BU-789-M
9/82

On the Geometric Motivation of Basic Regression Theory

CALVIN BERRY*

ABSTRACT

This paper is concerned with the geometric motivation of regression theory in introductory statistics courses. The development presented does not require any degree of geometric sophistication or familiarity with matrix computations. It is suggested that such a presentation may be easily and beneficially included in an introductory course.

KEY WORDS: Geometry; Orthogonal basis; Projection; Regression.

* Calvin Berry is a student in the field of statistics at Cornell University, 337 Warren Hall, Ithaca, NY 14853. The author wishes to thank Professor David G. Herr for introducing him to the coordinate-free approach to linear model theory. A note of thanks is also extended to Professor Foster B. Cady for letting the author present a lecture based on this paper to his regression methods course. This paper is BU-789-M in the Biometrics Unit Series, Cornell University, Ithaca, NY 14853.

On the Geometric Motivation of Basic Regression Theory

CALVIN BERRY*

1. INTRODUCTION

The statistical literature includes numerous papers and books which present regression theory and linear models from a geometric or coordinate-free approach. For a survey of the use of geometry in linear models the interested reader is referred to Herr (1980).

Introductory statistics courses often present regression theory from an algebraic point of view with little or no mention of the underlying geometric properties. Instructors who take this approach argue that students at this level do not possess the necessary background to understand a geometric approach to regression. The aim of this paper is to demonstrate that this is not the case. It will be shown that geometric motivation can be given in an introductory statistics course via ideas no more complex than one-dimensional projections and orthogonal bases.

At the intermediate or advanced level there are many textbooks and papers which present linear models and regression from a geometric point of view (Burdick *et al*., 1974; Seber, 1977; Kruskal, 1961). In his 1979 *American Statistician* article, Margolis advocates and demonstrates the use of perpendicular projections in basic statistics. The presentation in Margolis (1979) requires some familiarity with matrix operators in the context of multidimensional projections. This approach may not be accessible to introductory courses where the students are unfamiliar with matrices or linear algebra. In fact, the use of matrix operators may discourage teachers from presenting geometric motivation in introductory courses. Hence this paper presents a simple approach to geometric motivation more in line with the basic algebraic approach.

The material presented herein is hardly novel. The justification for this paper is the widespread lack of geometric motivation in introductory statistic courses. Hopefully, teachers will be motivated to reconsider the applicability of geometric motivation in introductory statistics.

The presentation of regression theory from a purely algebraic point of view may lead to algebraic derivations which seem mysterious to some students. Simple geometric motivation in the teaching of regression theory gives the student a better grasp of the basic concepts of regression theory. In particular, the student will be more satisfied with what may otherwise appear as obscure algebraic results. The basic geometric ideas involved are quite simple and do not require any particular degree of geometric sophistication. The material presented in this paper can be easily presented in one or two lectures.

2. BASIC FACTS

2.1 Definitions

This paper will assume some familiarity with the basic concepts of the n-dimensional Euclidean space, R^n. The concepts of vectors, vector addition and scalar multiplication will suffice. We begin with some definitions.

The usual Euclidean inner product of the vectors $\underset{\sim}{x} = (x_1, \cdots, x_n)'$ and $\underset{\sim}{y} = (y_1, \cdots, y_n)'$, denoted by $\underset{\sim}{x}'\underset{\sim}{y}$, is defined to be the sum of the componentwise products of the elements of $\underset{\sim}{x}$ and $\underset{\sim}{y}$, i.e., $\underset{\sim}{x}'\underset{\sim}{y} = \Sigma_{i=1}^{n} x_i y_i$.

The squared length of the vector $\underset{\sim}{x} = (x_1, \cdots, x_n)'$, denoted by $\|\underset{\sim}{x}\|^2$, is defined to be the inner product of $\underset{\sim}{x}$ with itself, i.e., $\|\underset{\sim}{x}\|^2 = \underset{\sim}{x}'\underset{\sim}{x} = \Sigma_{i=1}^{n} x_i^2$. The length of the vector $\underset{\sim}{x}$, denoted by $\|\underset{\sim}{x}\|$, is defined as the principal square root of the squared length of $\underset{\sim}{x}$.

The distance between the vectors $\underset{\sim}{x} = (x_1, \cdots, x_n)'$ and $\underset{\sim}{y} = (y_1, \cdots, y_n)'$, denoted by $\|\underset{\sim}{x} - \underset{\sim}{y}\|$, is defined as the length of their difference, i.e., $\|\underset{\sim}{x} - \underset{\sim}{y}\| = \sqrt{(\underset{\sim}{x} - \underset{\sim}{y})'(\underset{\sim}{x} - \underset{\sim}{y})} = \sqrt{\Sigma_{i=1}^{n}(x_i - y_i)^2}$.

