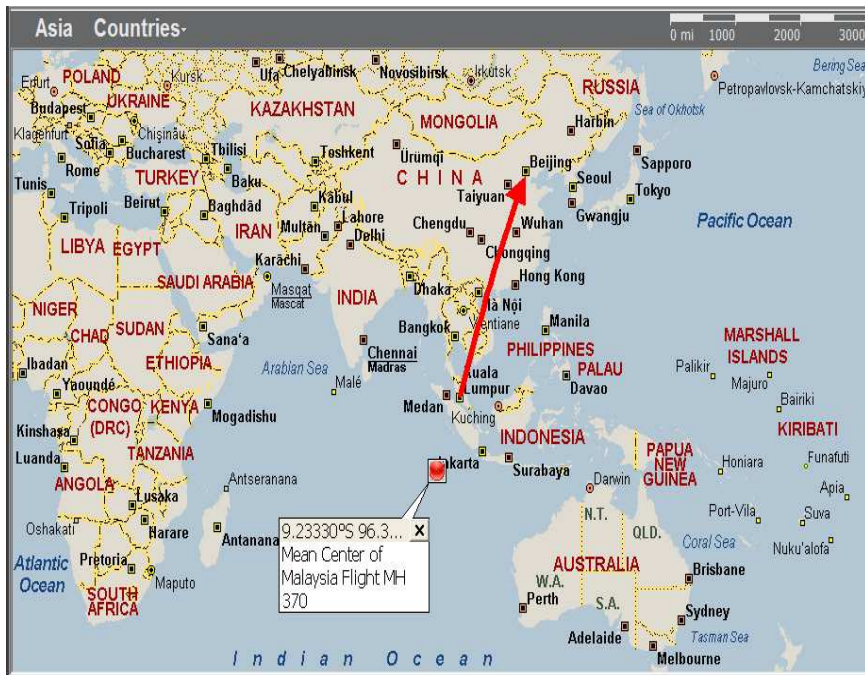


Figure 5.4 This figure shows the plot of the mean center for the Malaysia Flight MH 370 data.

Table 5.7 2008 Results on Gross Domestic Product of OECD Countries

Description	Measurement
Center	(44.95865069, 173.1929)
Axes	$a = 4,308.807649$, $b = 695.7350122$
Area	9,417,829.832 sq. mi.
Standard Deviations	$\delta_x = 20.69202965$, $\delta_y = 19.93139537$
Rotation	X-Axis
Orientation	$\theta_x = 1.307691476^\circ$, $\theta_y = -88.69230852^\circ$
Eccentricity	0.986877931

5.9.2 Analyses B

This section shows the weighted standard deviational ellipse and the statistics for the Gross Domestic Product of OECD Countries data set. Since this is a global data set, the quantities tend to be enormous compared to the other two examples. Tables 5.9 and 5.10 summarize the standard deviational ellipse for the GDP data for 2008 and 2009. We see that the ellipses are highly eccentric. The total area contracted from



Figure 5.5 This figure shows the plot of the mean centers of global GDP on a map. As the years progress, the mean center measurements move closer to the Chinese coastline.

Table 5.8 2009 Results on Gross Domestic Product of OECD Countries

Description	Measurement
Center	(45.10440685, 172.9595)
Axes	$a = 4, 259.806334, b = 700.0081681$
Area	9,367,912.709 sq. mi.
Standard Deviations	$\delta_x = 21.04761702, \delta_y = 20.27737248$
Rotation	X-Axis
Orientation	$\theta_x = 1.328357537, \theta_y = -88.67164246$
Eccentricity	0.986405646

2008 to 2009. This is the period where the Great Recession took place. The mean center changed two degrees in the longitude direction.

The GDP data for 2010 gives the center of gravity as (37.3520, 19.08634). The GDP data for 2006 gives the center of gravity as (37.748583, 12.018697) and for 2007 as (37.647639, 14.081689). As we can see, the trend shows that the global GDP is moving away from the U.S. and toward the Greece coastline before and after the Great Recession.



Figure 5.6 This figure shows the plot of the mean centers of global GDP on a map. As the years progress, the mean center measurements move closer to the Greece coastline.

Table 5.9 2008 Results on Gross Domestic Product of OECD Countries

Description	Measurement
Center	(37.7659, 15.9541)
Axes	$a = 104,975.14, b = 29,828.06$
Area	9,836,971,633.82 sq. mi.
Standard Deviations	$\delta_x = 86.5802, \delta_y = 86.6132$
Rotation	Rotate on Y-Axis
Orientation	$\theta_x = 89.1754^\circ, \theta_y = -0.8246^\circ$
Eccentricity	0.9588

The advantage of the weighted standard deviational ellipse lies in the fact that the calculations can accommodate both positive and negative data. This is not the case for the other ellipses where there is an assumed underlying distribution. In the next two chapters, the data must be positive.

Table 5.10 2009 Results on Gross Domestic Product of OECD Countries

Description	Measurement
Center	(37.3888, 18.0689)
Axes	$a = 104,769.8, b = 19,486.6$
Area	6,413,914,177.8 sq. mi.
Standard Deviations	$\delta_x = 16.8746, \delta_y = 16.0726$
Rotation	X-Axis
Orientation	$\theta_x = 1.7355, \theta_y = -88.2645$
Eccentricity	0.9826

5.10 Exercises

1. Show that if there are no predictor variables in Table 3.1 and $w_i = \frac{1}{n}$, then the total sum of squares equals to the standard distance formula. HINT: $\hat{x}_i = \bar{x}$ and $\hat{y} = \bar{y}$.

Solution: If $\hat{x}_i = \bar{x}$ and $\hat{y}_i = \bar{y}$, then the sum of squares due to the model equals to zero.

$$\sum_{i=1}^n (\hat{x}_i - \bar{x})^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (\bar{x} - \bar{x})^2 + \sum_{i=1}^n (\bar{y} - \bar{y})^2 = 0$$

Then, the total sums of squares due to error,

$$\sum_{i=1}^n w_i (x_i - \bar{x})^2 + \sum_{i=1}^n w_i (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

The standard deviation is the square root of the sums of squares which gives Equation (5.1).

2. Derive Equation (5.17) using Equations (5.11), (5.12) and $F = \pi ab$.
3. What are two ways to obtain un-weighted estimates and statistics from the given Excel spreadsheets? **Soluton:**
 - Put $\frac{1}{n}$ into column 3.
 - Copy and paste the VBA code to a new module, modify it to perform un-weighted calculations.
4. Show that the linear regression least squares sums of squares due to error does not equal to the standard deviational ellipse error.
5. Perform a chi-square test for normality on the Kentucky data in Table 4.1.

6. Perform a chi-square test for normality on the violent crime data in Table 4.4.
7. Assign weights to the Air France Flight data using sampling without replacement. Re-calculate the weighted standard deviation ellipse statistics in Table 5.5.
8. Assign weights to the Malaysia Flight MH 370 data using sampling without replacement. Re-calculate the weighted standard deviation ellipse statistics in Table 5.6.

5.11 VBA Code

To calculate the weights w_i , $i = 1, 2, \dots, n$, we first create the VBA subroutine called `SDE_Stats(AVar)`. This subroutine does not require any input parameters. The first `FOR-NEXT` loop (2) calculates the sum of the total weights $\sum_{i=1}^n w_i$. The second `FOR-NEXT` loop (3) calculates the observation weights

$$w'_i = \frac{w_i}{\sum_{i=1}^n w_i}, \quad 0 \leq w'_i \leq 1.$$

The observation weights are stored in column `D` of the active spreadsheet. Upon running this subroutine, column `D` will always sum to one. The use of the `WITH-END-WITH` statement is optional. It is a short-cut to not having to keep typing in the Excel object — in this case, the spreadsheet last referenced "Activesheet."

To calculate the mean center of gravity in Equation (5.9), the weights must be set on the active spreadsheet first. The subroutine named `SDE_Stats(AVar)` calculates the weighted mean center of the standard deviational ellipse. The `FOR-NEXT` loop (2) multiplies the latitude and longitude observations by the weights. It sums the observations and puts the sums into the variables `mean_latitude` and `mean_longitude`.

The cell reference is always row, then column. For example, the code `.Cell(3, 2)` refers to row 3 and column 2. Column 2 refers to column B.

The subroutine `Angle_of_Rotation()` calculates the angle of rotation θ in Equation (5.10). It also gives the axis of rotation. The first `FOR-NEXT` loop (3) calculates the individual sums in Equation (5.10) and stores the values in the variables `x`, `y`, and `xy`. The next four VBA statements calculate the angle of rotation with the plus and minus sign accounted for. The angle of rotation θ is stored in the variables `theta1` and `theta2`. The `IF-THEN-ELSE` statement (13) determines the axis of rotation using the values `theta1` and `theta2`. If `theta1 > theta2` then the axis of rotation is the Y-axis because the square root term is positive. Otherwise, the square root term is negative (or equal) and the axis of rotation is the X-axis. The `WITH-END-WITH` statement (14) puts the results onto the spreadsheet called `STATS`.

The subroutine `SDE_Stats(AVar)` calculates the standard error terms along the axes in Equations (5.11) and (5.12). The first `FOR-NEXT` loop (4) calculates

the sums and stores the values in variables `delta_x` and `delta_y`. After the `FOR-NEXT` loop (4), the two VBA statements divide by the total sum of the weights and take the square weights of the sums. The final values of the standard error terms along the axes are stored in the variables `delta_x` and `delta_y`. The `WITH-END-WITH` statements (16) save the results to the output spreadsheet called `STATS`.

The subroutine `Eccentricity()` calculates the eccentricity using Equation (5.18). Avoiding taking the square root of a negative number, if $a_1 > b_1$, then a_1 is the semi-major axis and b_1 is the semi-minor axis. Otherwise, if $a_1 < b_1$, then a_1 is the semi-minor axis and b_1 is the semi-major axis.

We use the function `Area()` to calculate the area of the standard deviational ellipse. It calculates the area using the formula $F = \pi a_1 b_1$. We can write special code to calculate the area using Equation (5.17). It does need the sums x , y , and xy from the subroutine `Angle_of_Rotation()`. If $\sum_{i=1}^n x_i'^2 w_i \geq \sum_{i=1}^n y_i'^2 w_i$, then the lengths must be calculated using the two equations:

$$a_1^2 - \frac{1}{n} \sum_{i=1}^n y_i'^2 w_i =$$

$$+ \frac{1}{n} \frac{2 (\sum_{i=1}^n x_i' y_i' w_i)^2}{- (\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i) + \sqrt{(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i)^2 + 4 (\sum_{i=1}^n x_i' y_i' w_i)^2}}$$

and

$$b_1^2 - \frac{1}{n} \sum_{i=1}^n x_i'^2 w_i =$$

$$- \frac{1}{n} \frac{2 (\sum_{i=1}^n x_i' y_i' w_i)^2}{- (\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i) + \sqrt{(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i)^2 + 4 (\sum_{i=1}^n x_i' y_i' w_i)^2}}$$

If $\sum_{i=1}^n x_i'^2 w_i < \sum_{i=1}^n y_i'^2 w_i$, then the lengths must be calculated using the two equations:

$$a_1^2 - \frac{1}{n} \sum_{i=1}^n x_i'^2 w_i =$$

$$+ \frac{1}{n} \frac{2 (\sum_{i=1}^n x_i' y_i' w_i)^2}{- (\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i) + \sqrt{(\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i)^2 + 4 (\sum_{i=1}^n x_i' y_i' w_i)^2}}$$

and

$$b_1^2 - \frac{1}{n} \sum_{i=1}^n y_i'^2 w_i =$$

$$\frac{1}{n - (\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i) + \sqrt{(\sum_{i=1}^n y_i'^2 w_i - \sum_{i=1}^n x_i'^2 w_i)^2 + 4 (\sum_{i=1}^n x_i' y_i' w_i)^2}}$$

This ensures we do not take the square root of a negative number in the denominator.

Avoiding taking the square root of a negative number, if $a_1 > b_1$, then a_1 is the semi-major axis and b_1 is the semi-minor axis. Otherwise, if $a_1 < b_1$, then a_1 is the semi-minor axis and b_1 is the semi-major axis.

The subroutine `SDE_Likelihood(AVar)` calculates the log likelihood function of the regression model from Equation (5.8) of the weighted standard deviational ellipse for comparison to other mathematical models later. The subroutine `Put_SDEOutput_In_Spreadsheet(AVar)` outputs the variables and labels of the standard deviational ellipse to the spreadsheet called `STATS`.

Finally, we output the area, the axes lengths, and the eccentricity to the spreadsheet called `STATS` using the `WITH-END-WITH` statement (18).

```

Sub SDE_Stats(AVar)
'calculate the mean center for the latitude and the longitude
mean_latitude = 0
mean_longitude = 0
n = Selection.Rows.Count
  For i = 2 To n
    mean_latitude = mean_latitude + ActiveSheet.Cells(i, 1).Value* ActiveSheet.Cells(i, 3).Value
    mean_longitude = mean_longitude + ActiveSheet.Cells(i, 2).Value* ActiveSheet.Cells(i, 3).Value
  Next
mean_latitude = mean_latitude / sum_weight
mean_longitude = mean_longitude / sum_weight
'calculate the rotation from the axes
x = 0
y = 0
xy = 0
  For i = 2 To n
    x = x + (ActiveSheet.Cells(i, 1).Value- mean_latitude) ^ 2 * ActiveSheet.Cells(i, 3).Value
    y = y + (ActiveSheet.Cells(i, 2).Value- mean_longitude) ^ 2 * ActiveSheet.Cells(i, 3).Value
    xy = xy + (ActiveSheet.Cells(i, 1).Value- mean_latitude) * (ActiveSheet.Cells(i, 2).Value- mean_longitude) * _
      ActiveSheet.Cells(i, 3).Value
  Next
Call Angle_of_Rotation(x, y, xy, note, atheta, itheta)
'calculate the deltas
delta_x = 0
delta_y = 0
  For i = 2 To n
    delta_x = delta_x + ((ActiveSheet.Cells(i, 2).Value- mean_longitude) * Sin(atheta / (180 / WorksheetFunction.Pi()))- _
      (ActiveSheet.Cells(i, 1).Value - mean_latitude) * Cos(atheta / (180 / WorksheetFunction.Pi())))^ 2 * ActiveSheet.Cells(i, 3).Value
    delta_y = delta_y + ((ActiveSheet.Cells(i, 2).Value- mean_longitude) * Cos(itheta / (180 / WorksheetFunction.Pi())) - _
      (ActiveSheet.Cells(i, 1).Value- mean_latitude) * Sin(itheta / (180 / WorksheetFunction.Pi())))^ 2 * ActiveSheet.Cells(i, 3).Value
  Next
delta_x = Sqr(delta_x / sum_weight)
delta_y = Sqr(delta_y / sum_weight)
'calculate the area, axes lengths, and eccentricity
m = n - 1
f = Area(m, x, y, xy) 'area
Call Axes_Length(m, x, y, xy, a, b)
e = Eccentricity(a, b)
area_check2 = Area_Check(a, b)
End Sub

```

```

Sub SDE_Likelihood(AVar)
Dim sum_latitude, sum_longitude As Double 'Z^Tw
Dim alpha, beta As Double 'parameter estimates
'calculate the mean center for the latitude and the longitude
mean_latitude = 0
mean_longitude = 0
n = Selection.Rows.Count
  For i = 2 To n
6   mean_latitude = mean_latitude + ActiveSheet.Cells(i, 1).Value * ActiveSheet.Cells(i, 3).Value
    mean_longitude = mean_longitude + ActiveSheet.Cells(i, 2).Value * ActiveSheet.Cells(i, 3).Value
  Next
mean_latitude = mean_latitude / sum_weight
mean_longitude = mean_longitude / sum_weight
'calculate Z^TZ and store it in the active spreadsheet
5   For j = 1 To 2
      For i = 2 To n
        If j = 1 Then
          ActiveSheet.Cells(j, j + 20).Value = ActiveSheet.Cells(j, j + 20).Value + _
            (ActiveSheet.Cells(i, j).Value - mean_latitude) * (ActiveSheet.Cells(i, j).Value - mean_latitude)
7 8       Else
          ActiveSheet.Cells(j, j + 20).Value = ActiveSheet.Cells(j, j + 20).Value + _
            (ActiveSheet.Cells(i, j).Value - mean_longitude) * (ActiveSheet.Cells(i, j).Value - mean_longitude)
        End If
      Next i
    Next j

  For i = 2 To n
9   ActiveSheet.Cells(2, 21).Value = ActiveSheet.Cells(2, 21).Value + _
    (ActiveSheet.Cells(i, 1).Value - mean_latitude) * (ActiveSheet.Cells(i, 2).Value - mean_longitude)
  Next ' off diagonals

```



```

ActiveSheet.Cells(1, 22).Value= ActiveSheet.Cells(2, 21).Value
'keep a copy for checks
ActiveSheet.Cells(9, 21).Value= ActiveSheet.Cells(1, 21).Value
ActiveSheet.Cells(10, 21).Value = ActiveSheet.Cells(2, 21).Value
ActiveSheet.Cells(9, 22).Value = ActiveSheet.Cells(1, 22).Value
ActiveSheet.Cells(10, 22).Value = ActiveSheet.Cells(2, 22).Value
'initialize the identity matrix
ActiveSheet.Cells(1, 23).Value = 1
ActiveSheet.Cells(2, 24).Value = 1
ActiveSheet.Cells(1, 24).Value = 0
ActiveSheet.Cells(2, 23).Value = 0
'perform gaussian elimination to find the inverse of Z^TZ
m = 24
  Do While m >= 21
    ActiveSheet.Cells(1, m).Value = ActiveSheet.Cells(1, m).Value / ActiveSheet.Cells(1, 21).Value
10  ActiveSheet.Cells(2, m).Value= ActiveSheet.Cells(2, m).Value / ActiveSheet.Cells(2, 21).Value
    m = m - 1
  Loop
5 'r1 - r2 ----> r2
m = 24
  Do While m >= 21
11  ActiveSheet.Cells(2, m).Value = ActiveSheet.Cells(1, m).Value - ActiveSheet.Cells(2, m).Value
    m = m - 1
  Loop
'r2 / a_22 ----> r2
m = 24
  Do While m >= 21
12  ActiveSheet.Cells(2, m).Value = ActiveSheet.Cells(2, m).Value / ActiveSheet.Cells(2, 22).Value
    m = m - 1
  Loop
'a_12 r2 - r1 ----> r1
m = 24
  Do While m >= 21
13  ActiveSheet.Cells(1, m).Value = ActiveSheet.Cells(1, 22).Value * ActiveSheet.Cells(2, m).Value -
    ActiveSheet.Cells(1, m).Value
    m = m - 1
  Loop

```

```

'-1*r1 ----> r1
m = 24
  Do While m >= 21
14   ActiveSheet.Cells(1, m).Value = -1 * ActiveSheet.Cells(1, m).Value
      m = m - 1
  Loop
'end of gaussian elimination. (Z^TZ)^-1 resides in (23,1), (24,1), (23,2), (24,2)
'(Z^TZ) resides in (21,9), (22,9), (21,10), (22,10)
'calculate Z^Tw
sum_latitude = 0
sum_longitude = 0
n = Selection.Rows.Count
  For i = 2 To n
5   sum_latitude = sum_latitude + ActiveSheet.Cells(i, 1).Value * ActiveSheet.Cells(i, 3).Value 'w_i
15  sum_longitude = sum_longitude + ActiveSheet.Cells(i, 2).Value * ActiveSheet.Cells(i, 3).Value 'w_i
  Next
'calculate alpha and beta
alpha = ActiveSheet.Cells(1, 23).Value * sum_latitude + ActiveSheet.Cells(1, 24).Value * sum_longitude
beta = ActiveSheet.Cells(2, 23).Value * sum_latitude + ActiveSheet.Cells(2, 24).Value * sum_longitude
'calculate log L
L_reg = 0
n = Selection.Rows.Count
  For i = 2 To n
16  L_reg = L_reg + (ActiveSheet.Cells(i, 3).Value - alpha * (ActiveSheet.Cells(i, 1).Value - mean_latitude) -
      beta * (ActiveSheet.Cells(i, 2).Value - mean_longitude))
  Next
End Sub

```

```

Sub Put_SDEOutput_In_Spreadsheet(AVar)
'put the values in STATS spreadsheet
  With Worksheets("STATS")
    'center of gravity
    .Cells(1, 1).Value = "Standard Deviatonal Ellipse Statistics"
    .Cells(2, 1).Value = "Latitude Mean Value"
    .Cells(3, 1).Value = "Longitude Mean Value"
    .Cells(2, 2).Value = mean_latitude
    .Cells(3, 2).Value = mean_longitude
    'rotation from axes
    .Cells(4, 1).Value = "Rotation"
    .Cells(4, 2).Value = note
    .Cells(5, 1).Value = "Major Axis Rotation"
    .Cells(5, 2).Value = atheta
    .Cells(6, 1).Value = "Minor Axis Rotation"
    .Cells(6, 2).Value = itheta
    'deltas along the axes
    .Cells(7, 1).Value = "Delta X"
    .Cells(7, 2).Value = delta_x
    .Cells(8, 1).Value = "Delta Y"
    .Cells(8, 2).Value = delta_y
    'area, axis length, eccentricity
    .Cells(9, 1).Value = "Area (F)"
    .Cells(9, 2).Value = f
    .Cells(10, 1).Value = "Major Axis Length (a)"
    .Cells(10, 2).Value = a
    .Cells(11, 1).Value = "Minor Axis Length (b)"
    .Cells(11, 2).Value = b
    .Cells(12, 1).Value = "Area Check"
    .Cells(12, 2).Value = area_check2
    .Cells(13, 1).Value = "Eccentricity (e)"
    .Cells(13, 2).Value = e
    'likelihood function
    .Cells(14, 1).Value = "Likelihood L(x,y)"
    .Cells(14, 2).Value = L_reg
  End With

```

CHAPTER 6

THE EXPONENTIAL ELLIPSE

Most of the theory in this section follows the text in [Lee, (36)]. [Lee, (36), pages 164-165] demonstrates how to prove that a set of data has a particular distribution using graph paper. In the absence of graph paper, modern graphing software will suffice. The technique involves transforming the data to a linear regression model and plotting the transformed data. This section deviates from the traditional linear regression literature in that the data are not directly observed, then modeled.

6.1 Mean Latitude

We consider fitting the weighted exponential distribution to the data. If the fit is a success, then the result is a linear relationship between the random variable and the cumulative distribution function. If the fit fails to result in a linear relationship, then we can add a location parameter and a shape parameter to the distribution. Before we can transform the data, we need to find the maximum likelihood estimator for the scale parameter of the weighted exponential distribution. We summarize the latitude data using the maximum likelihood estimator.

The exponential distribution function considered for the latitude is

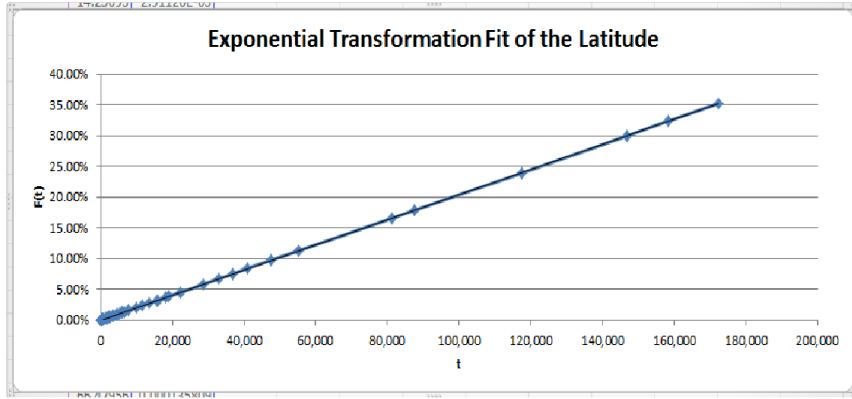


Figure 6.1 This figure shows the probability plot of the weighted latitude transformation. The exponential distribution fits the weighted latitude data since it shows a straight line.

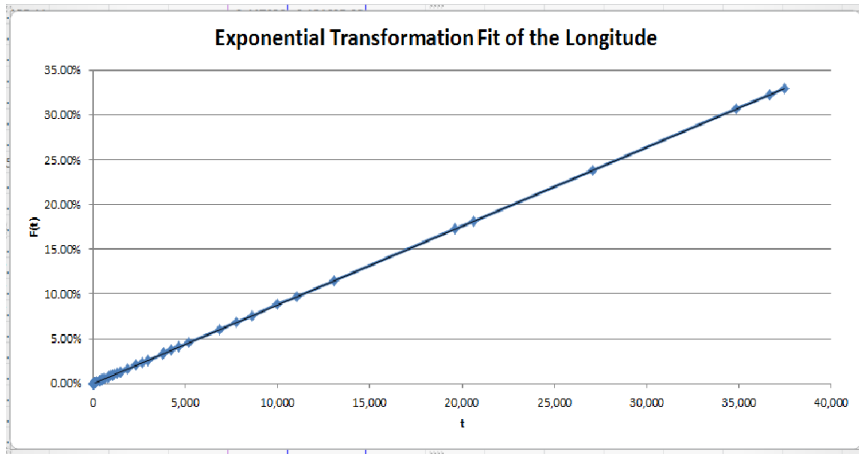


Figure 6.2 This figure shows the probability plot of the weighted longitude transformation. The exponential distribution fits the absolute value of the weighted longitude data since it shows a straight line. It was necessary to take absolute values since longitudinal data is negative.

$$f_x(w_i, x_i) = \lambda_x w_i e^{-\lambda_x x_i}, \quad 0 \leq x_i \leq 180, \quad w_i \geq 0, \quad (6.1)$$

where λ_x is the scale parameter. We estimate λ_x using the sample data. Equation (6.1) is a probability density function only if $\sum_{i=1}^n w_i = 1$. Why is this? Because the cumulative distribution function is

$$F_x(w, x) = \int_0^t f(w, x) dx = \int_0^t \lambda_x w e^{-\lambda_x x} dx = -w e^{-\lambda_x x} \Big|_0^t = w - w e^{-\lambda_x t} = F_t(t_x).$$

1. As $t \rightarrow 0$, $F_t \rightarrow 0$.
2. As $t \rightarrow 180$, $F_t \rightarrow w$.
3. F_t is right continuous.

Thus, for F_t to be a true cumulative distribution function, $w = \sum_{i=1}^n w_i = 1$ is a necessary condition. We let each observation be weighted by $w_i = \frac{w'_i}{\sum_{i=1}^n w'_i}$ where w'_i is the observed value of the random variable with the latitude x_i .

The maximum likelihood estimator for the parameter λ_x occurs at the first derivative of the likelihood function $L = \prod_{i=1}^n f_x(w, x_i)$ taken with respect to λ_x .

$$L = \prod_{i=1}^n f_x(w, x_i) = \prod_{i=1}^n \lambda_x w \exp\{-\lambda_x x_i\} = (\lambda_x w)^n \exp\left\{-\lambda_x \sum_{i=1}^n x_i\right\},$$

$$\log L = n \log(\lambda_x w) - \lambda_x \sum_{i=1}^n x_i, \quad (6.2)$$

$$\frac{d \log L}{d \lambda_x} = \frac{n}{\lambda_x w} - \sum_{i=1}^n x_i = 0 \Rightarrow \hat{\lambda}_x = \frac{n}{w \sum_{i=1}^n x_i}. \quad (6.3)$$

The maximum likelihood estimator appears in Equation (6.3).

As in [Lee, pages 162-164,(36)], we wish to fit a linear relationship by solving the following equality:

$$e^{-\lambda_x w t_x} = 1 - F_t(t_x),$$

$$\lambda_x w t_x = \log_e \left[\frac{1}{1 - F_t(t_x)} \right],$$

$$\bar{t}_x = \frac{1}{\lambda_x} \log_e \left[\frac{1}{1 - F_t(t_x)} \right] / w. \quad (6.4)$$

Using Equation (6.4), we plot our transformed data points $[t_{(x,i)}, F(t_{x,i})]$, to obtain the graph in Figure 6.1 on page 84. As one can see, it shows a linear relation that means the weighted latitude is exponentially distributed. We repeat the same procedure for the longitude.

For the joint distribution, we do not need to prove linearity. We do need the relationship between the two transformed variables, though.

6.2 Mean Longitude

A similar set of equations follow for the longitude from the latitude. The parameter estimator is different from that for the latitude since the data can be negative and the sample space for exponential distribution only allows for positive numbers (including zero). The exponential distribution function considered for the longitude is

$$f_y(w_i, y_i) = \lambda_y w_i e^{-\lambda_y y_i}, 0 \leq y_i \leq 360, w_i > 0. \quad (6.5)$$

We derive the cumulative distribution function as follow.

$$\begin{aligned} F_y(w, y) &= \int_0^u f(w, y) dy = \int_0^u \lambda_y w e^{-\lambda_y y} dy = \\ &= -w e^{-\lambda_y y} \Big|_0^u = w - w e^{-\lambda_y u} = F_u(u_y). \end{aligned}$$

1. As $u \rightarrow 0$, $F_u \rightarrow 0$.
2. As $u \rightarrow 360$, $F_u \rightarrow w$.
3. F_u is right continuous.

Thus, for F_u to be a true cumulative distribution function, $w = \sum_{i=1}^n w_i = 1$ is a necessary condition. We let each observation be weighted by $w_i = \frac{w'_i}{\sum_{i=1}^n w'_i}$ where w'_i is the observed value of the random variable with the longitude y_i .

We obtain the maximum likelihood estimator for λ_y for the longitude by taking the derivative of the likelihood function $L = \prod_{i=1}^n f(w, y_i)$ with respect to λ_y ; setting it to zero; and solving for λ_y .

$$\begin{aligned} L &= \prod_{i=1}^n f(w, y_i) = \prod_{i=1}^n \lambda_y w \exp\{-\lambda_y y_i\} = (\lambda_y w)^n \exp\left\{-\lambda_y \sum_{i=1}^n y_i\right\} \\ \log L &= n \log(\lambda_y w) - \lambda_y \sum_{i=1}^n y_i, \end{aligned} \quad (6.6)$$

$$\frac{d \log L}{d \lambda_y} = \frac{n}{\lambda_y w} - \sum_{i=1}^n y_i = 0.$$

This results in Equation (6.7).

$$\hat{\lambda}_y = \frac{n}{w \sum_{i=1}^n y_i}. \quad (6.7)$$

We wish to see if we can obtain a linear relationship using the exponential distribution. We transform the data as follow:

$$\begin{aligned}
e^{-\lambda_y w u_y} &= 1 - F_u(u_y), \\
\lambda_y w u_y &= \log_e \left[\frac{1}{1 - F_u(u_y)} \right], \\
\bar{u}_y &= \frac{1}{\lambda_y} \log_e \left[\frac{1}{1 - F_u(u_y)} \right] / w.
\end{aligned} \tag{6.8}$$

Using Equation (6.8), we plot our transformed data points $[u_{(y,i)}, F(u_{y,i})]$, to obtain the graph in Figure 6.2 on page 84. As one can see, it shows a linear relation that means the weighted longitude is exponentially distributed.

6.3 Joint Distribution

Finally, the joint distribution between two weighted exponential models needs to be determined. Given the two probability density functions in Equations (6.1) and (6.5), the joint likelihood function is

$$\begin{aligned}
L &= \prod_{i=1}^n f_x(w, x_i) f_y(w, y_i) = \\
&\prod_{i=1}^n \lambda_x w \exp \{-\lambda_x x_i\} \lambda_y w \exp \{-\lambda_y y_i\}.
\end{aligned}$$

Let $\lambda_x = \lambda_y = \lambda_{xy}$. Then, the joint likelihood function becomes,

$$L(w, x_i, y_i) = \prod_{i=1}^n \lambda_{xy}^2 w^2 \exp \{-\lambda_{xy}(x_i + y_i)\}.$$

Next, we find the maximum likelihood estimator of λ_{xy} .

$$\log L(w, x_i, y_i) = 2n \log(\lambda_{xy} w) - \lambda_{xy} \sum_{i=1}^n (x_i + y_i). \tag{6.9}$$

$$\frac{d \log L}{d \lambda_{xy}} = \frac{2n}{\lambda_{xy} w} - \sum_{i=1}^n (x_i + y_i) \Rightarrow$$

$$\hat{\lambda}_{xy} = \frac{2n}{w \sum_{i=1}^n (x_i + y_i)}. \tag{6.10}$$

Equation (6.10) gives the maximum likelihood estimator for λ_{xy} for the joint distribution of the latitude and longitude. The joint probability density function of the latitude and longitude is

$$f_{xy}(w_i, x_i, y_i) = \lambda_{xy}^2 w_i^2 \exp \{-\lambda_{xy}(x_i + y_i)\}.$$

We need to find the cumulative distribution function $F_{xy}(w_i, x_i, y_i)$ for the joint distribution, next. We need the cumulative distribution function in the *exponential* standard deviational ellipse formula.

$$\begin{aligned} F_{xy}(w, x, y) &= \\ & \int_0^t \int_0^u (w\lambda_{xy})^2 \exp \{-\lambda_{xy}(x+y)\} dy dx = \\ & \int_0^t -w^2 \lambda_{xy} \exp \{-\lambda_{xy}(x+y)\} \Big|_0^u dx = \\ & \int_0^t -\lambda_{xy} w^2 \exp \{-\lambda_{xy}(x+u)\} + \lambda_{xy} w^2 \exp \{-\lambda_{xy}x\} dx = \\ & w^2 \exp \{-\lambda_{xy}(x+u)\} - w^2 \exp \{-\lambda_{xy}x\} \Big|_0^t = \\ & w^2 \exp \{-\lambda_{xy}(t+u)\} - w^2 \exp \{-\lambda_{xy}t\} - w^2 \exp \{-\lambda_{xy}u\} + w^2 \quad (6.11) \\ & = F_{tu}(w, t, u) \end{aligned}$$

Is Equation (6.11) a valid cumulative distribution function?

