

Less than the Least: An Alternative Method of Least Squares Linear Regression

THESIS

Presented to the Honors Committee of
McMurry University in Partial
Fulfillment of the Requirements

For Undergraduate Departmental Honors
in Mathematics

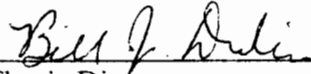
By
Sarah B. Martin

Abilene, Texas

April 2, 1998

Less than the Least: An Alternative Method of Least Squares Linear Regression

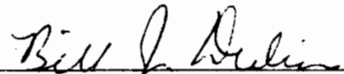
Approved:



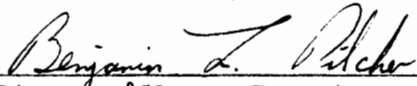
Thesis Director



Thesis Director

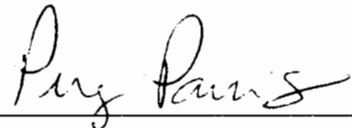


Department Chair



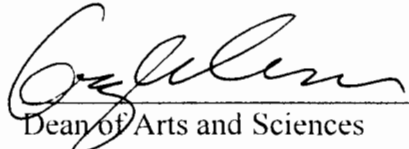
Director of Honors Committee



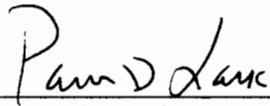




Honors Committee Members



Dean of Arts and Sciences



Vice President for Academic Affairs

JOURNAL OF THESIS ABSTRACTS

THESIS SUBMISSION FORM

AUTHOR: Sarah Beth Martin
THESIS TITLE: Less than the Least:
An Alternative Method of Least Squares Linear Regression
THESIS ADVISORS: Dr. Bill Dulin and Dr. Kelly McCoun
ADVISORS' DEPT.: Mathematics
DATE: April 2, 1998
HONORS PROGRAM: Departmental
NAME OF COLLEGE: McMurry University
PAGE LENGTH: 48
BIBLIOGRAPHY: Yes
ILLUSTRATED: Yes
COPIES AVAILABLE: Hard Copy

IS YOUR THESIS OR ANY PART BEING SUBMITTED FOR PUBLICATION? Yes
**IF ANY PART HAS BEEN ACCEPTED FOR PUBLICATION, PLEASE INDICATE
WHERE:** N/A

SUBJECT HEADINGS: Linear Regression, Least Squares, History of Statistics, Minimization
Techniques, Numerical Analysis

ABSTRACT:

The thesis presents and investigates an alternative method of least squares linear regression, which defines the error for a point to be the perpendicular distance from the data point to the regression line, rather than the traditional vertical distance.

The thesis is divided into six chapters. The first presents background history on the method of least squares. The second presents and derives the slope and y -intercept formulas for the traditional method least squares linear regression. The third presents and derives the slope and y -intercept formulas for the alternative least squares regression method. The fourth discusses the advantages and disadvantages of the alternative method as compared with the traditional least squares linear regression method. The fifth presents a method for comparing the accuracy of the alternative and traditional methods. The final chapter includes concluding thoughts and information on other references to a similar method of linear regression. Appendices include examples comparing the results of the alternative and traditional least squares method, the template program for the calculation of these results, and the accuracy comparison programs for data sets with error uniformly distributed in both variables and in only the y -variable.

Acknowledgments

I would like to extend many thanks to my advisors, Dr. Kelly McCoun and Dr. Bill Dulin. They have helped me immensely throughout this project, from the initial idea to the final product. This thesis would have been impossible without their constant support, innovative ideas, and endless assistance. They have taught me more than I ever expected, and helped me accomplish more than I ever imagined.

I also wanted to greatly thank Dr. Doris Miller, who willingly proofread this work through various stages of progress. She provided a wealth of structural and technical suggestions, greatly improving the final version.

I especially want to thank my future husband, Dan Meyer. He has offered endless support, assurance, and encouragement, giving me the strength necessary to complete this endeavor. When frustration began to set in, he was always there to rescue me.

Countless thanks belong to so many other faculty members, friends, and family. Their belief in my ability to complete this task far surpassed my own; their unconditional faith and confidence have been invaluable.

TABLE OF CONTENTS

I. History of the Method of Least Squares	1
II. Traditional Least Squares Linear Regression Method	9
III. Alternative Method of Least Squares Linear Regression	14
IV. Advantages and Disadvantages of Alternative Method	26
V. Accuracy Comparison	28
VI. Conclusion	31
Appendix A: Maple Template Program	34
Appendix B: Example 1	36
Example 2	39
Appendix C: Accuracy Comparison Program for Data With Error In Both Variables	42
Appendix D: Accuracy Comparison Program for Data With Error In Only y-variable	45
Appendix E: Accuracy Comparison Results	48
Appendix F: Bibliography	49
Appendix G: Vita	51

Chapter I

History of the Method of Least Squares

Linear regression is a method that allows a linear equation to be developed that “best” describes a collection of points. The “best” equation is the one “that makes the data appear most likely when taken as a whole” (Daniels 6); the term “best” refers to the solution that has the smallest total error, or sum of individual errors. The method of least squares is the most common method of linear regression. The least squares method takes the individual error for a point to be the distance between the point and the theoretical line. The concept of linear regression, including the method of least squares, has had a long history. The actual method of least squares did not appear in print until the early nineteenth century, but mathematical, statistical, and scientific advancements throughout the eighteenth century provided the incentive and structure for the development of the least squares method of linear regression.

In the mid 1700s the statistical method of taking arithmetic means of small collections of measurements was popular in astronomy and navigation. Only measures considered to be equivalent, those measurements made by the same observer with the same instrument at the same time in the same place, were averaged. The combination of measurements not considered to be of equivalent accuracy were rare before 1750. One exception is the work of Roger Cotes. His work was published posthumously in 1722. “Cotes’s rule can be (and has been) read as recommending a weighted mean, or even as an early appearance of the method of least squares” (Stigler 16). However, Cotes did not provide examples to clear up the meaning and extent of his rule. “To understand the genesis of the method of least squares, we must look not just at what

the investigators say they are doing (and how statements might be most charitably interpreted in the light of later developments) but also at what was actually done. Cotes's rule had little or no influence on Cotes's immediate posterity" (Stigler 16).

Three major scientific problems of the eighteenth century are closely associated with the development of the method of least squares: the attempt to mathematically determine and represent the motion of the moon; the desire to account for the inequalities in the motions, or apparently nonperiodic motions, of Jupiter and Saturn; and the effort to determine the shape of the earth. "These problems all involved astronomical observations and the theory of gravitational attraction, and they all presented intellectual challenges that engaged the attention of many of the ablest mathematical scientists of the period" (Stigler 16-17). Two widely read works that greatly influenced later works and, statistically, formed "a unique and dramatic contrast in the handling of observational evidence" were Leonhard Euler's *Recherches sur la question des inégalités du mouvement de Saturne et de Jupiter*, published in 1749, and Tobias Mayer's *Abhsndlung über die Umwälzung des Monds um seine Axe und die scheinbare Bewegung der Mondsflecken*, published in 1750 (Stigler 17). Statistically, the work of Mayer, an astronomer, proved to be successful, while mathematician Euler's statistical work was, comparatively, a failure. "They show why the discovery of the method of least squares was not possible in the intellectual climate of 1750, and they highlight the conceptual barriers that had to be crossed before this climate became sufficiently tropical to support the later advances of Legendre, Gauss, and Laplace" (Stigler 17).

In 1750 Johann Tobias Mayer showed how his numerous observations of the moon explained the eccentricities of the moon's orbit. From April 1748 to March 1749, Mayer

recorded the positions of several prominent lunar features, especially the crater Manilius.

Mayer's method consisted of developing equations based on his observations and trigonometric identities. He had twenty-seven equations, based on twenty-seven days of observation, and only three unknowns. He grouped these equations into three groups of nine, added them, and solved the remaining three equations for his unknowns. Although this method seems relatively simple today, in the mid-eighteenth century it was revolutionary. His method of handling this data was innovative, and it quickly became widely circulated. "Mayer's method, unlike least squares, was not 'best' in the sense that it appeared as the solution to a mathematically posed problem of finding the 'best' combination of inconsistent equations. It was an ad hoc method, and its acceptance depended upon its reputation for past successful use, its ease of application, and the investigator's intuitive feeling that by combining the equations in such a way that the coefficients of the unknowns are successfully maximized, a mechanically stable (and hence reliable) solution would result" (Stigler 39).

In 1787 Pierre Simon Laplace proposed, in the course of writing a memoir on the inequalities in the motions of Saturn and Jupiter, another method of dealing with inconsistent linear equations. Laplace's method is somewhat similar to Mayer's, but Laplace reduced twenty-four equations into four equations, and solved these equations for his four unknowns. Although he did not provide an explanation, Laplace's method of selection of the four groups of equations seems to be based on the relationships of the coefficients of the unknowns in the twenty-four original equations. Mayer had only grouped his equations into disjoint groups, but Laplace grouped the same equations in several different ways. "Another investigator using Laplace's method would not require the same degree of 'good luck' required by a follower of

Mayer and usually would be rewarded by greater accuracy as well” (Stigler 37). By 1800 the idea of combining observational equations, made popular through the work of Mayer and Laplace, became a convenient and widespread procedure.

In March of 1805, Adrien Marie Legendre, a French mathematical scientist, first published the method of least squares in an eighty page book under the title *Nouvelles méthodes pour la détermination des orbites des comètes*. In January 1806, it gained a fifty-five page supplement, and in August of 1820, a second eighty page supplement. The actual method of least squares was presented in a nine page appendix entitled “Sur la méthode des moindres carré.” Legendre’s presentation of the method of least squares is said to be unsurpassed in its stark clarity, and considered to be “one of the clearest and most elegant introductions of a new statistical method in the history of statistics” (Stigler 13). Legendre wrote in 1805 of the method of least squares:

Of all the principles that can be proposed . . . I think there is none more general, more exact, and more easy of application, than that of which we have made use . . . which consists of rendering the sum of the squares of the errors a *minimum*. By this means there is established among the errors a sort of equilibrium which, preventing the extremes from exerting an undue influence, is very well fitted to reveal that state of the system which most nearly approaches the truth.

(Legendre 577)

Legendre added several other points in defense of his least squares method: if a perfect fit were possible, his method would find it; and if the error of a point was later decided to be too large, it would be simple to revise the equations by subtracting the appropriate terms. In conclusion of

his presentation, Legendre wrote that the method of least squares “reveals to us, in a fashion, the center around which all the results furnished by experiments tend to distribute themselves, in such a manner to make their deviations from it as small as possible” (Legendre 579).

Initially, Legendre accepted the statistical ideas that evolved from Mayer’s early work: that a balance should be struck between measurements. Legendre reconsidered these ideas in his 1805 preparation of a memoir on the determination of cometary orbits. Legendre worked the equations according to Mayer’s method and wrote that the errors were “quite tolerable in the theory of comets. But it is possible to reduce them further by seeking the *minimum* of the sum of squares.” He then reworked the equations in accordance with his idea of least squares. There is speculation that Legendre discovered the method of least squares while his memoir was in the late stages of preparation, since it is not used early in the memoir. However, it is clear that Legendre immediately realized the importance and extent of the least squares method. In his two-page explanation of the least squares method, the word “minimum” made five italicized appearances, revealing his apparent excitement. Legendre concluded the appendix with an example, revealing the depth of his understanding. He translated the use of the least squares method from determining the orbits of a comet, to another problem of the day, determining the figure of the earth using the 1795 measurements of the French meridian arc from Montjouy to Dunkirk. By using this data, Legendre’s example was not just a simple explanation, but a significant advance in geodesy.