The (linear) span of the set of vectors $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_k$, denoted by $[\{\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_k\}]$, is defined as the set of all linear combinations of the vectors $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_k$, i.e., $[\{\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_k\}] = \{\underset{\sim}{z} : \underset{\sim}{z} = \Sigma_{i=1}^{k}\lambda_i \underset{\sim}{z}_i \text{ for } \lambda_i \in R\}$. For $k = 1$ the span $[\{\underset{\sim}{z}_1\}]$ is a line and for $k = 2$ the span $[\{\underset{\sim}{z}_1, \underset{\sim}{z}_2\}]$ is the plane determined by the vectors $\underset{\sim}{z}_1$ and $\underset{\sim}{z}_2$. The ideas of length and distance are illustrated in Figure 1.

INSERT FIG. 1 HERE

2.2 Angles and Projections

A simple application of the law of cosines in Figure 1 yields

$$\|\underset{\sim}{x} - \underset{\sim}{y}\|^2 = \|\underset{\sim}{x}\|^2 + \|\underset{\sim}{y}\|^2 - 2\|\underset{\sim}{x}\|\|\underset{\sim}{y}\|\cos\theta \quad , \tag{1}$$

where θ is the angle between $\underset{\sim}{x}$ and $\underset{\sim}{y}$. Furthermore, on expanding the quadratic terms in $\|\underset{\sim}{x} - \underset{\sim}{y}\|^2$ we have

$$\|\underset{\sim}{x} - \underset{\sim}{y}\|^2 = \|\underset{\sim}{x}\|^2 + \|\underset{\sim}{y}\|^2 - 2\underset{\sim}{x}'\underset{\sim}{y} \quad . \tag{2}$$

Hence the following relationship between the angle θ and the inner product $\underset{\sim}{x}'\underset{\sim}{y}$ is established,

$$\underset{\sim}{x}'\underset{\sim}{y} = \|\underset{\sim}{x}\|\|\underset{\sim}{y}\|\cos\theta \quad . \tag{3}$$

Consider the vectors $\underset{\sim}{x} = (x_1, \cdots, x_n)'$ and $\underset{\sim}{y} = (y_1, \cdots, y_n)'$. Let $\bar{x} = (1/n)\Sigma_{i=1}^{n}x_i$, $\bar{y} = (1/n)\Sigma_{i=1}^{n}y_i$ and with some abuse of notation let $(\underset{\sim}{x} - \bar{\underset{\sim}{x}}) = (x_1 - \bar{x}, \cdots, x_n - \bar{x})'$ and $(\underset{\sim}{y} - \bar{\underset{\sim}{y}}) = (y_1 - \bar{y}, \cdots, y_n - \bar{y})'$. Now we can readily observe a useful relationship between ϕ, the angle between $(\underset{\sim}{x} - \bar{\underset{\sim}{x}})$ and $(\underset{\sim}{y} - \bar{\underset{\sim}{y}})$, and $r(x,y)$, the sample correlation coefficient of $\underset{\sim}{x}$ with $\underset{\sim}{y}$. From equation (3), it follows that

$$\cos\phi = \frac{(\underset{\sim}{x}-\bar{\underset{\sim}{x}})'(\underset{\sim}{y}-\bar{\underset{\sim}{y}})}{\|\underset{\sim}{x}-\bar{\underset{\sim}{x}}\|\|\underset{\sim}{y}-\bar{\underset{\sim}{y}}\|} = \frac{\Sigma_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i-\bar{x})^2\Sigma_{i=1}^{n}(y_i-\bar{y})^2}} = r(x,y) \quad . \qquad (4)$$

Taking advantage of the relationship given in equation (4), many properties of $r(x,y)$ follow immediately from the properties of the cosine function. For example, the function $\cos\phi$ ranges from one to negative one as ϕ ranges from 0° to 180°. Hence, $|r(x,y)| \le 1$ and if $r(x,y)$ is close to one in magnitude, then $(\underset{\sim}{x} - \bar{\underset{\sim}{x}})$ and $(\underset{\sim}{y} - \bar{\underset{\sim}{y}})$ are nearly colinear.