1. As $t \rightarrow 0$, and $u \rightarrow 0$, then $F_{tu} \rightarrow 0$.
2. As $t \rightarrow 180$, and $u \rightarrow 360$, then $F_{tu} \rightarrow w$.
3. F_{tu} is right continuous.

This is a valid cumulative distribution function only if the weights w sum to 1 as $t \rightarrow 180$ and $u \rightarrow 360$. We choose the weights

$$w'_i = \sqrt{\frac{w_i}{\sum_{i=1}^n w_i}}$$

where w_i equals to the observed value of the random variable. Then it is a trivial fact that

$$w^2 = \sum_{i=1}^n w_i'^2 = \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} = 1.$$

6.4 Exponential Ellipse

We really do not know how the ellipse in Section 5 got its name. The sums come from the method of least squares in regression. Both the general linear means model and the linear regression model assume the random variable has a normal distribution. The normal distribution has a location parameter typically named the mean μ and a scale parameter typically named the variance σ^2 . The term really does not apply to the exponential distribution since it does not have a location parameter. This Section will derive the *exponentially distributed, weighted, deviational ellipse*.

6.4.1 Mean Center

The weighted mean estimates (\bar{x}, \bar{y}) under the general linear means model in Section 5 and the exponential distribution are exactly the same — except possibly the sign on the longitude. Equation (6.12) gives the relationship for the exponential mean center for the latitude. Equation (6.13) gives the relationship for the exponential mean center for the longitude.

$$\frac{n}{\widehat{\lambda}_x} = \bar{x}. \quad (6.12)$$

$$\frac{n}{\widehat{\lambda}_y} = \bar{y}. \quad (6.13)$$

The exponential ellipse uses the following formulas.

$$t_x = \sum_{i=1}^n \frac{1}{\widehat{\lambda}_x} \log_e \left[\frac{1}{1 - F_t(t_i)} \right] / w_i. \quad (6.14)$$

$$u_y = \sum_{i=1}^n \frac{1}{\widehat{\lambda}_y} \log_e \left[\frac{1}{1 - F_u(u_i)} \right] / w_i. \quad (6.15)$$

$$t_{xy} = \sum_{i=1}^n \frac{1}{\widehat{\lambda}_{xy}} \log_e \left[\frac{1}{1 - F_{tu}(t_i, u_i)} \right] / w_i. \quad (6.16)$$

The two ellipses differ from the sums of squares between the original data $(w_i x_i, w_i y_i)$ and the transformed data $(t_{(x,i)}, u_{(y,i)})$. Equation (6.16) gives the mean joint distribution estimate between the weighted latitude and weighted longitude estimates. The estimators for λ_x , λ_y , and λ_{xy} appear in Equations (6.3), (6.7) and (6.10).

6.4.2 Ellipse

Equation (7) in [Yuill, (68)] is general enough to obtain a weighted, exponential standard deviational ellipse. We need to substitute the original sums of squares with the new sums of squares and re-calculate the major and minor axes. These axes lengths will determine the total area from which we can compare the area and number

of weighted points inside the ellipse. Additionally, we can compare the standard deviations and the eccentricity of the ellipses.

Using equations (6.14), (6.15), and (6.16), we obtain our *exponentially distributed, weighted, deviational ellipse* in Equation (6.17).

$$\tan \theta = -\frac{(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2)}{2 \sum_{i=1}^n t_{x,y,i}} \pm \frac{\sqrt{(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2)^2 + 4 (\sum_{i=1}^n t_{x,y,i})^2}}{2 \sum_{i=1}^n t_{x,y,i}} \quad (6.17)$$

6.5 Axis Rotation

The two values of θ from Equation (6.17) determine the axis of rotation. If the sum $\sum_{i=1}^n t_{x,i}^2 < \sum_{i=1}^n u_{y,i}^2$, then the axis of rotation is about the Y-axis. Otherwise, if $\sum_{i=1}^n t_{x,i}^2 \geq \sum_{i=1}^n u_{y,i}^2$, then the rotation is about the X-axis.

6.6 Ellipse Properties

Equation (6.18) gives the formula for calculating the deviations along the X axis.

$$\delta_x = \sum_{i=1}^n (u_{y,i} \sin \theta_a - t_{x,i} \cos \theta_a)^2 w_i \quad (6.18)$$

where the subscript a denotes the major axis for the angle of rotation θ . Equation (6.19) gives the formula for calculating the deviations along the Y axis.

$$\delta_y = \sum_{i=1}^n (u_{y,i} \cos \theta_b - t_{x,i} \sin \theta_b)^2 w_i \quad (6.19)$$

where the subscript b denotes the minor axis for the angle of rotation θ .

Equation (6.20) gives the formula for calculating the major axis.

$$a^2 - \frac{\sum_{i=1}^n u_{y,i}^2}{n} = \quad (6.20)$$

$$+ \frac{2 (\sum_{i=1}^n t_{x,y,i})^2}{n (- (\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2) + \sqrt{(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2)^2 + 4 (\sum_{i=1}^n t_{x,y,i})^2})}$$

subject to

$$\sum_{i=1}^n t_{x,i}^2 \leq \sum_{i=1}^n u_{y,i}^2.$$

Equation (6.21) gives the formula for calculating the minor axis.

$$b^2 = \frac{\sum_{i=1}^n u_{y,i}^2}{n} = \quad (6.21)$$

$$\frac{2 \left(\sum_{i=1}^n t_{x,y,i} \right)^2}{n \left(- \left(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2 \right) + \sqrt{\left(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2 \right)^2 + 4 \left(\sum_{i=1}^n t_{x,y,i} \right)^2} \right)}$$

subject to

$$\sum_{i=1}^n t_{x,i}^2 \geq \sum_{i=1}^n u_{y,i}^2.$$

The difference between Equations (6.20) and (6.21) is the positive and negative sign after the first term. Since square roots are involved in the formulae, one must ensure that the conditions are true.

Equation (6.22) gives the formula for calculating the area.

$$F = \frac{\pi}{n} \sqrt{\sum_{i=1}^n t_{x,i}^2 \sum_{i=1}^n u_{y,i}^2 - \left(\sum_{i=1}^n t_{x,y,i} \right)^2} \quad (6.22)$$

It can be verified that $F = \pi ab$. The last equation is for the eccentricity of the ellipse. Equation (6.23) gives the formula for the eccentricity.

$$e = \frac{\sqrt{a^2 - b^2}}{a} \quad (6.23)$$

where Equation (6.20) gives the major axis length a and Equation (6.21) gives the minor axis length b .

Should the value $\sum_{i=1}^n t_{x,i}^2 < \sum_{i=1}^n u_{y,i}^2$, then we reverse the roles of the sums to calculate the lengths of the semi-major and semi-minor axes. This approach avoids the square root of a negative number in the denominator.

If the axis lengths $a \geq b$, then a is the semi-major axis. If $b > a$, then b is assumed to be the semi-major axis and Equation (6.23) is applied appropriately.

6.7 Combining Strata Measures

This section discusses the implications of multi-stratum data. Care must be taken when adding the ellipse properties. Before adding the two stratum area estimates together, we must:

1. Take the square of each stratum estimate.
2. Multiply by n_i where n_i represents the sample size in each stratum.
3. Add the (squared) stratum estimates.

4. Square root the result.

The above approach applies to the following measures:

- The semi-major and semi-minor axes in Equations (6.20), (6.21) for the exponential ellipse.
- The area in Equation (6.22) for the exponential ellipse.
- The standard error for the X -axis in Equation (6.18) for the exponential ellipse.
- The standard error for the Y -axis in Equation (6.19) for the exponential ellipse.
- The eccentricity in Equation (6.23) for the exponential ellipse.

6.8 Testing for Randomness

Consider the Air France 447 data set. Let the observations for the latitude and the longitude both be estimated using estimators from an exponential distribution. Then, the expected value of the latitude, \bar{x}^2 , equals to the variance. It is for this reason, $\bar{x} > 0$ under the exponential distribution. Equation (3.14) states that the null hypothesis as H_0 : the latitudes are sampled from an exponential distribution versus H_1 : the latitudes are not sampled from an exponential distribution. We use the transformed data to test the null hypothesis versus the alternative hypothesis.

For the latitude, $\bar{x} = 46.348875$.

$$v = \sum_{i=1}^8 \frac{[x_i - E(x_i)]^2}{Var(x_i)} = \frac{(178.35 - 46.348875)^2}{(46.348875)^2} + \dots + \frac{(3.103 - 46.348875)^2}{(46.348875)^2} = 21.82238071.$$

Since $\chi_{0.05}^2(7) = 14.0671$ and $v = 21.82238071 > 14.0671$, we reject the null hypothesis. The latitudes are not sampled from an exponential distribution.

Apply the test for randomness to the longitude observations in the Air France 447 data set using exponential parameter estimators. For the longitude, $\bar{y} = 328.59075$.

$$\sum_{i=1}^8 \frac{[y_i - E(y_i)]^2}{Var(y_i)} = \frac{(178.35 - 328.59075)^2}{(328.59075)^2} + \dots + \frac{(3.103 - 328.59075)^2}{(328.59075)^2} = 0.00005.$$

Since $\chi_{0.05}^2(7) = 14.0671$ and $v = 0.00005 < 14.0671$, we accept the null hypothesis. The longitudes are sampled from an exponential distribution.

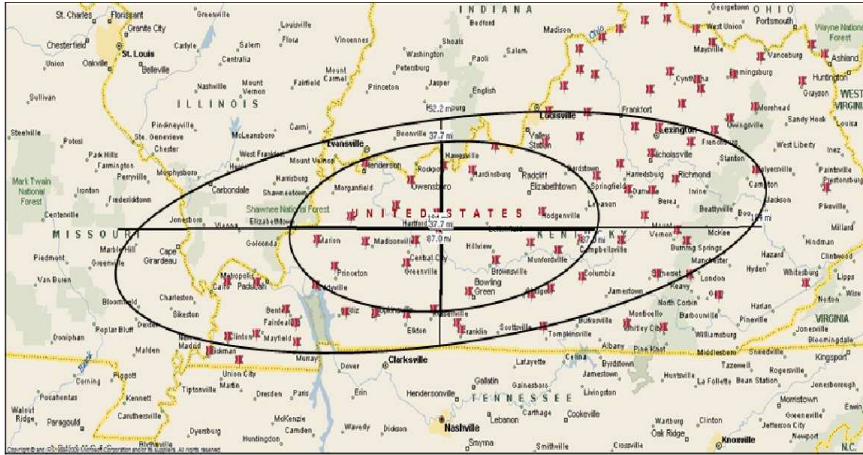


Figure 6.3 This figure shows sketches of both the standard deviational ellipse and the exponential deviational ellipse from the Kentucky data for the year 2003. The exponential ellipse is the smaller of the two ellipses. It covers one-third the area of the standard deviational ellipse, yet one-half of the same points.

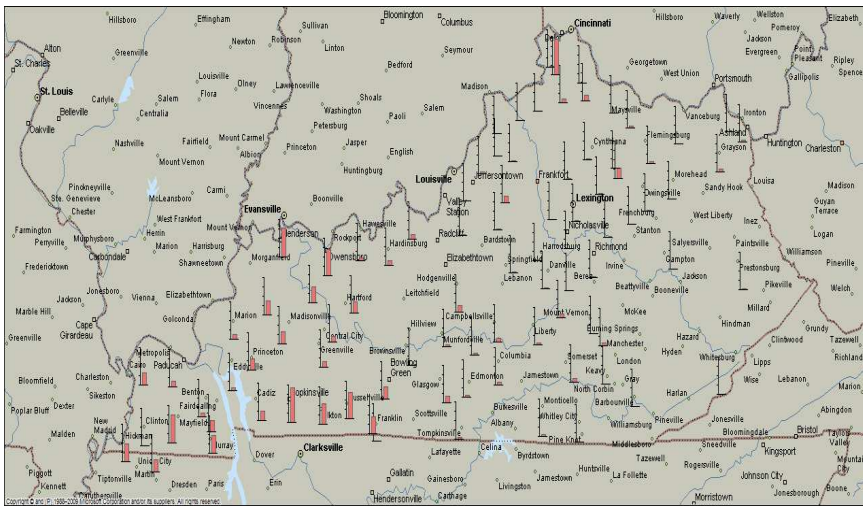


Figure 6.4 This figure shows a raised bar graph of the weights in their respective counties for the Kentucky data for the year 2003. The majority of the weighted data comes from the lower-left corner of the state of Kentucky. The majority of the un-weighted data comes from the center of the state and the right-hand side.

Table 6.1 2003 Results for the Exponential Model for Kentucky

Description	Measurement
Center	(37.38297399, -86.8129514)
Axes	$a = 273.2320657, b = 37.38920285$
Area	32,094.2879 sq. mi.
Standard Deviations	$\delta_x = 3.050790055, \delta_y = 3.050809397$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99867285^\circ, \theta_y = -0.001327443^\circ$
Eccentricity	0.990593099
Concentration	27/90

Table 6.2 2004 Results for the Exponential Model for Kentucky

Description	Measurement
Center	(37.40588805, -86.8066601)
Axes	$a = 273.2391008, b = 37.4120928$
Area	32,114.76312 sq. mi.
Standard Deviations	$\delta_x = 2.987069658, \delta_y = 2.987088203$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99870116^\circ, \theta_y = -0.001299137^\circ$
Eccentricity	0.99058201

6.9 Kentucky Example

We compare the weighted, standard deviational ellipse Tables 5.1 on page 67 to the exponential ellipse Tables 6.1 and 6.2. It shows that the area of the ellipses have been significantly reduced (within years). It shows that the exponential ellipse is contained inside the standard deviational ellipse.

Figure 6.3 on page 93 shows both the standard deviational ellipse and the exponential deviational ellipse using 2003 Kentucky data as an example. The gain in knowing the distribution of the data is in the reduction of the standard deviations and the reduction of area. What did the two exponential ellipses identify that the standard deviational ellipse did not? The bar graph in Figure 6.4 shows the weighted data in Kentucky. Looking at Figure 6.4, it found those land segments with the largest weights that are closest to those with smaller weights.

6.10 Crime Example

The example in this section comes from violent crime data in the U.S. for 2007 and 2008. Tables 6.3 and 6.3 summarize the results for the weighted exponential



Figure 6.5 This figure shows both the exponential deviational ellipse and the standard deviational ellipse for the violent crime data in the US for 2007. Both have the same center of gravity. The exponential ellipse is much smaller than the standard deviational ellipse.

deviational ellipse for violent crime in the U.S. for 2007 and 2008. The data includes the continental U.S.

Table 6.3 2007 Results for the Exponential Model on Violent Crime in the U.S.

Description	Measurement
Center	(36.8766, -92.1131260)
Axes	$a = 268.0294408, b = 36.89621172$
Area	31,068.06111 sq. mi.
Standard Deviations	$\delta_x = 5.511172335, \delta_y = 5.511237311$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99754656^\circ, \theta_y = -0.002453737^\circ$
Eccentricity	0.990479917

Figure 6.5 on page 95 shows the exponential ellipse inside of the standard deviational ellipse. The center of gravity is the same as that for the standard deviational ellipse (36.89957, -92.03222). There is a tremendous difference in the size of the two different ellipses.

Table 6.4 2008 Results for the Exponential Model on Violent Crime in the U.S.

Description	Measurement
Center	(36.88957333, -92.0322279)
Axes	$a = 268.1090262$, $b = 36.90903466$
Area	31,088.08671 sq. mi.
Standard Deviations	$\delta_x = 5.51140013$, $\delta_y = 5.511465094$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99754726^\circ$, $\theta_y = -0.002453032^\circ$
Eccentricity	0.990478949

6.11 Airline Examples

Table 6.5 summarizes the data for Air France Flight 447. Recall that this data falls into two strata. To obtain the correct mean center, take modulo 45° of the latitude.

$$\bar{x}' = \text{mod}(45, \bar{x}) = \text{mod}(45, 46.348875) = 1.3488750.$$

To obtain the correct longitude, take -1 modulo 360° of the longitude

$$\bar{y}' = \text{mod}(360, \bar{y}) = -\text{mod}(360, 328.59075) = -31.4092500.$$

Table 6.6 summarizes the data for Malaysia Flight MH 370. To obtain the correct mean center, take modulo arithmetic of the latitude as follow:

$$\bar{x}' = \text{mod}(90, \bar{x}) = -\text{mod}(90, 80.76666667) = -9.2333333.$$

The VBA code for the exponential ellipse calculates the same longitude as that of the standard deviational ellipse. This ellipse is not very eccentric. Hypothetically, when $a = b$, we have a circle. In this case b is more than half the length of a making the ellipse rounder.

Table 6.5 Results for the Exponential Model on Air France Flight 447

Description	Measurement
Center	(1.3488750, -31.4092500)
Axes	$a = 350.8914437$, $b = 28.05941647$
Area	30,931.52627 sq. mi.
Standard Deviations	$\delta_x = 43.86124605$, $\delta_y = 43.86161462$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99634792^\circ$, $\theta_y = -0.003652085^\circ$
Eccentricity	0.996797585

Table 6.6 Results for the Exponential Model on Malaysia Flight MH 370

Description	Measurement
Center	(-9.2333333, 96.34166667)
Axes	$a = 100.5931545, b = 66.78438297$
Area	21,105.38941 sq. mi.
Standard Deviations	$\delta_x = 8.377579294, \delta_y = 8.38793405$
Rotation	Y-Axis
Orientation	$\theta_x = 89.94319138^\circ, \theta_y = -0.056808616^\circ$
Eccentricity	0.747816176

6.12 GDP Example

Table 6.7 2008 Results for the Exponential Model on Gross Domestic Product

Description	Measurement
Center	(44.95865069, 173.19287)
Axes	$a = 162.1154426, b = 45.08062313$
Area	22,959.59217 sq. mi.
Standard Deviations	$\delta_x = 5.203392563, \delta_y = 5.203732811$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99301352^\circ, \theta_y = -0.006986772^\circ$
Eccentricity	0.960558702

Table 6.8 2009 Results for the Exponential Model on Gross Domestic Product

Description	Measurement
Center	(45.10440685, 172.9595353)
Axes	$a = 161.9197005, b = 45.22752649$
Area	23,006.59785 sq. mi.
Standard Deviations	$\delta_x = 5.198792719, \delta_y = 5.199135181$
Rotation	Y-Axis
Orientation	$\theta_x = 89.99299395, \theta_y = -0.007006348$
Eccentricity	0.960197864

. This section shows the weighted exponential ellipse and the statistics for the Gross Domestic Product of OECD Countries data set. Since this is a global data set, the quantities tend to be enormous compared to the other two examples. Tables 6.7 and 6.7 summarize the standard deviational ellipse for the 2008 and 2009 GDP

data. We see that the ellipses are highly eccentric. The total area contracted from 2008 to 2009. This is the period where the Great Recession took place. The mean center changed an entire degree in both the latitude and longitude directions. Where are the mean centers located? In 2008, the center of gravity for the OECD countries was in the Mediterranean Sea between the United States and Greece. In 2009, the center of gravity moved away from the United States coastline and closer to Greece's coastline. See Figure 5.6.

6.13 Comparison to SDE

We will derive a proof that shows the weighted exponential deviational ellipse always has a smaller area than the standard deviational ellipse. The proof assumes that the data does not cross the Prime Meridian or the Equator. We base the proof on showing that both the semi-major axis and the semi-minor axis are shorter than those of the standard deviational ellipse are.

Semi-major Axes Length Equation (6.24) gives the length of the semi-major axis under the standard deviational ellipse by the quantity a_1 .

$$a_1^2 - \frac{1}{n} \sum_{i=1}^n y_i'^2 w_i = \quad (6.24)$$

$$+ \frac{1}{n} \frac{2 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right) + \sqrt{\left(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i \right)^2 + 4 \left(\sum_{i=1}^n x_i' y_i' w_i \right)^2}}$$

where $x_i' = x_i - \bar{x}$, $-90 \leq x_i \leq 90$ and $y_i' = y_i - \bar{y}$, $-180 \leq y_i \leq 180$, $w_i \geq 0$. Both \bar{x} and \bar{y} are weighted means. Equation (6.24) is subject to the constraint $\sum_{i=1}^n x_i'^2 w_i > \sum_{i=1}^n y_i'^2 w_i$. Equation (6.25) gives the length of the semi-major axis under the weighted, exponential deviational ellipse.

$$a_2^2 - \frac{1}{n} \sum_{i=1}^n u_{y,i}^2 = \quad (6.25)$$

$$+ \frac{1}{n} \frac{2 \left(\sum_{i=1}^n u_{x,y,i} \right)^2}{\left(\sum_{i=1}^n u_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2 \right) + \sqrt{\left(\sum_{i=1}^n u_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2 \right)^2 + 4 \left(\sum_{i=1}^n u_{x,y,i} \right)^2}}$$

where

$$u_{x,i}^2 = \sum_{i=1}^n \left(\frac{1}{\hat{\lambda}_x} \log_e \left[\frac{1}{(1 - F_x(u_i))} \right] / w_i' \right)^2,$$

$$\hat{\lambda}_x = \frac{n}{\sum_{i=1}^n w_i' x_i}, \quad w_i' = \frac{w_i}{\sum_{i=1}^n w_i}, \quad 0 \leq x_i \leq 180.$$

$$\begin{aligned}
u_{y,i}^2 &= \sum_{i=1}^n \left(\frac{1}{\widehat{\lambda}_y} \log_e \left[\frac{1}{(1 - F_y(u_i))} \right] / w'_i \right)^2, \\
\widehat{\lambda}_y &= \frac{n}{\sum_{i=1}^n w'_i y_i}, \quad w'_i = \frac{w_i}{\sum_{i=1}^n w_i}, \quad 0 \leq y_i \leq 360. \\
u_{x,y} &= \sum_{i=1}^n \left(\frac{1}{\widehat{\lambda}_{xy}} \log_e \left[\frac{1}{(1 - F_{xy}(t_i, u_i))} \right] / w'_i \right), \\
\widehat{\lambda}_{xy} &= \frac{2n}{\sum_{i=1}^n w'_i x_i + w'_i y_i}, \quad w'_i = \frac{w_i}{\sum_{i=1}^n w_i}, \\
&\quad 0 \leq x_i \leq 180, \quad 0 \leq y_i \leq 360.
\end{aligned}$$

Under both models, $0 \leq \bar{x} \leq 90$ and $0 \leq \bar{y} \leq 180$. Now that the lengths of the semi-major axes have been defined under the weighted standard deviational ellipse and the weighted exponential deviational ellipse, it will be shown that the sums of squares for the latitude, longitude, and jointly are smaller for the exponential ellipse. If that is proven, then $n a_2^2 \leq n a_1^2$.

From Equation (6.24), we obtain the sum from the standard deviational ellipse in Equation (6.26)

$$n \sum_{i=1}^n y_i'^2 w_i = n \sum_{i=1}^n (y_i - \bar{y})^2 w_i \quad (6.26)$$

where \bar{y} is a weighted mean is positive definite. Using the relationship between $\widehat{\lambda}_y$ and \bar{y} in Equation (6.12), we can rewrite the sum from the exponential ellipse as in Equation (6.28).

$$n u_{y,i}^2 = n \sum_{i=1}^n \left(\frac{1}{\widehat{\lambda}_x} \log_e \left[\frac{1}{(1 - F_y(u_i))} \right] / w'_i \right)^2 = \quad (6.27)$$

$$\frac{1}{n} \sum_{i=1}^n \left(\bar{y} \log_e \left[\frac{1}{(1 - F_y(u_i))} \right] / \frac{w_i}{\sum_{i=1}^n w_i} \right)^2 \quad (6.28)$$

We make two observations.

1. In the standard ellipse, the sum $\sum_{i=1}^n (y_i - \bar{y})^2$ is positive definite.
2. In the exponential ellipse, an approximate formula for the $\log_{10}(1 - y)$ is $\log_{10}(1 - y) = -(1 - y) + \frac{(1-y)^2}{2} - \frac{(1-y)^3}{3} + \frac{(1-y)^4}{4} - \dots$, $-2 < y \leq 0$. The important part of this formula are the bounds.

So, from Equation (6.26) we write:

$$\sum_{i=1}^n (y_i - \bar{y})^2 w_i = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y}) w_i = \sum_{i=1}^n (y_i^2 - y_i \bar{y} - y_i \bar{y} + \bar{y}^2) w_i, \quad (6.29)$$

$$\sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2)w_i = \sum_{i=1}^n (y_i^2 - 2n\bar{y}^2 + n\bar{y}^2)w_i = \sum_{i=1}^n (y_i^2 - n\bar{y}^2)w_i. \quad (6.30)$$

So, from Equations (6.28) and (6.30) we write:

$$\sum_{i=1}^n (y_i^2 - n\bar{y}^2)w_i > \frac{1}{n} \sum_{i=1}^n \left(\bar{y} \ln \left[\frac{1}{1 - F_y(u_i)} \right] / w_i \right)^2 = \frac{\bar{y}^2}{n} \sum_{i=1}^n \left(\ln \left[\frac{1}{1 - F_y(u_i)} \right] / w_i \right)^2. \quad (6.31)$$

Then, from Equation (6.31) we write:

$$\begin{aligned} \sum_{i=1}^n y_i^2 w_i - n\bar{y}^2 w_i &= \frac{\bar{y}^2}{n} \sum_{i=1}^n \left(\ln \left[\frac{1}{1 - F_y(u_i)} \right] / w_i \right)^2, \\ n \sum_{i=1}^n \frac{y_i^2 w_i}{\bar{y}^2} - n^2 w_i &= \sum_{i=1}^n \left(\ln \left[\frac{1}{1 - F_y(u_i)} \right] / w_i \right)^2, \end{aligned}$$

$$\begin{aligned} \text{If } n \sum_{i=1}^n \frac{y_i^2 w_i}{\bar{y}^2} - n^2 w_i &\leq \sum_{i=1}^n \left(\ln \left[\frac{1}{1 - F_y(u_i)} \right] / w_i \right)^2, & \text{SDE major axis longer.} \\ \text{If } n \sum_{i=1}^n \frac{y_i^2 w_i}{\bar{y}^2} - n^2 w_i &> \sum_{i=1}^n \left(\ln \left[\frac{1}{1 - F_y(u_i)} \right] / w_i \right)^2, & \text{Exp major axis longer.} \end{aligned} \quad (6.32)$$

The derivation for the semi-minor axis is similar.

Axes Summary We showed that both the semi-major axis length and semi-minor axis length of the exponential ellipse are smaller than those of the standard deviational ellipse. The proof was conditional on the data strataling the Prime Meridian and the Equator. The proof was also conditional on the final inequality in Equation (6.32).

Given that both share the same mean center, then next obvious question is that, does the weighted standard deviational ellipse always encompass the weighted exponential deviational ellipse? Obviously, in the numeric example given, the answer is maybe. Under what conditions do the standard deviational ellipses encompass the weighted exponential deviational ellipses? If the axes relationship holds true, $a_1 \leq b_2$, then the standard deviational ellipse encompasses the exponential ellipse regardless of the angle of rotation θ . Let θ_1 be the expression for the angle of rotation for the standard deviational ellipse given in Equation (6.33).

$$\tan \theta_1 = - \frac{\left(\sum_{i=1}^n x_i^2 w_i - \sum_{i=1}^n y_i^2 w_i \right)}{2 \sum_{i=1}^n x_i y_i w_i} \pm \quad (6.33)$$

$$\frac{\sqrt{(\sum_{i=1}^n x_i'^2 w_i - \sum_{i=1}^n y_i'^2 w_i)^2 + 4(\sum_{i=1}^n x_i' y_i' w_i)^2}}{2 \sum_{i=1}^n x_i' y_i' w_i},$$

$-1 \leq \tan \theta_1 \leq 1$. Let θ_2 be the angle of rotation for the exponential ellipse given in Equation (6.34).

$$\tan \theta_2 = - \frac{(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2)}{2 \sum_{i=1}^n t_{x,y,i}} \pm \quad (6.34)$$

$$\frac{\sqrt{(\sum_{i=1}^n t_{x,i}^2 - \sum_{i=1}^n u_{y,i}^2)^2 + 4(\sum_{i=1}^n t_{x,y,i})^2}}{2 \sum_{i=1}^n t_{x,y,i}},$$

$-1 \leq \tan \theta_2 \leq 1$. Tables 6.9 and 6.10 summarize the relationship between θ_1 and θ_2 for different values of the plus/minus sign in the middle of θ_1 and θ_2 .

Table 6.9 Relationship of angles of rotation for large x_i' s.

Bounds	Plus/Minus	$\tan()$	Angle	Inference
$\sum_{i=1}^n x_i'^2 w_i > \sum_{i=1}^n y_i'^2 w_i$	+	Positive	$\theta_1 > 0$	$0 < \theta_1 < 90$
$\sum_{i=1}^n t_{x,i}^2 > \sum_{i=1}^n u_{y,i}^2$	+	Positive	$\theta_2 > 0$	$0 < \theta_2 < \theta_1 < 90$
$\sum_{i=1}^n x_i'^2 w_i > \sum_{i=1}^n y_i'^2 w_i$	-	Negative	$\theta_1 < 0$	$0 < \theta_1 < 90$
$\sum_{i=1}^n t_{x,i}^2 > \sum_{i=1}^n u_{y,i}^2$	-	Negative	$\theta_2 < 0$	$0 < \theta_1 < \theta_2 < 90$

Table 6.10 Relationship of angles of rotation for small x_i' s.

Bounds	Plus/Minus	$\tan()$	Angle	Inference
$\sum_{i=1}^n x_i'^2 w_i < \sum_{i=1}^n y_i'^2 w_i$	+	Positive	$\theta_1 > 0$	$0 < \theta_1 < 90$
$\sum_{i=1}^n t_{x,i}^2 < \sum_{i=1}^n u_{y,i}^2$	+	Positive	$\theta_2 > 0$	$0 < \theta_2 < \theta_1 < 90$
$\sum_{i=1}^n x_i'^2 w_i < \sum_{i=1}^n y_i'^2 w_i$	-	Negative	$\theta_1 < 0$	$0 < \theta_1 < 90$
$\sum_{i=1}^n t_{x,i}^2 < \sum_{i=1}^n u_{y,i}^2$	-	Negative	$\theta_2 < 0$	$0 < \theta_1 < \theta_2 < 90$

Upon examining Tables 6.9 and 6.10, we would expect that for extreme values of θ_1 that the exponential ellipse would fall inside the standard deviational ellipse.

6.14 Exercises

1. Consider the pdf $f(y) = \lambda e^{-\lambda y/w}$.

(a) Derive the mle for λ_y .

(b) Find the cdf. Show that it is a true cdf for $\sum_{i=1}^n w_i = 1$. Solution:

$$f(y) = \lambda e^{-\lambda y/w} \Rightarrow$$

$$F(y) = \int_0^u \lambda e^{-\lambda y/w} dy = w e^{-\lambda y/w} \Big|_0^u = w - w e^{-\lambda u/w}$$

As $u \rightarrow \infty$, $F(y) \rightarrow w$. So, for $F(y)$ to be a cdf, $w = \sum_{i=1}^n w_i = 1$.