Before the end of 1805, Legendre’s method of least squares had already appeared in another publication, Puissant’s *Traité de géodésie*, and in August 1806 it was presented to a German audience by von Lindenau in von Zach’s astronomical journal, *Monatliche*

Correspondenz. The multiple publications of Legendre's least squares method caught the attention of Carl Friedrich Gauss, who mentioned in *Theoria Motus Corporum Coelestium in Sectionibus Solum Ambientium*, published in 1809, that he had been using the principle of least squares since 1795, when he was eighteen years old, and "referring to the method of least squares as 'our principle'" (Stigler 145). The matter of who first developed the method of least squares, Legendre or Gauss, proved to be an ongoing dispute. "Legendre appended a semianonymous attack on Gauss to the 1820 version of his *Nouvelles méthodes pour la détermination des orbites des comètes*, and Gauss solicited reluctant testimony from friends that he had told them of the method before 1805" (Stigler 145). In 1812 Gauss wrote to Laplace, who was acting as a diplomatic mediator between Legendre and Gauss, "that he had used the method almost daily ever since 1802 in his calculations concerning the new planets, and that he had communicated the method to several of his friends" (Hall 74-75). Legendre, understandably upset by Gauss's claim, wrote a letter to him "practically accusing him of dishonesty and complaining that Gauss, so rich in discoveries, might have had the decency not to appropriate the method of least squares" (Newman 331).

Later research showed that Gauss actually had developed and utilized the method of least squares prior to Legendre's publication of it in 1805. "Gauss's correspondence and the papers found after his death proved that he was certainly first to make the discovery, but since Legendre was first to publish it, priority rights belong to the latter. Honor belongs to both of them, since they reached the result independent of each other" (Hall 75). However, Gauss "was unsuccessful in whatever attempts he made to communicate it before 1805. In addition, there is no indication that he saw great potential before he learned of Legendre's work" (Stigler 146).

Traditionally, Legendre is recognized as the originator of the method of least squares. Gauss however did have impact on the subject. In his 1823 work, *Theoria combinationis observationum erroribus minimis obnoxiae*, "he develops the method of least squares with mathematical rigor as, in general, the most suitable way of combining observations" (Hall 78).

By 1815, ten years after Legendre's first publication of the method of least squares, "it was a standard tool in astronomy and geodesy in France, Italy, and Prussia" (Stigler 15). The earliest English translation of Legendre's presentation of the method of least squares was by George Harvey in 1822, which he entitled "On the Method of Minimum Squares." Before 1825, when the least squares method became a standard in England, it was referred to as the method of "minimum squares" or "small squares" (Stigler 15).

People were initially unwilling to reject the methods of Mayer and Laplace and accept the notion of least squares. One reason for the continued popularity of Laplace's method of reducing systems of inconsistent linear equations was its simplicity; the method required only addition and subtraction, while still providing a measure of accuracy close to that of the least squares method. Their reluctance is understandable because, "Who could not feel confident using a method that had reconciled the motions of Jupiter and Saturn with Newtonian gravitational theory?" (Stigler 39). Despite the public's reluctance to change, the least squares method gradually became more accepted. As the years went by, "more experience and ways of simplifying computation made least squares easier to use and as succeeding generations of mathematicians made no successful attempt to give formal statement to the vague intuitive notions of the reasonableness of Mayer's method, it faded from view--or rather moved from the workshop of the practitioner to the display case of the statistical museum" (Stigler 39). As time

passed, "the vivid impression of these past triumphs [of Laplace's method] faded, however, to be replaced by stories of the triumphs of least squares" (Stigler 39).

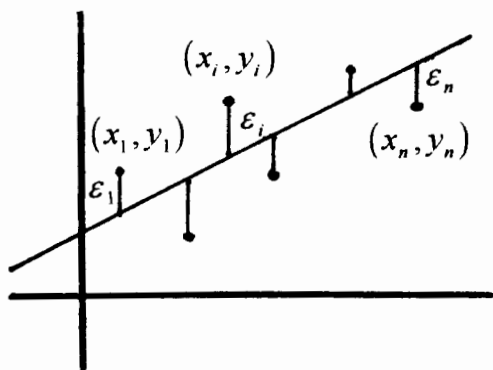
The development concerning the method of least squares was far from being completed. Further work by Gauss and Laplace investigated the relationship of least squares with formal probability. Without this connection, "no assessment of the method's results was possible. The method of least squares produced results that could be called 'best,' as they minimized the sum of squared errors and produced an appealing mechanical equilibrium," but the question, "how good is the best?" could not be answered without further development (Stigler 140). Additional work on the method of least squares was conducted to determine the accuracy of its results and to develop method to simplify the computations. "Many hands and minds would be involved in the process-- Laplace and Gauss, Poisson and Bienaymé and Cauchy, and a host of other astronomers, mathematicians and geodesists. The technical complexity was such that it is hard to appreciate what was accomplished and impossible to gauge accurately the depth of contemporary understanding by looking further at the works of the period describing purely mathematical development" (Stigler 148).

The method of least squares has had great impact on the world of mathematics as well as the observational sciences. "Its efficiency was the most blatant demonstration of the fact that natural phenomena could efficiently be investigated by mathematical means. Gauss would have expressed this fact by a much stronger statement-- for him, mathematics governed the workings of nature, and the mathematical penetration of the natural sciences showed to what degree they had been understood" (Bühler 140). As a tool, the method of least squares is very powerful, capable of producing reliable and accurate descriptions of observations.

Chapter II

Traditional Least Squares Linear Regression

The traditional method of least squares linear regression defines the error, ϵ , for a point to be the correction between the actual and predicted y-values, or the vertical distance between a point and the line.



$$\begin{aligned} y_1 &= mx_1 + b + \epsilon_1 \\ y_2 &= mx_2 + b + \epsilon_2 \\ &\vdots \\ y_i &= mx_i + b + \epsilon_i \\ &\vdots \\ y_n &= mx_n + b + \epsilon_n \end{aligned}$$

The error term for each point is determined to be $\epsilon_i = y_i - mx_i - b$. We want to minimize the sum of squared errors, S , so the minimizing slope, m , and y-intercept, b , values must be found.

$$S(m, b) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - mx_i - b)^2$$

To minimize $S(m, b)$ the partial derivatives with respect to m and b are found.

$$\frac{\partial S}{\partial m} = \frac{\partial}{\partial m} \left(\sum_{i=1}^n (y_i - mx_i - b)^2 \right) = -2 \sum_{i=1}^n x_i (y_i - mx_i - b)$$

$$\frac{\partial S}{\partial b} = \frac{\partial}{\partial b} \left(\sum_{i=1}^n (y_i - mx_i - b)^2 \right) = -2 \sum_{i=1}^n (y_i - mx_i - b)$$

The critical point, or possible extreme values of $S(m, b)$, is found by setting the first partial derivatives equal to zero and solving the system.

$$\begin{aligned}
\frac{\partial S}{\partial m} &= -2 \sum_{i=1}^n x_i (y_i - mx_i - b) = 0 \\
\sum_{i=1}^n x_i (y_i - mx_i - b) &= 0 \\
\sum_{i=1}^n x_i y_i - m \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i &= 0 \\
\sum_{i=1}^n x_i y_i &= m \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i
\end{aligned} \tag{2.1}$$

$$\begin{aligned}
\frac{\partial S}{\partial b} &= -2 \sum_{i=1}^n (y_i - mx_i - b) = 0 \\
\sum_{i=1}^n (y_i - mx_i - b) &= 0 \\
\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i - bn &= 0 \\
\frac{1}{n} \sum_{i=1}^n y_i - \frac{m}{n} \sum_{i=1}^n x_i - b &= 0 \\
\text{so } \bar{y} - m\bar{x} - b &= 0, \text{ and}
\end{aligned} \tag{2.2}$$

$$\begin{aligned}
\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i - bn &= 0 \\
\sum_{i=1}^n y_i &= m \sum_{i=1}^n x_i + bn
\end{aligned} \tag{2.3}$$

Cramer's rule is used to solve the system of equations 1 and 3 for the critical m value.

$$m = \frac{\begin{bmatrix} \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i & n \end{bmatrix}}{\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix}} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \tag{2.4}$$

The y -intercept, b , can be found by solving equation 2.2 for b and substituting the slope value found by equation 2.4.

$$b = \bar{y} - m\bar{x} \quad (2.5)$$

To prove that the values of m and b given by formulas 2.4 and 2.5 minimize the sum of squared errors, S , the second partial derivatives must be determined. The second partial derivatives of S with respect to both m and b , regardless of the order, are equal, so only one needs to be found.

$$\begin{aligned} \frac{\partial^2 S}{\partial m^2} &= \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n x_i (y_i - mx_i - b) \right) \\ &= \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n x_i y_i + 2m \sum_{i=1}^n x_i^2 + 2b \sum_{i=1}^n x_i \right) \\ &= 2 \sum_{i=1}^n x_i^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 S}{\partial b \partial m} &= \frac{\partial^2 S}{\partial m \partial b} = \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n (y_i - mx_i - b) \right) \\ &= \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n y_i + 2m \sum_{i=1}^n x_i + 2bn \right) \\ &= 2 \sum_{i=1}^n x_i \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 S}{\partial b^2} &= \frac{\partial}{\partial b} \left(-2 \sum_{i=1}^n (y_i - mx_i - b) \right) \\ &= \frac{\partial}{\partial b} \left(-2 \sum_{i=1}^n y_i + 2m \sum_{i=1}^n x_i + 2bn \right) \\ &= 2n \end{aligned}$$

The second partial derivatives of $S(m, b)$ is the following 2×2 matrix.

$$D^2(S(m, b)) = \begin{bmatrix} \frac{\partial^2 S}{\partial b^2} & \frac{\partial^2 S}{\partial m \partial b} \\ \frac{\partial^2 S}{\partial b \partial m} & \frac{\partial^2 S}{\partial m^2} \end{bmatrix} = \begin{bmatrix} 2n & 2 \sum_{i=1}^n x_i \\ 2 \sum_{i=1}^n x_i & 2 \sum_{i=1}^n x_i^2 \end{bmatrix} = 2 \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

For the critical point (m, b) to minimize the sum of squared errors, the second derivative matrix must be positive definite. A matrix is defined as positive definite if the determinant of all submatrices is positive. The upper left submatrix of $D^2(S(m, b))$ is always positive, because the number of points, n , is always positive. The determinant of the second partial derivative matrix of $S(m, b)$ is found to be the following:

$$|D^2(S(m, b))| = 4n \sum_{i=1}^n x_i^2 - 4 \left(\sum_{i=1}^n x_i \right)^2 \quad (2.6)$$

Since we are concerned with values that will make the matrix positive definite, so equation 2.6 is set greater than zero, which reduces to inequality 2.7, and can be shown true for all values of

x_i .

$$n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 > 0 \quad (2.7)$$

$$\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 > 0$$

$$\sum_{i=1}^n x_i^2 - n \left(\frac{1}{n^2} \right) \left(\sum_{i=1}^n x_i \right)^2 > 0$$

$$\sum_{i=1}^n x_i^2 - n \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 > 0$$

$$\sum_{i=1}^n x_i^2 - n \bar{x}^2 > 0$$

$$\sum_{i=1}^n x_i^2 + (-2\bar{x}^2 n + n\bar{x}^2) > 0$$

$$\begin{aligned}
\sum_{i=1}^n x_i^2 - 2\bar{x}n\left(\frac{1}{n}\sum_{i=1}^n x_i\right) + n\bar{x}^2 &> 0 \\
\sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n\bar{x}^2 &> 0 \\
\sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) &> 0 \\
\sum_{i=1}^n (x_i - \bar{x})^2 &> 0
\end{aligned} \tag{2.8}$$

Inequality 2.8 is true if and only if there exists some $x_i \neq \bar{x}$. It is valid to assume that not all x_i 's are equal, because the line describing the points would be vertical, of the form $x = c$, where c is a constant. The traditional method of least squares linear regression can not produce vertical equations.