Next consider the projection of one vector onto the span of another. To avoid trivial cases assume that both vectors are non-null. The projection of $\underset{\sim}{y}$ onto the span of $\underset{\sim}{x}$, denoted by $\hat{\underset{\sim}{y}}(\underset{\sim}{x})$ or $P(\underset{\sim}{y} : [\{\underset{\sim}{x}\}])$, is illustrated in Figure 2 and is represented as $\hat{\underset{\sim}{y}}(\underset{\sim}{x}) = a\underset{\sim}{x}/\|x\|$, where $a = \|\hat{\underset{\sim}{y}}(\underset{\sim}{x})\|$.

INSERT FIG. 2 HERE

The triangle in Figure 2 is a right triangle, hence $\cos\theta = a/\|\underset{\sim}{y}\|$. Therefore, from equation (3) we have $\underset{\sim}{x}'\underset{\sim}{y} = \|\underset{\sim}{x}\|a$ and

$$\hat{\underset{\sim}{y}}(\underset{\sim}{x}) = \frac{\underset{\sim}{x}'\underset{\sim}{y}}{\|\underset{\sim}{x}\|^2}\underset{\sim}{x} \quad . \qquad (5)$$

2.3 Orthogonalization

Two non-null vectors $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are said to be orthogonal (perpendicular) if $\underset{\sim}{x}'\underset{\sim}{y} = 0$. Since $\underset{\sim}{x}'\underset{\sim}{y} = 0$ implies that $\cos\theta = 0$, where θ is the angle between $\underset{\sim}{x}$ and $\underset{\sim}{y}$, it follows that $\theta = 90°$ and hence $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are orthogonal. Thus for non-null $\underset{\sim}{x}$ and $\underset{\sim}{y}$ we have the following orthogonal decomposition of $\underset{\sim}{y}$,

$$y = \hat{y}(x) + [y - \hat{y}(x)] \quad . \tag{6}$$

This decomposition is illustrated in Figure 3. It is easily verified that $\hat{y}(x)$ is orthogonal to $y - \hat{y}(x)$.

INSERT FIG. 3 HERE

For a given set of linearly independent vectors, $\{z_1, \cdots, z_k\}$, it is often desirable to have an orthogonal basis for the span of these vectors. By repeated application of equation (6) we arrive at the orthogonal basis $\{z_1, z_2 - \hat{z}_2(z_1), \cdots, z_k - \hat{z}_k(z_1, \cdots, z_{k-1})\}$ for $[\{z_1, \cdots, z_k\}]$. The ith vector in this basis is the difference between z_i and the projection of z_i onto the span of the preceding vectors in the basis. The ith vector is obtained sequentially as

$$\begin{aligned} & z_i - \hat{z}_i(z_1, \cdots, z_{i-1}) \\ & \quad = z_i - \hat{z}_i(z_1) - \hat{z}_i[z_2 - \hat{z}_2(z_1)] - \cdots - \hat{z}_i[z_{i-1} - \hat{z}_{i-1}(z_1, \cdots, z_{i-2})] \quad . \end{aligned} \tag{7}$$

A bit of concentration on Figure 3 should convince the reader that this is an orthogonal set of vectors. Furthermore, each vector in this set is a linear combination of the vectors $z_1, \cdots, z_k$, hence it follows that

$$[\{z_1, \cdots, z_k\}] = [\{z_1, z_2 - \hat{z}_2(z_1), \cdots, z_k - \hat{z}_k(z_1, \cdots, z_{k-1})\}] \quad . \tag{8}$$

It is instructive to notice that for $k = 2$, equation (8) merely states that the plane determined by z_1 and z_2 is the same as the plane determined by z_1 and $z_2 - \hat{z}_2(z_1)$, as is clear from Figure 3.

3. REGRESSION AND LEAST SQUARES

3.1 Estimation

For the purposes of this paper it will suffice to consider a two-variable regression model. In vector form the general two-variable regression model is given by $\underset{\sim}{y} = \beta_1 \underset{\sim}{x}_1 + \beta_2 \underset{\sim}{x}_2 + \underset{\sim}{\epsilon}$, where y is a vector of n observations, β_1 and β_2 are fixed but unknown constants, $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are known vectors and $\underset{\sim}{\epsilon}$ is a random vector with expectation zero. Thus $\underset{\sim}{y}$ is observed as a vector in the span $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2, \underset{\sim}{\epsilon}\}]$ and we assume that the vector $E(\underset{\sim}{y})$, the expectation of $\underset{\sim}{y}$, lies in $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$. This regression set-up is illustrated in Figure 4.