2. Consider the pdf $f = \lambda e^{-\lambda_y w y}$.
 - (a) Derive the mle for λ_y .
 - (b) Find the cdf. Show that it is not necessarily a cdf for $\sum_{i=1}^n w_i = 1$.
3. Programming Exercise. Verify the cdfs $F(x)$, $F(y)$, and $F(x, y)$ sum to 1. This should take no more than three or four lines of additional VBA code.
4. Explicitly show that

$$\sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} = 1.$$

Solution: Let $w = \sum_{i=1}^n w_i$, $w_i \geq 0$, $\forall i$. Then,

$$\sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} = \frac{1}{w} \sum_{i=1}^n w_i = \frac{w}{w} = 1.$$

5. Show that

$$\frac{n}{\widehat{\lambda}_x} = \bar{x} \text{ in Chapter 5.}$$

Solution: Let $w = \sum_{i=1}^n \frac{w_i}{w_i}$, $w_i \geq 0$, $\forall i$. Then,

$$\widehat{\lambda}_x = \frac{n}{w \sum_{i=1}^n x_i} = \frac{n}{\left(\sum_{i=1}^n \frac{w_i x_i}{\sum_{i=1}^n w_i} \right)} \Rightarrow$$

$$\frac{n}{\widehat{\lambda}_x} = \sum_{i=1}^n \frac{w_i x_i}{\sum_{i=1}^n w_i} = \bar{x}, \quad 0 \leq x_i \leq 180, \quad -90 \leq \bar{x} \leq 90.$$

6. Show that

$$\frac{n}{\widehat{\lambda}_y} = \bar{y} \text{ in Chapter 5.}$$

Solution: This is very similar to Question 2, but the bounds are different. Let $w = \frac{w_i}{\sum_{i=1}^n w_i}$, $w_i \geq 0, \forall i$. Then,

$$\hat{\lambda}_y = \frac{n}{w \sum_{i=1}^n y_i} = \frac{n}{\left(\sum_{i=1}^n \frac{w_i y_i}{\sum_{i=1}^n w_i} \right)} \Rightarrow$$

$$\frac{n}{\hat{\lambda}_y} = \sum_{i=1}^n \frac{w_i y_i}{\sum_{i=1}^n w_i} = \bar{y},$$

$$0 \leq y_i \leq 360, \quad -180 \leq \bar{y} \leq 180.$$

7. Derive Equation (6.22) using Equations (6.18), (6.19) and $F = \pi ab$.

6.15 VBA Code

The subroutine `EXP_Stats(AVar)` (1) calculates the parameter estimates $\hat{\lambda}_x$, $\hat{\lambda}_y$, and $\hat{\lambda}_{xy}$ using Equations (6.3), (6.7), and (6.10). It also calculates the mean latitude and the mean longitude (\bar{x}, \bar{y}) using Equations (6.12) and (6.13). The `FOR-NEXT` loop (3) calculates the sums for the maximum likelihood estimates. The VBA statements that follow the `FOR-NEXT` loop (3) calculate the final values for the parameter estimates and store them in the variables `lambda_x` for the latitude, `lambda_y` for the longitude, and `lambda_xy` for the joint distribution. We need the parameter estimates to calculate the distribution functions.

The subroutine `EXP_Stats(AVar)` (1) calculates the exponential distributions for the latitude using Equation (6.1) and the longitude using Equation (6.5). The subroutine calculates the cumulative joint distribution function in Equation (6.11) and stores the value in the variable `F_xy`. The `FOR-NEXT` loop (4) calculates the sums from Equations (6.14), (6.15), and (6.16) and stores the values in the variables `t_x`, `t_y`, and `t_xy`, respectively. This `FOR-NEXT` loop also calculates the log likelihood functions in Equations (6.2), (6.6), and (6.9). It stores those values in the variables `L_x`, `L_y`, and `L_xy`, respectively.

The subroutine `EXP_Stats(AVar)` (1) calculates the deviations on the X-axis and the Y-axis. The `FOR-NEXT` loop (18) calculates the values t_x and u_y using Equations (6.14) and (6.15), respectively. In addition to t_x and u_y , using the two values for θ we previously obtained using the plus and minus sign, we can calculate the sums for the deviations on the X-axis and Y-axis using Equations (6.18) and (6.19). We store the final values in the variables `delta_x` and `delta_y`. The source code after the `FOR-NEXT` loop calculates the standard errors on the X-axis and Y-axis. We simply divide by the sum of the weights and then take the square root of the quantity. The final values for the standard errors in Equations (6.18) and (6.19) reside in the variables `error_x` and `error_y`.

The `EXP_Stats(AVar)` subroutine calculates the major and minor axes lengths, the area, and the eccentricity. One of the first VBA statements calculates the area of the exponential ellipse before the `IF-THEN-ELSE` statement using Equation (6.22) and saves the value in the variable `f`. The `IF-THEN-ELSE` statement (20) calculates the semi-major axis length a in Equation (6.20) and the semi-minor axis length b in Equation (6.21). Should the value $\sum_{i=1}^n t_{x,i}^2 < \sum_{i=1}^n u_{y,i}^2$, then we reverse the roles of the sums to calculate the lengths of the semi-major and semi-minor axes. This approach avoids the square root of a negative number in the denominator.

The second `IF-THEN-ELSE` statement (21) calculates the eccentricity. If the axis lengths $a \geq b$, then a is the semi-major axis. The code uses Equation (6.23) to calculate the eccentricity. If $b > a$, then b is assumed to be the semi-major axis and Equation (6.23) is applied appropriately.

The `WITH-END-WITH` statement (7) saves the remaining exponential statistics to the `STATS` spreadsheet in column E (column 5). We add appropriate labels in column D (column 4).

```

Sub EXP_Stats(AVar)
n = Selection.Rows.Count
  For i = 2 To n
2   ActiveSheet.Cells(i, 5) = ActiveSheet.Cells(i, 2).Value 'copy values to column E
  Next
wx = 0
wy = 0
  For i = 2 To n
3   wx = wx + ActiveSheet.Cells(i, 1).Value * ActiveSheet.Cells(i, 4).Value
    wy = wy + ActiveSheet.Cells(i, 5).Value * ActiveSheet.Cells(i, 4).Value
  Next
lambda_x = (n - 1) / wx
lambda_y = (n - 1) / wy
lambda_xy = 2 * (n - 1) / (wx + wy)
'calculate the cumulative distribution functions for the latitude, longitude, and the joint distribution
'calculate the values t_x, t_y, and t_xy, tx^2, ty^2, and t_xy^2
t_x = 0
t_y = 0
t_xy = 0
t_xsq = 0
t_ysq = 0
1 t_xysq = 0
L_x = 0 'used for the log L calculation
L_y = 0 'used for the log L calculation
L_xy = 0 'used for the log L calculation
  For i = 2 To n
    f_x = ActiveSheet.Cells(i, 4).Value - ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_x * ActiveSheet.Cells(i, 1).Value)
    F_y = ActiveSheet.Cells(i, 4).Value - ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_y * ActiveSheet.Cells(i, 5).Value)
    F_xy = ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_xy * (ActiveSheet.Cells(i, 1).Value + ActiveSheet.Cells(i, 5).Value)) - _
    ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_xy * ActiveSheet.Cells(i, 1).Value) - _
    ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_xy * ActiveSheet.Cells(i, 5).Value) + _
    ActiveSheet.Cells(i, 4).Value
    t_x = t_x + 1 / lambda_x * Log (1 / (1 - f_x)) / ActiveSheet.Cells(i, 4).Value
    t_y = t_y + 1 / lambda_y * Log (1 / (1 - F_y)) / ActiveSheet.Cells(i, 4).Value
4   t_xy = t_xy + 1 / lambda_xy * Log (1 / (1 - F_xy)) / ActiveSheet.Cells(i, 4).Value
    t_xsq = t_xsq + ((n - 1) / lambda_x * Log (1 / (1 - f_x)) / ActiveSheet.Cells(i, 4).Value) ^ 2
    t_ysq = t_ysq + ((n - 1) / lambda_y * Log (1 / (1 - F_y)) / ActiveSheet.Cells(i, 4).Value) ^ 2
    t_xysq = t_xysq + ((n - 1) / lambda_xy * Log (1 / (1 - F_xy)) / ActiveSheet.Cells(i, 4).Value) ^ 2
    L_x = L_x + (n - 1) * Log (lambda_x * ActiveSheet.Cells(i, 4).Value) - lambda_x * ActiveSheet.Cells(i, 1).Value
    L_y = L_y + (n - 1) * Log (lambda_y * ActiveSheet.Cells(i, 4).Value) - lambda_y * ActiveSheet.Cells(i, 5).Value
    L_xy = L_xy + 2 * (n - 1) * Log (lambda_xy * ActiveSheet.Cells(i, 4).Value) - _
    lambda_xy * (ActiveSheet.Cells(i, 1).Value + ActiveSheet.Cells(i, 5).Value)
  Next

```

```

'calculate the angle of rotations about the axes
Call Angle_of_Rotation(t_xsqr, t_ysqr, t_xy, note, atheta, itheta)
'calculate the deviations on the axes
delta_x = 0
delta_y = 0
For i = 2 To n
  f_x = ActiveSheet.Cells(i, 4).Value - ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_x * ActiveSheet.Cells(i, 1).Value)
  F_y = ActiveSheet.Cells(i, 4).Value - ActiveSheet.Cells(i, 4).Value * EXP(-1 * lambda_y * ActiveSheet.Cells(i, 5).Value)
  t_x_ = 1 / lambda_x * Log(1 / (1 - f_x)) / ActiveSheet.Cells(i, 4).Value
  t_y_ = 1 / lambda_y * Log(1 / (1 - F_y)) / ActiveSheet.Cells(i, 4).Value
  delta_x = delta_x + (t_y_ * Sin(atheta / 57.2957795) - t_x_ * Cos(atheta / 57.2957795)) ^ 2 * ActiveSheet.Cells(i, 3).Value
  delta_y = delta_y + (t_y_ * Cos(itheta / 57.2957795) - t_x_ * Sin(itheta / 57.2957795)) ^ 2 * ActiveSheet.Cells(i, 3).Value
Next
error_x = Sqr(delta_x / sum_weight) 'standard errors on the x axis
error_y = Sqr(delta_y / sum_weight) 'standard errors on the y axis
'calculate the area and eccentricity
f = Area(n - 1, t_xsqr, t_ysqr, t_xy)
Call Axes_Length(n - 1, t_xsqr, t_ysqr, t_xy, a, b)
e = Eccentricity(a, b)
area_check2 = Area_Check(a, b)
End Sub

Sub Put_EXPOutput_In_Spreadsheet(AVar)
  With Worksheets("STATS")
    'center of gravity
    .Cells(1, 4).Value = "Exponential Ellipse Statistics"
    .Cells(2, 4).Value = "Mean Latitude"
    .Cells(3, 4).Value = "Mean Longitude"
    .Cells(2, 5).Value = (n - 1) / lambda_x
    .Cells(3, 5).Value = (n - 1) / lambda_y
    'rotations about the axes
    .Cells(4, 4).Value = "Rotation"
    .Cells(5, 4).Value = "Major Axis Rotation"
    .Cells(6, 4).Value = "Minor Axis Rotation"
    .Cells(4, 5).Value = note
    .Cells(5, 5).Value = atheta
    .Cells(6, 5).Value = itheta
  End With

```

```

' the deltas
.Cells(7, 4).Value = "Delta X"
.Cells(8, 4).Value = "Delta Y"
.Cells(7, 5).Value = error_x
.Cells(8, 5).Value = error_y
' area, axes lengths, eccentricity, likelihood function
.Cells(9, 4).Value = "Area (F)"
.Cells(10, 4).Value = "Major Axis Length (a)"
.Cells(11, 4).Value = "Minor Axis Length (b)"
.Cells(12, 4).Value = "Area Check"
.Cells(13, 4).Value = "Eccentricity (e)"
6 7 .Cells(14, 4).Value = "Likelihood Function L(x)"
.Cells(15, 4).Value = "Likelihood Function L(y)"
.Cells(16, 4).Value = "Likelihood Function L(xy)"
.Cells(9, 5).Value = f
.Cells(10, 5).Value = a
.Cells(11, 5).Value = b
.Cells(12, 5).Value = area_check2
.Cells(13, 5).Value = e
.Cells(14, 5).Value = L_x
.Cells(15, 5).Value = L_y
.Cells(16, 5).Value = L_xy
End With
End Sub

```


CHAPTER 7

THE WEIBULL ELLIPSE

7.1 Latitude Mean Center

Equation (7.1) gives the weighted probability density function for the latitude.

$$f_x(w_i, x_i) = \gamma_x (w_i \lambda_x)^{\gamma_x} (x_i)^{\gamma_x - 1} \exp \{ - (\lambda_x x_i)^{\gamma_x} \}, \quad (7.1)$$

$$0 \leq x_i \leq 180, \lambda_x > 0, \gamma_x > 0.$$

We need the cumulative distribution function $F_x(w, x_i)$. Therefore, we integrate the probability density function $f_x(w, x_i)$.

$$\begin{aligned} F_x(w, x_i) &= \int_0^t f(w, x) dx = \\ &= \int_0^t \gamma_x (w \lambda_x)^{\gamma_x} (x)^{\gamma_x - 1} \exp \{ - (\lambda_x x)^{\gamma_x} \} dx = \\ &= -w^{\gamma_x} \exp \{ - (\lambda_x x)^{\gamma_x} \} \Big|_0^t = w^{\gamma_x} - w^{\gamma_x} \exp \{ - (\lambda_x t)^{\gamma_x} \} = \end{aligned}$$

$$F_t(w, t).$$

The properties of this cumulative distribution function are:

1. As $t \rightarrow 0$, $F_t \rightarrow 0$.
2. As $t \rightarrow 180$, $F_t \rightarrow w^{\gamma_x}$.
3. F_t is right continuous.

As $t \rightarrow 180$, $F_t \rightarrow w^{\gamma_x} \Rightarrow w^{\gamma_x}$ must sum to 1. We choose

$$w'_i = \left(\frac{w_i}{\sum_{i=1}^n w_i} \right)^{1/\gamma_y}$$

where w_i is the observed value of the random variable for the latitude.

The parameter estimation for the exponential ellipse was straightforward because there was only one. The Weibull ellipse has two parameters and we have to solve for these parameters simultaneously. The procedure is similar. The maximum likelihood estimator for λ_x occurs at the first derivative of $L = \prod_{i=1}^n f_x(w, x_i)$ taken with respect with λ_x . The maximum likelihood estimator for γ_x occurs at the first derivative of L taken with respect with γ_x . x_1, x_2, \dots, x_n is the random sample. Equation (7.2) gives the likelihood function.

$$L(w, x_i) = \gamma_x^n (w\lambda_x)^{n\gamma_x} \left(\prod_{i=1}^n x_i \right)^{\gamma_x - 1} \exp \left\{ -\lambda_x^{\gamma_x} \sum_{i=1}^n x_i^{\gamma_x} \right\} \quad (7.2)$$

(x_1, x_2, \dots, x_n) is the minimum set of sufficient statistics. Equation (7.3) gives the logarithm of the likelihood function.

$$\log L = n \log \gamma_x + n\gamma_x \log(w\lambda_x) + \sum_{i=1}^n (\gamma_x - 1) \log x_i - \lambda_x^{\gamma_x} \sum_{i=1}^n x_i^{\gamma_x}. \quad (7.3)$$

Taking the usual derivatives with respect to the parameters, $\frac{d \log L}{d \lambda_x} = 0$, and $\frac{d \log L}{d \gamma_x} = 0$ yields the following two equations.

$$\begin{aligned} \frac{d \log L}{d \lambda_x} &= 0 \Rightarrow \\ \frac{n\gamma_x}{w\lambda_x} - \gamma_x \lambda_x^{\gamma_x - 1} \sum_{i=1}^n x_i^{\gamma_x} &= 0 \\ n\gamma_x - w\gamma_x \lambda_x^{\gamma_x} \sum_{i=1}^n x_i^{\gamma_x} &= 0 \end{aligned}$$

$$n - w \lambda_x^{\gamma_x} \sum_{i=1}^n x_i^{\gamma_x} = 0 \quad (7.4)$$

and

$$\begin{aligned} \frac{d \log L}{d \gamma_x} = 0 \Rightarrow \\ \frac{n}{\gamma_x} + n \log(w \lambda_x) + \sum_{i=1}^n \log x_i - \\ \lambda_x^{\gamma_x} \sum_{i=1}^n x_i^{\gamma_x} (\log \lambda_x + \log x_i) = 0 = L'(w, x_i). \end{aligned} \quad (7.5)$$

Solving Equation (7.4) for λ_x yields

$$\lambda_x = \left(\frac{n}{w \sum_{i=1}^n x_i^{\gamma_x}} \right)^{\frac{1}{\gamma_x}}. \quad (7.6)$$

We must solve for γ_x using numerical algorithms which use Equations (7.3), (7.5), (7.6). Using [Lee 36, Appendix A], we apply the Newton-Raphson method. Instead of working with the Jacobian matrix of derivatives, we substitute Equation (7.6) into Equation (7.5).

Then we wish to iterate through the following expression using the variable k .

$$\gamma_{(x,k+1)} = \hat{\gamma}_{(x,k)} - \frac{L(w, x_i)}{L'(w, x_i)}.$$

An alternative to using the Newton-Raphson method is to apply the Secant method. This only involves using Equation (7.3). We apply it as follow:

$$\gamma_{(k+1)} = \quad (7.7)$$

$$\gamma_{(k)} - L \left[\hat{\lambda}_{(k)}, \gamma_{(k)} \right] \left(\frac{\gamma_{(k)} - \gamma_{(k-1)}}{L \left[\hat{\lambda}_{(k)}, \gamma_{(k)} \right] - L \left[\hat{\lambda}_{(k-1)}, \gamma_{(k-1)} \right]} \right)$$

We need to add an additional subscript x to denote the latitude to identify it from the longitude y . So, the equation for the Secant method becomes:

$$\begin{aligned} \gamma_{(x,k+1)} = \gamma_{(x,k)} - \\ L \left[\hat{\lambda}_{(x,k)}, \gamma_{(x,k)} \right] \left(\frac{\gamma_{(x,k)} - \gamma_{(x,k-1)}}{L \left[\hat{\lambda}_{(x,k)}, \gamma_{(x,k)} \right] - L \left[\hat{\lambda}_{(x,k-1)}, \gamma_{(x,k-1)} \right]} \right) \end{aligned} \quad (7.8)$$

where k denotes the iteration.

With the Secant method, we must choose two initial values for γ_x instead of one. The reader will find that working with either method is difficult to work with for

finding the value of λ_x . Knowledge of the functions L , $\log L$, and the derivatives are paramount. The Secant method is the easiest way out of these derivations and initialization. However, more initial values are required on the reader's part for the Secant method.

The relationship of the parameter estimate $\hat{\lambda}_x$ in Equations (7.6) and (7.8) to the weighted mean center is

$$\bar{x} = \frac{n}{\hat{\lambda}_x} \quad (7.9)$$

Notice that it depends on the range of λ_x . This textbook provides Excel 2010 Visual Basic for Applications (VBA) code to calculate the weighted means based these derivations. We must state the initial values of γ_x in addition to the maximum number of iterations. Some trial and error is required to obtain valid results. Other authors' recommendations include setting the two initial values of γ_x fairly far apart to obtain convergence. Both values must be greater than one. If the reader is in doubt about the number of iterations, then try 1; then 2; then 3; and so on. It only took seven tries to obtain valid results in the Kentucky example.

7.2 Longitude Mean Center

The longitude data can range from -180° to 180° . The Weibull distribution does not allow negative numbers. However, if all of the longitude observations are negative, then it is simple to change the sign to positive to perform parameter estimation. Afterwards, we can change the sign on the mean center. If this is not the case, some data are negative, and some are positive, then convert the observations to circular coordinate system where all of the observations are positive.

Equation (7.10) gives the weighted probability density function for the longitude.

$$f_y(w_i, y_i) = \gamma_y (w_i \lambda_y)^{\gamma_y} (y_i)^{\gamma_y - 1} \exp \{ - (\lambda_y y_i)^{\gamma_y} \} \quad (7.10)$$

$$0 \leq y_i \leq 360, \lambda_y > 0, \gamma_y > 0.$$

We need the cumulative distribution function $F_y(w, y_i)$. Therefore, we integrate the probability density function $f_y(w, y_i)$.

$$F_y(w, y) = \int_0^u f(w, y) dy =$$

$$\int_0^u \gamma_y (w \lambda_y)^{\gamma_y} (y_i)^{\gamma_y - 1} \exp \{ - (\lambda_y y_i)^{\gamma_y} \} dy =$$

$$-w^{\gamma_y} \exp \{ - (\lambda_y y)^{\gamma_y} \} \Big|_0^u = w^{\gamma_y} - w^{\gamma_y} \exp \{ - (\lambda_y u)^{\gamma_y} \} =$$

$$F_u(w, u).$$

The properties of this cumulative distribution function are:

1. As $u \rightarrow 0$, $F_u \rightarrow 0$.
2. As $u \rightarrow 360$, $F_u \rightarrow w^{\gamma_y}$.
3. F_u is right continuous.

As $u \rightarrow 360$, $F_u \rightarrow w^{\gamma_y} \Rightarrow w^{\gamma_y}$ must sum to 1. We choose

$$w'_i = \left(\frac{w_i}{\sum_{i=1}^n w_i} \right)^{1/\gamma_y}$$

where w_i is the observed value of the random variable for the longitude.

The parameter estimation for the exponential ellipse was straightforward because there was only one. The Weibull ellipse has two parameters and we have to solve for the parameters simultaneously. The procedure is similar. The maximum likelihood estimator for λ_y occurs at the first derivative of $L = \prod_{i=1}^n f_y(w, y_i)$ taken with respect with λ_y . The maximum likelihood estimator for γ_y occurs at the first derivative of L taken with respect with γ_y . y_1, y_2, \dots, y_n is the random sample. Equation (7.11) gives the likelihood function.

$$L(w, y_i) = \gamma_y^n (w\lambda_y)^{n\gamma_y} \left(\prod_{i=1}^n y_i \right)^{\gamma_y - 1} \exp \left\{ -\lambda_y^{\gamma_y} \sum_{i=1}^n y_i^{\gamma_y} \right\} \quad (7.11)$$

(y_1, y_2, \dots, y_n) is the minimum set of sufficient statistics.

Equation (7.12) gives the logarithm of the likelihood function.

$$\log L = n \log \gamma_y + n\gamma_y \log(w\lambda_y) + \sum_{i=1}^n (\gamma_y - 1) \log y_i - \lambda_y^{\gamma_y} \sum_{i=1}^n y_i^{\gamma_y}. \quad (7.12)$$

Taking the usual derivatives with respect to the parameters, $\frac{d \log L}{d \lambda_y} = 0$, and $\frac{d \log L}{d \gamma_y} = 0$ yields the following two equations.

$$\begin{aligned} \frac{d \log L}{d \lambda_y} &= 0 \Rightarrow \\ \frac{n\gamma_y}{w\lambda_y} - \gamma_y \lambda_y^{\gamma_y - 1} \sum_{i=1}^n y_i^{\gamma_y} &= 0 \\ n\gamma_y - w\gamma_y \lambda_y^{\gamma_y} \sum_{i=1}^n y_i^{\gamma_y} &= 0 \\ n - w\lambda_y^{\gamma_y} \sum_{i=1}^n y_i^{\gamma_y} &= 0 \end{aligned} \quad (7.13)$$

and

$$\begin{aligned} \frac{d \log L}{d \gamma_y} = 0 \Rightarrow \\ \frac{n}{\gamma_y} + n \log(w \lambda_y) + \sum_{i=1}^n \log y_i - \\ \lambda_y^{\gamma_y} \sum_{i=1}^n y_i^{\gamma_y} (\log \lambda_y + \log y_i) = 0 = L'(w, y_i). \end{aligned} \quad (7.14)$$

Solving Equation (7.13) for λ_y yields

$$\lambda_y = \left(\frac{n}{w \sum y_i^{\gamma_y}} \right)^{\frac{1}{\gamma_y}}. \quad (7.15)$$

γ_y must be solved numerically using Equations (7.12), (7.14), (7.15). Using [Lee 36, Appendix A], we apply the Newton-Raphson method. Instead of working with the Jacobian matrix of derivatives, we substitute Equation (7.15) into Equation (7.14).

Then we wish to iterate through the following expression using the variable k .

$$\gamma_{(y,k+1)} = \hat{\gamma}_{(y,k)} - \frac{L(w, y_i)}{L'(w, y_i)}.$$

An alternative to using the Newton-Raphson method is to apply the Secant method. This only involves using Equation (7.12). We apply it as follow:

$$\begin{aligned} \gamma_{(k+1)} = \gamma_{(k)} - \log L \left[\hat{\lambda}_{(k)}, \gamma_{(k)} \right] \\ \left(\frac{\gamma_{(k)} - \gamma_{(k-1)}}{\log L \left[\hat{\lambda}_{(k)}, \gamma_{(k)} \right] - \log L \left[\hat{\lambda}_{(k-1)}, \gamma_{(k-1)} \right]} \right) \end{aligned} \quad (7.16)$$

We need to add an additional subscript y to denote the longitude to identify it from the latitude x . So, the equation for the Secant method becomes:

$$\begin{aligned} \gamma_{(y,k+1)} = \gamma_{(y,k)} - \log L \left[\hat{\lambda}_{(y,k)}, \gamma_{(y,k)} \right] \\ \left(\frac{\gamma_{(y,k)} - \gamma_{(y,k-1)}}{\log L \left[\hat{\lambda}_{(y,k)}, \gamma_{(y,k)} \right] - \log L \left[\hat{\lambda}_{(y,k-1)}, \gamma_{(y,k-1)} \right]} \right) \end{aligned} \quad (7.17)$$

where k denotes the iteration.

With the Secant method, we must choose two initial values for γ_y instead of one. The reader will find that working with either method is difficult to work with for finding the value of λ_y . Knowledge of the functions L , $\log L$, and the derivatives are paramount. The Secant method is the easiest way out of these derivations and initialization. However, more initial values are required on the reader's part for the Secant method.

Equations (7.15) and (7.8) give the relationship of the parameter estimate $\widehat{\lambda}_y$ to the weighted mean center.

$$\bar{y} = \frac{n}{\widehat{\lambda}_y} \quad (7.18)$$

Notice that the relationship depends on the range of λ_y if we do not use circular coordinates. If we use circular coordinates, then we can remove the absolute value operator on the mean center \bar{y} .

7.3 Joint Distribution

The joint distribution between two weighted Weibull models need to be determined. Equations (7.1) and (7.10) gives the two probability density functions for the joint likelihood function L .

$$\begin{aligned} L &= \prod_{i=1}^n f_x(w, x_i) f_y(w, y_i) = \\ & \prod_{i=1}^n \gamma_x (w \lambda_x)^{\gamma_x} (x_i)^{\gamma_x - 1} \exp \{ - (\lambda_x x_i)^{\gamma_x} \} \\ & \quad \gamma_y (w \lambda_y)^{\gamma_y} (y_i)^{\gamma_y - 1} \exp \{ - (\lambda_y y_i)^{\gamma_y} \}. \end{aligned}$$

Let $\lambda_x = \lambda_y = \lambda_{xy}$, and let $\gamma_x = \gamma_y = \gamma_{xy}$. Then, the joint likelihood function becomes,

$$\begin{aligned} L(w, x_i, y_i) &= \\ & \prod_{i=1}^n \gamma_{xy} (w \lambda_{xy})^{\gamma_{xy}} (x_i)^{\gamma_{xy} - 1} \exp \{ - (\lambda_{xy} x_i)^{\gamma_{xy}} \} \\ & \quad \gamma_{xy} (w \lambda_{xy})^{\gamma_{xy}} (y_i)^{\gamma_{xy} - 1} \exp \{ - (\lambda_{xy} y_i)^{\gamma_{xy}} \} = \\ & \prod_{i=1}^n \gamma_{xy}^2 (w \lambda_{xy})^{2\gamma_{xy}} (x_i y_i)^{\gamma_{xy} - 1} \exp \{ - (\lambda_{xy} x_i)^{\gamma_{xy}} - (\lambda_{xy} y_i)^{\gamma_{xy}} \}, \end{aligned}$$

$$\lambda_{xy} > 0, \quad \gamma_{xy} > 0.$$

Next, we find the maximum likelihood estimator of λ_{xy} .

$$\begin{aligned} \log L(w, x_i, y_i) &= \\ & 2 \sum_{i=1}^n \log \gamma_{xy} + 2n \gamma_{xy} \log(w \lambda_{xy}) + (\gamma_{xy} - 1) \sum_{i=1}^n \log x_i y_i - \end{aligned} \quad (7.19)$$

$$\sum_{i=1}^n (\lambda_{xy} x_i)^{\gamma_{xy}} - \sum_{i=1}^n (\lambda_{xy} y_i)^{\gamma_{xy}}.$$

$$\frac{d \log L}{d \lambda_{xy}} = 0 + \frac{2n\gamma_{xy}}{w\lambda_{xy}} + 0 + 0 -$$

$$\gamma_{xy} \sum_{i=1}^n (\lambda_{xy} x_i)^{\gamma_{xy}-1} - \gamma_{xy} \sum_{i=1}^n (\lambda_{xy} y_i)^{\gamma_{xy}-1} = 0,$$

$$2n\gamma_{xy} - w\lambda_{xy}^{\gamma_{xy}} \gamma_{xy} \sum_{i=1}^n (x_i)^{\gamma_{xy}} - w\lambda_{xy}^{\gamma_{xy}} \gamma_{xy} \sum_{i=1}^n (y_i)^{\gamma_{xy}} = 0$$

Solving for λ_{xy} yields the following expressions.

$$-w\gamma_{xy}\lambda_{xy}^{\gamma_{xy}} \sum_{i=1}^n x_i^{\gamma_{xy}} - w\gamma_{xy}\lambda_{xy}^{\gamma_{xy}} \sum_{i=1}^n y_i^{\gamma_{xy}} = -2n\gamma_{xy} =$$

$$\lambda_{xy}^{\gamma_{xy}} \left[w \left(\sum_{i=1}^n x_i^{\gamma_{xy}} + \sum_{i=1}^n y_i^{\gamma_{xy}} \right) \right] = 2n$$

$$\hat{\lambda}_{xy} = 2n / \left[w \left(\sum_{i=1}^n x_i^{\gamma_{xy}} + \sum_{i=1}^n y_i^{\gamma_{xy}} \right) \right]^{1/(\gamma_{xy})}. \quad (7.20)$$

Equation (7.20) gives the maximum likelihood estimator for λ_{xy} for the joint distribution of the latitude and longitude. Using the Secant algorithm, we find the maximum likelihood estimates for λ_{xy} and γ_{xy} as before. We apply the Secant algorithm to Equation (7.21) to find $\hat{\gamma}_{xy}$.

$$\gamma_{(xy,k+1)} = \gamma_{(xy,k)} - \log L \left[\hat{\lambda}_{(xy,k)}, \gamma_{(xy,k)} \right] \quad (7.21)$$

$$\left(\frac{\gamma_{(xy,k)} - \gamma_{(xy,k-1)}}{\log L \left[\hat{\lambda}_{(xy,k)}, \gamma_{(xy,k)} \right] - \log L \left[\hat{\lambda}_{(xy,k-1)}, \gamma_{(xy,k-1)} \right]} \right)$$

where k denotes the iteration, Equation (7.19) gives $\log L$, and Equation (7.20) gives $\hat{\lambda}_{xy}$.

The joint probability density function of the latitude and longitude is

$$f_{xy}(w_i, x_i, y_i) =$$

$$(\gamma_{xy})^2 (w_i \lambda_{xy})^{2\gamma_{xy}} (x_i y_i)^{\gamma_{xy}-1} \exp \{ -(\lambda_{xy} x_i)^{\gamma_{xy}} - (\lambda_{xy} y_i)^{\gamma_{xy}} \}.$$

We need to find the cumulative distribution function $F_{xy}(w_i, x_i, y_i)$ for the joint distribution, next. We need the joint distribution function so that we can derive the *Weibull* standard deviational ellipse.

$$\begin{aligned}
F_{xy}(w, x, y) &= \\
&\int_0^t \int_0^u (\gamma_{xy})^2 (w\lambda_{xy})^{2\gamma_{xy}} (xy)^{\gamma_{xy}-1} \times \\
&\exp\{-(\lambda_{xy}x)^{\gamma_{xy}} - (\lambda_{xy}y)^{\gamma_{xy}}\} dy dx = \\
&\int_0^t -(\gamma_{xy}) (w\lambda_{xy})^{\gamma_{xy}} (x)^{\gamma_{xy}-1} \times \\
&\exp\{-(\lambda_{xy}x)^{\gamma_{xy}} - (\lambda_{xy}y)^{\gamma_{xy}}\} \Big|_0^u dx = \\
&\int_0^t (\gamma_{xy}) (w\lambda_{xy})^{\gamma_{xy}} (x)^{\gamma_{xy}-1} \exp\{-(\lambda_{xy}x)^{\gamma_{xy}} - (\lambda_{xy}u)^{\gamma_{xy}}\} - \\
&(\gamma_{xy}) (w\lambda_{xy})^{\gamma_{xy}} (x)^{\gamma_{xy}-1} \exp\{-(\lambda_{xy}x)^{\gamma_{xy}}\} dx = \\
&w^{\gamma_{xy}} \exp\{-(\lambda_{xy}x)^{\gamma_{xy}} - (\lambda_{xy}u)^{\gamma_{xy}}\} - \\
&w^{\gamma_{xy}} \exp\{-(\lambda_{xy}x)^{\gamma_{xy}}\} \Big|_0^t = \\
&w^{\gamma_{xy}} \exp\{-(\lambda_{xy}t)^{\gamma_{xy}} - (\lambda_{xy}u)^{\gamma_{xy}}\} - \\
&w^{\gamma_{xy}} \exp\{-(\lambda_{xy}t)^{\gamma_{xy}}\} - w^{\gamma_{xy}} \exp\{-(\lambda_{xy}u)^{\gamma_{xy}}\} + w^{\gamma_{xy}} \\
&= F_{tu}(w, t, u)
\end{aligned} \tag{7.22}$$

Is Equation (7.22) a valid cumulative distribution function?