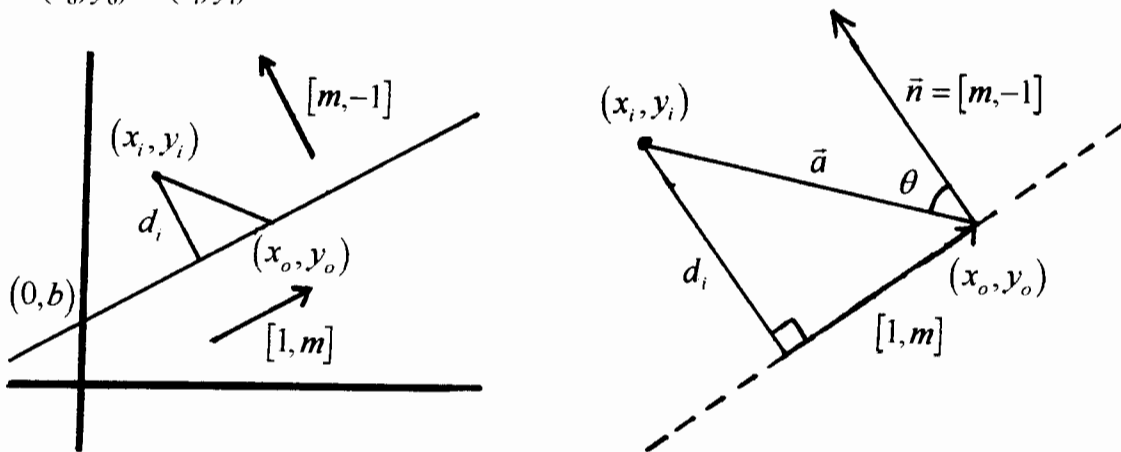
Since the determinant of the second derivative matrix is always positive, the values of slope and y -intercept produced by equations 2.4 and 2.5 minimize S , the sum of square errors. According to the definition of error as the vertical distance between the data point and the regression line, the traditional method of least square linear regression will produce the error-minimizing line.

Chapter III

Alternative Method of Least Squares Linear Regression

The alternative method of least squares linear regression is derived in a similar manner to the derivation of the traditional method presented in the preceding chapter. The same steps are followed, but the definition of error, or distance from a point to the regression line, is changed. The formulas derived will find the “best” line according to the alternative definition of error.

A line with slope m and y -intercept b is parallel to the vector $[1, m]$, and perpendicular to the vector $[m, -1]$. The perpendicular distance, d_i , between a point (x_i, y_i) and the line $y = mx + b$ can be found in the following manner. Suppose (x_o, y_o) is a point on the line and \vec{a} is the vector from (x_o, y_o) to (x_i, y_i) .



$$d_i = (\|\vec{a}\| \cos \theta) \left(\frac{\|\vec{n}\|}{\|\vec{n}\|} \right)$$

$$d_i = \frac{\vec{a} \cdot \vec{n}}{\|\vec{n}\|}$$

$$d_i = \frac{[x_i - x_o, y_i - y_o] \cdot [m, -1]}{\sqrt{m^2 + 1}}$$

$$d_i = \frac{m(x_i - x_o) - (y_i - y_o)}{\sqrt{m^2 + 1}}$$

As in the traditional method of linear regression, we will use the squared distance, d_i^2 , to find the sum of squared errors, S , or total error of the regression line.

$$d_i^2 = \frac{(m(x_i - x_0) - (y_i - y_0))^2}{m^2 + 1}$$

$$d_i^2 = \frac{(mx_i - y_i + (y_0 - mx_0))^2}{m^2 + 1}$$

$$d_i^2 = \frac{(mx_i - y_i + b)^2}{m^2 + 1}$$

$$S = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \left(\frac{(mx_i - y_i + b)^2}{m^2 + 1} \right)$$

$$S = \frac{1}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b)^2$$

To minimize the sum of squared errors, S , the first step is to take the partial derivatives of S with respect to m and b .

$$\frac{\partial S}{\partial b} = \frac{1}{m^2 + 1} \sum_{i=1}^n (2)(mx_i - y_i + b)$$

$$\frac{\partial S}{\partial b} = \frac{2}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b)$$

$$\frac{\partial S}{\partial m} = \frac{-2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 + \frac{1}{m^2 + 1} \sum_{i=1}^n (2x_i)(mx_i - y_i + b)$$

$$\frac{\partial S}{\partial m} = \frac{2}{m^2 + 1} \sum_{i=1}^n x_i(mx_i - y_i + b) - \frac{2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2$$

To solve for the possible error minimizing values of m and b , the first partial derivatives are set equal to zero and solved for m or b . The points of form (m, b) are obtained by solving the

resulting system of equations. These points are called the critical points of $S(m, b)$. On the three dimensional graph of $S(m, b)$ the critical points are visible as turning points on the graph.

$$\frac{\delta S}{\delta b} = \frac{2}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b) = 0$$

$$\sum_{i=1}^n (mx_i - y_i + b) = 0 \quad (3.1)$$

$$m \sum_{i=1}^n x_i - \sum_{i=1}^n y_i + nb = 0$$

$$b = \frac{1}{n} \sum_{i=1}^n y_i - \frac{m}{n} \sum_{i=1}^n x_i$$

$$b = \bar{y} - m\bar{x} \quad (3.2)$$

$$\frac{\delta S}{\delta m} = \frac{2}{(m^2 + 1)} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 = 0$$

$$(m^2 + 1) \sum_{i=1}^n x_i (mx_i - y_i + b) - m \sum_{i=1}^n (mx_i - y_i + b)^2 = 0$$

$$(m^2 + 1) \sum_{i=1}^n x_i (mx_i - y_i + b) = m \sum_{i=1}^n (mx_i - y_i + b)^2 \quad (3.3)$$

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

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

Define $X_i = x_i - \bar{x}$ and $Y_i = y_i - \bar{y}$.

$$(m^2 + 1) \sum_{i=1}^n (X_i + \bar{x})(mX_i - Y_i) = m \sum_{i=1}^n (mX_i - Y_i)^2$$

$$(m^2 + 1) \sum_{i=1}^n (mX_i^2 + m\bar{x}X_i - X_iY_i - \bar{x}Y_i) = m \sum_{i=1}^n (m^2X_i^2 - 2mX_iY_i + Y_i^2)$$

$$m^3 \sum_{i=1}^n X_i^2 + m \sum_{i=1}^n X_i^2 + m^3 \bar{x} \sum_{i=1}^n X_i + m\bar{x} \sum_{i=1}^n X_i - m^2 \sum_{i=1}^n X_iY_i - \sum_{i=1}^n X_iY_i - m^2 \bar{x} \sum_{i=1}^n Y_i + \bar{x} \sum_{i=1}^n Y_i$$

$$= m^3 \sum_{i=1}^n X_i^2 - 2m^2 \sum_{i=1}^n X_iY_i + m \sum_{i=1}^n Y_i^2 \quad (3.4)$$

Equation 3.4 can be simplified through cancellation and by the following property of the summation of centered coordinates.

$$\begin{aligned}
 \sum_{i=1}^n X_i &= \sum_{i=1}^n (x_i - \bar{x}) & \sum_{i=1}^n Y_i &= \sum_{i=1}^n (y_i - \bar{y}) \\
 &= \sum_{i=1}^n x_i - n\bar{x} & &= \sum_{i=1}^n y_i - n\bar{y} \\
 &= \sum_{i=1}^n x_i - n\left(\frac{1}{n} \sum_{i=1}^n x_i\right) & &= \sum_{i=1}^n y_i - n\left(\frac{1}{n} \sum_{i=1}^n y_i\right) \\
 &= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i & &= \sum_{i=1}^n y_i - \sum_{i=1}^n y_i \\
 &= 0 & &= 0
 \end{aligned}$$

With this substitution, equation 3.4 can be solved for m by using the quadratic equation, giving two slope equations for the alternative regression line.

$$\begin{aligned}
 m \sum_{i=1}^n X_i^2 - m^2 \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i Y_i &= -2m^2 \sum_{i=1}^n X_i Y_i + m \sum_{i=1}^n Y_i^2 \\
 m^2 \left(\sum_{i=1}^n X_i Y_i \right) + m \left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2 \right) &+ \left(-\sum_{i=1}^n X_i Y_i \right) = 0 \\
 m = \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 \pm \sqrt{\left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2 \right)^2 + 4 \left(\sum_{i=1}^n X_i Y_i \right)^2}}{2 \sum_{i=1}^n X_i Y_i} & \quad (3.5)
 \end{aligned}$$

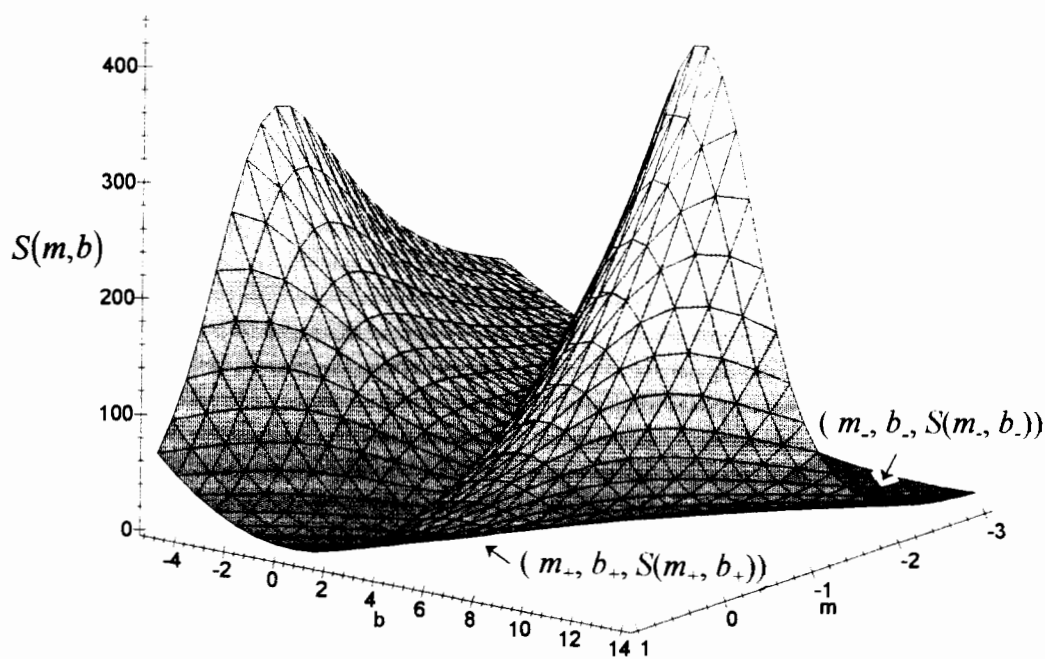
Two equations for the slope are produced, one with the square root term added, the other with the square root subtracted. Let m_+ refer to the equation with the addition option and m_- to the equation with the subtraction option.