INSERT FIG. 4 HERE

A first aim is to find a reasonable estimate of $E(\underset{\sim}{y})$. For reasons that will become clear we denote this estimate by $\hat{\underset{\sim}{y}}$. The criterion of least squares chooses $\hat{\underset{\sim}{y}}$ as the estimate which minimizes the squared distance between the vector of observations $\underset{\sim}{y}$ and the estimate $\hat{\underset{\sim}{y}}$, i.e., minimizes $||\underset{\sim}{y} - \hat{\underset{\sim}{y}}||^2$ over all values of $\hat{\underset{\sim}{y}}$.

Since we desire that $\hat{\underset{\sim}{y}}$ lie in $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$, there is a unique solution to this minimization problem. The solution is the vector $\hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$, the projection of $\underset{\sim}{y}$ onto $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$, since any other vector in $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$ must lie farther from $\underset{\sim}{y}$ in terms of squared distance. To see that $\hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ is the least squares estimate, consider Figure 5.

INSERT FIG. 5 HERE

Here $\hat{\underset{\sim}{y}} = \hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ and $\underset{\sim}{w}$ is an arbitrary vector in $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$. Now $||\underset{\sim}{y} - \underset{\sim}{w}||^2 = ||\underset{\sim}{y} - \hat{\underset{\sim}{y}}||^2 + ||\hat{\underset{\sim}{y}} - \underset{\sim}{w}||^2$, which is at least as large as $||\underset{\sim}{y} - \hat{\underset{\sim}{y}}||^2$ with equality if, and only if, $\underset{\sim}{w} = \hat{\underset{\sim}{y}}$. Hence $\hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ is the unique least squares estimate of $E(\underset{\sim}{y})$.

The estimate $\hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2) = P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}])$ can be represented by

$$\hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2) = \hat{\beta}_1 \underset{\sim}{x}_1 + \hat{\beta}_2 \underset{\sim}{x}_2 \tag{9}$$

for some constants $\hat{\beta}_1$ and $\hat{\beta}_2$ as is illustrated in Figure 6.

INSERT FIG. 6 HERE

A second aim is to determine the relationship between the regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ and the projection of $\underset{\sim}{y}$ onto $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$. Recall from equation (8) that $\{\underset{\sim}{x}_1, \underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}$ is an orthogonal basis for $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$, i.e., $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}] = [\{\underset{\sim}{x}_1, \underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}]$. Hence we have

$$\hat{\underset{\sim}{y}}(\underset{\sim}{x}_1, \underset{\sim}{x}_2) = b_1 \underset{\sim}{x}_1 + b_{2\cdot 1}[\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)] \tag{10}$$

for some constants b_1 and $b_{2\cdot 1}$ as is illustrated in Figure 7.

INSERT FIG. 7 HERE

Some reflection on Figure 7 will convince the reader that

$$P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]) = P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1\}]) + P(\underset{\sim}{y} : [\{\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}) \quad , \tag{11}$$

since $\{\underset{\sim}{x}_1, \underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}$ is an orthogonal basis for $[\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}]$.

Now from equations (10) and (11) we see that

$$b_1 \underset{\sim}{x}_1 = P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1\}]) = \frac{\underset{\sim}{y}' \underset{\sim}{x}_1}{\|\underset{\sim}{x}_1\|^2} \underset{\sim}{x}_1 \tag{12}$$

and

$$b_{2\cdot 1}[\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)] = P(\underset{\sim}{y} : [\{\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}]) = \frac{\underset{\sim}{y}'[\underset{\sim}{x}_2-\hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)]}{\|\underset{\sim}{x}_2-\hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\|^2}[\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)] \quad . \qquad (13)$$

Therefore, $b_1 = \underset{\sim}{y}'\underset{\sim}{x}_1/\|\underset{\sim}{x}_1\|^2$ and $b_{2\cdot 1} = \{\underset{\sim}{y}'[\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)]\}/\|\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\|^2$. Notice that $\hat{\beta}_1 = b_1 - b_{2\cdot 1}(\underset{\sim}{x}_2'\underset{\sim}{x}_1/\|\underset{\sim}{x}_1\|^2)$ and $\hat{\beta}_2 = b_{2\cdot 1}$.