1. As $t \rightarrow 0$ and $u \rightarrow 0$, $F_{tu} \rightarrow 0$.
2. As $t \rightarrow 180$ and $u \rightarrow 360$, $F_{tu} \rightarrow w^{\gamma_{xy}}$.
3. F_{tu} is right continuous.

This is a valid cumulative distribution function only if the weights $w^{\gamma_{xy}}$ sum to 1 as $t \rightarrow 180$ and $u \rightarrow 360$. We choose the weights

$$w'_i = \left(\frac{w_i}{\sum_{i=1}^n w_i} \right)^{1/\gamma_{xy}}$$

where w_i is the observed value of the random variable. It does not change for latitude and longitude. Then it is a trivial fact that

$$w^{\gamma_{xy}} = \sum_{i=1}^n w'_i{}^{\gamma_{xy}} = \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} = 1.$$

7.4 Weibull Ellipse

This Section will derive the *Weibull distributed, weighted, deviational ellipse*. The weighted mean estimates (\bar{x}, \bar{y}) under the general linear means model in Section 5 and the Weibull distribution are exactly the same — except possibly the sign on the longitude only if γ_x and γ_y both equal to 1. Generally, this is not the case. The weighted mean estimates (\bar{x}, \bar{y}) under the general linear means model and the Weibull distribution do not equal each other.

Graphically, if the data has a straight line with the following transformation, then it comes from a Weibull distribution.

$$\log t = \log(1/\lambda) + (1/\gamma) \log \left\{ \log_e \left[\frac{1}{1-F} \right] \right\}$$

[Lee (36), p 167] shows a sheet of graph paper for estimating the parameters for the Weibull distribution. It is possible to Xerox that page, whiteout her example, and plot your data. That is what this author did in graduate school. For small data sets, this may seem plausible at first. It is best to run the VBA code to estimate the parameters.

The Weibull ellipse uses the following formulas.

$$\log t_x = \sum_{i=1}^n \log \left[\frac{1}{\hat{\lambda}_x} \right] + \left[\frac{1}{\hat{\gamma}_x} \right] \log \left\{ \log_e \left[\frac{1}{1-F_t(t_i)} \right] \right\}. \quad (7.23)$$

$$\log u_y = \sum_{i=1}^n \log \left[\frac{1}{\hat{\lambda}_y} \right] + \left[\frac{1}{\hat{\gamma}_y} \right] \log \left\{ \log_e \left[\frac{1}{1-F_u(u_i)} \right] \right\}. \quad (7.24)$$

$$\log t_{xy} = \sum_{i=1}^n \log \left[\frac{1}{\hat{\lambda}_{xy}} \right] + \left[\frac{1}{\hat{\gamma}_{xy}} \right] \log \left\{ \log_e \left[\frac{1}{1-F_{tu}(t_i, u_i)} \right] \right\}. \quad (7.25)$$

The two ellipses differ (Exponential and Weibull) differ between the sums of squares of the original data $(w_i x_i, w_i y_i)$ and the transformed data $(\log t_{(x,i)}, \log u_{(y,i)})$. Equation (7.25) gives the mean joint distribution estimate between the weighted latitude and the weighted longitude estimates. The estimators for γ_x , γ_y , and γ_{xy} appear in Equations (7.8), (7.17) and (7.21). We had to perform Secant iterations to find these. The λ 's are already weighted. Therefore, the term w_i is absent from the formulas. The estimators for λ_x , λ_y , and λ_{xy} appear in Equations (7.6), (7.15), and (7.20). These estimators had closed forms *given their respective values for gamma*. Thus, we did not need to perform additional Secant iterations for these values.

7.4.1 Ellipse Angle

Equation (7) in [Yuill, (68)] is general enough to obtain a weighted, Weibull standard deviational ellipse. We need to substitute the original sums of squares with the new

sums of squares and re-calculate the major and minor axes. These axes lengths will determine the total area from which we can compare the area and number of weighted points inside the ellipse. Additionally, we can compare the standard deviations and the eccentricity of the ellipses.

Using equations (7.23), (7.24), and (7.25), we get our *Weibull distributed, weighted, deviational ellipse* in Equation (7.26).

$$\tan \theta = -\frac{(\sum_{i=1}^n \log t_{x,i}^2 - \sum_{i=1}^n \log u_{y,i}^2)}{2 \sum_{i=1}^n \log t_{x,y,i}} \pm \frac{\sqrt{(\sum_{i=1}^n \log t_{x,i}^2 - \sum_{i=1}^n \log u_{y,i}^2)^2 + 4 (\sum_{i=1}^n \log t_{x,y,i})^2}}{2 \sum_{i=1}^n \log t_{x,y,i}} \quad (7.26)$$

This textbook provides an Excel 2010 Visual Basic for Applications (VBA) program that calculates the weighted, Weibull ellipse. In addition to the angles of rotation, it calculates the mean center, the area, the standard deviations, and the eccentricity.

7.5 Ellipse Properties

Equation (7.27) gives the formula for the deviations on the X axis.

$$\delta_x = \sum_{i=1}^n [\log_e(u_{y,i}) \sin \theta_a - \log_e(t_{x,i}) \cos \theta_a]^2 w_i^{\gamma_x} \quad (7.27)$$

where the subscript a denotes the angle of rotation θ about the major axis. Equation (7.28) gives the formula for the deviations on the Y axis.

$$\delta_y = \sum_{i=1}^n [\log_e(u_{y,i}) \cos \theta_b - \log_e(t_{x,i}) \sin \theta_b]^2 w_i^{\gamma_y} \quad (7.28)$$

where the subscript b denotes the angle of rotation θ about the minor axis. Equation (7.27) raises the weights w_i to the power γ_x . Thus, the weights in Equation (7.28) differ from those in Equation (7.27) because Equation (7.28) raises the weights to the power of γ_y .

Equation (7.29) gives the formula for calculating the semi-major axis length a .

$$a^2 = \frac{\sum_{i=1}^n \log u_{y,i}^2}{n} = \frac{2 (\sum_{i=1}^n \log t_{x,y,i})^2}{n[-1(\sum_{i=1}^n \log t_{x,i}^2 - \sum_{i=1}^n \log u_{y,i}^2) + \sqrt{(\sum_{i=1}^n \log t_{x,i}^2 - \sum_{i=1}^n \log u_{y,i}^2)^2 + 4 (\sum_{i=1}^n \log t_{x,y,i})^2}]} \quad (7.29)$$

Equation (7.30) gives the formula for calculating the semi-minor axis length b .

$$b^2 - \frac{\sum_{i=1}^n \log u_{y,i}^2}{n} = \quad (7.30)$$

$$\frac{2 \left(\sum_{i=1}^n \log t_{x,y,i} \right)^2}{n \left[-1 \left(\sum_{i=1}^n \log t_{x,i}^2 - \sum_{i=1}^n \log u_{y,i}^2 \right) + \sqrt{\left(\sum_{i=1}^n \log t_{x,i}^2 - \sum_{i=1}^n \log u_{y,i}^2 \right)^2 + 4 \left(\sum_{i=1}^n \log t_{x,y,i} \right)^2} \right]}$$

subject to the condition

$$\sum_{i=1}^n \log t_{x,i}^2 \geq \sum_{i=1}^n \log u_{y,i}^2.$$

Equation (7.31) gives the formula for calculating the area F .

$$F = \frac{\pi}{n} \sqrt{\left(\sum_{i=1}^n \log t_{x,i}^2 \right) \left(\sum_{i=1}^n \log u_{y,i}^2 \right) - \left(\sum_{i=1}^n \log t_{x,y,i} \right)^2} \quad (7.31)$$

We can always use the area can always use the formula $F = \pi ab$ to check the area. Finally, Equation (7.32) gives the eccentricity e .

$$e = \frac{\sqrt{a^2 - b^2}}{a} \quad (7.32)$$

where Equation (7.29) gives the semi-major axis length a and Equation (7.30) gives the semi-minor axis length b .

7.6 Combining Strata Measures

This section discusses the implications of multi-stratum data. Care must be taken when adding the ellipse properties. Before adding the two stratum area estimates together, we must:

1. Take the square of each stratum estimate.
2. Multiply by the number of observations in each stratum n_i .
3. Add the (squared) stratum estimates.
4. Square root the result.

The above approach applies to the following measures:

- The semi-major and semi-minor axes in Equations (7.29), (7.30) for the Weibull ellipse.
- The area in Equation (7.31) for the Weibull ellipse.

- The standard error for the X -axis in Equation (7.27) for the Weibull ellipse.
- The standard error for the Y -axis in Equation (7.28) for the Weibull ellipse.
- The eccentricity in Equation (7.32) for the Weibull ellipse.

7.7 Kentucky Example

This textbook provides two Secant programs because the likelihood functions for the joint distribution and the individual distributions are different. The code for the two individual distributions is general enough so that it will work on both the latitude and longitude data.

The input parameters to the Secant algorithm are as follow.

```
Call secant1(variable=1, iterations = 10, gamma1=1,
            gamma2=5)
```

- VARIABLE — either the column number to the latitude or the longitude. Be sure to use circular coordinates.
- ITERATIONS — the maximum number of iterations to perform.
- GAMMA1 — the first guess for γ_x or γ_y . Be consistent when working with latitude and longitude.
- GAMMA2 — the second guess for γ_x or γ_y . Again, be consistent when working with latitude and longitude.

The input parameters to the Secant algorithm for the joint distribution are very similar to those for the individual distributions.

```
Call secant2(latitude =1, longitude =5,
            iterations = 4, gamma1=1, gamma2=5);
```

The input parameters to the Weibull program are as follow.

This textbook contains the VBA code for the likelihood functions in Section ???. The Secant programs will return the values for $\hat{\lambda}_x$, $\hat{\lambda}_y$, and $\hat{\lambda}_{xy}$. They will also give the difference between the last two iterations for convergence analysis.

It took much trial and error to find the number of iterations and to set the initial parameter estimates for the Secant method. Figure 7.1 shows a map of the Weibull deviational ellipse for the Kentucky data for the year 2003. Tables 7.1 and 7.2 show the Weibull ellipse statistics for the Kentucky example for the years 2003 and 2004.

Here are some tips on whether the Secant algorithm found credible parameter estimates.

1. The literature states to halt the Secant algorithm when two iterations of the algorithm results in a parameter difference of $|\log L(\lambda, \gamma, k) - \log L(\lambda, \gamma, k - 1)| \leq \epsilon |\log L(\lambda, \gamma, k)|$. Arbitrarily choose $\epsilon = 10^{-6}$.

Table 7.1 2003 Results for the Weibull Model for Kentucky

Description	Measurement
Center	(37.38297005, -86.81297917)
Axes	$a = 3.749083127, b = 2.34002055$
Area	27.56097734 sq. mi.
Standard Deviations	$\delta_x = 1.774503013, \delta_y = 1.783855806$
Rotation	X-Axis
Orientation	$\theta_x = 89.89609928^\circ, \theta_y = -0.103900704^\circ$
Eccentricity	0.781298076

Table 7.2 2004 Results for the Weibull Model for Kentucky

Description	Measurement
Center	(37.40588282, -86.80669635)
Axes	$a = 3.842076104, b = 2.392385246$
Area	28.87665946 sq. mi.
Standard Deviations	$\delta_x = 1.78033221, \delta_y = 1.789024565$
Rotation	X-Axis
Orientation	$\theta_x = 89.90357593^\circ, \theta_y = -0.096424046^\circ$
Eccentricity	0.782476324

2. Choose two initial values for γ that are not too close to each other.
3. The weighted mean values (\bar{x}, \bar{y}) for the latitude and longitude are functions of the parameter estimates $\hat{\lambda}_x$ and $\hat{\lambda}_y$. The values (\bar{x}, \bar{y}) should be within the range of the sample data. This is a hard, must have rule.
4. When running the Secant programs, begin the iterations slowly, working up to more iterations until convergence occurs. According to the literature, performing too many iterations may indeed cause the two functions $\log L(\lambda, \gamma, k)$ and $\log L(\lambda, \gamma, k - 1)$ to converge to division by zero.

Should the reader get an error message while running the VBA code, a number of reasons can cause this.

1. Mouse over the variables `L_X` and `L_X_Minus`. Check if the two values are the same. If so, set the number of iterations to a lower number.
2. We must highlight the cells with the data in the spreadsheet before going into the VBA Editor.

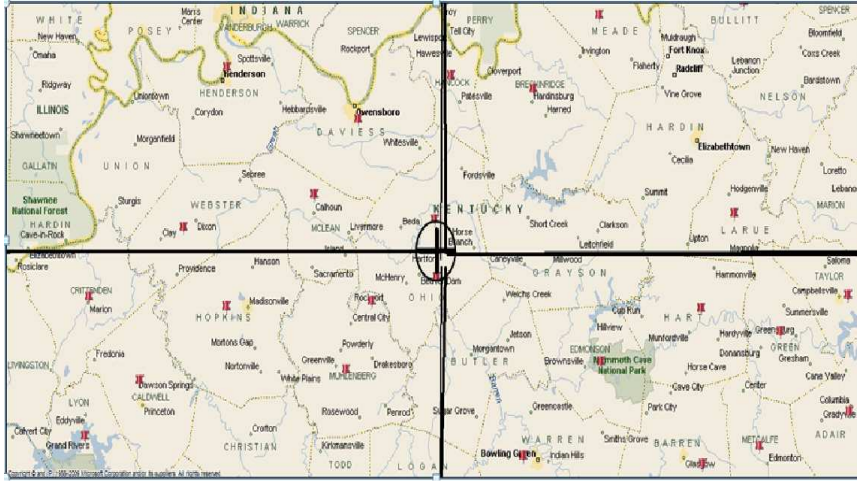


Figure 7.1 This figure shows both the Weibull deviational ellipse for the Kentucky data for the year 2003. The zoom level is 200 miles above the earth. The standard deviational ellipse and the exponential deviational ellipse were too large to fit on the map. It has the same center of gravity as the other ellipses. The Weibull ellipse is much smaller than the other two.

7.8 Crime Example

The example in this section comes from violent crime in the U.S. for 2007 and 2008. Tables 7.3 and 7.4 summarize the results for the weighted standard deviational ellipse for violent crime in the U.S. for the years 2007 and 2008. The data includes the continental U.S. Figure 7.2 shows a graph of both the Weibull ellipse and the exponential ellipse for the Violent Crime data in the U.S. for the year 2007.

Table 7.3 2007 Results for the Weibull Model on Violent Crime in the U.S.

Description	Measurement
Center	(36.87659982, -92.11312648)
Axes	$a = 2.684221693, b = 1.421680712$
Area	11.98865122 sq. mi.
Standard Deviations	$\delta_x = 1.305334214, \delta_y = 1.320237824$
Rotation	X-Axis
Orientation	$\theta_x = 89.79107974^\circ, \theta_y = -0.208920244^\circ$
Eccentricity	0.848220318

Table 7.4 2008 Results for the Weibull Model on Violent Crime in the U.S.

Description	Measurement
Center	(36.88957315, -92.03222761)
Axes	$a = 2.653413305$, $b = 1.407222952$
Area	11.73053137 sq. mi.
Standard Deviations	$\delta_x = 1.306416749$, $\delta_y = 1.322062007$
Rotation	X-Axis
Orientation	$\theta_x = 89.78095302^\circ$, $\theta_y = -0.219046959^\circ$
Eccentricity	0.847782278

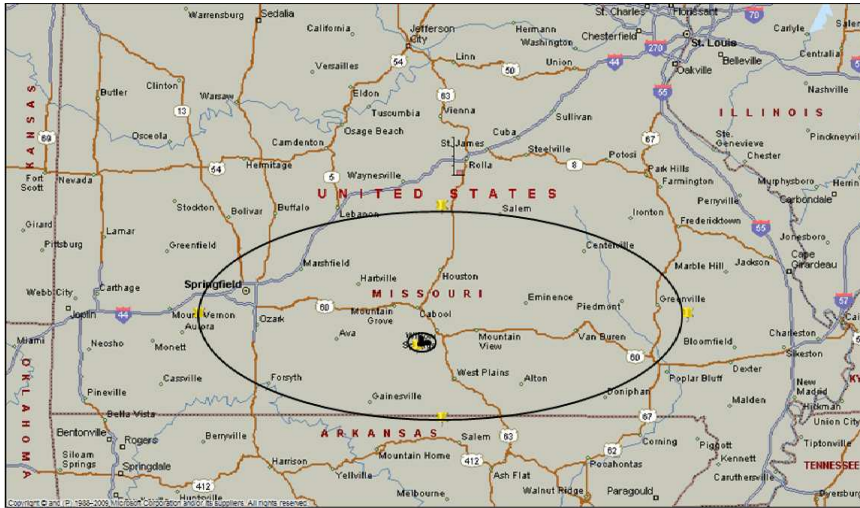


Figure 7.2 This figure shows both the Weibull deviational ellipse for the Crime data for the year 2007. The zoom level is 550 miles above the earth. The standard deviational ellipse was too large to fit on the map. It has the same center of gravity as the other ellipses. The Weibull ellipse is much smaller than the other two.

7.9 Airline Examples

Table 7.5 summarizes the data for Air France Flight 447. This data falls into two strata. To obtain the correct mean center, take modulo 45° of the latitude.

$$\bar{x}' = \text{mod}(45, \bar{x}) = \text{mod}(45, 46.348875) = 1.3488750.$$

To obtain the correct longitude, take -1 modulo 360° of the longitude

$$\bar{y}' = \text{mod}(360, \bar{y}) = -\text{mod}(360, 328.59075) = -31.4092500.$$

Table 7.6 summarizes the data for Malaysia Flight MH 370. To obtain the correct mean center, take modulo arithmetic of the latitude as follow:

$$\bar{x}' = \text{mod}(90, \bar{x}) = -\text{mod}(90, 80.76666667) = -9.2333333.$$

The VBA code for the Weibull ellipse calculates the same longitude as that of the standard deviational ellipse. This ellipse is not very eccentric. The areas of the Weibull ellipses are smaller than those of the exponential ellipses. This, of course, indicates that the semi-major axis a and the semi-minor axis b will be shorter for the Weibull ellipses. Notice that for the MH 370 data, the axis of rotation changed from the Y-axis to the X-axis from the previous examples.

Table 7.5 Results for the Weibull Model on Air France Flight 447

Description	Measurement
Center	(1.3488750, -31.4092500)
Axes	$a = 3.838408409, b = 2.278996317$
Area	27.48176858 sq. mi.
Standard Deviations	$\delta_x = 3.826401076, \delta_y = 3.851996075$
Rotation	Y-Axis
Orientation	$\theta_x = 89.62804212^\circ, \theta_y = -0.371957877^\circ$
Eccentricity	0.804660854

Table 7.6 Results for the Weibull Model on Malaysia Flight MH 370

Description	Measurement
Center	(-9.2333333, 96.34166667)
Axes	$a = 1.698669509, b = 1.182651775$
Area	6.311253896 sq. mi.
Standard Deviations	$\delta_x = 1.144499793, \delta_y = 1.174490655$
Rotation	X-Axis
Orientation	$\theta_x = 89.38210464^\circ, \theta_y = -0.617895361^\circ$
Eccentricity	0.717826135

7.10 GDP Example

Tables 7.7 and 7.8 summarize the Weibull ellipse for the GDP data for 2008 and 2009. We notice that the mean center is not *exactly the same* for the exponential ellipse and the standard deviational ellipse. This is because the values for the parameters γ_x and γ_y did not equal to exactly 1. Therefore, we can find some variation there.

Table 7.7 2008 Results for the Weibull Model on Gross Domestic Product

Description	Measurement
Center	(44.95865049, 173.1927484)
Axes	$a = 1.876920973$, $b = 1.48828381$
Area	8.777 sq. mi.
Standard Deviations	$\delta_x = 1.5549$, $\delta_y = 1.5947$
Rotation	X-Axis
Orientation	$\theta_x = 89.28^\circ$, $\theta_y = -0.72^\circ$
Eccentricity	0.61

Table 7.8 2009 Results for the Weibull Model on Gross Domestic Product

Description	Measurement
Center	(45.10440668, 172.959405)
Axes	$a = 1.872381589$, $b = 1.493600132$
Area	8.786 sq. mi.
Standard Deviations	$\delta_x = 1.5588$, $\delta_y = 1.5997$
Rotation	X-Axis
Orientation	$\theta_x = 89.26$, $\theta_y = -0.74$
Eccentricity	0.60

7.11 Axes Length Comparison

To compare the lengths of the axes to those of the exponential ellipse, we use the relationships of the mean centers. When the parameters $\gamma_x = 1$ and $\gamma_y = 1$, then the mean center (\bar{x}, \bar{y}) are the same for both ellipses. The sums of squares are exactly the same also.

1. $0 < \gamma_x < 1$ (Roots): The exponential ellipse is smaller.
2. $\gamma_x > 1$ (Powers): The Weibull ellipse is smaller.

7.12 Sample Distribution Fitting

Three underlying distributions have been presented, the normal distribution (the weighted standard deviational ellipse), the exponential distribution (the weighted exponential deviational ellipse), and the Weibull distribution (the weighted Weibull deviational ellipse). We presented three data sets usually over multiple years, the USDA Kentucky corn example, the Department of Justice violent crime example, and OECD gross domestic product example. The examples expand from a local to a global perspective. We can use graphs to show the distribution of the data. What

happens when all of the graphs show a straight line? What is the distribution of the sample?

We can continue to use the likelihood functions derived in previous sections. Since we already calculated the parameters for the samples, it is an easy task to calculate the likelihood function. Using the maximum likelihood theory, the distribution with the maximum value of the likelihood function L (and $\log L$) is the better fit.

Example	Model	$2k - 2 \ln L(w, x, y)$	Comment
KY 2003 Corn Acres	SDE	-11,436.74	Min AIC
KY 2003 Corn Acres	Exp	222,911.00	
KY 2003 Corn Acres	Weibull	274,909.42	
KY 2004 Corn Acres	SDE	-2,110.16	Min AIC
KY 2004 Corn Acres	Exp	235,242.64	
KY 2004 Corn Acres	Weibull	291,173.28	
US Crime 2007	SDE	5,697.60	Min AIC
US Crime 2007	Exp	64,790.55	
US Crime 2007	Weibull	78,359.93	
US Crime 2008	SDE	5,707.33	Min AIC
US Crime 2008	Exp	64,623.60	
US Crime 2008	Weibull	78,111.50	
Global GDP 2008	SDE	4,976.84	Min AIC
Global GDP 2008	Exp	39,282.18	
Global GDP 2008	Weibull	47,388.74	
Global GDP 2009	SDE	4,906.12	Min AIC
Global GDP 2009	Exp	39,391.99	
Global GDP 2009	Weibull	47,560.40	

Given the following summary tables, for each of the data sets, the standard deviational distribution usually fits the sample data best.

When comparing the two probability models, the exponential model consistently outperforms the Weibull model. Adding the shape parameter using the Weibull distribution did not help in this particular analysis.

Example	Parameter Estimates (Rounded)
KY 2003 Corn Acres	$\hat{\alpha} = -1147830.91, \hat{\beta} = 1087263.36$
KY 2004 Corn Acres	$\hat{\alpha} = -1030547.49, \hat{\beta} = 1020882.76$
KY 2003 Corn Acres	$\hat{\lambda}_x = 2.40751, \hat{\lambda}_y = 0.32944, \hat{\lambda}_{xy} = 0.57958$
KY 2004 Corn Acres	$\hat{\lambda}_x = 2.45951, \hat{\lambda}_y = 0.33676, \hat{\lambda}_{xy} = 0.59240$
KY 2003 Corn Acres	$\hat{\lambda}_x = 2.40751, \hat{\lambda}_y = 0.32944, \hat{\lambda}_{xy} = 0.57958$
	$\hat{\gamma}_x = 0.76857, \hat{\gamma}_y = 0.76237, \hat{\gamma}_{xy} = 1.00000$
KY 2004 Corn Acres	$\hat{\lambda}_x = 2.45951, \hat{\lambda}_y = 0.33675, \hat{\lambda}_{xy} = 0.59240$
	$\hat{\gamma}_x = 0.76640, \hat{\gamma}_y = 0.76018, \hat{\gamma}_{xy} = 1.00000$
US Crime 2007	$\hat{\alpha} = 42344.54, \hat{\beta} = 35730.76$
US Crime 2008	$\hat{\alpha} = 41952.14, \hat{\beta} = 35152.29$
US Crime 2007	$\hat{\lambda}_x = 1.32875, \hat{\lambda}_y = 0.18291, \hat{\lambda}_{xy} = 0.32156$
US Crime 2008	$\hat{\lambda}_x = 1.38289, \hat{\lambda}_y = 0.18286, \hat{\lambda}_{xy} = 0.32146$
US Crime 2007	$\hat{\lambda}_x = 1.32876, \hat{\lambda}_y = 0.18291, \hat{\lambda}_{xy} = 0.32156$
	$\hat{\gamma}_x = 0.765003, \hat{\gamma}_y = 0.74781, \hat{\gamma}_{xy} = 1.00000$
US Crime 2008	$\hat{\lambda}_x = 1.32829, \hat{\lambda}_y = 0.18286, \hat{\lambda}_{xy} = 0.32146$
	$\hat{\gamma}_x = 0.765406, \hat{\gamma}_y = 0.75763, \hat{\gamma}_{xy} = 1.00000$
Global GDP 2008	$\hat{\alpha} = 44620.24, \hat{\beta} = 13643.50$
Global GDP 2009	$\hat{\alpha} = 44258.72, \hat{\beta} = 13570.98$
Global GDP 2008	$\hat{\lambda}_x = 0.84687, \hat{\lambda}_y = 0.25678, \hat{\lambda}_{xy} = 0.39406$
Global GDP 2009	$\hat{\lambda}_x = 0.84915, \hat{\lambda}_y = 0.25969, \hat{\lambda}_{xy} = 0.39774$
Global GDP 2008	$\hat{\lambda}_x = 0.84687, \hat{\lambda}_y = 0.25677, \hat{\lambda}_{xy} = 0.39405$
	$\hat{\gamma}_x = 0.77847, \hat{\gamma}_y = 0.76382, \hat{\gamma}_{xy} = 1.00000$
Global GDP 2009	$\hat{\lambda}_x = 0.84915, \hat{\lambda}_y = 0.25969, \hat{\lambda}_{xy} = 0.39773$
	$\hat{\gamma}_x = 0.77787, \hat{\gamma}_y = 0.76407, \hat{\gamma}_{xy} = 1.00000$

A word of caution about interpreting the parameters in the above table. For example, the parameters under the SDE models for the Violent Crime data survey and the Global GDP data survey have approximately the same north-south parameter estimate $\hat{\alpha}$, but a vastly different east-west parameter estimate $\hat{\beta}$. Does this mean that the United States capital (money) went east-ward, while the crime stayed in the U.S? No, because the two data surveys intentionally represent two different levels of geography to illustration the calculations.

7.13 Exercises

1. The GDP data is bimodal. Perform a data analysis similar to Section 5.9 for the Greece end of the data.
2. Show that $\sum_{i=1}^n w_i^{\gamma_y} = 1$ where

$$w_i = \left(\frac{w'_i}{\sum_{i=1}^n w'_i} \right)^{1/\gamma_y} = 1$$

Solution: Let $\sum_{i=1}^n w'_i = w$. Then,

$$w_i = \left(\frac{w'_i}{\sum_{i=1}^n w'_i} \right)^{1/\gamma_y} = \left(\frac{w'_i}{w} \right)^{1/\gamma_y} \Rightarrow \sum_{i=1}^n w_i^{\gamma_y} = \sum_{i=1}^n \frac{w'_i}{w} = \frac{\sum_{i=1}^n w'_i}{w} = \frac{w}{w} = 1.$$

3. Show that

$$\frac{n}{\hat{\lambda}_x} = \bar{x} \text{ in Chapter 5.}$$

Solution:

$$\hat{\lambda} = \left(\frac{n}{w \sum_{i=1}^n x_i^{\gamma_x}} \right)^{1/\gamma_x}, \quad 0 \leq x_i \leq 180.$$

Let $w = \sum_{i=1}^n \frac{w_i}{w_i}$.

$$\hat{\lambda} = \left(\frac{n}{\sum_{i=1}^n \left(\frac{w_i x_i^{\gamma_x}}{\sum_{i=1}^n w_i} \right)} \right)^{1/\gamma_x}$$

4. Show that

$$\frac{n}{\hat{\lambda}_y} = \bar{y} \text{ in Chapter 5.}$$

5. For what value of γ_x and γ_y in the Weibull ellipse makes it equal to the exponential ellipse [special case]? Show this algebraically.

7.14 VBA Code

The function `Log10(X)` calculates and returns the logarithm with the base 10 of the variable `X`. By default, VBA calculates natural logarithms. The programmer must write his own function for non-natural logarithms.

The `Weibull()` subroutine (15) calculates the remaining measures for the Weibull ellipse. The first two `FOR-NEXT` loops (16, 17) calculate the weights and save them to columns `D`, `F`, and `G` in the active spreadsheet. The source code that estimates $\hat{\lambda}_x$, $\hat{\lambda}_y$, $\hat{\lambda}_{xy}$, $\hat{\gamma}_x$, $\hat{\gamma}_y$, and $\hat{\gamma}_{xy}$ need these values initialized. The subroutines `secant1()` and `secant2()` only needed two iterations of a single weight initialized.

The `FOR-NEXT` loop (19) calculates the *sums* used in Equation (7.26), the axes lengths, the area, and the eccentricity.

The subroutine `Angle_of_Rotation()` determines which axis to rotate the ellipse. Since θ from Equation (7.26) contains a plus and minus sign, both values are calculated. If $\sum_{i=1}^n \log t_{x,i}^2 > \sum_{i=1}^n \log u_{y,i}^2$, then rotate the ellipse about the Y-axis. Otherwise, rotate the ellipse about the X-axis.

The subroutine `Area()` calculates the *sums* used for the area. The program calculates the area, F , in Equation (7.31) and saves it in the variable `f`.

We use Equation (7.27) to calculate the sums for the standard error for the X-axis. We use Equation (7.28) to calculate the sums for the standard error for the Y-axis. The VBA statements following the `FOR-NEXT` loop finish the calculations and save the results in the variables `error_x` and `error_y`.

The subroutine `Axes_Length()` determines the major and minor axis lengths. If $\sum_{i=1}^n \log t_{x,i}^2 \geq \sum_{i=1}^n \log u_{y,i}^2$, then we can calculate the semi-major axis length directly with Equation (7.29) and the semi-minor axis length with Equation (7.30). Otherwise, we must reverse the roles of the sums in the two equations.

We calculate the eccentricity using the subroutine `Eccentricity()`. If the axis length $a \geq b$, then the formula in Equation (7.32) can be used directly used. Otherwise, we reverse the roles of a and b so that we do not take the square root of a negative number.