$$m_+ = \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 + \sqrt{\left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2\right)^2 + 4\left(\sum_{i=1}^n X_i Y_i\right)^2}}{2\sum_{i=1}^n X_i Y_i}$$

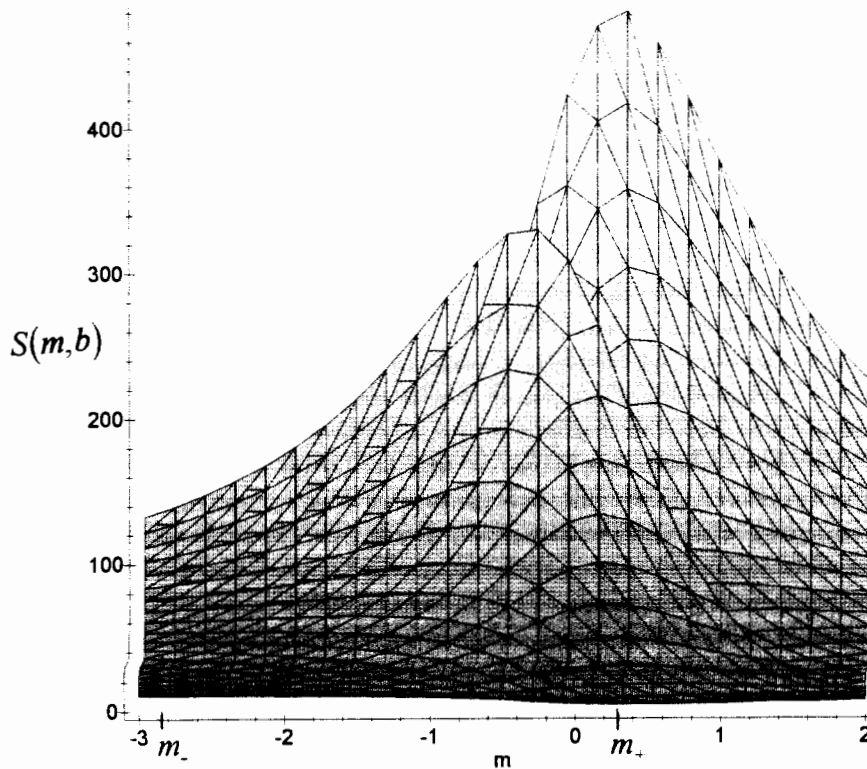
$$m_- = \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 - \sqrt{\left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2\right)^2 + 4\left(\sum_{i=1}^n X_i Y_i\right)^2}}{2\sum_{i=1}^n X_i Y_i}$$

The slope values, m_+ and m_- , produce two y -intercepts values, b_+ and b_- , respectively.

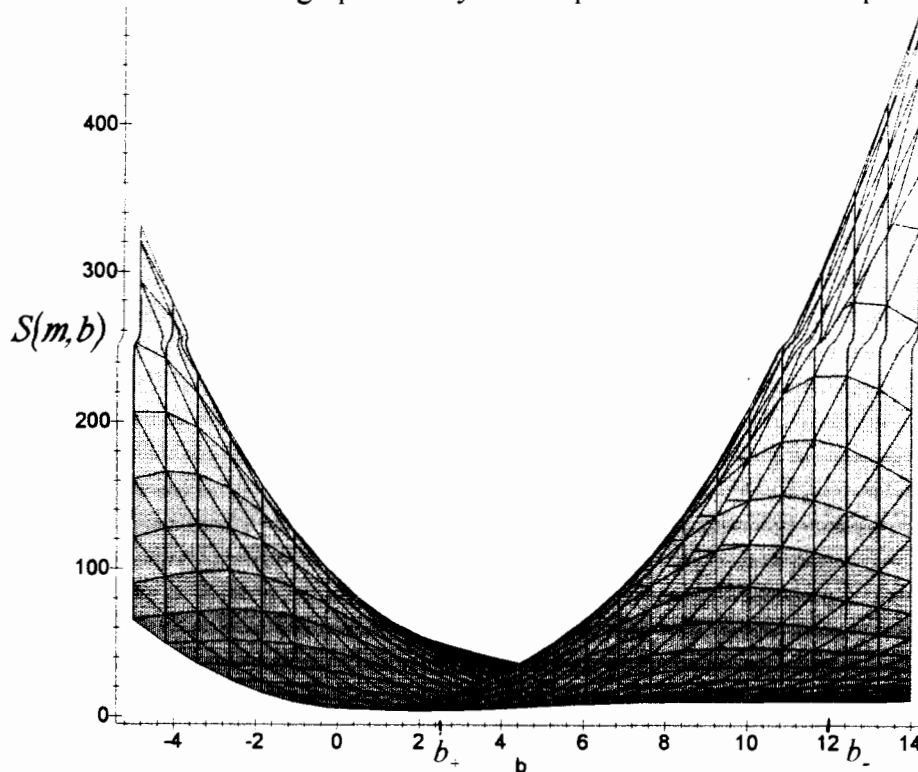
On the graph of $S(m,b)$ in three-space, two low points, or apparent minimums, are visible. These low points have coordinates $(m_+, b_+, S(m_+, b_+))$ and $(m_-, b_-, S(m_-, b_-))$. These are the critical points of the sum of squared errors function. These critical points are possible minimizing values of the sum of squared errors function. Graphs of the sum of squared errors function are possible when numerical values are given for the x and y -values. Although the following graphs are from the data set used in Example 1, presented in Appendix B, graphs for other examples are similarly shaped, but with the critical points in different locations.



The critical slope values are more apparent on the graph of the slope versus sum of squared errors. From this view, it is obvious that one of the slope values, either m_+ or m_- , has a smaller sum of squared errors.



The same affect occurs when the graph of the y-intercept versus the sum of squared errors.



To theoretically find the values of m and b that minimize the sum of squared errors, $S(m, b)$, the matrix of second partial derivatives, or Hessian matrix, of S must be determined.

$$D^2(S(m, b)) = \begin{bmatrix} \frac{\partial^2 S}{\partial b^2} & \frac{\partial^2 S}{\partial m \partial b} \\ \frac{\partial^2 S}{\partial b \partial m} & \frac{\partial^2 S}{\partial m^2} \end{bmatrix} = \begin{bmatrix} S_{bb} & S_{bm} \\ S_{mb} & S_{mm} \end{bmatrix} \quad (3.5)$$

We must find the values of m and b that will make this 2×2 symmetric matrix positive definite, but first the second partial derivatives of the sum of squared errors function must be found.

They can be determined by taking partial derivatives of the first partial derivatives of $S(m, b)$ with respect to m and b , found earlier. Since only the critical m and b values are possible extremas, we can use equalities 3.1 and 3.3 from the derivations of the critical points to make further simplifications.

$$\begin{aligned} S_{bb} &= \frac{\partial^2 S}{\partial b^2} = \frac{\partial}{\partial b} \left(\frac{2}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b) \right) \\ &= \left(\frac{2}{m^2 + 1} \right) \sum_{i=1}^n \left(\frac{\partial}{\partial b} (mx_i - y_i + b) \right) \\ &= \frac{2}{m^2 + 1} \sum_{i=1}^n 1 \\ &= \frac{2n}{m^2 + 1} \end{aligned} \quad (3.6)$$

$$\begin{aligned} S_{mm} &= \frac{\partial^2 S}{\partial m^2} = \frac{\partial}{\partial m} \left(\frac{2}{m^2 + 1} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 \right) \\ &= \frac{2}{m^2 + 1} \sum_{i=1}^n x_i^2 - \frac{4m}{(m^2 + 1)^2} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{4m}{(m^2 + 1)^2} \sum_{i=1}^n x_i (mx_i - y_i + b) \\ &\quad + \left(\frac{8m^2}{(m^2 + 1)^3} - \frac{2}{(m^2 + 1)^2} \right) \sum_{i=1}^n (mx_i - y_i + b)^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{8m}{(m^2+1)^2} \sum_{i=1}^n x_i (mx_i - y_i + b) + \left(\frac{8m^2}{(m^2+1)^3} - \frac{2}{(m^2+1)^2} \right) \sum_{i=1}^n (mx_i - y_i + b)^2 \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{8m^2}{(m^2+1)^3} \sum_{i=1}^n (mx_i - y_i + b)^2 + \left(\frac{8m^2}{(m^2+1)^3} - \frac{2}{(m^2+1)^2} \right) \sum_{i=1}^n (mx_i - y_i + b)^2 \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{(m^2+1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 \quad \text{above by (3.2)} \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \left(\frac{2}{(m^2+1)^2} \right) \left(\frac{(m^2+1)}{m} \right) \sum_{i=1}^n x_i (mx_i - y_i + b) \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{m(m^2+1)} \left(m \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n x_i \right) \quad \text{by (3.2)} \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{m(m^2+1)} \left(- \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n x_i \right) \\
&= \frac{2}{m(m^2+1)} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) \quad (3.7)
\end{aligned}$$

The second partial derivatives with respect to both m and b are equal, $S_{bm} = S_{mb}$ so only one of the two must be found. The first partial derivative of S with respect to b is simpler, so S_{mb} will be found.

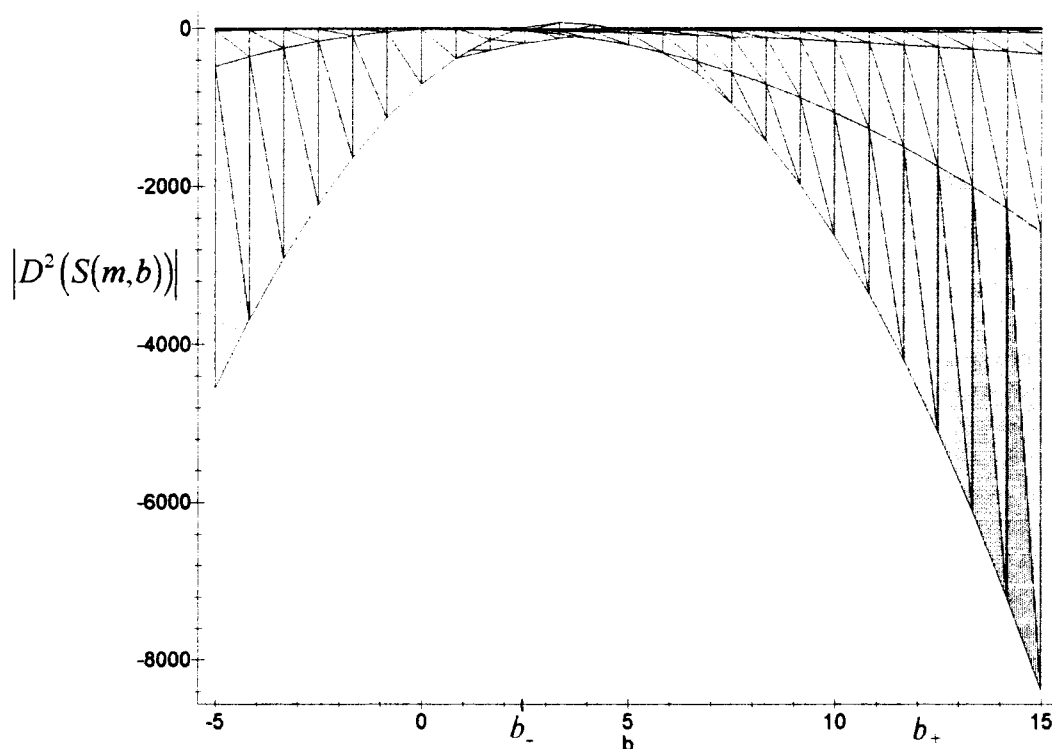
$$\begin{aligned}
S_{mb} &= \frac{\partial^2 S}{\partial m \partial b} = \frac{\partial}{\partial m} \left(\frac{2}{m^2+1} \sum_{i=1}^n (mx_i - y_i + b) \right) \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i - \frac{4m}{(m^2+1)^2} \sum_{i=1}^n (mx_i - y_i + b) \\
&= \frac{2}{m^2+1} \sum_{i=1}^n x_i \quad \text{by (3.1)} \quad (3.8)
\end{aligned}$$

These values make the second partial derivative matrix of $S(m, b)$:

$$D^2(S(m, b)) = \frac{2}{m^2 + 1} \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \frac{1}{m} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) \end{bmatrix} \quad (3.9)$$

For the values of m and b to minimize $S(m, b)$, the matrix must be positive definite at the values of m and b . The determinant of the first upper left-hand submatrix is always positive, because n , the number of data points, is always positive. The second condition requires that the determinant of the second partial derivative matrix be positive, $S_{bb}S_{mm} - S_{mb}^2 > 0$.