In particular, consider the case when $\underset{\sim}{x}_1 = (1,\cdots,1)'$. Then

$$\hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1) = [(1/n)\Sigma x_{2i}]\underset{\sim}{x}_1 = (\bar{x}_2,\cdots,\bar{x}_2)' \quad ,$$

$$b_1 = \frac{\underset{\sim}{y}'\underset{\sim}{x}_1}{\|\underset{\sim}{x}_1\|^2} = (1/n)\sum_{i=1}^{n} y_i = \bar{y}$$

and

$$b_{2\cdot 1} = \frac{\underset{\sim}{y}'[\underset{\sim}{x}_2-\hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)]}{\|\underset{\sim}{x}_2-\hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\|^2} = \frac{\Sigma y_i(x_{2i}-\bar{x}_2)}{\Sigma(x_{2i}-\bar{x}_2)^2} \quad ,$$

which are easily recognized as the usual algebraic formulae for the sequential regression coefficients in the straight line regression model. Similarly, $\hat{\beta}_1 = b_1 - b_{2\cdot 1}(\underset{\sim}{x}_2'\underset{\sim}{x}_1/\|\underset{\sim}{x}_1\|^2) = \bar{y} - b_{2\cdot 1}[(1/n)\Sigma x_{2i}] = \bar{y} - b_{2\cdot 1}\bar{x}$ and $\hat{\beta}_2 = b_{2\cdot 1}$, which are the usual algebraic formulae for the partial regression coefficients in the straight line regression model.

3.2 Analysis of Variance

The analysis of variance that corresponds to the regression model considered above is merely a decomposition of the squared length of $\underset{\sim}{y}$, $\|\underset{\sim}{y}\|^2$. Notice that $\|\underset{\sim}{y}\|^2 = \Sigma_{i=1}^{n} y_i^2$ is the usual uncorrected total sum of squares.

Now $\underset{\sim}{x}_1$ is orthogonal to $\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)$; hence, it can be shown that equation (11) implies that

$$\|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1,\underset{\sim}{x}_2\}])\|^2 = \|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1\}])\|^2 + \|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}])\|^2 \quad . \qquad (14)$$

From this we derive the analysis of variance table which follows:

INSERT TABLE 1 HERE

Notice that the sum of the first two sums of squares is equal to $\|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1, \underset{\sim}{x}_2\}])\|^2$ by equation (14) and this is the usual model sum of squares with two d.f. provided $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are linearly independent.

Finally, for the case where $\underset{\sim}{x}_1 = (1, \cdots, 1)'$ we see that $\|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1\}])\|^2$ $= \|\bar{y}\underset{\sim}{x}_1\|^2 = n\bar{y}^2$ and

$$\|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}])\|^2 = \frac{[\Sigma y_i(x_{2i} - \bar{x}_2)^2]^2}{\Sigma(x_{2i} - \bar{x}_2)^2} ,$$

which are the usual algebraic formulae for the sums of squares for the mean and regression, respectively.

4. CONCLUSION

In this paper it has been demonstrated that regression theory can be motivated geometrically at an introductory level. The methods and concepts involved are no more complex than those used in the usual algebraic approach. This sort of geometric motivation need not be confined to regression theory and may be useful in the presentation of many statistical methods.

REFERENCES

BURDICK, DONALD S., HERR, DAVID G., O'FALLON, W. MICHAEL, and O'NEILL, BERNARD V. (1974), "Exact Methods in the Unbalanced, Two-Way Analysis of Variance – A Geometric View," Communications in Statistics, 3, 581-595.

HERR, DAVID G. (1980), "On the History of the Use of Geometry in the General Linear Model," The American Statistician, 34, 43-47.

KRUSKAL, WILLIAM H. (1961), "The Coordinate-Free Approach to Gauss-Markov Estimation and Its Application to Missing and Extra Observations," 4th Berkeley Symposium on Mathematical Statistics and Probability, 1, 435-451.

MARGOLIS, MARVIN S. (1979), "Perpendicular Projections and Elementary Statistics," The American Statistician, 33, 131-135.

SEBER, G. A. F. (1977), Linear Regression Analysis, New York: John Wiley and Sons.