The set of VBA statements inside the `WITH-END-WITH` (20) save the results to the spreadsheet called `STATS` in column `L` (column 12). We write appropriate text documentation to column `K` (column 11).

```

Sub secant1(iterations, variable, Gamma1, gamma2)
j = 1
dx = 1 'trivial value
condition = 0 'trivial value
  While (j <= iterations)And (dx > condition)
    'set the weights
    n = Selection.Rows.Count
    sum_weight1 = 0
    sum_weight2 = 0
    For i = 2 To n
      sum_weight1 = sum_weight1 + (ActiveSheet.Cells(i, 3).Value)
      sum_weight2 = sum_weight2 + (ActiveSheet.Cells(i, 3).Value)
    Next
    ActiveSheet.Cells(1, 4).Value = "Lambda 1 Weights"
    ActiveSheet.Cells(1, 6).Value = "Lambda 2 Weights"
    For i = 2 To n
      ActiveSheet.Cells(i, 4).Value = ((ActiveSheet.Cells(i, 3).Value) / sum_weight1) ^ (1 / Gamma1) 'previous iteration
      ActiveSheet.Cells(i, 6).Value = ((ActiveSheet.Cells(i, 3).Value) / sum_weight2) ^ (1 / gamma2) 'next iteration
    Next
    'end of setting the weights
    sumw_x2 = 0
    sumw_x1 = 0
    sum_x2 = 0
    sum_x1 = 0
    sum_logx = 0
    For i = 2 To n
      sumw_x2 = sumw_x2 + ( ActiveSheet.Cells(i, 6) ^ gamma2 * ActiveSheet.Cells(i, variable).Value ) ^ gamma2 'weighted
      sumw_x1 = sumw_x1 + ( ActiveSheet.Cells(i, 4) ^ Gamma1 * ActiveSheet.Cells(i, variable).Value ) ^ Gamma1 'weighted
      sum_x2 = sum_x2 + ActiveSheet.Cells(i, variable).Value ^ gamma2 'unweighted
      sum_x1 = sum_x1 + ActiveSheet.Cells(i, variable).Value ^ Gamma1 'unweighted
      sum_logx = sum_logx + Log(Abs( ActiveSheet.Cells(i, variable).Value )) 'unweighted
    Next
    lambda2 = ((n - 1) / sumw_x2) ^ (1 / gamma2) 'weighted
    lambda1 = ((n - 1) / sumw_x1) ^ (1 / Gamma1) 'weighted
    L_x = 0
    L_X_Minus = 0
    L_x = L_x + (n - 1) * Log(gamma2) + (n - 1) * gamma2 * _
    Log(sum_weight2 * lambda2) + (gamma2 - 1) * sum_logx - lambda2 ^ gamma2 * sum_x2
    L_X_Minus = L_X_Minus + (n - 1) * Log(Gamma1) + (n - 1) * Gamma1 * _
    Log(sum_weight1 * lambda1) + (Gamma1 - 1) * sum_logx - lambda1 ^ Gamma1 * sum_x1
  End While
End Sub

```

```

1  [
    [
      gamma3 = gamma2 - L_x * (gamma2 - Gamma1) / (L_x - L_X_Minus)
      Gamma1 = gamma2 'correct assignment
      gamma2 = Abs(gamma3) 'correct assignment, must be positive
    2  dx = Abs(L_x - L_X_Minus)
      condition = 10 ^ (-6) * Abs(L_x)
      j = j + 1
      Wend
    ]
  ]

1  [
    [
      With Worksheets("STATS")
      .Cells (1, 7).Value = "Secant Algorithm"
      .Cells (2, 7).Value = "dx"
      .Cells (2, 8).Value = dx
      .Cells (2, 9).Value = condition
      .Cells (3, 7).Value = "Lambda"
    6  .Cells (3, 8).Value = lambda2
      .Cells (4, 7).Value = "Gamma"
      .Cells (4, 8).Value = gamma2
      .Cells (5, 7).Value = "Mean"
      .Cells(5, 8).Value = (n - 1) / lambda2
    ]
  ]
End Sub

```

```

Sub secant2(iterations, Gamma1, gamma2)
j = 1
dx = 1 'trivial value
condition = 0 'trivial value
  While (j <= iterations) And (dx > condition)
    'set the weights
    n = Selection.Rows.Count
    sum_weight1 = 0
    sum_weight2 = 0
    For i = 2 To n
      sum_weight1 = sum_weight1 + (ActiveSheet.Cells (i, 3).Value )
      sum_weight2 = sum_weight2 + (ActiveSheet.Cells (i, 3).Value )
    Next
    ActiveSheet.Cells (1, 4).Value = "Lambda 1 Weights"
    ActiveSheet.Cells (1, 6).Value = "Lambda 2 Weights"
    For i = 2 To n
      ActiveSheet.Cells (i, 4).Value = ((ActiveSheet.Cells (i, 3).Value ) / sum_weight1) ^ (1 / Gamma1) 'previous iteration
      ActiveSheet.Cells (i, 6).Value = ((ActiveSheet.Cells (i, 3).Value ) / sum_weight2) ^ (1 / gamma2) 'next iteration
    Next
  End While
  'end of setting the weights
  sumw_x2 = 0
  sumw_y2 = 0
  sum_x1 = 0
  sum_y1 = 0
  log_gamma1 = 0
  log_gamma2 = 0
  log_xy1 = 0
  log_xy2 = 0
  sum_Lambdax2 = 0
  sum_Lambdax1 = 0
  sum_Lambday2 = 0
  sum_Lambday1 = 0
  sumw_Lambda2 = 0
  sumw_Lambda1 = 0
  'calculate the sums used in the Secant Method

```

```

For i = 2 To n
sumw_x2 = sumw_x2 + (ActiveSheet.Cells(i, 6).Value ^ gamma2 * ActiveSheet.Cells(i, 1).Value) ^ gamma2 'weighted
sumw_y2 = sumw_y2 + (ActiveSheet.Cells(i, 6).Value ^ gamma2 * Abs(ActiveSheet.Cells(i, 2).Value)) ^ gamma2 'weighted
sumw_x1 = sumw_x1 + (ActiveSheet.Cells(i, 4).Value ^ Gamma1 * ActiveSheet.Cells(i, 1).Value) ^ Gamma1 'weighted
sumw_y1 = sumw_y1 + (ActiveSheet.Cells(i, 4).Value ^ Gamma1 * Abs(ActiveSheet.Cells(i, 2).Value)) ^ Gamma1 'weighted
11 log_gamma2 = log_gamma2 + 2 * Log(gamma2)
log_gamma1 = log_gamma1 + 2 * Log(Gamma1)
log_xy2 = log_xy2 + (gamma2 - 1) * Log(ActiveSheet.Cells(i, 1).Value * Abs(ActiveSheet.Cells(i, 2).Value))
log_xy1 = log_xy1 + (Gamma1 - 1) * Log(ActiveSheet.Cells(i, 1).Value * Abs(ActiveSheet.Cells(i, 2).Value))
Next
lambda2 = (2 * (n - 1) / (sumw_x2 + sumw_y2)) ^ (1 / gamma2) 'weighted
lambda1 = (2 * (n - 1) / (sumw_x1 + sumw_y1)) ^ (1 / Gamma1) 'weighted
'calculate the sums involving lambda
For i = 2 To n
sum_LambdaX2 = sum_LambdaX2 + (lambda2 * ActiveSheet.Cells(i, 1).Value) ^ gamma2
sum_LambdaX1 = sum_LambdaX1 + (lambda1 * ActiveSheet.Cells(i, 1).Value) ^ Gamma1
12 sum_LambdaY2 = sum_LambdaY2 + (lambda2 * Abs(ActiveSheet.Cells(i, 2).Value)) ^ gamma2
sum_LambdaY1 = sum_LambdaY1 + (lambda1 * Abs(ActiveSheet.Cells(i, 2).Value)) ^ Gamma1
sumw_Lambda2 = sumw_Lambda2 + 2 * (n - 1) * gamma2 * Log(ActiveSheet.Cells(i, 6).Value * lambda2)
sumw_Lambda1 = sumw_Lambda1 + 2 * (n - 1) * Gamma1 * Log(ActiveSheet.Cells(i, 4).Value * lambda1)
Next
'estimate gamma using the Secant Method
L_x = 0
L_X_Minus = 0
For i = 2 To n
13 L_x = L_x + log_gamma2 + sumw_Lambda2 + log_xy2 - sum_LambdaX2 - sum_LambdaY2
L_X_Minus = L_X_Minus + log_gamma1 + sumw_Lambda1 + log_xy1 - sum_LambdaX1 - sum_LambdaY1
Next
gamma3 = gamma2 - L_x * (gamma2 - Gamma1) / (L_x - L_X_Minus)
Gamma1 = gamma2 'correct assignment
gamma2 = Abs(gamma3) 'correct assignment, must be positive
dx = Abs(L_x - L_X_Minus)
condition = 10 ^ (-6) * Abs(L_x)
j = j + 1
Wend

```

```

7  [ With Worksheets("STATS")
    [ .Cells(1, 7).Value = "Secant Algorithm"
      [ .Cells(2, 7).Value = "dx"
        [ .Cells(2, 8).Value = dx
          [ .Cells(2, 9).Value = condition
            [ .Cells(3, 7).Value = "Lambda XY"
              [ .Cells(3, 8).Value = lambda2
                [ .Cells(4, 7).Value = "Gamma XY"
                  [ .Cells(4, 8).Value = gamma2
                    [ .Cells(5, 7).Value = "Mean XY"
                      [ .Cells(5, 8).Value = (n - 1) / lambda2
                        End With
                    End Sub
                End Sub
            End Sub
        End Sub
    End With
    Sub Weibull(gamma_x, gamma_y, gamma_xy, lambda_x, lambda_y, lambda_xy)
    'set the weights
    n = Selection.Rows.Count
    m = n - 1
    sum_weight1 = 0
    sum_weight2 = 0
    sum_weight3 = 0
    [ For i = 2 To n
      [ sum_weight1 = sum_weight1 + (ActiveSheet.Cells(i, 3).Value) 'single x
        [ sum_weight2 = sum_weight2 + (ActiveSheet.Cells(i, 3).Value) 'single y
          [ sum_weight3 = sum_weight3 + (ActiveSheet.Cells(i, 3).Value) 'joint xy
            Next
          ActiveSheet.Cells(1, 4).Value = "Lambda X Weights"
          ActiveSheet.Cells(1, 6).Value = "Lambda Y Weights"
          ActiveSheet.Cells(1, 7).Value = "Lambda XY Weights"
        [ For i = 2 To n
          [ ActiveSheet.Cells(i, 4).Value = ((ActiveSheet.Cells(i, 3).Value) / sum_weight1) ^ (1 / gamma_x) 'lambda_x
            [ ActiveSheet.Cells(i, 6).Value = ((ActiveSheet.Cells(i, 3).Value) / sum_weight2) ^ (1 / gamma_y) 'lambda_y
              [ ActiveSheet.Cells(i, 7).Value = ((ActiveSheet.Cells(i, 3).Value) / sum_weight3) ^ (1 / gamma_xy) 'lambda_xy
                Next
              'end of setting the weights
              f_x = 0
              F_y = 0
              F_xy = 0
              L_xy = 0 'used for the log L calculation
              logt_x = 0
              logt_y = 0
              logt_xy = 0
              logt_xsqr = 0
              logt_ysqr = 0
              logt_xysqr = 0

```



```

For i = 2 To n
f_x = ActiveSheet.Cells(i, 4).Value ^ gamma_x - ActiveSheet.Cells(i, 4).Value ^ gamma_x * _
EXP(-1 * lambda_x * (ActiveSheet.Cells(i, 1).Value) ^ gamma_x)
F_y = ActiveSheet.Cells(i, 6).Value ^ gamma_y - ActiveSheet.Cells(i, 6).Value * _
EXP(-1 * lambda_y * Abs(ActiveSheet.Cells(i, 2).Value) ^ gamma_y)
F_xy = ActiveSheet.Cells(i, 7).Value ^ gamma_xy * _
EXP(-1 * lambda_xy * ActiveSheet.Cells(i, 1).Value ^ gamma_xy - lambda_xy * Abs(ActiveSheet.Cells(i, 2).Value) ^ gamma_xy) -
- ActiveSheet.Cells(i, 7).Value ^ gamma_xy * EXP(-1 * lambda_xy * ActiveSheet.Cells(i, 1).Value ^ gamma_xy) -
- ActiveSheet.Cells(i, 7).Value ^ gamma_xy * EXP(-1 * lambda_xy * Abs(ActiveSheet.Cells(i, 1).Value) ^ gamma_xy) +
- ActiveSheet.Cells(i, 7).Value ^ gamma_xy
logt_x = logt_x + (Log10(1 / lambda_x) + (1 / gamma_x) * _
Log10(Log(1 / (1 - f_x)))) * ActiveSheet.Cells(i, 4).Value ^ gamma_x
logt_y = logt_y + (Log10(1 / lambda_y) + (1 / gamma_y) * _
18 Log10(Log(1 / (1 - F_y)))) * ActiveSheet.Cells(i, 6).Value ^ gamma_y
logt_xy = logt_xy + (Log10(1 / lambda_xy) + (1 / gamma_xy) * Log10(Log(1 / (1 - F_xy)))) * _
ActiveSheet.Cells(i, 7).Value ^ gamma_xy
15 logt_xsqr = logt_xsqr + ((m * (Log10(1 / lambda_x) + (1 / gamma_x) * Log10(Log(1 / (1 - f_x)))) * _
ActiveSheet.Cells(i, 4).Value ^ gamma_x) ^ 2
logt_ysqr = logt_ysqr + ((m * (Log10(1 / lambda_y) + (1 / gamma_y) * Log10(Log(1 / (1 - F_y)))) * _
ActiveSheet.Cells(i, 6).Value ^ gamma_y) ^ 2
logt_xysqr = logt_xysqr + ((m * (Log10(1 / lambda_xy) + (1 / gamma_xy) * Log10(Log(1 / (1 - F_xy)))) * _
ActiveSheet.Cells(i, 7).Value ^ gamma_xy) ^ 2
L_xy = L_xy + 2 * Log(gamma_xy) + 2 * (n - 1) * gamma_xy * Log(lambda_xy * _
ActiveSheet.Cells(i, 4).Value) + (gamma_xy - 1) * (Log(ActiveSheet.Cells(i, 1).Value * _
ActiveSheet.Cells(i, 2).Value)) - (lambda_xy * ActiveSheet.Cells(i, 1).Value) ^ (gamma_xy) - (lambda_xy * _
ActiveSheet.Cells(i, 2).Value) ^ (gamma_xy)
Next
'calculate the area statistics for the Weibull Ellipse
Call Angle_of_Rotation(logt_xsqr, logt_ysqr, logt_xy, note, atheta, itheta)
'calculate the deltas
delta_x = 0
delta_y = 0
eF_x = 0
eF_y = 0

```

```

15 19 For i = 2 To n
    eF_x = ActiveSheet.Cells(i, 4).Value ^ gamma_x - ActiveSheet.Cells(i, 4).Value ^ gamma_x * _
    EXP(-1 * lambda_x * (ActiveSheet.Cells(i, 1).Value))
    eF_y = ActiveSheet.Cells(i, 6).Value ^ gamma_y - ActiveSheet.Cells(i, 6).Value ^ gamma_y * _
    EXP(-1 * lambda_y * Abs(ActiveSheet.Cells(i, 2).Value))
    elogt_x = Log10(1 / lambda_x) + (1 / gamma_x) * Log10(Log(1 / (1 - eF_x)))
    elogt_y = Log10(1 / lambda_y) + (1 / gamma_y) * Log10(Log(1 / (1 - eF_y)))
    delta_x = delta_x + (elogt_y * Sin(atheta / 57.2957795) - elogt_x * Cos(atheta / 57.2957795)) ^ 2 * _
    ActiveSheet.Cells(i, 4).Value ^ gamma_x
    delta_y = delta_y + (elogt_y * Cos(itheta / 57.2957795) - elogt_x * Sin(itheta / 57.2957795)) ^ 2 * _
    ActiveSheet.Cells(i, 6).Value ^ gamma_y
Next

```

```

error_x = Sqr(delta_x)
error_y = Sqr(delta_y)
'calculate the area, eccentricity, and axes lengths
f = Area(m, logt_xsqr, logt_ysqr, logt_xy)
Call Axes_Length(m, logt_xsqr, logt_ysqr, logt_xy, a, b)
e = Eccentricity(a, b)
Area2 = Area_Check(a, b)
'write the stats to the STATS spreadsheet
  With Worksheets("STATS")
    .Cells(1, 11).Value = "Weibull Ellipse Statistics"
    .Cells(2, 11).Value = "Mean Latitude"
    .Cells(3, 11).Value = "Mean Longitude"
    .Cells(2, 12).Value = (n - 1) / lambda_x
    .Cells(3, 12).Value = (n - 1) / lambda_y
    .Cells(4, 11).Value = "Rotation"
    .Cells(5, 11).Value = "Major Axis Rotation"
    .Cells(6, 11).Value = "Minor Axis Rotation"
    .Cells(4, 12).Value = note
    .Cells(5, 12).Value = atheta
    .Cells(6, 12).Value = itheta
    .Cells(7, 11).Value = "Delta X"
    .Cells(8, 11).Value = "Delta Y"
    20 .Cells(7, 12).Value = error_x
    .Cells(8, 12).Value = error_y
    .Cells(9, 11).Value = "Area (F)"
    .Cells(9, 12).Value = f
    .Cells(10, 11).Value = "Major Axis Length (a)"
    .Cells(11, 11).Value = "Minor Axis Length (b)"
    .Cells(10, 12).Value = a
    .Cells(11, 12).Value = b
    .Cells(12, 11).Value = "Area Check"
    .Cells(13, 11).Value = "Eccentricity (e)"
    .Cells(14, 11).Value = "Likelihood Function L(x,y)"
    .Cells(12, 12).Value = Area2
    .Cells(13, 12).Value = e
    .Cells(14, 12).Value = L_xy
  End With
End Sub

```

CHAPTER 8

SPHERICAL STATISTICS

8.1 Introduction

So far, we have used the latitude and longitude coordinate system found in most geography literature. The statistical estimates gave meaningful results (e.g., the mean center landed in the survey area, the expected standard errors followed a known distribution, and two methods for calculating the area matched). In this Chapter, we introduce the concept of spherical statistics. For these analyses to work, we must convert the data from latitude and longitude coordinates to the Cartesian coordinate system (i.e. having an X , Y , and Z -axes each extending to infinity with a single pole at the point $(0, 0, 0)$).

This Chapter gives another view of the approach to spatial analysis. The spherical statistics may not match the other Chapter statistics, exactly. These are presently because they give additional information such as the "shape" of the data. The derivations and calculations of the spherical mean centers and standard deviations may not align well with previous Chapters. We follow mathematical and statistical theory carefully.

These calculations are unweighted, which means that we are analyzing the latitude and longitude observations in the absence of the random variable. Given this, multiple

year survey statistics give the same results since the latitudes and longitudes have been augmented to the data sets. New information not attainable before, will be presented.

This Section covers the multiple ellipses from a single population. The data shows more than one value with a high a frequency (probability) occurring in the survey. In directional statistics, we define the *mode* as the maximum a value of a given distribution and we define the *anti-mode* as the minimum of the distribution. For example, if we construct a histogram of a distribution for categorical ranges of the sample data and count the number of observations that fall into each category, then the category with the largest number of observations is the mode. The category with the smallest number of observations is the anti-mode.

The literature on multi-modal ellipses divides the techniques for measuring the ellipses into several broad categories:

- Multiple variable statistics.
- Multivariate statistics.
- Multi-modal statistics.

Multiple variable statistics includes such techniques as linear regression with one random variable and multiple explanatory variables [B. L. Bowerman and R. T. O'Connell, (3)]. We estimate the parameters using least squares. On the other hand, multivariate statistics linear regression with two or more random variables and can include one or more explanatory variables for each random variable [R. Khattree and D. N. Naik, (34)]. There may be correlation among the random variables. [R. G. Petersen (51); E. T. Lee, (36)] discuss more general analysis of variance models (ANOVA models) and specifically the distributions discussed in preceding Chapters for the univariate cases. [R. Khattree and D. N. Naik, (34)] give analogous multivariate analysis of variance models (MANOVA models).

Multimodal statistics is a property of a particular distribution albeit a univariate distribution or a multivariate distribution. [L. Cobb, P. Koppstein, and N. H. Chen (8)] contains a discussion on multimodal distributions for the exponential distribution, the normal distribution, and the beta distribution. The paper also contains a test for multimodality [(8, page 128]. Although the paper seems relevant, after some correspondence with L. Cobb, the paper does not discuss random variables or random distributions. [Mardia, (46)] contains extensive discussions on univariate and multivariate sampled distributions. Of interest in this text is [Mardia, (46), page 221], which shows how to calculate the mean center of gravity, using sines and cosines. The author has a simple test for multimodality [Mardia (46), page 209]. The author gives other, advanced tests also.

[D. J. Steffensmeier, E. A. Allan, M. D. Harer, and C. Streifel, (60)] provide an in-depth example of crime data linking crime type and age. This article discusses two distinct issues.

- Committing crime as a person ages (supported by Sociological studies).

- Shift in the age distribution by the type of crime (supported by the data).

The authors rely on percentages and quantiles throughout the paper for statistical comparisons. To show multimodality, the authors show the shift in the age distribution using quantiles of committing a specific type of crime over a period from 1940 to 1980. They also provide the usual skewness and kurtosis calculations for the shape of the distributions as well as chi-square tests.

8.2 Concepts

Directional statistics models probability densities and angles. In prior chapters, our data came in the triplet (x_i, y_i, w_i) for the latitude, the longitude, and the random variable and mapped onto the Earth's surface. In this chapter, the data comes from the sample of size n of points on a sphere $P(\phi_i, \theta_i)$. The latitude ϕ_i is the vertical line running north and south parallel of the Prime Meridian and has the range $[-90^\circ, 90^\circ]$. The longitude θ_i is the horizontal line running east and west parallel of the Equator and has the range $[-180^\circ, 180^\circ]$. We define the Equator as having latitude of 0° . We define the Prime Meridian as having longitude of 0° . We will need to convert decimal degrees into radians often in this Chapter because the calculations rely heavily on trigonometry functions. [Shelby (58), pages 389-392] gives an excellent discussion of projecting coordinates from the Earth onto a circle.

How do we convert our latitude and longitude data to directional data? The Equator divides the Earth in half into the Northern and Southern Hemispheres each having the same mass. The latitude of the Equator is $\phi^\circ = 0^\circ$.



Figure 8.1 This figure shows a map of the U.S. relative to the Equator (represented by the black line).

To obtain a scatter diagram in two dimensions on the surface of a sphere, use a projection on the sphere. Consider the polar coordinates (ρ, ψ) where ρ is the polar

distance and ψ is the polar angle. [Mardia, (46)] defines the *stereographic equal-angle projection* as $\rho = \tan\left(\frac{\theta}{2}\right)$, and $\psi = \phi$ where $0 < \rho < 1$ and $0 < \psi < 2\pi$. [Mardia, (46)] calls the plot a *Wulff-net* or *stereographic-net*. It is an "area preserving" plot compared to other plots.

Steps to mimic the previous chapters:

1. Convert the spatial data to circular data.
2. Calculate the quantities $(\bar{l}_0, \bar{m}_0, \bar{n}_0)$. These calculations result in radians when using Excel 2010 spreadsheets.
3. Convert the results back to degrees from radians.
4. Convert the mean latitude and mean longitude back to spatial coordinates.

8.3 \bar{X} -Axis Mean Center and Resultant Length

To find the mean direction North-South using directional statistics, we first calculate the cosines in Equation (8.1).

$$n_0 = \sum_{i=1}^n n_i = \sum_{i=1}^n \cos \phi_i. \quad (8.1)$$

where $P(\phi_i, \theta_i)$ is a point on the Earth; ϕ_i equals to the latitude and θ_i equals to the longitude for observation i . To find the average of n_0 , we divide by the sample size n to obtain

$$\bar{X} = \bar{n}_0 = \frac{1}{n} n_0. \quad (8.2)$$

We define the resultant length R next.

$$\bar{R} = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n l_i\right)^2 + \left(\frac{1}{n} \sum_{i=1}^n m_i\right)^2 + \left(\frac{1}{n} \sum_{i=1}^n n_i\right)^2}, \quad 0 < \bar{R} \leq 1 \quad (8.3)$$

where n is the sample size and l_i , m_i , and n_i are defined in this Section and the next. For now it suffices to say that R is between 0 and 1.

The diagram in Figure 8.2 shows the triangle with the longitude θ . It has a right angle (90°), the angle θ identified, the hypotenuse R and the two sides n_0 and m_0 . We will define m_0 in Section 8.4.

Equation (8.4) gives the weighted cosines and average latitude using the random variable W .

$$\bar{X}_w = \bar{n}_{w_0} = \sum_{i=1}^n \frac{w_i n_i}{\sum_{i=1}^n w_i} = \sum_{i=1}^n \frac{w_i \cos \phi_i}{\sum_{i=1}^n w_i}. \quad (8.4)$$

Accordingly, Equation (8.5) gives the weighted resultant length \bar{R} .

$$\bar{R}_w = \sqrt{\left(\sum_{i=1}^n \frac{w_i l_i}{\sum_{i=1}^n w_i}\right)^2 + \left(\sum_{i=1}^n \frac{w_i m_i}{\sum_{i=1}^n w_i}\right)^2 + \left(\sum_{i=1}^n \frac{w_i n_i}{\sum_{i=1}^n w_i}\right)^2}, 0 < \bar{R}_w \leq 1 \quad (8.5)$$

8.4 Y and Z-Axis Mean Centers

To find the means in the Y and Z-axis directions using directional statistics, we need both the latitude (ϕ_i) and the longitude (θ_i) observations for the points $P(\phi_i, \theta_i)$ on the Earth. We calculate the sums for the Y-axis in Equation (8.6) and the sums for the Z-axis in Equation (8.7).

$$l_0 = \sum_{i=1}^n l_i = \sum_{i=1}^n \sin \phi_i \cos \theta_i, \quad (8.6)$$

$$m_0 = \sum_{i=1}^n m_i = \sum_{i=1}^n \sin \phi_i \sin \theta_i. \quad (8.7)$$

$P(\phi_i, \theta_i)$ is a point on the Earth. ϕ_i is called the latitude and θ_i is called the longitude for observation i . For now, it suffices to say that, Equations (8.8) and (8.9) gives the unweighted means for l_0 and m_0 .

$$\bar{Y} = \bar{l}_0 = \frac{1}{n} l_0. \quad (8.8)$$

$$\bar{Z} = \bar{m}_0 = \frac{1}{n} m_0. \quad (8.9)$$

Equations (8.10) and (8.11) give the weighted averages for the Y and Z axes.

$$\bar{Y}_w = \bar{l}_{w_0} = \sum_{i=1}^n \frac{w_i \sin \phi_i \cos \theta_i}{\sum_{i=1}^n w_i}. \quad (8.10)$$

$$\bar{Z}_w = \bar{m}_{w_0} = \sum_{i=1}^n \frac{w_i \sin \phi_i \sin \theta_i}{\sum_{i=1}^n w_i}. \quad (8.11)$$

8.4.1 Spherical Variance

[Mardia and Jupp, (47)] give two definitions for the sample spherical variance. The first sample spherical variance is defined in Equation (8.12). This definition minimizes a matrix of the arithmetic mean of the Euclidean distances.

$$S^* = 2(1 - \bar{R}), 0 < S^* < 1 \quad (8.12)$$

where R is the resultant length. For small S^* , the data is uniform. There is no preferred direction. For large S^* , the data is directional and may be clustered or may not be clustered. When $\bar{R} \approx 1$, the data are heavily concentrated. \bar{R} measures clustering around the mean direction.

A second measure of the sample spherical variance is given in Equation (8.13). This definition is an analogue from a derivation of the matrix T . The matrix T will be used later.

$$S^* = 1 - \bar{R}^2, 0 < S^* < 1. \quad (8.13)$$

8.5 Test and Characterizations

The spherical distributions can have one of three visual characteristics:

- Uniform: The points have a uniformly dispersed pattern and have no "preferred" direction.
- Bimodal: The points have a clustered pattern about the ends of an axis (e.g. the north and south poles). The axis joins the two modes as well as the two anti-modes.
- Girdle: The points have a dispersed, elliptic pattern. The mode(s) are on a plane.

To detect which distribution we have, define the matrix B as:

$$B = nI - T = \begin{bmatrix} n - \sum_{i=1}^n l_i^2 & -\sum_{i=1}^n l_i m_i & -\sum_{i=1}^n l_i n_i \\ -\sum_{i=1}^n l_i m_i & n - \sum_{i=1}^n m_i^2 & -\sum_{i=1}^n m_i n_i \\ -\sum_{i=1}^n l_i n_i & -\sum_{i=1}^n m_i n_i & n - \sum_{i=1}^n n_i^2 \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad (8.14)$$

We find the eigenvalues of matrix B to determine the visual shape of the data. We define the vector \underline{b}_1 as the first row times the vector of unknowns:

$$\underline{b}_1 = \begin{bmatrix} n - \sum_{i=1}^n l_i^2 & -\sum_{i=1}^n l_i m_i & -\sum_{i=1}^n l_i n_i \\ -\sum_{i=1}^n l_i m_i & n - \sum_{i=1}^n m_i^2 & -\sum_{i=1}^n m_i n_i \\ -\sum_{i=1}^n l_i n_i & -\sum_{i=1}^n m_i n_i & n - \sum_{i=1}^n n_i^2 \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} b_1(n - \sum_{i=1}^n l_i^2) \\ -b_2 \sum_{i=1}^n l_i m_i \\ -b_3 \sum_{i=1}^n l_i n_i \end{pmatrix}$$

Vector \underline{b}_2 corresponds to row 2 and vector \underline{b}_3 corresponds to row 3.

Finding the eigenvalues of B involves two steps:

1. Find the roots λ to the determinant $Det(B - \lambda I)$ where the matrix I is the 3×3 identity matrix. There may be at most three unique roots to the determinant.
2. For each eigenvalue λ_j , solve the system $(B - \lambda_j I)\underline{v} = \underline{0}$.

Equation (8.15) gives the determinant for $B - \lambda_j I$.

$$\begin{aligned}
 Det(B - \lambda_j I) = & \tag{8.15} \\
 & \left(n - \sum_{i=1}^n l_i^2 - \lambda_j \right) \left(n - \sum_{i=1}^n m_i^2 - \lambda_j \right) \left(n - \sum_{i=1}^n n_i^2 - \lambda_j \right) - \\
 & \sum_{i=1}^n l_i m_i \sum_{i=1}^n m_i n_i \sum_{i=1}^n l_i n_i - \sum_{i=1}^n l_i n_i \sum_{i=1}^n l_i m_i \sum_{i=1}^n m_i n_i + \\
 & \sum_{i=1}^n l_i n_i \sum_{i=1}^n l_i n_i \left(n - \sum_{i=1}^n m_i^2 - \lambda_j \right) + \\
 & \sum_{i=1}^n m_i n_i \sum_{i=1}^n m_i n_i \left(n - \sum_{i=1}^n l_i^2 - \lambda_j \right) + \\
 & \sum_{i=1}^n l_i m_i \sum_{i=1}^n l_i m_i \left(n - \sum_{i=1}^n n_i^2 - \lambda_j \right).
 \end{aligned}$$

Arrange the eigenvalues λ_1, λ_2 , and λ_3 in ascending order such that $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq n$ where n is the sample size and $\sum_{i=1}^3 \lambda_i = n$. The trace of matrix B sums to $2n$. As in [Mardia (46), page 223-224], we interpret the diagonal elements of B and the eigenvectors.

$\beta_i, i = 1, 2, 3$ equal to the diagonal elements of B where $\beta_1 = b_{11}, \beta_2 = b_{22}$, and $\beta_3 = b_{33}$. Working with the matrix T , we define the following eigenvalues and eigenvectors.

$$\lambda_i = n - \beta_i, \quad \underline{t}_i = \underline{b}_i, \quad i = 1, 2, 3, \quad \lambda_1 + \lambda_2 + \lambda_3 = n.$$

The terms large and small are subjective. The reader may ask whether there is a statistical test to differentiate between the values for λ_i ? [Mardia (46), pages 276-283] shows such a test statistic. Let the null hypothesis be $H_0 : \lambda_1 = \lambda_2 = \lambda_3 = \frac{n}{3}$, which states that the eigenvalues have equal values. The test statistic is S_U

$$S_U = \frac{15}{2n} \sum_{i=1}^3 \left(\lambda_i - \frac{1}{3}n \right)^2 \sim \chi^2(5). \tag{8.16}$$

Equation (8.16) is asymptotically a chi-square distribution with five degrees of freedom. For large n , the chi-square test works best. For sufficiently large sample sizes, this should not be an issue. S_U tests for uniform λ_i 's. Other statistical tests test for girdle and rotational symmetry using the λ_i 's.

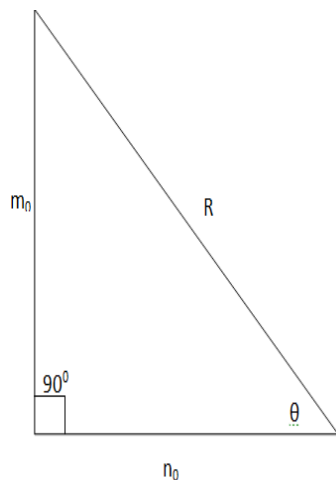


Figure 8.2 This figure shows the diagram of a triangle with the quantities for the longitude θ , n_0 , and the resultant length R .