Graphing either the slope or y -intercept versus the determinant of the second partial derivative matrix shows that only a limited interval of slope or y -intercept values will produce a positive determinant. As in the previous graphs, this graph is possible only when x and y values are supplied. This particular graph is for the Example 1 data set presented in Appendix B.



In attempt to theoretically determine which slope-intercept pair should be used to minimize the sum of squared errors, the inequality of the determinant of the second partial derivative matrix must be reduced.

$$\begin{aligned}
& \frac{2}{m^2 + 1} \left(\frac{n}{m} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) - \left(\sum_{i=1}^n x_i \right)^2 \right) > 0 \\
& \frac{n}{m} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) > \left(\sum_{i=1}^n x_i \right)^2 \\
& \frac{n}{m} \sum_{i=1}^n x_i y_i - \frac{bn}{m} \sum_{i=1}^n x_i > \left(\sum_{i=1}^n x_i \right)^2 \\
& \frac{n}{m} \sum_{i=1}^n x_i y_i > \frac{bn}{m} \sum_{i=1}^n x_i + \left(\sum_{i=1}^n x_i \right)^2 \\
& \frac{n}{m} \sum_{i=1}^n x_i y_i > \left(\sum_{i=1}^n x_i \right) \left(\frac{bn}{m} + \sum_{i=1}^n x_i \right) \\
& \frac{n}{m} \sum_{i=1}^n x_i y_i > n \left(\sum_{i=1}^n x_i \right) \left(\frac{b}{m} + \frac{1}{n} \sum_{i=1}^n x_i \right) \\
& \frac{n}{m} \sum_{i=1}^n x_i y_i > n \left(\sum_{i=1}^n x_i \right) \left(\frac{b}{m} + \bar{x} \right) \\
& \frac{1}{m} \sum_{i=1}^n x_i y_i > \frac{n}{m} \left(\sum_{i=1}^n x_i \right) (b + m\bar{x}) \\
& \frac{1}{m} \sum_{i=1}^n x_i y_i > \frac{n}{m} \left(\sum_{i=1}^n x_i \right) (\bar{y}) \\
& \frac{n}{m} \sum_{i=1}^n x_i y_i > \frac{1}{m} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \tag{3.10}
\end{aligned}$$

Further simplification of equation 3.10 can not be done until the nature of m is determined. The slopes m_+ and m_- are shown to be negative inverses of each other, so the lines they produce will always be perpendicular. Variables a , b , and c are used for simplification.

$$a = \sum_{i=1}^n X_i^2, \quad b = \sum_{i=1}^n Y_i^2, \quad \text{and} \quad c = \sum_{i=1}^n X_i Y_i$$

$$m_+ m_- = \left(\frac{b-a + \sqrt{(a-b)^2 + 4c^2}}{2c} \right) \left(\frac{b-a - \sqrt{(a-b)^2 + 4c^2}}{2c} \right)$$

$$m_+ m_- = \frac{(b-a)^2 - ((a-b)^2 + 4c^2)}{4c^2}$$

$$m_+ m_- = \frac{(a-b)^2 - (a-b)^2 - 4c^2}{4c^2}$$

$$m_+ m_- = -1$$

Since the product of the two real-valued slopes is negative, one must be positive and the other negative. This information is used to further simplify equation 3.10.

$$\begin{aligned} \text{If the slope is positive: } n \sum_{i=1}^n x_i y_i &> \sum_{i=1}^n x_i \sum_{i=1}^n y_i \\ \text{If the slope is negative: } n \sum_{i=1}^n x_i y_i &< \sum_{i=1}^n x_i \sum_{i=1}^n y_i \end{aligned} \quad (3.11)$$

All steps in the simplification are reversible, so the following statements are true:

$$\begin{aligned} \text{If } n \sum_{i=1}^n x_i y_i > \sum_{i=1}^n x_i \sum_{i=1}^n y_i, & \text{ then the positive slope minimizes } S(m, b). \\ \text{If } n \sum_{i=1}^n x_i y_i < \sum_{i=1}^n x_i \sum_{i=1}^n y_i, & \text{ then the negative slope minimizes } S(m, b). \end{aligned} \quad (3.12)$$

Since the upper left-hand submatrix of the second derivative matrix of $S(m, b)$ is always positive, the matrix of second partial derivatives will never be a negative-definite at the critical points, and therefore no critical values m and b can maximize $S(m, b)$. Logically this is true because a line can be set infinitely far from the data points, creating an infinitely large sum of squared errors. The second critical point (m, b) of the function S , which makes the determinant of $D^2(S(m, b))$

negative, will be a saddlepoint on the graph of $S(m,b)$, and give a sum of squared errors greater than the minimum.

Following the method presented will find the error-minimizing slope and y -intercept according the definition of the error from a data point to the regression line as the perpendicular distance. Using equations 3.2 and 3.5 and condition 3.12, the regression line that minimizes sum of squared errors according to the alternative definition can be found.

Working through the alternative method of least squares linear regression by hand may prove to be tedious. However, with the use of calculator or computer programs, this method is no more difficult than the traditional least squares method. A sample Maple V Release 4 program is presented in Appendix A, and examples with sample data sets are presented in Appendix B.

Chapter IV

Advantages and Disadvantages of the Alternative Method

The alternative method of linear regression presented in the preceding chapter has both its advantages and disadvantages when compared with the traditional least squares linear regression method. As seen in Appendix B, the two methods of least squares linear regression produce similar results, although the lines are visibly different. Several issues should first be considered before comparing the accuracy of the two regression methods

Unlike the traditional method of least squares linear regression, the alternative method cannot produce horizontal lines, lines of form $y = b$ that have a slope of zero. Neither the traditional method nor the alternative method of regression can produce vertical lines, lines of form $x = c$. When points lie exactly on a line, both regression methods produce the same line (the line that all the points are on) so both have a sum of squared errors equal to zero.

The traditional method of least squares linear regression is simpler; its slope formula is much shorter than the formula to calculate the slope with the alternative method. The traditional method is best to use if calculating the regression line by hand. However, with the use of calculator and computer programs, the alternative method is no more difficult.

The alternative method of least squares linear regression destroys independent-dependent variable correlation. In the traditional linear regression method, error is present only in the y -variable, which is referred to as the dependent variable. This assumption is valuable in many practical applications, where it is useful to set one variable as faultless, and consider all error to be in the dependent variable. The alternative linear regression method no longer assumes the x

variable to be error-free. The alternative method allows for the presence of error in both variables by measuring the distance from a point to the regression line perpendicularly; the distance line can range from nearly vertical, like the traditional method, to almost horizontal, depending on the arrangement of the data points.

The sum of squared errors of the alternative regression line is always less than or equal to that of the traditional line. This fact makes the alternative least squares linear regression method appear to be superior. The least squares regression line is often called the "best fit" line because it is the line that has the minimum sum of squared errors. If comparing the sum of squared errors for each method is the sole means of determining whether the alternative or traditional regression line is more accurate, it is easy to assume that the alternative regression line is "better" since it has a smaller sum of squared errors than the traditional regression line's minimum sum of squared errors. Depending on which qualities are considered to be most important, the alternative method of least squares linear regression may be a better choice. However, we have not yet shown that the alternative method of least squares linear regression is necessarily more accurate.

Chapter V

Accuracy Comparison

Although the alternative method of least squares linear regression gives a sum of squared errors, less than or equal to that of the traditional method, these two values are not actually comparable. Because the distance from a point to a line, or error, is defined differently in each method, each value of the sum of squared errors describes something different. Comparing the sum of squared errors for the alternative regression line and the traditional regression line can be likened to comparing apples and oranges. To find which regression line is more accurate, a different method must be employed.

A method to find the area between two lines or functions is presented in integral calculus. To find the area between two functions, $f(x)$ and $g(x)$, between the x -values of a and b , the integral from a to b is taken of the difference of $f(x)$ and $g(x)$:

$$\int_a^b (f(x) - g(x)) dx.$$

Precaution must be taken with this formula, because a negative area may be given if $f(x)$ is not always greater than $g(x)$ from a to b . One alternative to guarantee a positive area is to use the absolute value function inside the integral:

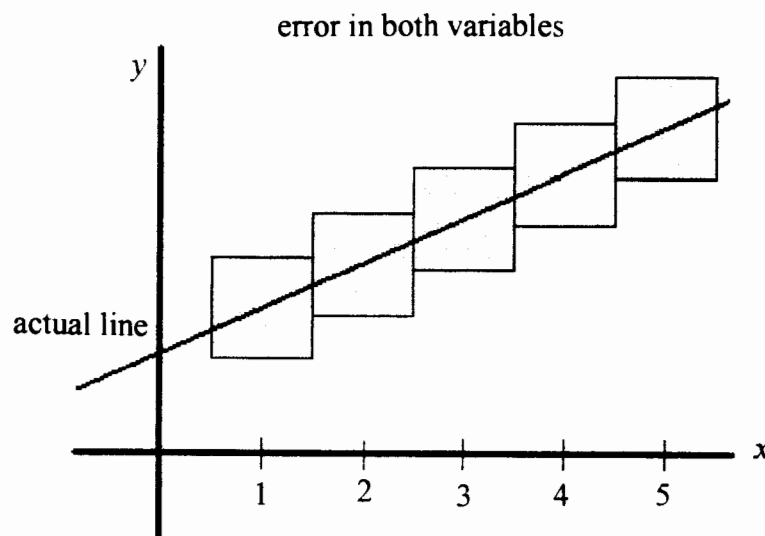
$$\int_a^b |f(x) - g(x)| dx.$$

Another alternative is to square the difference.

$$\int_a^b (f(x) - g(x))^2 dx.$$

The method of squaring the difference does not give the true area between the two functions, but can be used for comparison.

The method used to compare the accuracy of the alternative and the traditional regression line involves randomly choosing a line, which is referred to as the actual line, and plotting points randomly about this line. The slope of the actual line can range from $1/10$ to 3; the y -intercept can range from 0 to $3\frac{1}{3}$. These points are uniformly distributed within a unit square centered about a point on the actual line, allowing for error in both variables. These points are used to calculate both the alternative and the traditional regression lines. Using either integral method, the area between the traditional regression line and the actual line, and between the alternative regression line and the actual line is found and compared. The regression line with the smaller area difference is considered more accurate.