List of Titles for Table and Figures

Table 1. The ANOVA as a Decomposition of $\|\underset{\sim}{y}\|^2$

Source	df	SS
$\underset{\sim}{x}_1$	1	$\|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_1\}])\|^2$
$\underset{\sim}{x}_2$	1	$\|P(\underset{\sim}{y} : [\{\underset{\sim}{x}_2 - \hat{\underset{\sim}{x}}_2(\underset{\sim}{x}_1)\}])\|^2$
Residual	n - 2	$\|\underset{\sim}{y} - \hat{\underset{\sim}{y}}\|^2$
Total	n	$\|\underset{\sim}{y}\|^2$

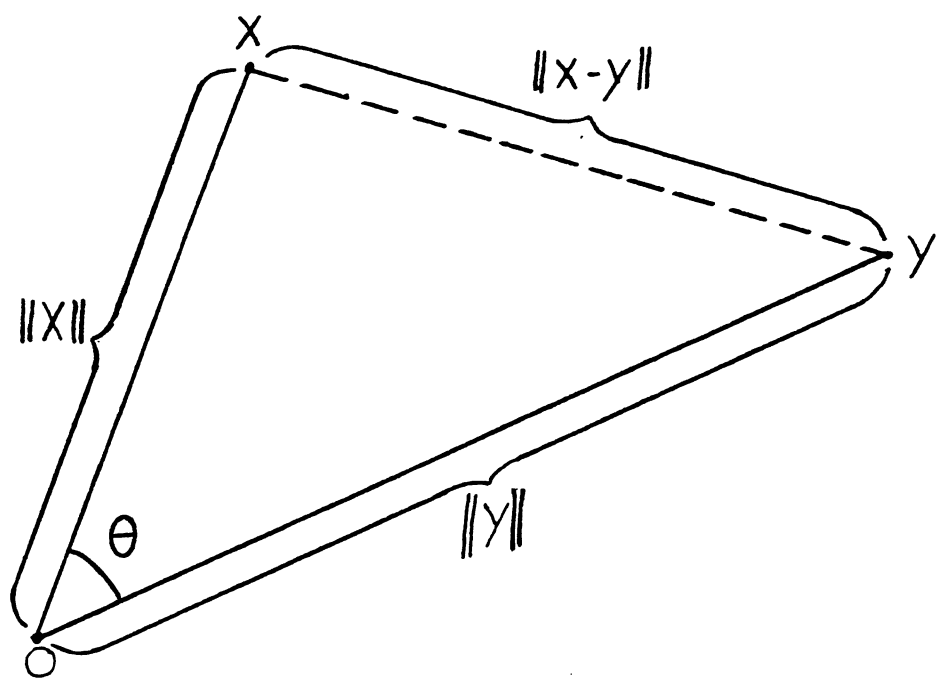

Figure 1. Length and Distance

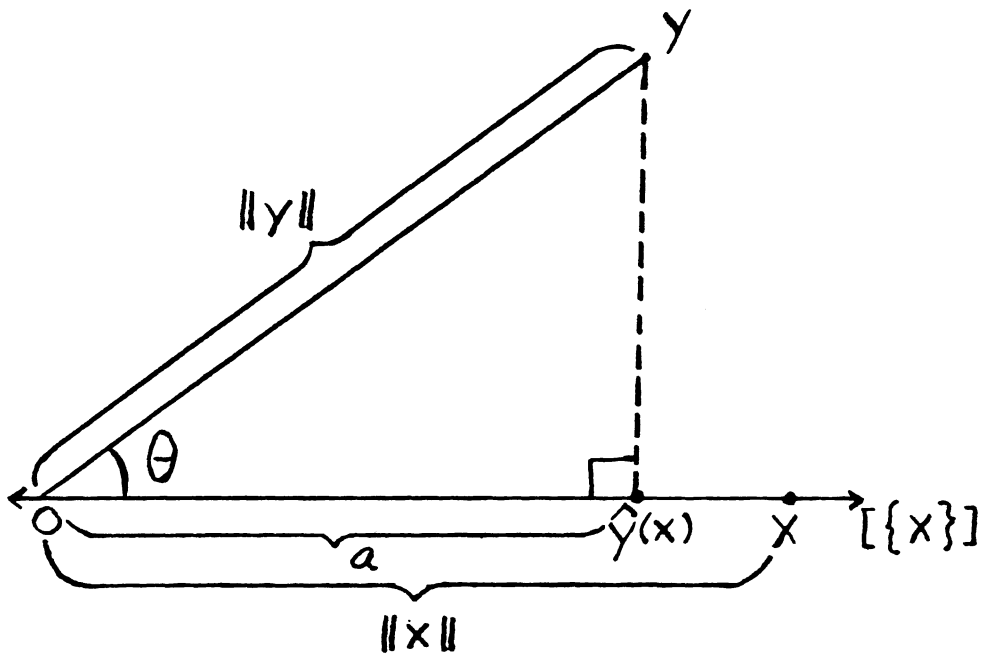

Figure 2. A One-Dimensional Projection

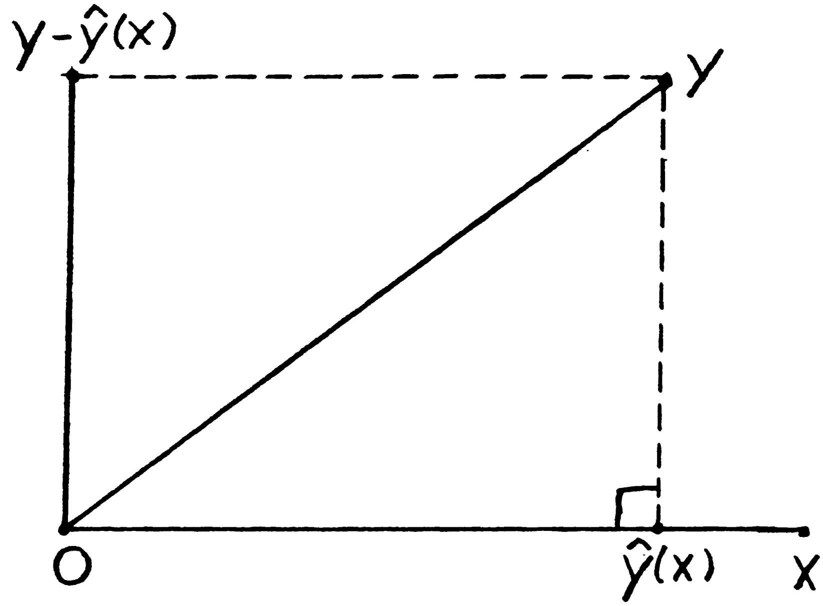

Figure 3. An Orthogonal Decomposition