Table 8.1 Interpretation of the Eigenvalues

Eigenvalues	Condition	Distribution	Build
$\lambda_1 \approx \lambda_2 \approx \lambda_3$		Uniform	Axes having no orientation
Small λ_1, λ_2 . Large λ_3	$\lambda_1 \neq \lambda_2$	Unimodal for large R . Bimodal otherwise.	If unimodal, the concentration at end of the eigenvector \underline{t}_3 . Otherwise, concentration at both ends of the eigenvector \underline{t}_3 .
Small λ_1, λ_2 . Large λ_3	$\lambda_1 \approx \lambda_2$	Unimodal for large R . Bipolar otherwise.	Rotational symmetry about \underline{t}_3 .
Small λ_1 . Large λ_2 , and λ_3 .	$\lambda_2 \neq \lambda_3$	Girdle	Girdle plane spanned by the the eigenvectors \underline{t}_2 and \underline{t}_3 .
Small λ_1 . Large λ_2 , and λ_3 .	$\lambda_2 \approx \lambda_3$	Symmetric girdle	Rotational symmetry about the eigenvector \underline{t}_1 .

8.5.1 Finding the Eigenvalues

This section presents a simple Newton algorithm for finding the eigenvalues λ_j , $j = 1, 2, 3$. The Gaussian elimination algorithm must be run to find the eigenvectors for each λ_j . The subroutine must be run multiple times. The routine needs to find the three values in the *determinant of the matrix B* :

$$\begin{bmatrix} a - \lambda & d & e \\ f & b - \lambda & g \\ h & i & c - \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

where the constants a, b, c, d, e, f, g, h , and i are given. The cubic equation must be solved for λ :

$$(a - \lambda)(b - \lambda)(c - \lambda) + dgh + efi - \quad (8.17)$$

$$he(b - \lambda) - ig(a - \lambda) - fd(c - \lambda).$$

The derivative of Equation (8.17) with respect to λ is given in Equation (8.18).

$$-(b - \lambda)(c - \lambda) - (a - \lambda)(c - \lambda) - (a - \lambda)(b - \lambda) + eh + ig + fd. \quad (8.18)$$

[Cheney and Kincaid, (6), pages 83-90] give an introduction and an algorithm for Newton's method. We wish to apply the method to find the eigenvalue λ_j , $j = 1, 2$ or 3. The recursive definition of Newton's method is

$$\lambda_{k+1} = \lambda_k - \frac{f(\lambda_k)}{f'(\lambda_k)}$$

where $k = 0, 1, 2, \dots$ represents the iteration (also called the index in the sequence). We use Equation (8.17) to calculate the function $f(\lambda_k)$. We use Equation (8.18) to calculate the function $f'(\lambda_k)$.

8.5.2 Finding the Eigenvectors

This book comes with a Gaussian elimination algorithm that performs row reduction on a 3×3 matrix. The subroutine only needs to be run once for each λ_j , $j = 1, 2, 3$. There is no trial and error as there was in finding the eigenvalues. [Grossman, (24), pages 277-281] gives the theory behind eigenvalues and eigenvectors and the strategy for finding each. Keep in mind that we are only solving a 3×3 matrix in Spherical Statistics and the code given reflects this simplicity. Setting the initial value for v_3 to back solve the system of equations is the reader's prerogative as stated in [Grossman, (24)]. There are an infinite number of solutions to the equations.

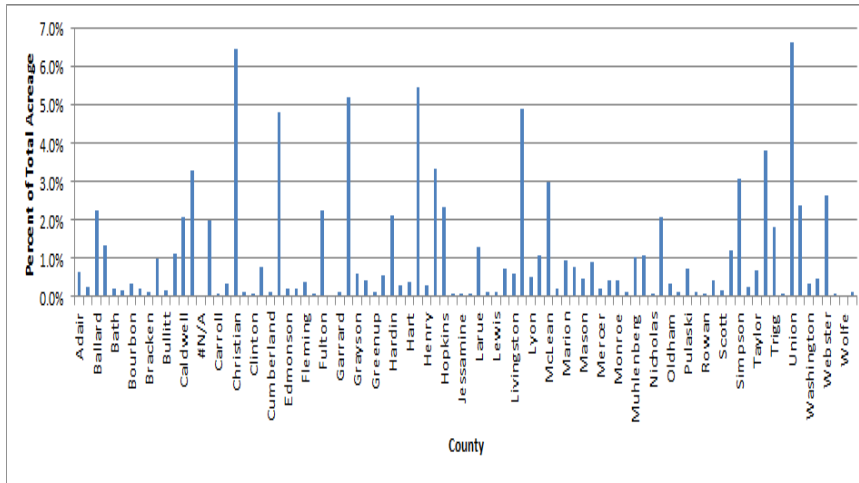


Figure 8.3 This figure shows the histogram for the Kentucky 2003 data.

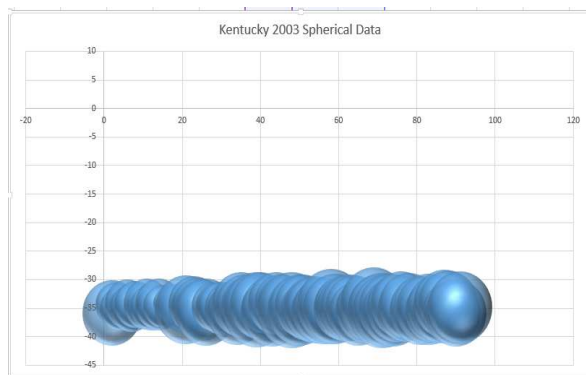


Figure 8.4 This figure shows the Kentucky 2003 converted data.

8.6 Kentucky Example

Figure 8.3 shows the histogram for the Kentucky 2003 data. There does not appear to be any one or two observations with a disproportional weight. This data looks fairly random. Figure 8.4 shows the Kentucky 2003 converted data. The Y_i and Z_i observations have been interchanged in the graph. This is because the Z -axis points directly at the reader. Had the the data not been interchanged, only one circle would have been shown. The 91 data markers, shown as circles, would lay on top of each other since the data is fairly linear.

We run the `Newton()` subroutine in Excel to calculate the eigenvalues. Since the equation of the determinant of B is triadic, we expect to find three roots to the determinant. The `Newton()` subroutine will find two of the three. The third λ_3 can be deduced by the equation $\lambda_3 = n - \lambda_1 - \lambda_2$ where λ_1 and λ_2 are known eigenvalues that have been found using the `Newton()` subroutine. Using a value close to n is a good guess for λ_3 . Here, there is no implied ordering between λ_1 , λ_2 , and λ_3 . Table 8.2 gives the eigenvalues and eigenvectors for the Kentucky data.

Matrix B for the 2003 Kentucky data is as follow:

$$B = \begin{pmatrix} 33.58777584 & -3.351383372 & 43.36764038 \\ -3.351383372 & 89.77014071 & 2.59054283 \\ 43.36764038 & 2.59054283 & 56.64208345 \end{pmatrix}$$

The reader can verify that the diagonal elements sum to $2n = 180$. Matrix T for the 2003 Kentucky data is as follow:

$$T = \begin{pmatrix} 56.41222416 & 3.351383372 & -43.36764038 \\ 3.351383372 & 0.229859286 & -2.59054283 \\ -43.36764038 & -2.59054283 & 33.35791655 \end{pmatrix}$$

Table 8.2 Eigenvalues and Eigenvectors for the 2003 Kentucky Data

	Value	Eigenvector \underline{t}'_i
λ_1	0.006629744	(0.734788415, 0.573334304, 1.0)
λ_2	0.034874073	(0.944911442, -2.955186475, 1.0)
λ_3	89.95849618	(-1.300507805, -0.077445098, 1.0)

The reader can verify that the property $T\underline{v}^T = \lambda_i\underline{v}^T$ where the vector \underline{v} is the eigenvector corresponding to the eigenvalue λ_i for $i = 1, 2, 3$. Just looking at the magnitude of the values of λ_1 , λ_2 , and λ_3 we can conclude that the eigenvalues are not equal. Hence, at a minimum the data is directional. We will calculate the test statistic S_U for illustrative purposes.

$$S_U = \frac{15}{2(90)} \sum_{i=1}^3 \left(\lambda_i - \frac{1}{3}(90) \right)^2 =$$

$$\frac{15}{180} \left((0.006629744 - 30)^2 + (0.034874073 - 30)^2 + \right.$$

$$\left. (89.95849618 - 30)^2 \right) = 449.378.$$

The cut-off value is $\chi^2(5) = 11.07$ at the 95% confidence level and $\chi^2(5) = 15.09$ at the 99% confidence level. Since $449 > 15.09$, we reject the hypothesis that the

Table 8.3 2003 Results for the Spherical Model for Kentucky

Description	Measurement
Center	(45.35963106, 2.701085798, -34.87836501)
(l_0, m_0, n_0)	(4082.366795, 243.0977218, -3139.052851)
Resultant Length	$R = 0.999769376$
Sample Size	$n = 90$
Data Type	Directional
Data Shape	Bipolar
Spherical Variance	$S = 0.000461194$

eigenvalues are equally valued. Thus, the data is directional. We need the resultant length R to determine the visual distribution of the points. This Section covers the derivation of R .

To obtain the results in Table 8.3, we ran the `Spherical_Dev_Ellipse()` routine. It does not have any input parameters, but does assume that the first three columns in the spreadsheet have the following values:

- Column 1 contains the latitude values.
- Column 2 contains the longitude values.
- Column 3 contains the weight values (random variable). This is not used in the computations.

In Table 8.2, we notice we have two small eigenvalues and one large eigenvalue. According to Table 8.1, these observations have a bipolar distribution because λ_3 is large; λ_1 and λ_2 are small. The concentration of the spherical data is at both ends of the vector \underline{t}_3 . We knew the data originated from a bipolar coordinate system with a north and south pole. The vector \underline{t}_3 in radians is $\underline{t}_3 = (-1.300507805, -0.077445098, 1.0)$.

8.7 Crime Example

We look at the histogram in Figure 8.5 and notice three states with approximately the same mode — California, Texas, and Florida. California accounts for approximately 13.6% of the total violent crime, Texas 8.7%, and Florida 9.4%. These three observations weigh the calculations disproportionately. One is on the west coast, one in the south west, and the other in the south. Figure 8.6 shows a graph of the converted data. The data is symmetric. One ball is separated from the rest of the data.

We need to determine the visual distribution of the data. To accomplish this, we first find the eigenvalues of matrix T (uses the 2007 data). The matrices B and T are as follow:

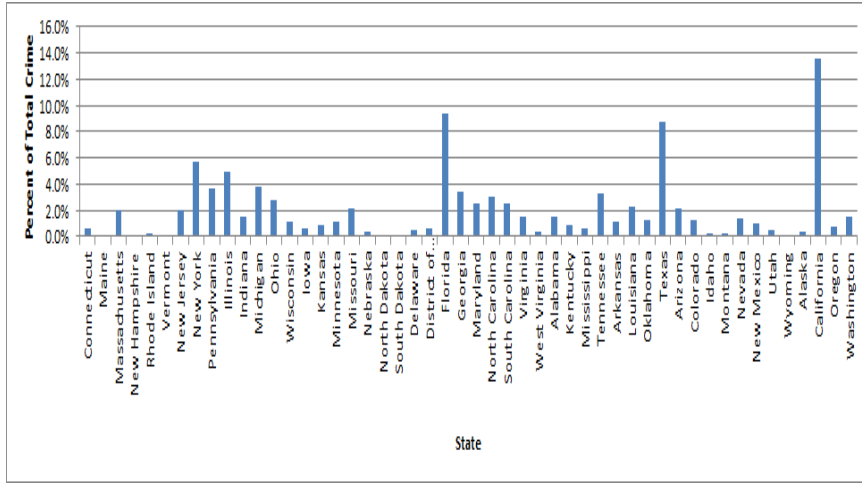


Figure 8.5 This figure shows the histogram for the Violent Crime 2008 data. Three possible modes appear at California, Texas, and Florida.

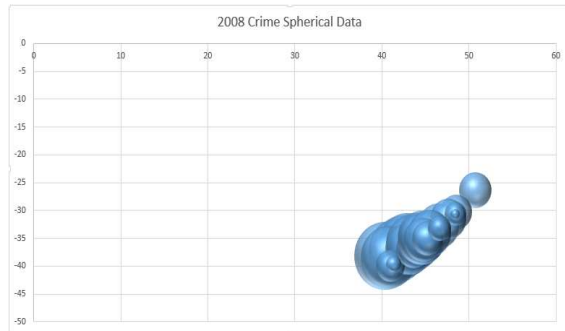


Figure 8.6 This figure shows the 2008 US Crime converted data.

$$B = \begin{pmatrix} 19.82660121 & 0.299933647 & 22.9389611 \\ 0.299933647 & 47.65668036 & -0.234303397 \\ 22.9389611 & -0.234303397 & 30.51671843 \end{pmatrix}$$

$$T = \begin{pmatrix} 29.17339879 & -0.299933647 & -22.9389611 \\ -0.299933647 & 1.343319644 & 0.234303397 \\ -22.9389611 & 0.234303397 & 18.48328157 \end{pmatrix}$$

Table 8.4 gives the eigenvalues and eigenvectors.

Table 8.4 Eigenvalues and Eigenvectors for the 2008 Vilonent Crime Data

Diagonal	Value	Eigenvector \underline{t}'_i
λ_1	0.27486963	(0.793812871, 0.003545129, 1.0)
λ_2	1.340181802	(-0.649158203, -136.7206467, 1.0)
λ_3	47.38494857	(-1.259802117, 0.013295803, 1.0)

We test the hypothesis that the eigenvalues are equal versus they are not equal using Equation (8.16) using the test statistic S_U .

$$S_U = \frac{15}{2(49)} \sum_{i=1}^3 \left(\lambda_i - \frac{1}{3}(49) \right)^2 =$$

$$\frac{15}{98} [(0.27486963 - 16.33)^2 + (1.340181802 - 16.33)^2 +$$

$$(47.38494857 - 16.33)^2] = 221.46.$$

We reject the null hypothesis $H_0 : \lambda_i = \frac{n}{3}, i = 1, 2, 3$ since $221.46 > 15.09$ at the 99% confidence level. We also reject the hypothesis that the 2008 crime data is uniform. The usual statistics for the sphere are given in Table 8.5.

Next, we determine the data shape by examining the λ'_i s. Using the λ'_i s in Table 8.4 and the interpretation in Table 8.1, we most likely have a unimodal distribution because λ_3 is large; λ_1 and λ_2 are small; R is close to 1. Since this is a unimodal distribution, the concentration of the data is at one end of the vector \underline{t}_3 where $\underline{t}_3 = (-1.259802117, 0.013295803, 1.0)$.

8.8 Airline Example

The plotted data for Malaysia Flight MH 370 appears in Figure 8.7. Since the data is unweighted, the plot follows the latitude and longitude observations more closely than the previous plots. The matrix B for Malaysia Flight MH 370 is as follow:

$$B = \begin{pmatrix} 1.112364265 & 0.182317512 & -2.559488235 \\ 0.182317512 & 11.9960797 & 0.061435538 \\ -2.559488235 & 0.061435538 & 10.89155604 \end{pmatrix}$$

Table 8.5 2008 Results for the Spherical Model for 2008 Crime

Description	Measurement
Center	(44.11484743, -0.535389408, -35.03421925)
(l_0, m_0, n_0)	(2161.627524, -26.23408097, -1716.676743)
Resultant Length	$R = 0.983257676$
Sample Size	$n = 48$
Data Type	Directional
Data Shape	Unimodal
Spherical Variance	$S = 0.033204342$

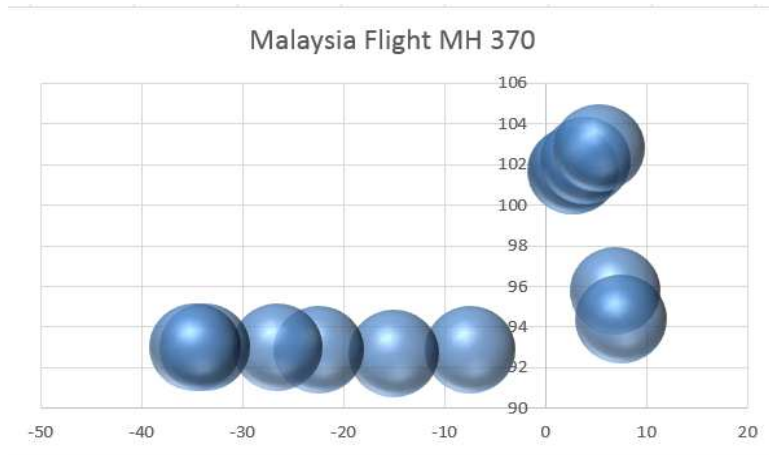


Figure 8.7 This figure shows the plotted, unweighted data for Malaysia Flight MH 370.

The sum of the diagonal elements sum to $2n = 2 \times 12 = 24$. The matrix T is as follow:

$$T = \begin{pmatrix} 10.88763573 & -0.182317512 & 2.559488235 \\ -0.182317512 & 0.003920305 & -0.061435538 \\ 2.559488235 & -0.061435538 & 1.10844396 \end{pmatrix}$$

The sum of the diagonal elements sum to $n = 12$. Table 8.6 gives the eigenvalues and eigenvectors. Since λ_1 and λ_2 are small; and λ_3 is large; and since $\lambda_1 \neq \lambda_2$, conclude that the data is unimodal. Furthermore, the data is concentrated at the end of of the vector t_2 . Since the eigenvalues are not approximatly equal, the data is directional.

Table 8.7 summarizes the data for Malaysia Flight MH 370.

Table 8.6 Eigenvalues and Eigenvectors for Malaysia Flight MH 370

Diagonal	Value	Eigenvector t'_i
λ_1	0.000186134	(0.221578377, 27.27062033, 1.0)
λ_2	0.479603641	(-0.246521967, -0.034666478, 1.0)
λ_3	11.52021023	(4.066236163, -0.069708353, 1.0)

Table 8.7 Results for the Spherical Model for Malaysia Flight MH 370

Description	Measurement
Center	(54.45338165, -0.941760218, 13.57286876)
(l_0, m_0, n_0)	(653.4405798, -11.30112262, 162.8744251)
Resultant Length	$R = 0.979607219$
Sample Size	$n = 12$
Data Type	Directional
Data Shape	Unimodal
Spherical Variance	$S = 0.040785563$

8.9 GDP Example

We begin by looking at the histogram of the observations in Figure 8.8. Two bars appear longer than the rest. These indicate modes in the distribution of the data. The United States accounts for approximately 25% of the total GDP and China accounts for approximately 14%. This will influence the shape of the data. Figure 8.9 shows the converted data. The data is spread out into the eight quadrants of the X , Y , and Z planes.

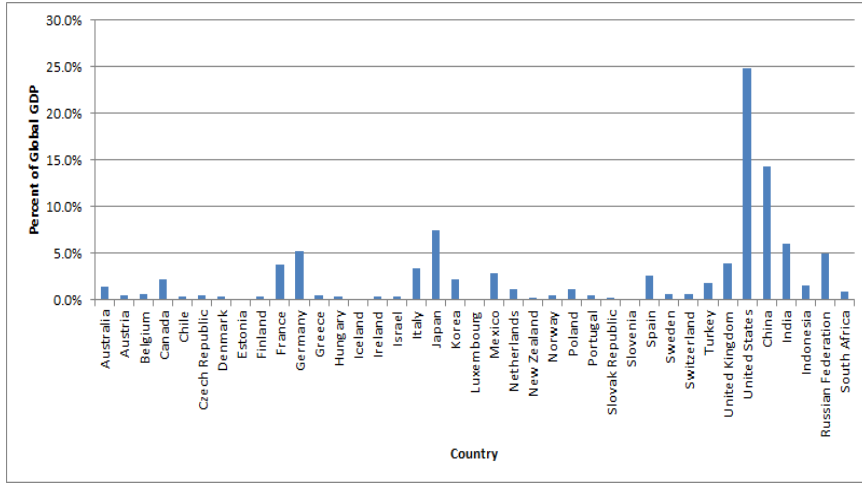


Figure 8.8 This figure shows the histogram for the GDP 2008 data. The U.S. accounts for approximately 25% of the total GDP and China accounts for approximately 14%.

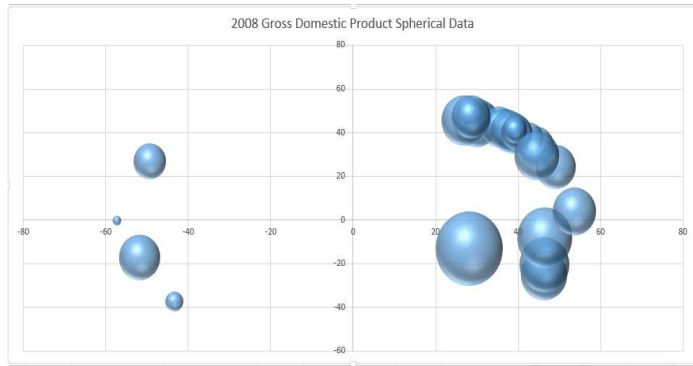


Figure 8.9 This figure shows the 2008 Gross Domestic Product converted data.

The matrices B and T for the GDP 2008 data is

$$B = \begin{pmatrix} 19.35173834 & -10.96157833 & -2.888240385 \\ -10.96157833 & 23.90143271 & -1.953298019 \\ -2.888240385 & -1.953298019 & 34.74682895 \end{pmatrix}$$

The diagonal of B sums to $2n = 2 \times 39 = 78$.

$$T = \begin{pmatrix} 19.64826166 & 10.96157833 & 2.888240385 \\ 10.96157833 & 15.09856729 & 1.953298019 \\ 2.888240385 & 1.953298019 & 4.253171052 \end{pmatrix}$$

The diagonal of T sums to $n = 39$. We test the hypothesis that the eigenvalues are equal versus they are not equal using Equation (8.16) using the test statistic S_U .

$$S_U = \frac{15}{2(39)} \sum_{i=1}^3 \left(\lambda_i - \frac{1}{3}(39) \right)^2 = \frac{15}{78} [(3.728832776 - 13)^2 + (6.216179462 - 13)^2 + (29.05498776 - 13)^2] = 74.95.$$

We reject the null hypothesis $H_0 : \lambda_i = \frac{n}{3}, i = 1, 2, 3$ since $74.95 > 15.09$ at the 99% confidence level. We also reject the hypothesis that the 2008 crime data is uniform. The usual statistics for the sphere are given in Table 8.9. Table 8.8 shows the eigenvalues and eigenvectors for the 2008 GDP example.

Table 8.8 Eigenvalues and Eigenvectors for the 2008 Gross Domestic Product

Diagonal	Value	Eigenvector t'_i
λ_1	3.728832776	(-0.187814653, 0.00927436, 1.0)
λ_2	6.216179462	(5.008334449, -6.400592894, 1.0)
λ_3	29.05498776	(5.546422255, 4.496198733, 1.0)

Table 8.9 2008 Results for the Spherical Model for GPD

Description	Measurement
Center	(27.14815799, 24.99879834, 6.46609507)
(l_0, m_0, n_0)	(1058.778161, 974.9531353, 252.1777077)
Resultant Length	$R = 0.653921742$
Sample Size	$n = 39$
Data Type	Directional
Data Shape	Bimodal
Spherical Variance	$S = 0.572386356$

Since the eigenvalue λ_3 is larger relative to λ_2 and λ_3 , and since R is not close to 1, this may be a bimodal distribution. The concentration is at both ends of the vector $t_3 = (5.546422255, 4.496198733, 1.0)$.

8.10 Exercises

1. Perform a test of equality on the eigenvalues for the MH 370 data.
2. Perform a data analysis on the Air France 447 data in Table 4.6.

8.11 VBA Code for Spherical Statistics

```

Dim n As Integer 'number of observations
Dim l_0, m_0, n_0, l0_sqrd, m0_sqrd, n0_sqrd, lm_sum, ln_sum, mn_sum As Double
Dim l_sum, m_sum, n_sum As Double
Dim R_length, R_length2 As Double
Dim bar_x, bar_y1, bar_y2 As Double
Dim TempSecSin As Double
Dim SumWeights As Double 'sum of the weight column
Dim R_zPrime As Double 'test statistic for the resultant R
Dim F_Theory As Double
Dimrm S As Double 'the spherical variance
Sub SphereSetWeights(AVar)
  n = Selection.Rows.Count
  'title the columns
  With ActiveSheet
    .Cells (1, 4).Value= "Latitude (Radians)"
    .Cells (1, 5).Value= "Abs Longitude (Radians)"
    .Cells (1, 6).Value= "Sin Latitude"
    2 .Cells (1, 7).Value= "Cos Latitude"
    .Cells (1, 8).Value= "Sin Longitude"
    .Cells (1, 9).Value= "l_i"
    .Cells (1, 10).Value= "m_i"
    .Cells (1, 11).Value= "n_i"
  End With
  SumWeights = 0
  'sum the weight column
  1 For i = 2 To n
    3 SumWeights = SumWeights + ActiveSheet.Cells(i, 3).Value 'unweighted calculations
  Next
  For i = 2 To n
    With ActiveSheet
      .Cells (i, 4).Value= Abs(ActiveSheet.Cells(i, 2).Value) * WorksheetFunction.Pi() / 180 'convert longitude to radians
      .Cells (i, 5).Value= Abs(ActiveSheet.Cells(i, 1).Value) * WorksheetFunction.Pi() / 180 'convert latitude to radians
      .Cells (i, 6).Value= Sin(ActiveSheet.Cells(i, 4).Value) 'sin of longitude in radians
      4 5 .Cells (i, 7).Value= Cos(ActiveSheet.Cells(i, 4).Value) 'cos of longitude in radians
      .Cells (i, 8).Value= Sin(ActiveSheet.Cells(i, 5).Value) 'sin of latitude in radians
      .Cells (i, 9).Value= ActiveSheet.Cells(i, 8).Value * ActiveSheet.Cells(i, 7).Value 'l_i
      .Cells (i, 10).Value= ActiveSheet.Cells(i, 8).Value * ActiveSheet.Cells(i, 6).Value 'm_i
      .Cells (i, 11).Value= Cos(ActiveSheet.Cells(i, 5).Value) 'n_i
    End With
  Next
End Sub

```



```

Sub Spherical_Dev_Ellipse()
  Dim trace As Double 'trace of the matrix T. use this instead of n.
  Dim AVar As Integer 'hides procedure names in pick list
  Call SphereSetWeights(AVar)
  Call Calculate_B(l_0, m_0, n_0, l0_sqr, lm_sum, ln_sum, m0_sqr, mn_sum, n0_sqr)
  Call Calculate_T(l_0, m_0, n_0, l0_sqr, m0_sqr, n0_sqr, lm_sum, ln_sum, mn_sum)
  R_length = Sqr((l_0 / (n - 1)) ^ 2 + (m_0 / (n - 1)) ^ 2 + (n_0 / (n - 1)) ^ 2)
  S = 2 * (1 - R_length) 'spherical variance
  l_sum = l_0 / (n - 1)
  m_sum = m_0 / (n - 1)
  n_sum = n_0 / (n - 1)
  bar_x = WorksheetFunction.Acos(n_sum) 'longitude mean
  TempSecSin = Sin(bar_x)
  bar_y1 = WorksheetFunction.Acos(l_sum / TempSecSin) 'latitude mean
  bar_y2 = WorksheetFunction.Asin(m_sum / TempSecSin) 'latitude mean
  'statistical testing
  barx_degrees = bar_x
  'put the stats in an excel worksheet
6  With Worksheets("STATS")
    .Cells (1, 14).Value = "Modal Ellipse Statistics"
    .Cells (2, 14).Value = "Mean X"
    .Cells (3, 14).Value = "Mean Y"
    .Cells (4, 14).Value = "Mean Z"
    .Cells (5, 14).Value = "l_0 (degrees)"
    .Cells (6, 14).Value = "m_0 (degrees)"
    .Cells (7, 14).Value = "n_0 (degrees)"
    .Cells (8, 14).Value = "Spherical Variance"
7  .Cells (9, 14).Value = "Resultant Length (R)"
    .Cells (18, 14).Value = "Matrix T"
    .Cells (2, 15).Value = n_sum * 180 / WorksheetFunction.Pi()
    .Cells (3, 15).Value = l_sum * 180 / WorksheetFunction.Pi()
    .Cells (4, 15).Value = m_sum * 180 / WorksheetFunction.Pi()
    .Cells (5, 15).Value = n_0 * 180 / WorksheetFunction.Pi()
    .Cells (6, 15).Value = l_0 * 180 / WorksheetFunction.Pi()
    .Cells (7, 15).Value = m_0 * 180 / WorksheetFunction.Pi()
    .Cells (8, 15).Value = S
    .Cells (9, 15).Value = R_length
  End With
End Sub

```

```
6 7  'matrix T  
    .Cells (19, 15).Value = l0_sqrd  
    .Cells (19, 16).Value = lm_sum  
    .Cells (19, 17).Value = ln_sum  
    .Cells (20, 16).Value = m0_sqrd  
    .Cells (20, 15).Value = lm_sum  
    .Cells (20, 17).Value = mn_sum  
    .Cells (21, 17).Value = n0_sqrd  
    .Cells (21, 15).Value = ln_sum  
    .Cells (21, 16).Value = mn_sum  
    End With  
End Sub
```

8.12 VBA Code for Numerical Techniques

```

Sub Gauss(Lambda)
'this subroutine performs simple row elimination as in Gaussian elimination
Dim T_Prime(1 To 3, 1 To 3) As Double
n = 3
'initialize the data structures A and B; the results get saved to X
  For i = 1 To n
    For j = 1 To n
      T_Prime(i, j) = Worksheets("STATS").Cells(i + 18, j + 14).Value
      If i = j Then
        T_Prime(i, j) = Worksheets("STATS").Cells(i + 18, j + 14).Value - Lambda
      End If
    Next
  Next
'row reduction
T_Prime(2, 2) = T_Prime(2, 2) / T_Prime(2, 1)
T_Prime(2, 3) = T_Prime(2, 3) / T_Prime(2, 1)
T_Prime(2, 1) = 1
T_Prime(3, 2) = T_Prime(3, 2) / T_Prime(3, 1)
8 T_Prime(3, 3) = T_Prime(3, 3) / T_Prime(3, 1)
T_Prime(3, 1) = 1 'subtract row 1 from rows 2 and 3
  For i = 1 To 3
    11 T_Prime(2, i) = T_Prime(2, i) * T_Prime(1, 1)
    T_Prime(3, i) = T_Prime(3, i) * T_Prime(1, 1)
  Next
  For i = 1 To 3
    12 T_Prime(2, i) = T_Prime(2, i) - T_Prime(1, i)
    T_Prime(3, i) = T_Prime(3, i) - T_Prime(1, i)
  Next
'end of subtraction
T_Prime(2, 3) = T_Prime(2, 3) / T_Prime(2, 2)
T_Prime(3, 3) = T_Prime(3, 3) / T_Prime(3, 2)
T_Prime(2, 2) = 1
T_Prime(3, 2) = 1
'subtract row 2 from 3
  13 For i = 1 To 3
    T_Prime(3, i) = T_Prime(3, i) - T_Prime(2, i)
  Next

```

```

'reduce row 1
T_Prime(1, 1) = T_Prime(1, 1) / T_Prime(1, 2)
T_Prime(1, 3) = T_Prime(1, 3) / T_Prime(1, 2)
T_Prime(1, 2) = 1
'subtract row 2 and row 1 to give row 1
  For i = 1 To 3
14   T_Prime(1, i) = T_Prime(1, i) - T_Prime(2, i)
  Next
8 'end of subtraction and row reductions
Worksheets("STATS").Cells(26, 14).Value = "T minus Lambda (I) row reduced"
  For i = 1 To n
    For j = 1 To n
15   16   Worksheets("STATS").Cells(25 + i, 14 + j).Value = T_Prime(i, j)
    Next 'j
  Next 'i
Worksheets("STATS").Cells(30, 15).Value = "The reader will have to determine the eigenvector from here."
End Sub

```

```

Function fx(Lambda)
'the determinate of the matrix T
Dim a, b, c, d, e, g As Double
'matrix T is symmetrix. not all of it needs to be read in.
a = Worksheets("STATS").Cells(19, 15).Value 'sum l_i^2
d = Worksheets("STATS").Cells(19, 16).Value 'sum l_i m_i
e = Worksheets("STATS").Cells(19, 17).Value 'sum l_i n_i
b = Worksheets("STATS").Cells(20, 16).Value 'sum m_i^2
17 g = Worksheets("STATS").Cells(20, 17).Value 'sum m_i n_i
c = Worksheets("STATS").Cells(21, 17).Value 'sum n_i^2
f = d
i = g
h = e
'fx_old = (a - Lambda) * (d - Lambda) * (g - Lambda) + 2 * b * e * c -
c ^ 2 * (d - Lambda) - e ^ 2 * (a - Lambda) - b ^ 2 * (g - Lambda)
fx = (a - Lambda) * (b - Lambda) * (c - Lambda) + d * g * h + e * f * i - h * e * (b - Lambda) -
i * g * (a - Lambda) - f * d * (c - Lambda)
End Function

```

```

Function gx(Lambda)
'the derivative of the determinate of matrix T
Dim a, b, c, d, e, g As Double
a = Worksheets("STATS").Cells(19, 15).Value 'sum l_i^2
d = Worksheets("STATS").Cells(19, 16).Value 'sum l_i m_i
e = Worksheets("STATS").Cells(19, 17).Value 'sum l_i n_i
b = Worksheets("STATS").Cells(20, 16).Value 'sum m_i^2
g = Worksheets("STATS").Cells(20, 17).Value 'sum m_i n_i
18 c = Worksheets("STATS").Cells(21, 17).Value 'sum n_i^2
f = d
i = g
h = e
gx = -1 * (b - Lambda) * (c - Lambda) - (a - Lambda) * (c - Lambda) - (a - Lambda) * _
(b - Lambda) + e * h + i * g + f * d
'gx = -1 * a * d - a * g + 2 * Lambda * a - d * g + 2 * Lambda * d - 2 * Lambda * g + _
3 * Lambda ^ 2 + c ^ 2 + e ^ 2 + b ^ 2
End Function

```

```
Sub Newton(Lambda)
Dim lambda0 As Double 'old value of lambda
Dim diff As Double 'termination condition
Dim i As Integer
diff = 1
i = 1
f_x = fx(Lambda)
  While (diff > 0.000000001)
    lambda0 = Lambda
    Lambda = Lambda - f_x / gx(Lambda)
20  f_x = fx(Lambda)
19  diff = Abs(lambda0 - Lambda)
    i = i + 1
  Wend
  With Worksheets("STATS")
    .Cells(23, 14).Value = "Eigenvalue"
    .Cells(24, 14).Value = "Newton Iterations"
21  .Cells(23, 15).Value = Lambda
    .Cells(24, 15).Value = i
    .Cells(20, 20).Value = Lambda
  End With
End Sub
```


APPENDIX A

STANDARD DEVIATIONAL ELLIPSE VBA DRIVER

The code for calculating the weighted standard deviational ellipse resides in the Developer Environment under the module named *Ch4StandardDeviationalEllipse*. In the Excel spreadsheet environment, only highlight the observations themselves using the <shift> ↓ key combination, not the entire column by clicking on the column headings. It is important that the reader only use columns 1 – 3 for his or her data. The VBA code in module *Ch4StandardDeviationalEllipse* uses the driver `Standard_Dev_Ellipse()` to calculate the following statistics.