A Maple V Release IV program to produce the random actual line and produce the random errors for both variables is given in Appendix D. Also included in Appendix D is a multiple iteration version of the program, which allows the program to be executed 50 times,

while recording which regression method produces a more accurate line. After one thousand iterations of the accuracy comparison program, the alternative method of least squares linear regression was shown to be more accurate in 85.2% of the executions.

A similar experiment was conducted for points with random errors included only for the y -variable. The traditional method of least squares linear regression assumes data to be of this type. The Maple V Release IV program for the accuracy comparison program for data with error in only the y -variable is given in Appendix E. As in the Appendix D, a multiple iteration version is included. Again, the slope of the actual line can range from $1/10$ to 3; the y -intercept can range from 0 to $3^{1/3}$. Points are plotted randomly about the line, but only uniformly distributed only along a unit interval on the y -axis, centered about the actual y -value. Surprisingly, after one thousand iterations, the alternative method of least squares linear regression was shown to be more accurate in 89.7% of the executions.

The method of randomly plotting points about an actual line and applying regression methods in attempt to return to the actual line as closely as possible is similar to an experimental situation where a regression line is found using observational data and compared to the theoretical equation. The definition of the "more accurate" line as the regression line with the smaller difference in area from the actual line is the only method of accuracy comparison investigated. By this definition of accuracy, the alternative method of linear regression has been shown to be more accurate a greater percent of the time.

Chapter VI

Conclusion

The alternative method of least square linear regression has been shown to be a legitimate method of linear regression, and a valid option in the search for a more accurate method of linear regression. The preceding chapter presents a method of measuring and comparing the accuracy of both the traditional and the alternative method of linear regression. In the two methods of measuring the accuracy, the alternative method is shown to be more accurate a greater percentage of the time. However, the method of measuring the area difference is only one definition of accuracy. Other possible definitions of accuracy have not been investigated.

Throughout most of my research of the alternative method of least squares linear regression, I was unable to find any mention of a similar regression method that defined the error for a data point to be the perpendicular distance from the point to the regression line. Recently, however, I found its mention in several books. The majority of the work presented in this thesis was completed, and none of these sources presented the alternative method in the same manner as it is presented here. None of these sources mentioned a method of accuracy comparison.

A similar method of least squares appears in a book entitled, Numerical Mathematical Analysis, Sixth Edition, by James B. Scarborough. He presents a method of least squares that allows for error in both variables and defines the error for a point to be the perpendicular distance from the point to the line. In Scarborough's three-and-a-half page article, he writes, "The assumption [that only one variable is subject to error] is legitimate in most cases, for it is

usually possible and practicable to obtain the values of one variable more accurately than the other. If both variables are subject to errors of the same magnitude, the problem of finding the best values for empirical constants is more complicated except in those cases in which data can be plotted as a straight-line graph" (Scarborough 551). This statement supports the experimental results found in the preceding chapter; the alternative method of linear regression is more accurate more frequently than the traditional method in situations with error present in both variables, but makes the results of the situations with error present in only the y -variable even more surprising. "The line which best fits these points will evidently be that for which the sum of squares of the *perpendicular distances* from the points to it is a minimum. . . . This symmetrical form being used because both x and y are equally subject to error" (Scarborough 551). He notes that the perpendicular distance is used because both variables are subject to the same amount of error. In the accuracy comparison program used in the preceding chapter, both variables are equally likely to have an error ranging from zero to one-half of a unit.

Scarborough's derivation of this alternative method of least squares linear regression differs from the derivation presented earlier because he begins with a version of the standard form of the line, $ax + by + 1 = 0$, rather than the slope-intercept form, $y = mx + b$. The variable b has a different value in each equation, so the b variable in Scarborough's equation will be denoted as b' . Due to the difference in the initial equations, the resulting equations for the distance and sum of squares errors formulas are different.

$$d_i = \frac{ax_i + b'y_i + 1}{\sqrt{a^2 + b'^2}}$$

$$\sum_{i=1}^n d_i^2 = \frac{1}{a^2 + b'^2} [(ax_1 + b'y_1 + 1)^2 + (ax_2 + b'y_2 + 1)^2 + \cdots + (ax_n + b'y_n + 1)^2]$$

Scarborough's initial line equation converted into slope-intercept form is:

$$y = -\frac{a}{b'}x - \frac{1}{b'}, \text{ so the slope, } m, \text{ is } -\frac{a}{b'} \text{ and the } y\text{-intercept, } b, \text{ is } -\frac{1}{b'}.$$

For substitution purposes, $b' = -b$ and $a = -mb' = m \cdot b$. Converting Scarborough's distance and sum of squares errors formulas, similar results are obtained. Scarborough does not completely solve for the formulas for a and b' , but inputs the data values after finding the critical points, and solves for a and b' algebraically.

The similar method is also mentioned in Prediction and Improved Estimation in Linear Models, by John Bibby, as an alternative to the traditional, or ordinary, method of least squares. This method is called the "orthogonal regression procedure" (Bibby 42-43). Bibby provides a reference to M. G. Kendall's A Course in Multivariate Analysis (1957). Kendall discusses a general linear regression method in which the error for a point is measured perpendicularly, but allows for multivariate data, not restricting the observational data to only two variables, as done in the derivation of the alternative method of least squares linear regression.

These three books are the only sources found with reference to a method of least squares linear regression similar to the alternative method presented here. None fully presented and derived the least squares linear regression method as done here. The alternative method of least squares is shown to be a valid method of linear regression and, according to the area definition of accuracy, it is more likely to more accurately describe linearly distributed observational data than the traditionally accepted method of least squares linear regression.

Appendix A
Maple V Release 4 Template Program

the data

```
> X:=[x1,x2,x3,x4];
> Y:=[y1,y2,y3,y4];
```

procedure and calculation for the alternative least squares line

```
> n:=nops(X);
> average:=proc(X::list)
> local n, i, total;
> n:=nops(X);
> if n=0 then ERROR(`empty list`) fi;
> total:=add(i,i=X);
> total/n;
> end;
> avgx:=average(X);
 $\bar{x}$ 
> avgy:=average(Y);
 $\bar{y}$ 
> b:=avgy-m*avgx;
 $\sum_{i=1}^n x_i y_i$ 
> sxy:=add(X[i]*Y[i],i=1..n);
 $\sum_{i=1}^n x_i$ 
> sx:=add(i,i=X);
 $\sum_{i=1}^n y_i$ 
> sy:=add(i,i=Y);
 $\sum_{i=1}^n X_i^2$ 
> sX2:=add((X[i]-avgx)^2,i=1..n);
 $\sum_{i=1}^n Y_i^2$ 
> sY2:=add((Y[i]-avgy)^2,i=1..n);
 $\sum_{i=1}^n X_i Y_i$ 
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n);
 $m_+$ 
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
 $m_-$ 
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
> if type(n*sxy>sx*sy,boolean) then m:=ma elif
type(n*sxy<sx*sy,boolean) then m:=ms fi;
> evalf(b);
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
```

procedure and calculation for the traditional least squares line

$$\sum_{i=1}^n x_i^2$$

```
> sx2:=add((X[i])^2,i=1..n);
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
> bt:=evalf(avgy-mt*avgx);
> St:=evalf(add((Y[i] - mt*X[i] - bt)^2,i=1..n));

> new_line:=evalf(m)*x+evalf(b);new_error:=S;
> trad_line:=evalf(mt)*x+evalf(bt);trad_error:=St;

> with(plots):
> p:=plot(new_line,x=0..10,y=0..10,color=blue):
> t:=plot(trad_line,x=0..10,y=0..10,color=red):
> q:=plot([[x1,y1],[x2,y2],[x3,y3],[x4,y4]],style=POINT,
color=black,title=`Title`):
> display(p,q,t);
>
```


Appendix B

Example 1

```

> X:=[1,2,4,5];
X:=[1,2,4,5]
> Y:=[3,4,2,5];
Y:=[3,4,2,5]
> n:=nops(X);
n:=4
> average:=proc(X::list)
> local n, i, total;
> n:=nops(X);
> if n=0 then ERROR(`empty list`) fi;
> total:=add(i,i=X);
> total/n;
> end;
> avgx:=average(X);
avgx:=3
> avgy:=average(Y);
avgy:= $\frac{7}{2}$ 
> sx:=add(i,i=X);
sx:=12
> sy:=add(i,i=Y);
sy:=14
> b:=avgy-m*avgx;
> sxy:=add(X[i]*Y[i],i=1..n);
sxy:=44
> sX2:=add((X[i]-avgx)^2,i=1..n);
sX2:=10
> sY2:=add((Y[i]-avgy)^2,i=1..n);
sY2:=5
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n);
sXY:=2
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
ma:=.3507810593
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
ms:=-2.850781060
> mp:=`if`(ma>0,ma,ms);mn:=`if`(ms<0,ms,ma);
mp:=.3507810593
mn:=-2.850781060

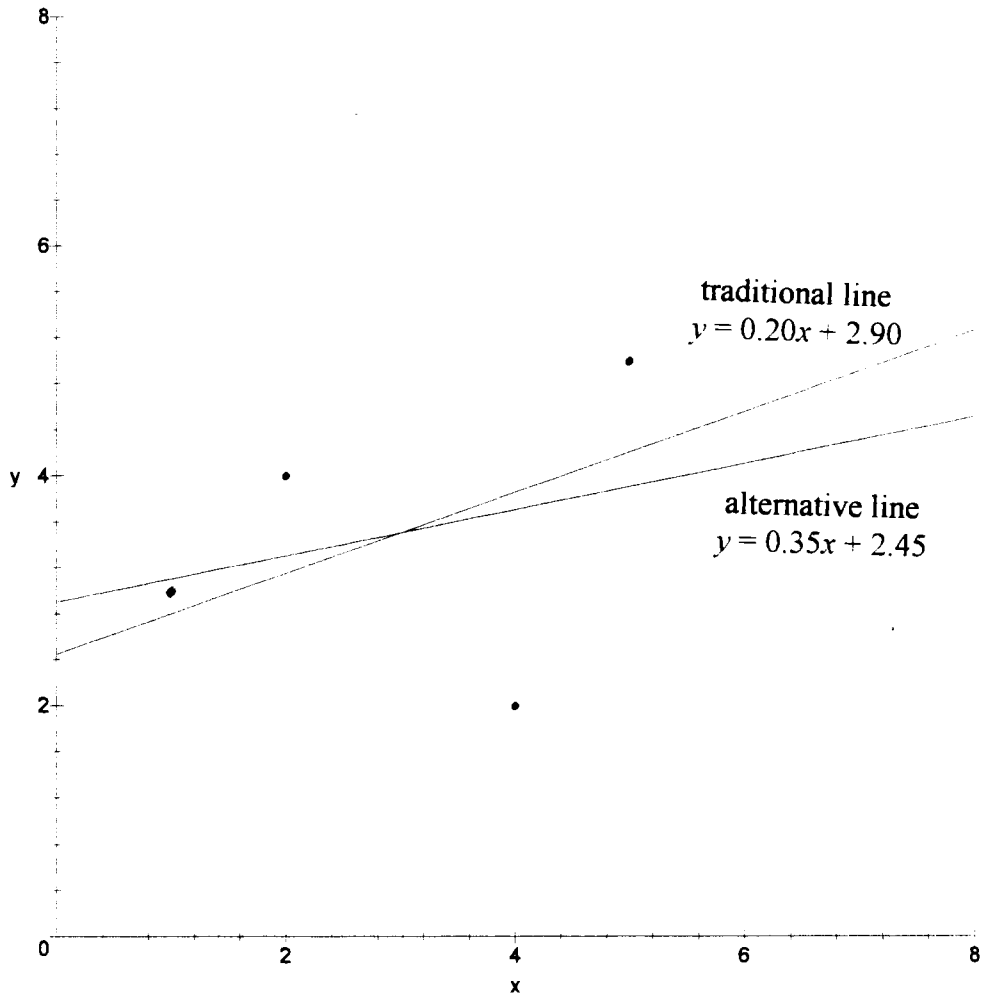
```

```

> m:=`if`(n*sxy>sx*sy,mp,mn);
                                     m := .3507810593
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
                                     S := 4.298437878
> b:=evalf(avgy-m*avgx);
                                     b := 2.447656822
> sx2:=add((X[i])^2,i=1..n);
                                     sx2 := 46
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
                                     mt := .2000000000
> bt:=evalf(avgy-mt*avgx);
                                     bt := 2.9000000000
> St:=evalf(add((Y[i] - mt*X[i] - bt)^2,i=1..n));
                                     St := 4.6000000000
> alt_line:=evalf(m)*x+evalf(b);alt_error:=S;
                                     alt_line := .3507810593 x + 2.447656822
                                     alt_error := 4.298437878
> trad_line:=evalf(mt)*x+evalf(bt);trad_error:=St;
                                     trad_line := .2000000000 x + 2.9000000000
                                     trad_error := 4.6000000000
> with(plots):
> p:=plot(alt_line,x=0..8,y=0..8,color = blue):
> r:=plot(trad_line,x=0..8,y=0..8,color = red):
> q:=plot([[1,3],[2,4],[4,2],[5,5]],style=POINT,color =
black,title=`Example 1 Graph`):
> display(p,r,q);