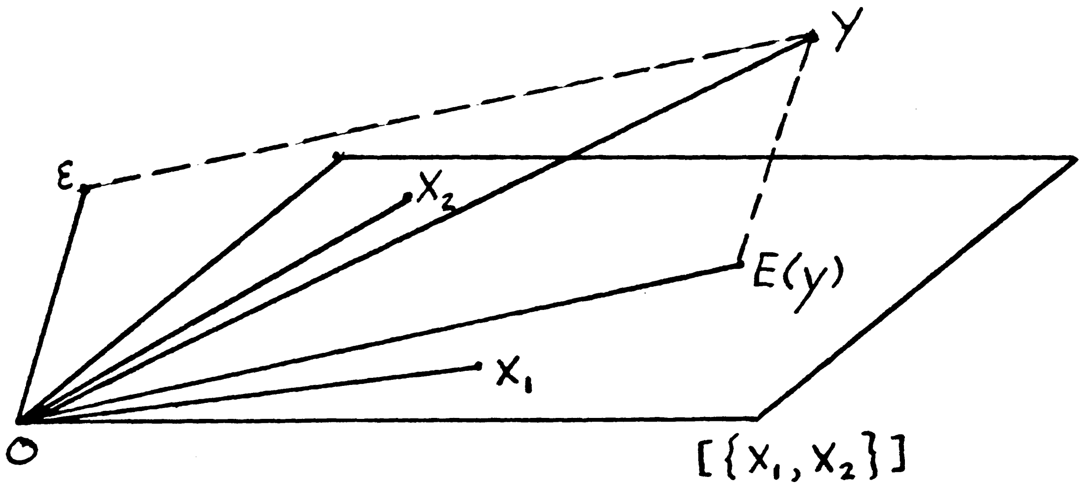

Figure 4. The Simple Linear Regression

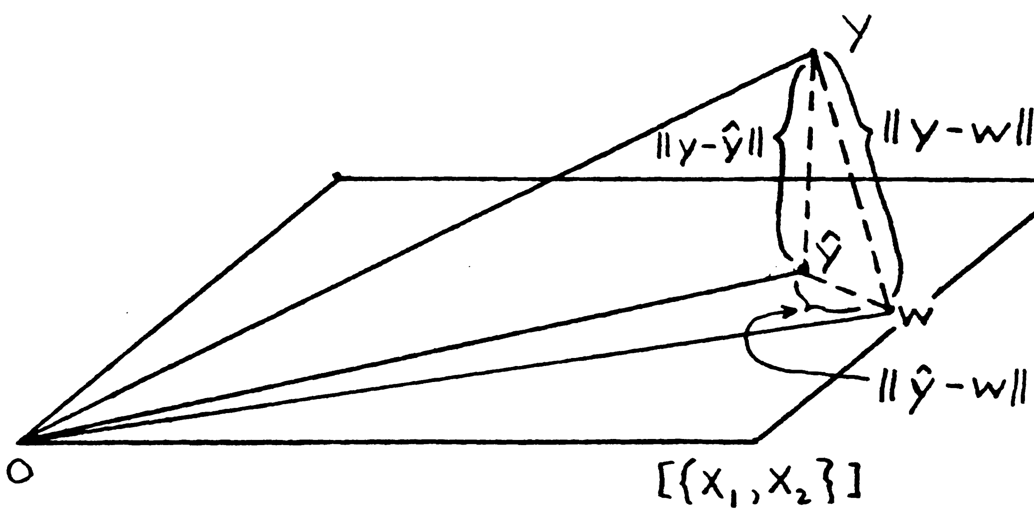

Figure 5. The Least Squares Estimator

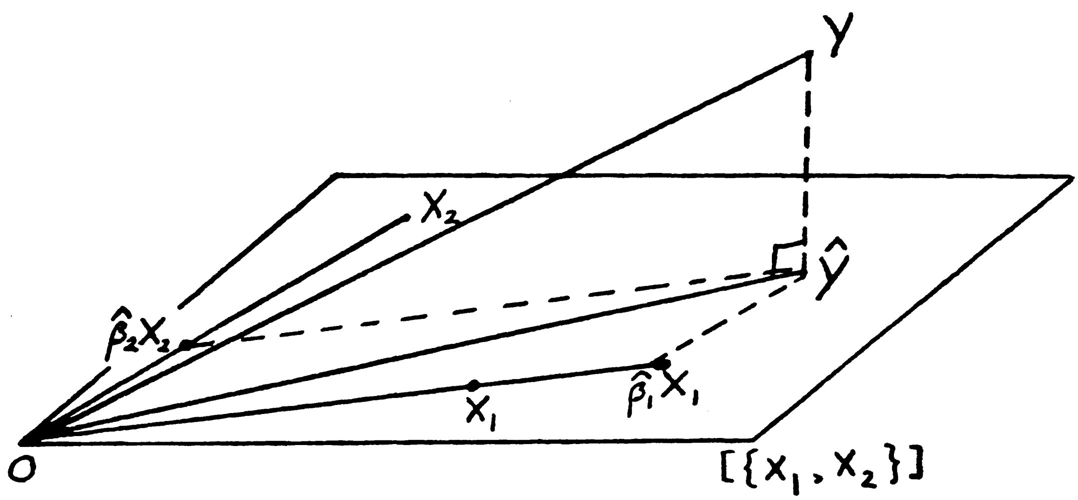

Figure 6. A Decomposition of the Least Squares Estimate

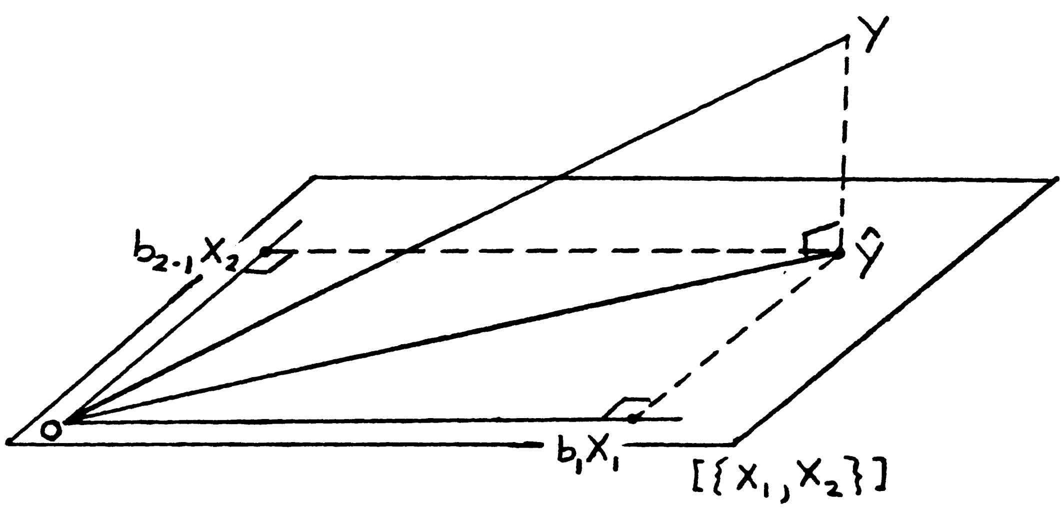

Figure 7. An Orthogonal Decomposition of the Least Squares Estimate