- The weighted mean center (\bar{x}, \bar{y}) from Equation (5.9).
- The angle of rotation on the axes θ_x , and θ_y from Equation (5.10).
- The error terms on the axes δ_x , and δ_y from Equations (5.11) and (5.12).
- The semi-major a and semi-minor b axes lengths from Equations (5.15) and (5.16).

- The area F from Equation (5.17).
- The eccentricity of the ellipse e from Equation (5.18).

To run the SDE code, follow these steps. Select and highlight the data.

1. Click in the spreadsheet tab that contains the data you wish to analyze.
2. Ensure that the first three columns of data (and only the data) are highlighted (selected). If not, follow these steps:
 - (a) Click on cell A1.
 - (b) While holding down the <SHIFT> key, arrow over to Column C. The headings of the data should be highlighted and selected.
 - (c) While still holding down the <SHIFT> key, hold down the <CTRL> key.
 - (d) Touch the down arrow key once. This should take you to the bottom of the data.

Figure A.2 shows the VBA environment for module *Ch4StandardDeviationalEllipse*. To enter the VBA environment and run the standard deviational ellipse code:

1. Click on the `Developer` ribbon.
2. Click on the `Visual Basic` button on the left side.
3. Click on *Ch4StandardDeviationalEllipse* on the left-side pane.

This module calls the following subroutines.

- `Set_Weights(sum_weight)` — This subroutine calculates the weights and saves them to the spreadsheet in the fourth column.
- `SDE_Stats(AVar)` — This subroutine calculates the following statistics:
 - The mean latitude, \bar{x} , and the mean longitude, \bar{y} .
 - The major axis rotation and minor axis rotation. The sum of the absolute value of both will always sum to 90° .
 - The δ_x for the latitude and the δ_y for the longitude.
 - The major axis length, minor axis length, the area, and the eccentricity.
- `SDE_Likelihood(AVar)` — This subroutine calculates the log likelihood function for the weighted regression model.
- `Put_SDEOutput_In_Spreadsheet(AVar)` — This subroutine saves the results to the STATS spreadsheet in columns A and B.

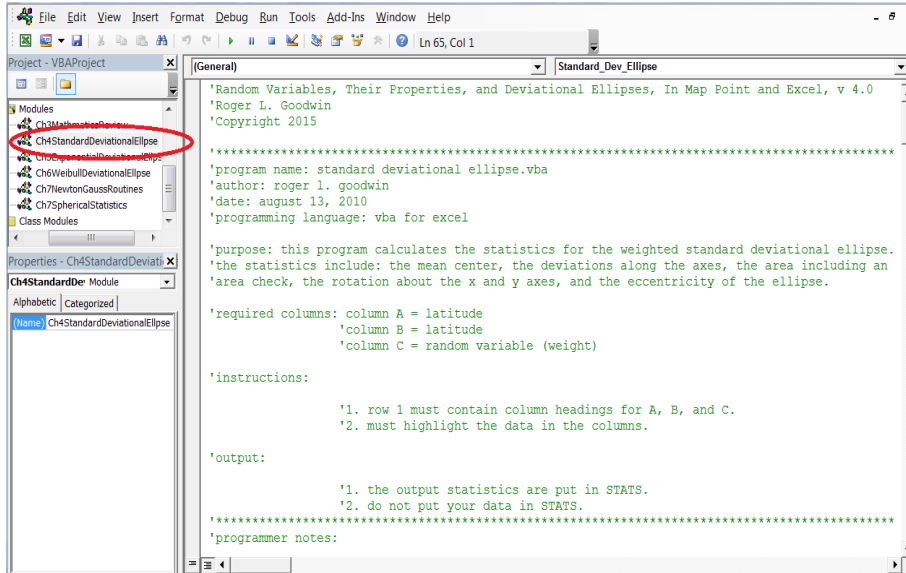


Figure A.1 In Excel, double click on the module named *Ch4StandardDeviationalEllipse* on the left-hand pane circled in red to display the VBA code for the Standard Deviational Ellipse. The subroutine called *Standard_Dev_Ellipse()* runs the programs to calculate the statistics in Chapter 5.

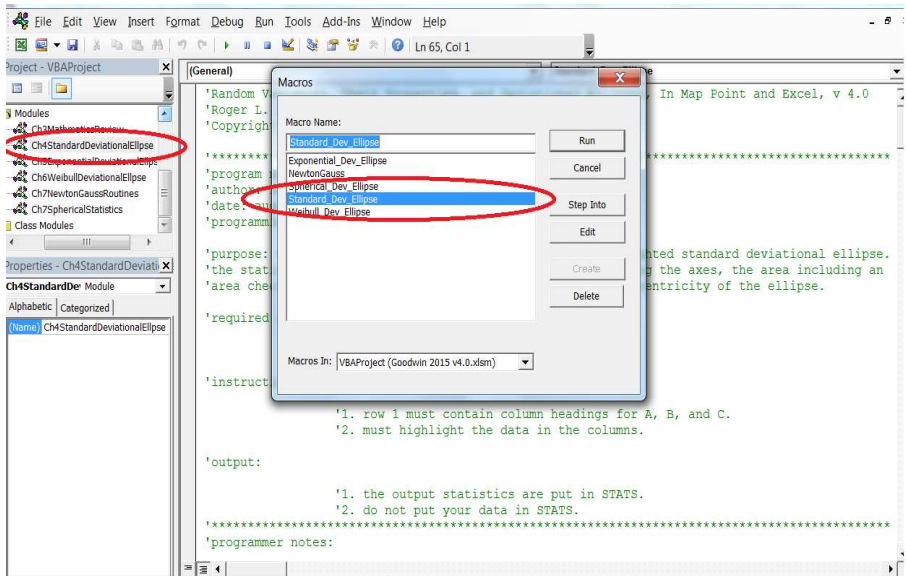


Figure A.2 This figure shows the VBA environment for the module named *Ch4StandardDeviationalEllipse*. It also highlights the driver subroutine *Standard_Dev_Ellipse()*.

	A	B	C
1	Standard Deviational Ellipse Statistics		
2	Latitude Mean Value =	37.38297399	
3	Longitude Mean Value =	273.1870486	=MOD(360,B3)
4	Rotation	Rotate on Y-Axis	
5	Major Axis Rotation	70.36257022	
6	Minor Axis Rotation	-19.63742978	
7	Delta X	1.29151776	
8	Delta Y	1.625551351	
9	Area (F)	30336.30686	
10	Major Axis Length (a)	184.8734659	
11	Minor Axis Length (b)	52.23219209	
12	Area Check	30336.30686	
13	Eccentricity (e)	0.959258636	
14	Likelihood L(x,y)	181,824,055.41	
15			
16			

Figure A.3 This figure shows the STATS worksheet and the results from the Standard_Dev_Ellipse() subroutine.

The Standard_Dev_Ellipse() subroutine(1) drives the standard deviational ellipse subroutines. The weighted standard deviational ellipse is one of the simplest concepts to implement, and thus, does not require input from the user. A circular coordinate system is not required, either. The Standard_Dev_Ellipse() routine will return the same results with either data set.

To run the Standard_Dev_Ellipse() subroutine, make sure you are in the upper-left corner of the VBA Editor (Line 1, Column 1).

1. Select the menu option Run .

2. Select the sub-menu item Run Sub/UserForm.

The subroutines called in the Standard_Dev_Ellipse() subroutine run quickly. As stated several times, the results will appear in the STATS spreadsheet.

Upon clicking on the STATS tab, the Standard_Dev_Ellipse() results appear in columns A and B. See Figure A.3. The SDE statistics and their locations in the STATS spreadsheet are as follow:

```

Dim n, m As Integer 'sample size
Dim sum_weight As Double 'sum of the weights
Dim mean_latitude, mean_longitude As Double 'mean latitude and longitude
Dim atheta, itheta As Double 'axes rotations
Dim x, y, xy As Double 'intermediate calculations used to calculate the area F
Dim lambda_x, lambda_y, lambda_xy As Double 'parameter estimates
Dim t_x, t_y, t_xy, t_xsq, t_ysq, t_xysq As Double 'exponential distribution ellipse variables
Dim a, b As Double 'semi-major and semi-minor axes
Dim note As String 'note on axis of rotation of ellipse
Dim delta_x, delta_y As Double 'deltas on the x and y axis
Dim f As Double 'area
Dim e As Double 'eccentricity
Dim area_check2 As Double 'area check
Dim L_reg As Double 'regression likelihood function value

```

```

Sub Standard_Dev_Ellipse()
  'run the routines that calculate the standard deviational ellipse statistics
  Dim AVar As Integer 'hides procedure names in pick list
  Call Set_Weights(sum_weight)
  Call SDE_Stats(AVar)
  Call SDE_Likelihood(AVar)
  Call Put_SDEOutput_In_Spreadsheet(AVar)
End Sub

```

- The weighted mean center from Equation (5.9), (\bar{x}, \bar{y}) , \bar{x} appears in cell B2, and \bar{y} appears in cell B3 .
- The axis of rotation in Equation (5.10) appears in cell B4 .
- The major axis angle of rotation θ appears in cell B5 .
- The minor axis angle of rotation θ appears in cell B6 .
- The standard error from Equation (5.11) in the latitude direction x appears in cell B7 .
- The standard error from Equation (5.12) in the longitude direction y appears in cell B8 .
- The area F of the weighted standard deviational ellipse from Equation (5.17) appears in cell B9 .
- The semi-major axis length a from Equation (5.15) appears in cell B10 .

- The semi-minor axis length b from Equation (5.16) appears in cell B11 .
- The area check using the formula $F = \pi ab$ appears in cell B12 .
- The eccentricity e of the standard deviational ellipse from Equation (5.18) appears in cell B13 .
- The logarithm of the likelihood function of the regression model from Equation (5.8) appears in cell B14 . This will be discussed in Section 7.12.

This application does not write in column C . Use column C to transform the mean center coordinates, if needed using Equation (3.3) or (3.5). See Figure A.3.

Implementation Issue 1: The implementation of the modulo function `mod()` differs in the VBA application versus the Excel spreadsheet. The VBA function returns an integer. The spreadsheet function returns a real number. Because of this difference, it is the responsibility of the user to convert the latitude and the longitude observations to circular coordinates on the input data set. This applies to every data set. It becomes more important for our global GNP data. If in doubt, it is never wrong to apply the formulas in Section 3.7 to the input coordinates and the output statistics for the mean center.

Implementation Issue 2: The implementation of the modulo function `mod()` differs in the VBA application versus the Excel spreadsheet. The VBA function returns an integer. The spreadsheet function returns a real number. Because of this difference, it is the responsibility of the user to convert the mean latitude and the mean longitude from circular coordinates to the original coordinate system.

APPENDIX B

EXPONENTIAL VBA DRIVER

The module `Ch5ExponentialDeviationalEllipse` provides an Excel 2010 VBA program to calculate the statistics for the weighted, exponential deviational ellipse. The VBA code calculates the following statistics.

To run the exponential ellipse code, follow these steps.

1. Click in the spreadsheet tab that contains the data you wish to analyze.
2. Select and highlight the data first. Follow these steps to select and highlight the data.
 - (a) Click on cell A1 .
 - (b) While holding down the <SHIFT> key, arrow over to column C . The headings of the data should be highlighted and selected.
 - (c) While still holding down the <SHIFT> key, hold down the <CTRL> key.
 - (d) Touch the down arrow key once. This should take you to the bottom of the data.

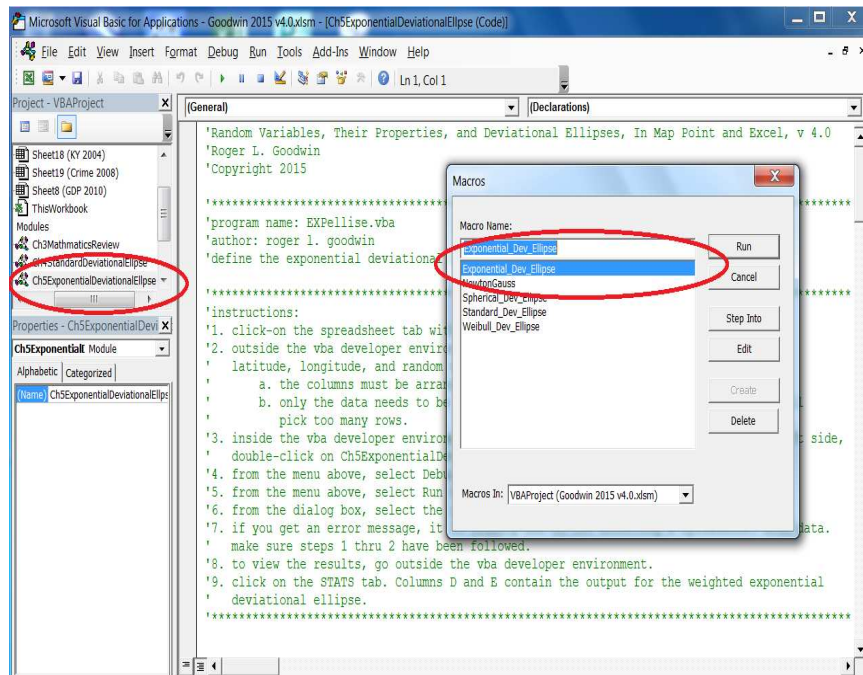


Figure B.1 This figure shows the VBA environment for the module `Ch5ExponentialDeviationalEllipse`. It also highlights the driver subroutine `Exponential_Dev_Ellipse()`.

Figure B.1 shows the VBA environment for the module `Ch5ExponentialDeviationalEllipse`. To enter the VBA environment and run the exponential ellipse code:

1. Click on the `Developer` ribbon.
2. Click on the `Visual Basic` button on the left side.
3. Click on the module `Ch5ExponentialDeviationalEllipse` on the left-side pane.

This module calls the following subroutines.

- `Set_Weights(sum_weight)` — This subroutine calculates the weights and saves them to the active data spreadsheet.
- `EXP_Stats(AVar)` — This subroutine calculates the parameter estimates, the major axis length, minor axis length, the area, and the eccentricity. It also writes the log likelihood value to the `STATS` spreadsheet.
- `Put_EXOutput_In_Spreadsheet(AVar)` — This subroutine puts the results of the calculates with labels for the weighted standard deviation ellipse to the `STATS` spreadsheet. The subroutine writes the results columns D and E.

D	E	F
Exponential Ellipse Statistics		
Mean Latitude	37.38297399	
Mean Longitude (Abs)	273.1870486	=MOD(360,E3)
Rotation	Rotate on Y-Axis	
Major Axis Rotation	89.9986581	
Minor Axis Rotation	-0.001342193	
Delta X	3.084687614	
Delta Y	3.084707388	
Area (F)	32450.8332	
Major Axis Length (a)	274.7458265	
Minor Axis Length (b)	37.59636008	
Area Check	32450.90232	
Eccentricity (e)	0.990593092	
Likelihood Function L(x)	-44211.86328	
Likelihood Function L(y)	-60300.24289	
Likelihood Function L(xy)	-111454.5019	

Figure B.2 This figure shows the STATS worksheet and the results from the Exponential_Dev_Ellipse() subroutine.

The Exponential_Dev_Ellipse() subroutine drives the exponential deviational ellipse subroutines. The weighted exponential deviational ellipse is another one of the simplest concepts to implement, and thus, does not require input from the user. A circular coordinate system is required.

To run the Exponential_Dev_Ellipse() subroutine, make sure you are in the upper-left corner of the VBA Editor (Line 1, Column 1).

1. Select the menu option Run .
2. Select the sub-menu item Run Sub/UserForm .

The subroutines called in the Exponential_Dev_Ellipse() subroutine run quickly. As previous stated, the results will appear in the STATS spreadsheet.

Upon clicking on the STATS tab, the Exponential_Dev_Ellipse() results appear in columns D and E . See Figure B.2. The Exponential_Dev_Ellipse() statistics and their locations in the STATS spreadsheet are as follow:

- Using Equations (6.12) and (6.13) for the weighted mean center (\bar{x}, \bar{y}) , \bar{x} appears in cell E2 , and \bar{y} appears in cell E3 .
- Using Equation (6.17), the axis of rotation appears in cell E4 .
- Using Equation (6.17), the major axis angle of rotation θ appears in cell E5 . The plus and minus operator gives two values.

- Using Equation (6.17), the minor axis angle of rotation θ appears in cell E6 . The plus and minus operator gives two values.
- Using Equation (6.18), the standard deviation in the latitude direction x appears in cell E7 .
- Using Equation (6.19), the standard deviation in the longitude direction y appears in cell E8 .
- Using Equation (6.20), the semi-major axis length a appears in cell E10 .
- Using Equation (6.21), the semi-minor axis length b appears in cell E11 .
- Using Equation (6.22), the area F of the weighted exponential deviational ellipse appears in cell E9 .
- The area check using the formula $F = \pi ab$ appears in cell E12 .
- Using Equation (6.23), the eccentricity e of the exponential deviational ellipse appears in cell E13 .
- Using Equation (6.2), the logarithm of the likelihood function of the exponential model for the latitude x appears in cell E14 .
- Using Equation (6.6), the logarithm of the likelihood function of the exponential model for the longitude y appears in cell E15 .
- Using Equation (6.9), the logarithm of the likelihood function of the exponential model for the joint distribution of the latitude and longitude (x, y) appears in cell E16 .

This application does not write in column F . If either of the mean center coordinates need to be transformed via Equation (3.3) or (3.5), then the reader can use column F for those calculations. See Figure B.2. Notice that the two values for the area (with-in the exponential ellipse statistics are equivalent). Also notice that the mean center for the exponential ellipse and for the standard deviational ellipse are the same.

Should the reader wish to calculate un-weighted exponential deviational ellipses, set the weight column to 1.0.

```

Dim wx, wy As Double
Dim lambda_x, lambda_y, lambda_xy As Double 'parameter estimates
Dim t_x, t_y, t_xy, t_xsqr, t_ysqr, L_xysqr As Double 'intermediate calculations
Dim sum_weight As Double 'sum of the weights
Dim atheta, itheta As Double 'angles of rotation
Dim L_x, L_y, L_xy As Double 'log likelihood functions
Dim n, m As Integer
Dim a, b As Double 'semi-major and semi-minor axes
Dim note As String 'note on axis of rotation of ellipse
Dim error_x, error_y As Double 'deviations on the x and y axis
Dim f As Double 'area
Dim e As Double 'eccentricity
Dim area_check2 As Double 'area check
1 Sub Exponential_Dev_Ellipse()
   Dim AVar As Integer 'hides procedure names in pick list
   Call Set_Weights(sum_weight)
   Call EXP_Stats(AVar)
   Call Put_EXPOutput_In_Spreadsheet(AVar)
End Sub

```


APPENDIX C

WEIBULL VBA DRIVER

Running the Weibull ellipse takes considerable more interaction and time than that for the previous two ellipses. The secant algorithm finds three parameters. Two initial guesses for each parameter must accompany each run. Two secant algorithms have been programmed — one for the individual, weighted variables (x_i, w_i) and (y_i, w_i) called `Secant1()` and a second algorithm for the joint distribution of (x_i, y_i, w_i) called `Secant2()`. This was necessary because the likelihood functions are different.

Figure C.1 shows the VBA environment for the module `Ch6 Weibull Deviatonal Ellipse`. This module has five subroutines.

- `Weibull_Dev_Ellipse()` — This subroutine is the driver that calls the other subroutines.
- `Secant1()` — This subroutine calculates the point estimates for the latitude $\hat{\lambda}_x$ and $\hat{\gamma}_x$ from Equation (7.8). It can also be used to calculate the point estimates for the longitude $\hat{\lambda}_y$, and $\hat{\gamma}_y$ from Equation (7.17)

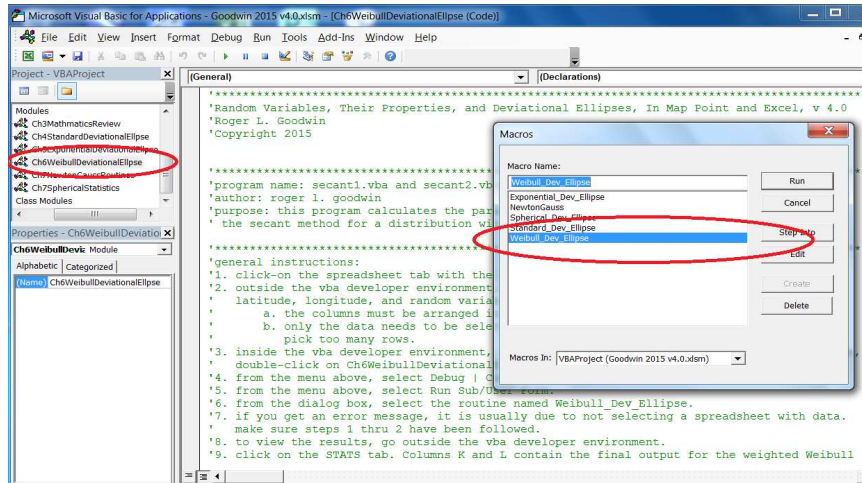


Figure C.1 This figure shows the VBA environment for the module `Ch6WeibullDeviationalEllipse`. It also highlights the driver subroutine `Weibull_Dev_Ellipse()`.

- `Secant2()` — This subroutine calculates the point estimates for the joint distribution $\hat{\lambda}_{xy}$ and $\hat{\gamma}_{xy}$ from Equation (7.21).
- `Log10(X)` — This subroutine calculates the logarithm of the given number X using base 10.
- `Weibull()` — This subroutine calculates all of the statistics for the Weibull ellipse. These statistics include the mean center (\bar{x}, \bar{y}) , the axis of rotation, the rotations θ_x and θ_y , the standard deviations on the axes δ_x and δ_y , the major and minor axes lengths a and b , the area F , the eccentricity e , and the log likelihood function $\log L_{xy}$.

The `Secant1()` algorithm has the following parameters:

- `Iterations` — This parameter is a termination condition. This is the maximum number of iterations that the secant algorithm should perform.
- `Variable` — This parameter is the column number of the variable (not the name in row 1). This will be either column 1 or column 2.
- `Gamma1` — This parameter is the first initial guess at γ_x or γ_y .
- `Gamma2` — This parameter is the second initial guess at γ_x or γ_y . Note that `Gamma1 < Gamma2`.

The `Secant1()` algorithm returns the values $\hat{\gamma}_x$ and $\hat{\lambda}_x$, or $\hat{\gamma}_y$ and $\hat{\lambda}_y$, depending on if `Variable = 1` or `Variable = 2`. All four point estimates have to be saved or recorded where they will not be over-written.

The `Secant2()` algorithm has the following parameters:

- **Iterations** — This parameter is a termination condition. This is the maximum number of iterations that the secant algorithm should perform.
- **Gamma1** — This parameter is the first initial guess at γ_x or γ_y .
- **Gamma2** — This parameter is the second initial guess at γ_x or γ_y . Note that `Gamma1 < Gamma2`.

Since the `Secant2()` subroutine returns the values $\hat{\gamma}_{xy}$ and $\hat{\lambda}_{xy}$, there is no need to specify the latitude or longitude variable. Save the point estimates so that they will not be over-written.

After finding the six point estimates of the parameters, run the `Weibull()` subroutine.

- $\hat{\gamma}_x$ — The point estimate for the latitude for the shape parameter.
- $\hat{\gamma}_y$ — The point estimate for the longitude for the shape parameter.
- $\hat{\gamma}_{xy}$ — The point estimate for the joint distribution for the shape parameter.
- $\hat{\lambda}_x$ — The point estimate for the latitude for the scale parameter from Equation (7.9).
- $\hat{\lambda}_y$ — The point estimate for the longitude for the scale parameter from Equation (7.18).
- $\hat{\lambda}_{xy}$ — The point estimate for the joint distribution for the scale parameter.

Upon clicking on the **STATS** tab, the `Weibull()` results appear in columns **K** and **L**. See Figure C.2. The `Weibull()` statistics and their locations in the **STATS** spreadsheet are as follow:

- Using Equations (7.9) and (7.18), for the weighted mean center (\bar{x}, \bar{y}) , \bar{x} appears in cell **L2**, and \bar{y} appears in cell **L3**.
- Using Equation (7.26), the axis of rotation appears in cell **L4**.
- Using Equation (7.26), the major axis angle of rotation θ appears in cell **L5**. The plus and minus sign in Equation (7.26) gives two values.
- Using Equation (7.26), the minor axis angle of rotation θ appears in cell **L6**. The plus and minus sign in Equation (7.26) gives two values.
- Using Equation (7.27), the standard deviation in the latitude direction x appears in cell **L7**.
- Using Equation (7.28), the standard deviation in the longitude direction y appears in cell **L8**.

J	K	L	M
	Weibull Ellipse Statistics		
	Mean Latitude	37.38297399	
	Mean Longitude (Abs)	273.1870490	
	Rotation	Rotate on X-Axis	
	Major Axis Rotation	89.89385427	
	Minor Axis Rotation	-0.10614571	
	Delta X	1.77010417	
	Delta Y	1.779642002	
	Area (F)	27.11384518	
	Major Axis Length (a)	3.720699854	
	Minor Axis Length (b)	2.319618704	
	Area Check	27.11384518	
	Eccentricity (e)	0.781874284	
	Likelihood Function L(x,y)	-137452.7083	

Figure C.2 This figure shows the STATS worksheet and the results from the Weibull () subroutine.

- Using Equation (7.31), the area F of the weighted standard deviational ellipse appears in cell L9 .

- Using Equation (7.29), the semi-major axis length a appears in cell L10 .

- Using Equation (7.30), the semi-minor axis length b appears in cell L11 .

- The area check using the formula $F = \pi ab$ appears in cell L12 .

- Using Equation (7.32), the eccentricity e of the standard deviational ellipse appears in cell L13 .

- The logarithm of the likelihood function of the regression model for the joint distribution of the latitude and longitude (x, y) appears in cell L14 .

```

Dim iterations As Integer
Dim Gamma1, gamma2, gamma3 As Double
Dim variable As Integer
Dim n, m As Integer
Dim logt_x, logt_y, logt_xy, logt_xsqr, logt_ysqr, logt_xysqr As Double 'intermediate calculations
Dim L_xy As Double 'likelihood function
Dim a, b As Double 'semi-major and semi-minor axes
Dim atheta, itheta As Double 'axes rotations
Dim note As String 'note on axis of rotation of ellipse
Sub Weibull_Dev_Ellipse()
    'latitude for 2003 kentucky data
    'Call secant1(9, 1, 1, 10)
    'longitude for 2003 kentucky data
    'Call secant1(9, 5, 1, 10)
    'joint distribution for 2003 kentucky data
    'Call secant2(9, 1, 11)
    'new data set
    'Call Weibull(0.768572408, 0.762370962, 1, 2.407513111, 0.329444607, 0.579579441)
    'old data set
    'Call Weibull(0.768634929, 0.762451336, 0.999998427, 2.434263513, 0#, 0.586020077)
    'latitude for 2004 kentucky data
1 'Call secant1(9, 1, 1, 10)
    'longitude for 2004 kentucky data
    'Call secant1(12, 5, 1, 12)
    'joint distribution for 2004 kentucky data
    'Call secant2(14, 1, 13)
    'Call Weibull(0.766401461, 0.760180292, 1, 2.459505837, 0.336757843, 0.592403275)
    'latitude for 2007 crime data
    'Call secant1(7, 1, 1, 11)
    'longitude for 2007 crime data
    'Call secant1(9, 5, 1, 14)
    'joint distribution for 2007 crime data
    'Call secant2(10, 1, 17)
    'Call Weibull(0.765003479, 0.747808984, 1, 1.328755906, 0.182913031, 0.321560844)

```



```

'latitude for 2008 crime data
'Call secant1(10, 1, 1, 13)
'longitude for 2008 crime data
'Call secant1(15, 5, 1, 17)
'joint distribution for 2008 crime data
'Call secant2(10, 1, 17)
'Call Weibull(0.765405522, 0.757628441, 1, 1.328288608, 0.18285781, 0.321461829)
'latitude for 2008 GDP data
'Call secant1(4, 1, 1, 10)
'longitude for 2008 GDP data
'Call secant1(10, 2, 1, 16)
'joint distribution for GDP data
'Call secant2(12, 1, 18)
1 'Call Weibull(0.778466889, 0.763823411, 1, 0.846866662, 0.256765441, 0.394055395)
'latitude for 2009 GDP data
'Call secant1(4, 1, 1, 10)
'longitude for 2008 GDP data
'Call secant1(10, 2, 1, 16)
'joint distribution for GDP data
'Call secant2(12, 1, 18)
'Call Weibull(0.777870853, 0.764074484, 1, 0.849145918, 0.259687836, 0.397738372)
'latitude for 2010 GDP data
'Call secant1(4, 1, 1, 10)
'longitude for 2010 GDP data 'Call secant1(10, 2, 1, 16)
'joint distribution for GDP data
'Call secant2(12, 1, 18)
'Call Weibull(0.777297498, 0.764470918, 1, 0.848898928, 0.260064435, 0.398152775)
End Sub

```

To obtain the results in this Section, this textbook provides three VBA programs for Excel 2010. Two of the programs perform the Secant estimation for the parameters γ_x , γ_y , and γ_{xy} . The other program calculates the usual statistics associated with an ellipse such as the mean center, the area and so on. The program `main` in `Ch6WeibullDeviationalEllipse` calculates the parameters λ_x , λ_y , γ_x and γ_y . The program `Weibull()` in `Ch6WeibullDeviationalEllipse` calculates the parameters λ_{xy} and γ_{xy} .

Double click on the module `Ch6 Weibull Deviatonal Ellipse` to display the VBA code for the Secant code and the Weibull Deviatonal Ellipse in Excel. The subroutines called `Secant1()`, `Secant2()`, and `Weibull_Dev_Ellipse()` run the programs to calculate the statistics in this chapter. As the reader may have

guessed, calculating the statistics in this chapter is more complex than those in previous chapters.

APPENDIX D

NEWTON AND GAUSSIAN ELIMINATION DRIVER

The statistics associated with the spherical distribution in Chapter 8 are complicated. They usually begin with finding the some statistic of a 3×3 matrix from either B or T . This Section describes the VBA code in `Ch7NewtonGaussRoutines` and `Ch7 Spherical Statistics`. It is best to start with `Ch7 Spherical Statistics`. Figure D.1 shows a screen capture of the VBA environment for `Ch7 Spherical Statistics`. The subroutines in `Ch7SphericalStatistics` calculate the matrix T and find the eigenvalues to matrix T . Since there are an infinite number of eigenvectors, some interaction and verification on the reader's part is required when finding the eigenvectors. `Ch7 Spherical Statistics` contains the following six subroutines:

1. `NewtonGauss()` — This is a driver subroutine.
2. `Spherical_Dev_Ellipse()` — This is a driver subroutine.