```

Example 1 Graph



Example 2

```

> X:=[0,1,2,3,4,5];
X:=[0, 1, 2, 3, 4, 5]
> Y:=[6,4,3,4,2,1];
Y:=[6, 4, 3, 4, 2, 1]
> n:=nops(X);
n:=6
> average:=proc(X::list)
> local n, i, total;
> n:=nops(X);
> if n=0 then ERROR(`empty list`) fi;
> total:=add(i,i=X);
> total/n;
> end;
> avgx:=average(X);
avgx:= $\frac{5}{2}$ 
> avgy:=average(Y);
avgy:= $\frac{10}{3}$ 
> sx:=add(i,i=X);
sx:=15
> sy:=add(i,i=Y);
sy:=20
> b:=avgy-m*avgx;
> sxy:=add(X[i]*Y[i],i=1..n);
sxy:=35
> sX2:=add((X[i]-avgx)^2,i=1..n);
sX2:= $\frac{35}{2}$ 
> sY2:=add((Y[i]-avgy)^2,i=1..n);
sY2:= $\frac{46}{3}$ 
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n);
sXY:=-15
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
ma:=-.9303824105
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
ms:=1.074826856

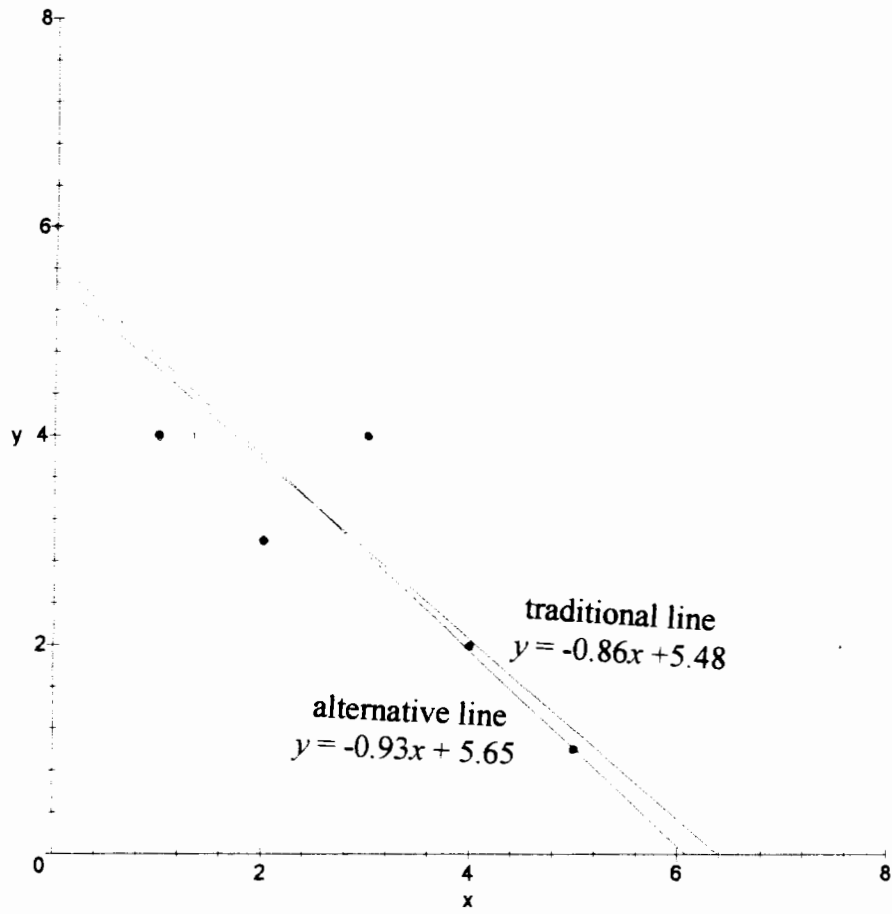
```

```

mp := 1.074826856
mn := -.9303824105
> m := `if` (n*sxy>sx*sy, mp, mn) ;
m := -.9303824105
> b := evalf (avgy - m*avgx) ;
b := 5.659289359
> S := evalf (add ( (m*X[i] - Y[i] + b)^2 / (m^2 + 1), i=1..n) ) ;
S := 1.377597178
> sx2 := add ( (X[i])^2, i=1..n) ;
sx2 := 55
> mt := evalf ( (n*sxy - sx*sy) / (n*sx2 - sx^2) ) ;
mt := -.8571428571
> bt := evalf (avgy - mt*avgx) ;
bt := 5.476190476
> St := evalf (add ( (Y[i] - mt*X[i] - bt)^2, i=1..n) ) ;
St := 2.476190476
> alt_line := evalf (m) *x + evalf (b) ; alt_error := S ;
alt_line := -.9303824105 x + 5.659289359
alt_error := 1.377597178
> trad_line := evalf (mt) *x + evalf (bt) ; trad_error := St ;
trad_line := -.8571428571 x + 5.476190476
trad_error := 2.476190476
> with (plots) :
> p := plot (alt_line, x=0..8, y=0..8, color = blue) :
> r := plot (trad_line, x=0..8, y=0..8, color = red) :
> q := plot ([ [0, 6], [1, 4], [2, 3], [3, 4], [4, 2], [5, 1] ],
style=POINT, color=black, title=`Example 2 Graph`):
> display (p, r, q) ;

```

Example 2 Graph



Appendix D

Accuracy Comparison Program for Data With Error in Both Variables

```
> err:=(rand(0..10))/20:
> m:=rand(1..30)/10:
> b:=rand(0..10)/3:
> ytrue:=m()*x+b();
```

$$y_{true} := \frac{7}{10}x + \frac{10}{3}$$

```
> X:=[1+(err())*(-1)^(rand()),2+(err())*(-1)^(rand()),3+(err())*(-1)^(rand()),4+(err())*(-1)^(rand()),5+(err())*(-1)^(rand())];
```

$$X := \left[\frac{17}{20}, \frac{21}{10}, \frac{7}{2}, \frac{18}{5}, 5 \right]$$

```
> Y:=[subs(x=1,ytrue)+(err())*(-1)^(rand()),subs(x=2,ytrue)+(err())*(-1)^(rand()),subs(x=3,ytrue)+(err())*(-1)^(rand()),subs(x=4,ytrue)+(err())*(-1)^(rand()),subs(x=5,ytrue)+(err())*(-1)^(rand())];
```

$$Y := \left[\frac{227}{60}, \frac{151}{30}, \frac{16}{3}, \frac{92}{15}, \frac{41}{6} \right]$$

```
> n:=nops(X);
```

$$n := 5$$

```
> sx:=add(i,i=X):
> avgx:=sx/n:
> sy:=add(i,i=Y):
> avgy:=sy/n:
> b:=avgy-m*avgx:
> sxy:=add(X[i]*Y[i],i=1..n):
> sX2:=add((X[i]-avgx)^2,i=1..n):
> sY2:=add((Y[i]-avgy)^2,i=1..n):
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n):
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mp:=`if`(ma>0,ma,ms):mn:=`if`(ms<0,ms,ma):
> m:=`if`(n*sxy>sx*sy,mp,mn);
```

$$m := .7216616120$$

```
> b:=evalf(avgy-m*avgx);
```

$$b := 3.251131881$$

```
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
```

$$S := .2340791103$$

```
> sx2:=add((X[i])^2,i=1..n):
```

```
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
```

```

                                mt := .7048396734
> bt:=evalf(avgy-mt*avgx);
                                bt := 3.301765916
> St:=evalf(add((Y[i] - mt*X[i] - bt)^2,i=1..n));
                                St := .3531447921
> alt_line:=evalf(m)*x+evalf(b);alt_error:=S;
                                alt_line := .7216616120 x + 3.251131881
                                alt_error := .2340791103
> trad_line:=evalf(mt)*x+evalf(bt);trad_error:=St;
                                trad_line := .7048396734 x + 3.301765916
                                trad_error := .3531447921

                                accuracy comparison against actual line
> errorAlt2:=int((alt_line-ytrue)^2,x=0..5);
                                errorAlt2 := .008821054346
> errorTrad2:=int((trad_line-ytrue)^2,x=0..5);
                                errorTrad2 := .002139044314
> errorAlt:=int(abs(alt_line-ytrue),x=0..5);
                                errorAlt := .1717008340
> errorTrad:=int(abs(trad_line-ytrue),x=0..5);
                                errorTrad := .09734116750
> most_accurate:=`if`(errorAlt<errorTrad,alternative,traditional);
                                most_accurate := traditional
>

```


**Accuracy Comparison Program
Multiple Iteration Version**

```

> numA:=0:numT:=0:
> for i from 1 to 50 do
> err:=(rand(0..10))/20:
> m:=rand(1..30)/10:
> b:=rand(0..10)/3:
> ytrue:=m()*x+b():
> n:=nops(X);
> sx:=add(i,i=X):
> avgx:=sx/n:
> sy:=add(i,i=Y):
> X:=[1+(err()*(-1)^(rand()),2+(err()*(-1)^(rand()),3+(err()*(-1)
  ^ (rand()),4+(err()*(-1)^(rand()),5+(err()*(-1)^(rand()))]:
> Y:=[subs(x=1,ytrue)+(err()*(-1)^(rand()),subs(x=2,ytrue)+(err()*
  (-1)^(rand()),subs(x=3,ytrue)+(err()*(-1)^(rand()),subs(x=4,ytrue
  )+(err()*(-1)^(rand()),subs(x=5,ytrue)+(err()*(-1)^(rand()))]:
> avgy:=sy/n:
> b:=avgy-m*avgx:
> sxy:=add(X[i]*Y[i],i=1..n):
> sX2:=add((X[i]-avgx)^2,i=1..n):
> sY2:=add((Y[i]-avgy)^2,i=1..n):
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n):
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mp:='if`(ma>0,ma,ms):mn:='if`(ms<0,ms,ma):
> m:='if`(n*sxy>sx*sy,mp,mn):
> b:=evalf(avgy-m*avgx):
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n)):
> sx2:=add((X[i])^2,i=1..n):
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2)):
> bt:=evalf(avgy-mt*avgx):
> St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n)):
> alt_line:=evalf(m)*x+evalf(b):alt_error:=S:
> trad_line:=evalf(mt)*x+evalf(bt):trad_error:=St:
> errorAlt2:=int((alt_line-ytrue)^2,x=0..5):
> errorTrad2:=int((trad_line-ytrue)^2,x=0..5):
> errorAlt:=int(abs(alt_line-ytrue),x=0..5):
> errorTrad:=int(abs(trad_line-ytrue),x=0..5):
> most_accurate:='if`(errorAlt<errorTrad,alternative,traditional):
> numA:='if`(most_accurate=alternative,numA+1,numA):
> numT:='if`(most_accurate=traditional,numT+1,numT):
> od:

```

Appendix E

Accuracy Comparison Program for Data With Error in Only y-variable

```
> err:=(rand(0..10))/20:
> m:=rand(1..30)/10:
> b:=rand(0..10)/3:
> ytrue:=m()*x+b();
```

$$y_{true} := \frac{9}{10}x + 3$$

```
> X:=[1,2,3,4,5];
```

$$X := [1, 2, 3, 4, 5]$$

```
> Y:=[subs(x=1,ytrue)+(err()*(-1)^(rand() ), subs(x=2,ytrue)+(err()*(-1)^(rand() ), subs(x=3,ytrue)+(err()*(-1)^(rand() ), subs(x=4,ytrue)+(err()*(-1)^(rand() ), subs(x=5,ytrue)+(err()*(-1)^(rand() ) ];
```

$$Y := \left[\frac{83}{20}, \frac{53}{10}, \frac{109}{20}, 7, \frac{36}{5} \right]$$

```
> n:=nops(X);
```

$$n := 5$$

```
> sx:=add(i,i=X):
```

```
> avgx:=sx/n:
```

```
> sy:=add(i,i=Y):
```

```
> avgy:=sy/n:
```

```
> b:=avgy-m*avgx:
```

```
> sxy:=add(X[i]*Y[i],i=1..n):
```

```
> sX2:=add((X[i]-avgx)^2,i=1..n):
```

```
> sY2:=add((Y[i]-avgy)^2,i=1..n):
```

```
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n):
```

```
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
```

```
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
```

```
> mp:=`if`(ma>0,ma,ms):mn:=`if`(ms<0,ms,ma):
```

```
> m:=`if`(n*sxy>sx*sy,mp,mn);
```

$$m := .8001501110$$

```
> b:=evalf(avgy-m*avgx);
```

$$b := 3.419549667$$

```
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
```

$$S := .2518291344$$

```
> sx2:=add((X[i])^2,i=1..n):
```

```
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
```

$$mt := .7800000000$$

```
> bt:=evalf(avgy-mt*avgx);
```

```

                bt := 3.480000000
> St:=evalf(add((Y[i] - mt*X[i] - bt)^2,i=1..n));
                St := .4090000000
> alt_line:=evalf(m)*x+evalf(b);alt_error:=S;
                alt_line := .8001501110 x + 3.419549667
                alt_error := .2518291344
> trad_line:=evalf(mt)*x+evalf(bt);trad_error:=St;
                trad_line := .7800000000 x + 3.480000000
                trad_error := .4090000000

                accuracy comparison against actual line
> errorAlt2:=int((alt_line-ytrue)^2,x=0..5);
                errorAlt2 := .2482266040
> errorTrad2:=int((trad_line-ytrue)^2,x=0..5);
                errorTrad2 := .3120000000
> errorAlt:=int(abs(alt_line-ytrue),x=0..5);
                errorAlt := .9132407633
> errorTrad:=int(abs(trad_line-ytrue),x=0..5);
                errorTrad := 1.020000000
> most_accurate:=`if`(errorAlt<errorTrad,alternative,traditional);
                most_accurate := alternative
>

```

Accuracy Comparison Program for Data With Error in Only y-variable
Multiple Iteration Version

```

> numA:=0:numT:=0:
> for i from 1 to 50 do
> err:=(rand(0..10))/20:
> m:=rand(1..30)/10:
> b:=rand(0..10)/3:
> ytrue:=m()*x+b():
> n:=nops(X);
> sx:=add(i,i=X):
> avgx:=sx/n:
> sy:=add(i,i=Y):
> X:=[1,2,3,4,5]:
> Y:=[subs(x=1,ytrue)+(err()*(-1)^(rand() ), subs(x=2,ytrue)+(err()*
(-1)^(rand() ), subs(x=3,ytrue)+(err()*(-1)^(rand() ), subs(x=4,ytrue
)+(err()*(-1)^(rand() ), subs(x=5,ytrue)+(err()*(-1)^(rand() ) ]:
> avgy:=sy/n:
> b:=avgy-m*avgx:
> sxy:=add(X[i]*Y[i],i=1..n):
> sX2:=add((X[i]-avgx)^2,i=1..n):
> sY2:=add((Y[i]-avgy)^2,i=1..n):
> sXY:=add((X[i]-avgx)*(Y[i]-avgy),i=1..n):
> ma:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> ms:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mp:=`if`(ma>0,ma,ms):mn:=`if`(ms<0,ms,ma):
> m:=`if`(n*sxy>sx*sy,mp,mn):
> b:=evalf(avgy-m*avgx):
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n)):
> sx2:=add((X[i])^2,i=1..n):
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2)):
> bt:=evalf(avgy-mt*avgx):
> St:=evalf(add((Y[i] - mt*X[i] - bt)^2,i=1..n)):
> alt_line:=evalf(m)*x+evalf(b):alt_error:=S:
> trad_line:=evalf(mt)*x+evalf(bt):trad_error:=St:
> errorAlt2:=int((alt_line-ytrue)^2,x=0..5):
> errorTrad2:=int((trad_line-ytrue)^2,x=0..5):
> errorAlt:=int(abs(alt_line-ytrue),x=0..5):
> errorTrad:=int(abs(trad_line-ytrue),x=0..5):
> most_accurate:=`if`(errorAlt<errorTrad,alternative,traditional);
> numA:=`if`(most_accurate=alternative,numA+1,numA):
> numT:=`if`(most_accurate=traditional,numT+1,numT):
> od:

```

Appendix E

Method More Accurate When Error is Present in Both Variables

Alternative	Traditional
42	8
41	9
45	5
42	8
42	8
41	9
45	5
42	8
45	5
43	7
39	11
44	6
41	9
45	5
40	10
42	8
41	9
45	5
42	8
45	5

Method More Accurate When Error is Present in Only y-variable

Alternative	Traditional
45	5
46	4
43	7
45	5
44	6
45	5
42	8
44	6
42	8
49	1
48	2
45	5
48	2
46	4
46	4
43	7
45	5
45	5
42	8
44	6

Appendix F**Bibliography**

Works Cited

- Bibby, John. Prediction and Improved Estimation in Linear Models. New York: John Wiley & Sons, Inc., 1977.
- Bühler, Walter K. Gauss: A Biographical Study. New York: Springer-Verlag, 1981.
- Daniel, Cuthbert and Fred S. Wood. Fitting Equations to Data. New York: John Wiley & Sons, Inc., 1971.
- Hall, Tord. Carl Friedrich Gauss. Trans. Albert Froderberg Cambridge, Massachussets: The MIT Press, 1970.
- Kendall, M. G. A Course in Multivariate Analysis. New York: Hafner Publishing Company, 1961.
- Legendre, Adrien Marie. "On the Method of Least Squares." Trans. Henry A. Ruger and Helen M. Walker. A Source Book in Mathematics. Ed. David Eugene Smith. New York: McGraw-Hill Book Company, Inc., 1929.
- Newman, James R. The World of Mathematics. Vol. 1. New York: Simon and Schuster, 1956.
- Scarborough, James B. Numerical Mathematical Analysis. 6th ed. Baltimore: The Johns Hopkins Press, 1966.
- Stigler, Stephen M. The History of Statistics. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1986.

Works Consulted

Anton, Howard. Elementary Linear Algebra. 7th ed. New York: John Wiley & Sons, Inc., 1994.

Marsden, Jerrold E. and Anthony J. Tromba. Vector Calculus. New York: W. H. Freeman and Company, 1996.

Taylor, Angus E. and W. Robert Mann. Advanced Calculus. 3rd ed. New York: John Wiley & Sons, Inc., 1983.

Thomas, George B., Jr. and Ross L. Finney. Calculus and Analytic Geometry. 8th ed. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1993.

Weiss, Neil A. Elementary Statistics, 3rd ed. Reading, Massachusetts: Addison Wesley Publishing Company, Inc., 1996.

Appendix H

Vita

Permanent Address

9670 Stonehaven
El Paso, TX 79925
(915) 592-6820

Degree

Bachelor of Science in Mathematics; May 1998
Minor: Physics

Educational Institutes Attended

Eastwood High School, El Paso, Texas
McMurry University, Abilene, Texas

Honors Tutorials

Spring 1997, MATH 4196*: Honors Tutorial: Calculus Based Statistics
Summer II 1997, MATH 4396*: Honors Tutorial: Various Methods of Linear Regression
Fall 1997, MATH 4397*: Senior Thesis: Applied Calculus Methods
Spring 1998, MATH 4397*: Senior Thesis: History of Mathematics

Organizations, Activities and Honors

Alpha Chi Honor Society, Fall 1996-Spring 1998; Kappa Mu Epsilon Mathematics Honor Society, Spring 1996 - Spring 1998, President, 1997-98, Secretary, 1996-97; McMurry Math Club, Fall 1995-Spring 1998; Society of Physics Students, Fall 1996-Spring 1998, Secretary-Treasurer 1997-98; Sigma Pi Sigma Physics Honor Society, Spring 1998; Gamma Sigma Women's Social Club, Spring 1996-Spring 1998, Secretary, 1996-97, Song Leader, 1996-97, 1997-98; Chanters, Spring 1996-Spring 1997; Students Against Drunk Driving, Fall 1995-Spring 1998, Treasurer, 1996-97, 1997-98; Galleon Literary Magazine Staff, 1996-98; Alpha Chi Quiz Bowl, winning student team 1996-97, 1997-98; Who's Who Among America's Colleges and Universities, 1997-98; Outstanding Junior in Mathematics, 1996-97; Outstanding Sophomore in Mathematics, 1995-96; Outstanding Physics Student, 1996-97; Outstanding Senior in Mathematics, 1997-98; Alpha Chi Region I scholarship recipient; Texas Tech University Graduate School Chancellor's Fellowship.

Experience

Academic Enrichment Center Tutor in Mathematics, Physics, and Core courses, Fall 1994-Spring 1998; AEC Calculus Study Sessions, Fall 1997-Spring 1998; National Mathematics Association of America Conference presentation, August 1997; Region I Alpha Chi Convention presentation, February 1998; Texas Section of Mathematics Association of America Conference presentation, March 1998.