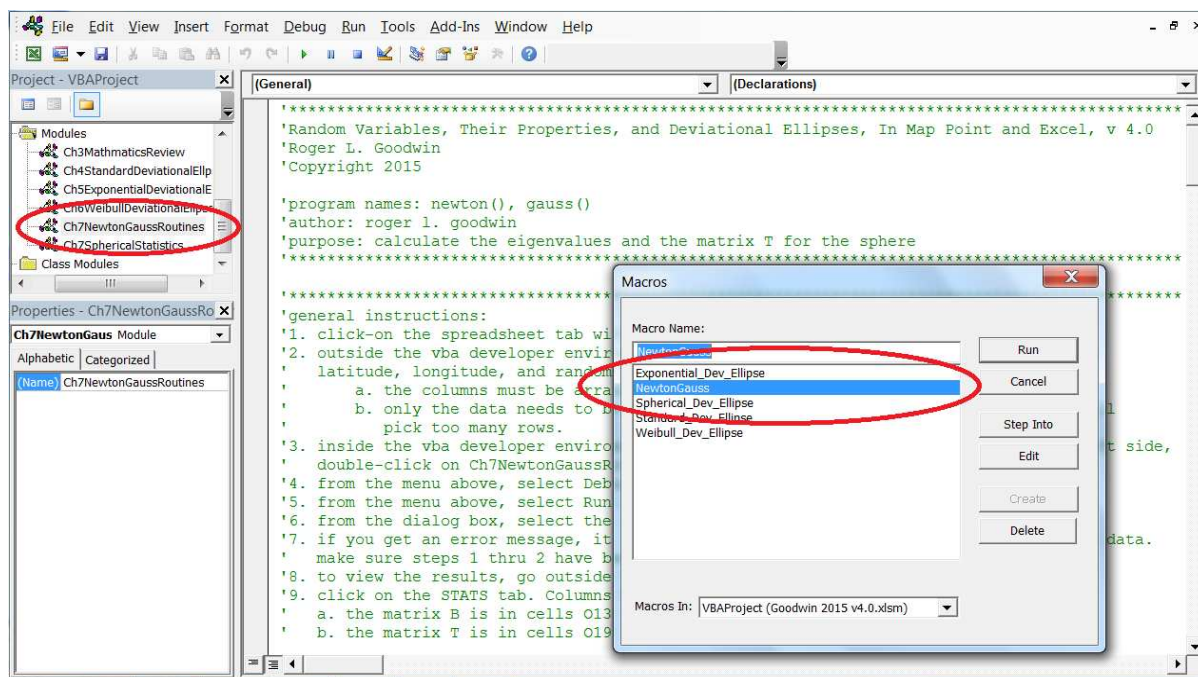


Figure D.1 This figure shows the VBA environment for the module Ch7NewtonGaussRoutines. It also highlights the driver subroutine NewtonGauss().

3. `Calculate_B()` — This subroutine calculates the matrix B from which the diagonal elements $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ are determined. We can re-use the subroutine in `Ch7SphericalStatistics` for calculating this matrix.
4. `Calculate_T()` — This subroutine calculates the matrix T . The upper left-hand corner of the matrix resides in cell `O19` in the `STATS` spreadsheet and cell `Q21` in the lower-right corner.

The `FOR-NEXT` loop (10) calculates the matrix T using Equation (8.14) where $T = nI - B$. The `WITH-END-WITH` statement (11) saves the matrix T below matrix B in the `STATS` spreadsheet. The upper-left corner is cell `O19` and the lower-right corner is cell `Q21`. The remaining code writes descriptive labels to the `STATS` spreadsheet for the eigenvectors in columns `R` and `V`.

5. `Gauss(Lambda)` — This subroutine performs row elimination on matrix T given the eigenvalue `Lambda`. This finds a possible eigenvector and puts it into cells `V19`, `V20`, and `V21`. The value in cell `V21` will always be equal to 1. This allows the reader to scale the other values in the vector up or down (usually be a factor of 10).

The `FOR-NEXT` loop (12) reads the matrix T into the array `T_prime`. Since T will always be a 3×3 matrix, the code in the `Gauss()` subroutine performs row reduction as one would do on paper.

6. `fx(Lambda)` — This subroutine finds the determinant of matrix T given the eigenvalue `Lambda`.
7. `gx(Lambda)` — This subroutine finds the derivative of the determinant of matrix T given the eigenvalue `Lambda`.
8. `Newton(Lambda)` — This subroutine finds the eigenvalues of matrix T using the subroutines `fx()` and `gx()`. It puts the eigenvalue in cell `T20`.

When trying to find the eigenvalues and the eigenvectors, the reader only needs to run the `Spherical_Dev_Ellipse()` subroutine and ensure that it contains the proper calls and parameters to the `Newton()` and `Gauss()` subroutines. Ordering is important. The subroutines in `Ch7SphericalStatistics`, particularly the `Newton()` subroutine, must be run before the `Gauss()` can be run. The statements in the `Spherical_Dev_Ellipse()` subroutine clearly show the ordering of the other subroutine calls.

Dim Lambda As Double '1 to 3 possible eigenvalues

Dim l_0, m_0, n_0, l0_sqrd, m0_sqrd, n0_sqrd, lm_sum, ln_sum, mn_sum As Double

Dim l_sum, m_sum, n_sum As Double

```

Sub NewtonGauss()
  Dim AVar As Integer 'hides procedure names in pick list
  'syntax
  Sub Newton(Lambda)
  Sub Gauss(Lambda)
  sub Calculation_T()
  Call SphereSetWeights(AVar)
  Call Calculate_B(l_0, m_0, n_0, l0_sqrd, lm_sum, ln_sum, m0_sum, mn_sum, n0_sqrd)
  Call Calculate_T(l_0, m_0, n_0, l0_sqrd, m0_sqrd, n0_sqrd, lm_sum, ln_sum, mn_sum)
  'kentucky 2004
  'lambda_2
  Call Newton(0)
  Call Gauss(0.006533004)
  'lambda_1
  Call Newton(2)
  Call Gauss(0.035928191)
  'final step — round the chosen shape to an integer for interpretation
1 Call Newton(92)
  Call Gauss(91.95753881)
  'crime 2008
  'lambda_1
  Call Newton(1)
  Call Gauss(0.047899513)
  'lambda_2
  Call Newton(30)
  Call Gauss(0.27486963)
  'lambda 3
  Call Newton(50)
  Call Gauss(47.38494857)
  'final step — round the chosen shape to an integer for interpretation
  Call Gauss(100)
  'testing
  Call Newton(30)
  Call Gauss(8.58033582016791E-02)

```

```
1  'GDP 2010  
   'lambda 1  
   'Call Newton(0)  
   'Call Gauss(3.728832776)  
   'lambda 2  
   'Call Newton(10)  
   'Call Gauss(6.216179462)  
   'lambda 3  
   Call Newton(39)  
   Call Gauss(29.05498776)  
   End Sub
```


APPENDIX E

REFERENCES

1. C. Ashton, A. S. Bruce, G. Colledge and M. Dickinson, "The Search for MH370," *The Journal of Navigation*, 68, 122, 2015.
2. R. Beran, "Exponential Models for Directional Data," *The Annals of Statistics*, vol. 7, no. 6 (Nov., 1979), p. 1162-1178.
3. B. L. Bowerman and R. T. O'Connell, *Linear Statistical Models, an Applied Approach*, Duxbury Press, 1990.
4. G. Brassard and P. Bratley (1988), *Algorithmics: Theory and Practice*, Prentice Hall, NJ.
5. M. B. Brown and A. B. Forsythe (1974), "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, vol. 69, no. 346, p. 364-367.
6. W. Cheney and D. Kincaid (1985), *Numerical Mathematics and Computing*, Wadsworth, Inc., Belmont, California.

7. R. S. Chhikara and L. Deng, "Using Multiyear Rotation Design Sampling in Agricultural Surveys," *Journal of the American Statistical Association*, Vol. 87, No. 420 (Dec., 1992), pp. 924-932.
8. L. Cobb, P. Keppstein, and N. H. Chen, "Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family," *Journal of the American Statistical Association*, Vol 78, No. 381 (Mar., 1983), pp. 124-130.
9. W. J. Conover (1999), *Practicing Non-Parametric Statistics, Third Edition* John Wiley & Sons, Inc., New York, New York.
10. B. R. Crain, "Estimation of Distributions Using Orthogonal Expansions," *The Annals of Statistics*, vol. 2, no. 3 (May, 1974), p. 454-463.
11. B. R. Crain, "More on Estimation of Distributions Using Orthogonal Expansions," *Journal of the American Statistical Association*, vol. 71, no. 355 (Sep., 1976), p. 741-745.
12. N. Cressie, *Statistics for Spatial Data (Revised Edition)*, John Wiley & Sons, Inc., New York, 1993.
13. P. Davis, *Data Description and Presentation*, Oxford University Press, Ely House, London, 1974.
14. D. Ebdon (1985), *Statistics in Geography, Second Edition*, Basil Blackwell, New York, NY.
15. J. Fagan, "Generating Implied Ratio Edits," unpublished Census manuscript. October 1999.
16. G. D. Faulkenberry and A. Garoui, "Estimating a Population Total Using an Area Frame," *Journal of the American Statistical Association*, Vol. 86, No. 414 (Jun., 1991), pp. 445-449.
17. I. P. Fellegi and D. Holt, "A Systematic Approach to Automatic Edit and Imputation," *The Journal of the American Statistical Association*, No. 71, 1976.
18. L. M. Friedman, C. D. Furberg and D. L. DeMets (1996), *Fundamentals of Clinical Trials, Third Edition*, Mosby-Year Book, Inc., St. Louis, Missouri.
19. J. Gong, "Clarifying the Standard Deviation Ellipse," *Geographical Analysis*, Vol. 34, No. 2 (April 2002), pp. 155-167.
20. M. Garcia and R. Goodwin, "Developing SAS Software for Generating a Complete Set of Ratio Edits," *Statistical Research Division, US Census Bureau, Research Report Series, Statistics # 2002-06*, November 8, 2002.

21. R. Goodwin, M. Garcia and K. Thompson, "The GenBounds Software for Generating a Complete Set of Ratio Edits: %Implied User's Guide," Statistical Research Division Report SS-Statistics # 2002-01.
22. B. Greenberg and T. Petkunas, "SPEER (Structured Program for Economic Editing and Referrals)," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1990.
23. S. Gregory, *Statistical Methods and the Geographer, Second Edition*, Longmans, Green, and Co. Ltd, London and Harlow, 1969.
24. S. I. Grossman (1984), *Elementary Linear Algebra*, Wadsworth Publishing Company, Belmont, California.
25. J. Hajek and Z. Sidak (1967), *Theory of Rank Tests*, Academic Press, New York, New York.
26. J. Hajek, Z. Sidak and P. K. Sen (1999), *Theory of Rank Tests*, Academic Press, San Diego, California.
27. Hammond, *Student Notebook Atlas*, Hammond World Atlas Corporation, 2010.
28. F. S. Hillier and G. J. Lieberman (1990), *Introduction to Operations Research*, McGraw-Hill Publishing Company, New York, New York.
29. R. V. Hogg and A. T. Craig (1995), *Introduction to Mathematical Statistics, Fifth Edition* Prentice Hall, Upper Saddle River, New Jersey.
30. R. V. Hogg and E. A. Tanis (1993), *Probability and Statistical Inference*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
31. M. Hollander and D. A. Wolfe (1999), *Nonparametric Statistical Methods, Second Edition*, John Wiley & Sons, Inc., New York, New York.
32. R. J. Johnston, *Multivariate Statistical Analysis in Geography*, Longman Group Limited, Longman, Inc., New York, 1978.
33. B. W. Kelly, "Probability Sampling in Collecting Farm Data," *Journal of Farm Economics*, Vol. 45, No. 5, Dec. 1963, pp. 1515-1520.
34. R. Khattree and D. N. Naik, *Applied Multivariate Statistics with SAS Software*, SAS Institute, Inc., Cary, North Carolina, 1995.
35. A. R. Kokan and S. Khan, "Optimum Allocation in Multivariate Surveys: An Analytical Solution," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 29, No. 1, 1967, pp. 115-125.
36. E.T. Lee, *Statistical Methods for Survival Data Analysis, Second Edition*, John Wiley & Sons, Inc., New York, 1992.

37. J. Lee and D. W. S. Wong, *Statistical Analysis with Arc View GIS*, John Wiley & Sons, Inc., New York, 2001.
38. E. L. Lehmann and H. J. M. D'Abrera (1975), *Nonparametrics, Statistical Methods Based on Ranks*, Holden-Day, Inc., Oakland, California.
39. P. Leong and S. Carlile, "Methods for Spherical Data Analysis and Visualization," *Journal of Neuroscience Methods*, vol. 80, 1998, pp. 191200.
40. P. A. W. Lewis (1966), "A Computer Program for the Statistical Analysis of Series of Events," *IBM Systems Journal*, vol. 5, no. 4.
41. D. W. Lefever, "Measuring Geographic Concentration by Means of the Standard Deviation Ellipse," *The American Journal of Sociology*, Vol. 32, No. 1. Jul., 1926, pp. 88-94.
42. R. V. Lenth, "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, Vol. 55, No. 3 (Aug., 2001), pp. 187-193.
43. L. Magee, "Improving Survey-Weighted Least Squares Regression," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Blackwell Publishing for the Royal Statistical Society, Vol. 60, No. 1, 1998, pp 115-126.
44. K. V. Mardia, "Statistics of Directional Data," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 37, No. 3 (1975), p. 349-393.
45. mardia2
46. K. V. Mardia, *Statistics of Directional Data*, Academic Press, New York, New York, 1972, pp. 52-58.
47. K. V. Mardia and P. E. Jupp, *Directional Statistics*, John Wiley & Sons, West Sussex, England, 2000.
48. A. Matei and Y. Tille, "Maximal and Minimal Sample Co-Ordination," *Sankhya: The Indian Journal of Statistics (2003)*, Vol. 67, No. 3 (Aug., 2005), pp. 590-612.
49. A. M. Mood (1954), "On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests," *The Annals of Mathematical Statistics*, vol. 25, no. 3, p. 514-522.
50. L. E. Moses (1963), "Rank Tests for Dispersion," *The Annals of Mathematical Statistics*, vol 34, no. 3, , p. 973-983.
51. R. G. Petersen, *Design and Analysis of Experiments* Marcel Dekker, Inc., New York, New York, 1985.
52. P. W. Porter, "What is the Point of Minimum Aggregate Travel?," *Annals of the Association of American Geographers*, Vol. 53, No. 2, Taylor & Francis,

- Ltd. on behalf of the Association of American Geographers, Jun., 1963, pp. 224-232.
53. N. D. Pratt, A. C. Bird, J. C. Taylor, and R. C. Carter, "Estimating Areas of Land Under Small-Scale Irrigation Using Satellite Imagery and Ground Data for a Study in N.E. Nigeria," *The Geographical Journal*, Vol. 163, No. 1 (Mar., 1997), pp. 65-77.
 54. M. L. Puri (1964), "Asymptotic Efficiency of a Class of c -sample Tests," *The Annals of Mathematical Statistics*, vol 35, no. 1, p. 102 - 121.
 55. V. F. Rickey and P. M. Tuchinsky, "An Application of Geography to Mathematics: History of the Integral of the Secant," *Mathematics Magazine*, Vol. 53, No. 3 (May, 1980), pp. 162-166.
 56. D. B. Rubin, "Composite Points in Weighted Least Squares Regressions," *Technometrics*, Vol. 22, No. 3, Aug. 1980, pp. 343-348.
 57. C. E. Sarndal, B. Swensson and J. Wretman (1999), *Model Assisted Survey Sampling*, Springer-Verlag New York, Inc., New York, New York.
 58. B. Shelby, "Girdle Distributions on the Sphere," *Biometrika Trust*, Vol. 51, No. 3/4, Dec. 1964, pp. 381-392.
 59. F. Solomon (1987), *Probability and Stochastic Processes*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
 60. D. J. Steffensmeier, E. A. Allan, M. D. Harer and C. Streifel, "Age and the Distribution of Crime," *The American Journal of Sociology*, Vol. 94, No. 4 (Jan., 1989), pp. 803-831.
 61. K. J. Thompson and R. Sigman, "Statistical Methods for Developing Ratio Edit Tolerances for Economic Data," *Journal of Official Statistics*, 2000.
 62. G. B. Thomas and R. L. Finney, *Calculus and Analytic Geometry, Fifth Edition*, Addison-Wesley Publishing Company, Reading Massachusetts, 1979.
 63. W. Winkler and L. Draper, "The SPEER Edit System," *Proceedings of the Conference of European Statisticians*, Section on Statistical Editing, UNECE, 1997.
 64. C. A. Wolf and D. A. Sumner, "Are Farm Size Distributions Bimodal? Evidence from Kernel Density Estimates of Dairy Farm Size Distributions," *American Journal of Agricultural Economics*, vol. 83, no. 1 (Feb., 2001), p. 77-88.
 65. A. Wood, "A Bimodal Distribution on the Sphere," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 31, No. 1 (1982), pp. 52-58.

66. R. F. Woolson and W. R. Clarke, "Analysis of Categorical Incomplete Longitudinal Data," *Journal of the Royal Statistical Society, Series A (General)*, Blackwell Publishing for the Royal Statistical Society, Vol. 147, No. 1, 1984, pp 87-99.
67. J. Yu, Z. Yang and X. Zhang (2002), "A Class of Nonlinear Stochastic Volatility Models and its Implications on Pricing Currency Options," www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2002/wp17-02.pdf.
68. R. S. Yuill, "The Standard Deviation Ellipse: An Updated Tool for Spatial Description," *Geografiska Annaler, Series B, Human Geography, Vol 53, No. 1*, Blackwell Publishing on behalf of the Swedish Society for Anthropology and Geography, 1971, pp. 28-39.

Index

Arctan function, 62
R, 142
Ch5ExponentialDeviationalEllipse
 , 174
Developer, 1
Dim statement, 3

weighted mean center, 43

Active.Cells, 9
altitude, 43
angle of rotation, 37, 61, 62, 66, 90, 100,
 119, 167
Angle_of_Rotation(), 168
anti-mode, 140, 144
approximate formula, 99
Arc GIS, 41
area, xxvii, 64, 91, 93, 94, 119, 120, 168,
 171, 172, 176, 182
area frame, xxv
area frame survey, xxv
area preserving plot, 142
Area(), 168
area, check, 120
area, total, 89
array, 4

arrays, 4
asymptotically, 145
average, 8, 142
averages, 29
axes lengths, 89, 119
axes relationship, 100
axis, 19
axis length, semi-major, 37
axis length, semi-minor, 37
axis of rotation, 20, 171, 175, 181
axis, semi-major, 14
axis, semi-minor, 14

bar graph, 94
bars, 19
beginning row, 9
bimodal, 144
browse, 16, 18

Calculate_B(), 189
Calculate_T(), 189
Cartesian coordinate system, 36, 37
cell, 7
cell contents, 7
cell, properties, 7
cental point, xxvii

- center of gravity, xxvi, 29, 37, 61, 95, 123, 124
- center of system, xxvii
- Ch4StandardDeviationalEllipse, 168
- Ch6WeibullDeviationalEllipse, 179
- Ch7NewtonGaussRoutines, 187
- Ch7SphericalStatistics, 187, 189
- chi-square distribution, 145
- circle, xxvii, 59, 64
- circular coordinate system, 112, 170
- circular coordinates, 115
- circular data, 37, 142
- closed form, 118
- clustered, 144
- clustered pattern, 144
- column chart, 18
- column names, 18
- columns, 8
- common syntax, 8
- concentration, 64
- concentration of the system, xxvii
- condition variable, 7
- continental U.S., 95
- conventions, 8
- convergence, 112, 122
- convergence analysis, 121
- coordinate pairs, 61
- coordinates, 13, 15, 42
- copying and pasting, 13
- correlation, 140
- correlation coefficient, 59
- county, 42
- county level, 44
- county-level, 44
- credible parameter estimates, 121
- crime data, 94, 95
- Crimestat, 41
- cross hairs, 21
- cubic equation, 147
- cumulative distribution function, xxviii, 84, 85, 112, 116, 117
- cumulative distribution function, joint, 88
- currency, 4

- data collection, xxvi
- Data mapping wizard, 16, 18
- data transformation, xxviii, 59
- data types, 4
- Data wizard, 16
- data, character, 8
- data, graph it, 18
- data, numeric, 8
- data, scale, 18
- data, shape, 139, 154

- data, transform, 83, 86
- data, transforming, 83
- date, 3, 4
- decimal degrees, 42
- default names, 8
- degrees, 142
- derivatives, 24, 112
- determinant, 145, 147, 149
- Developer ribbon, 9, 168, 174
- development environment, 21
- deviation, xxvi
- deviations, 90, 119
- diagonal elements, 145, 149
- direction, xxvii
- directional data, 141
- directional statistics, 140–142
- directional, data, 144, 149
- dispersed pattern, 144
- dispersion, 59
- disproportional weight, 148
- distance, 36
- distribution, 83
- distribution of the data, 94
- distribution of the points, xxvii
- distribution, sample, 127
- distribution, underlying, 126
- double, 3
- draw ellipses, 14
- drawing lines, 13

- Earth, 13, 15, 16, 20, 37
- eccentricity, xxvii, 64, 90, 91, 119, 168, 172, 176, 182
- eigenvalue, 145
- eigenvalues, 144, 145, 152, 156, 187
- eigenvectors, 145, 152, 156, 187
- ellipse, xxvii, 64, 90, 91
- ellipse, center, 37
- ellipse, draw, 19, 37
- ellipse, multi-modal, 140
- ellipse, rotate, 20, 37
- elliptic pattern, 144
- elliptical formula, 26
- ELSEIF, 6
- ending row, 9
- Equator, 141
- error messages, reasons, 122
- error term, 167
- error terms, 62
- estimates, 180
- estimator, 29, 62
- Excel, 3
- Excel 2010, VBA, 42
- Excel workbook, 16

- exp, 9
- exponential deviational ellipse, 93–95
- Exponential distribution, 24
- exponential distribution, 86, 89
- exponential distribution function, 83, 86
- exponential ellipse, 89, 93–95, 99–101, 110, 113, 126
- exponential ellipse code, 173
- exponential mean center, 89
- exponential model, 62
- Exponential_Dev_Ellipse(), 175
- exponentially distributed, 85, 87
- exponentially distributed, weighted, deviational ellipse, 89
- exponential model, 176
- EXPstats(), 174
- extreme values, 101

- first derivative, 110, 113
- font, 7
- font size, 7
- FOR-NEXT, 6, 7
- format shape, 20
- formulas, 8
- FORTRAN, 41
- fx(Lambda), 189

- Gamma1, 180
- Gamma2, 180
- Gauss(Lambda), 189
- Gaussian elimination, 147
- general linear means model, 62, 89
- geo-code, 13, 41
- geo-coding, 37
- geo-statistics, 59
- geographic data, 37
- girdle, 144, 145
- girdle distribution, 150
- global variables, 8
- Google Earth, 13, 41, 42
- graph, 87
- graph paper, 83
- graphing, 13
- graphs, 127
- gx(Lambda), 189

- histogram, 140, 148, 154
- hoop coordinate system, 36

- IF-ELSE-IF, 5
- IF-THEN-ELSE, 5
- image classification, xxv
- imaginary axis, 37
- Import data wizard, 15, 18
- inital values, 111
- initial guess, 181
- initial guesses, 179
- initial value, 122, 147
- initial values, 112, 114
- initialization, 112, 114
- input parameters, Weibull, 121
- integrals, 25
- iteger numbers, 3
- iteration, 24, 111, 116
- Iterations, 180
- iterations, 112, 122, 181
- iterations, number of, 112

- Jacobian matrix, 111, 114
- joint density function, 115
- joint distribution, 85, 87, 115, 116, 176, 181, 182
- joint likelihood function, 87
- joint probability density function, 87, 116
- June Area Survey, 42, 43

- kilometers, 19
- kurtosis, 141

- label text, 18
- land segments, 94
- largest weights, 94
- latitude, xxvi, 9, 24, 37, 83, 86, 87, 89, 99, 109, 116, 121, 122, 141, 150, 176
- latitude observation, 62
- least squares, 89, 140
- length, 13, 19
- likelihood function, 85, 86, 110, 121, 127, 172, 176, 182
- likelihood function, Weibull, 110, 113
- likelihood functions, 179
- linear means model, 118
- linear regression, 140
- linear regression model, 62, 83, 89
- linear relation, 85, 87
- linear relationship, 83, 85, 86
- local variables, 9
- location parameter, 83, 89
- location problem, xxvi
- log, 9
- Log10(X), 180
- logical, 4
- long integer, 4
- longitude, xxvi, 9, 24, 37, 63, 86, 87, 89, 99, 113, 116, 121, 122, 141, 150, 176
- longitude data, 112
- longitude observations, 112
- loop, For-Next, 9
- loop, While-Wend, 9

- major axis, 90, 119, 171
- major axis length, 91
- major axis of rotation, 175, 181
- map, 37
- Map Point, 13, 19, 20
- Map Soft, 41
- matrix B , 144, 149, 155
- matrix T , 145, 149, 150, 155, 187
- maximum likelihood estimates, 116
- maximum likelihood estimator, 83, 85–87, 110, 113, 115, 116
- maximum likelihood theory, 127
- mean center, xxvi, 13, 19, 59, 60, 62, 100, 112, 115, 119, 125, 139, 176
- mean center coordinates, 176
- mean center, transform, 172
- mean direction, 142
- mean joint distribution, 89
- mean joint distribution estimate, 118
- mean latitude, 142, 143
- mean longitude, 142
- measure lines, 37
- menu, Insert, 12
- merged cells, 49
- miles, 19
- minor axis, 90, 91, 119, 171
- minor axis length, 91
- minor axis of rotation, 176, 181
- mod, 8
- mode, 140, 144
- modes, 154
- module, 4
- modules, 9
- MS Word, 20
- multimodal statistics, 140

- Newton algorithm, 147
- Newton's method, 147
- Newton() subroutine, 149
- Newton(Lambda), 189
- Newton-Raphson method, 111, 114
- NewtonGauss(), 187
- no fill, 20
- non-uniform distribution, xxvii
- normal distribution, 89
- Northern Hemisphere, 141
- null hypothesis, 152, 156
- number of iterations, 112
- numerical algorithms, 111

- observation, 86
- observed value, 88, 110
- one pole, 36
- original data, 118

- oval, 20

- parameter, 24, 85
- parameter estimate, 110, 112, 113, 115
- parameter estimates, 121
- parameter estimation, 112
- parameter estimator, 86
- parameters, 110, 126, 180, 181, 189
- percentage, 141
- pie chart, 18
- plane, 144
- plot, 14, 85, 87
- plotting, 13
- plotting the transformed data, 83
- point estimates, 181
- point pattern, 59
- points, 37
- points, distribution, 150
- polar coordinate, 142
- polar coordinate system, 36
- positive definite, 99
- predicted values, 62
- preferred direction, no, 144
- Prime Meridian, 141
- probability density function, 84, 87, 109, 112, 115
- program code, 8
- properties, 7
- pushpins, 16
- Pythagoras' theorem, 59

- quadratic equation, 26
- quadratic equation, roots, 26, 61
- quantile, 141
- quantity, 44

- radians, 142
- random, 148
- random sample, 110, 113
- random variable, xxvi, 9, 24, 37, 41, 62, 63, 85, 86, 88, 89, 113, 117, 150
- random variable w_i , 110
- range, 115
- real number, 4
- real numbers, 3
- record, 4
- records, 4
- regional data, 49
- regional level, 44, 49
- regression, 89
- regression estimator, 59
- regression line, xxviii
- relationship, 89, 99
- relative concentration, xxvii
- resultant length, 144

- roots, 145
- rotate, 14
- rotational symmetry, 145
- row reduction, 147
- ruler, 37
- ruler, anchor, 19

- sample data, 84
- sample size, 24, 142
- sample space, 86
- satellite images, 42
- saving work, 13
- scale parameter, 24, 84, 89, 181
- scatter diagram, 141
- SDE code, 168
- SDE statistics, 170
- SDE.Likelihood(), 168
- SDEdeltas(), 168
- Secant algorithm, 116, 179, 181
- Secant estimation, 184
- Secant iterations, 118
- Secant method, 111, 112, 114
- Secant programs, 121
- Secant1(), 179, 180
- Secant2(), 180, 181
- segment, xxvi
- Selection.Rows.Count, 8
- semi-major axes, 99
- semi-major axis, 37, 62, 98
- semi-major axis length, 63, 66, 100, 119, 120, 167, 171, 176, 182
- semi-minor axis, 37, 62
- semi-minor axis length, 63, 66, 100, 119, 120, 167, 172, 176, 182
- series column chart, 18
- Set_weights(), 168, 174
- shaded circles, 16
- shape parameter, 83, 181
- shapes, 20
- single pole, 36
- sized circles, 16
- sized pie chart, 18
- skewness, 141
- smaller weights, 94
- Southern Hemisphere, 141
- spatial analysis, 139
- spatial data, 43, 142
- spatial statistics, xxvi
- specialty software, 37
- spherical distribution, 144, 187
- spherical statistics, 36, 37, 139, 147
- spherical variance, 143
- Spherical_Dev_Ellipse(), 150, 187
- spreadsheet, 49

- standard deviation, xxvi, 139, 171, 176, 181
- standard deviational ellipse, xxvi, 59, 63, 66, 93–95, 98–101, 116, 123, 124, 168
- standard deviational ellipse formula, 88
- standard deviations, 90, 94, 119
- standard distance, 59
- standard ellipse, 99
- standard error terms, 62
- state name, 42
- STATS, 9, 170, 175, 181
- stereographic-net, 142
- straight line, 84, 118
- string characters, 3
- subroutine, 4
- sufficient statistics, 110, 113
- sum, 8
- sums of squares, 89, 99, 119, 126
- survey, 43
- survey area, 19, 37
- survey data, 13

- termination condition, 181
- test statistic, 149, 152, 156
- THEN, 6
- time, 3
- trace, 145
- transformation, weighted latitude, 84
- transformed data, 85, 89, 118
- transformed variable, 85
- trend of the system, xxvii
- triadic, 149
- trial and error, 112, 147
- two dimensions, 36
- TYPE-END-TYPE, 4

- uniform, 144, 145, 156
- uniform distribution, xxvi, xxvii
- units, 19
- unweighted exponential deviational ellipse, 176
- unweighted, calculations, 139

- Variable, 180
- VBA code, 112
- VBA Development environment, 1
- VBA Editor, 12, 170
- VBA editor, 4
- VBA environment, 8, 168, 187
- VBA functions, 9
- VBA numeric calculations, 9
- VBA program, 173
- VBA statements, 8
- VBA, Standard_Dev_Ellipse(), 167
- vector \underline{b}_i , 144
- vectors, 36

violent crime, 67, 150
 violent crime, survey, 28
 visible, 4
 Visual Basic button, 12, 174
 Visual Basic Editor, 173
 Visual Basic for Applications (VBA), 42
 Visual Basic ribbon, 168

 Weibull deviational ellipse, 118, 121, 123, 124
 Weibull distribution, 24, 112, 118
 Weibull distribution, sufficient statistics, 110, 113
 Weibull ellipse, 110, 113, 123–126, 179
 Weibull models, weighted, 115
 Weibull(), 180, 181
 Weibull.Dev_Ellipse(), 179
 weight, 24, 29, 62, 150, 168
 weight column, 176
 weight, disproportionately, 150
 weighted data, 28, 94
 weighted ellipse, 62
 weighted exponential deviational ellipse, 95, 98–100
 weighted exponential distribution, 83
 weighted exponential model, 87
 weighted exponential standard deviational ellipse, 89
 weighted latitude, 84, 85, 89
 weighted longitude, 87, 89
 weighted mean, 98, 99
 weighted mean center, xxvii, 59, 60, 63, 115, 167, 171, 175, 181
 weighted mean estimate, 89
 weighted mean response, 62
 weighted means estimate, 118
 weighted points, 90
 weighted standard deviational ellipse, xxvii, 43, 60, 61, 99, 100, 167
 weighted sum of squares, 62
 weighted Weibull models, 115
 weighted, Weibull ellipse, 119
 weighted, Weibull standard deviational ellipse, 118
 weights, 117
 WHILE-WEND, 6, 7
 whole numbers, 3
 WITH-END-WITH, 7
 worksheet, 8
 worksheetfunction.pi(), 9
 Worksheets().Cells().Value, 9
 Wulff-net, 142