A Geometric Approach to Compare Variables in a Regression Model

Johan BRING

Geometry is a very useful tool for illustrating regression analysis. Despite its merits the geometric approach is seldom used. One reason for this might be that there are very few applications at an elementary level. This article gives a brief introduction to the geometric approach in regression analysis, and then geometry is used to shed some light on the problem of comparing the "importance" of the independent variables in a multiple regression model. Even though no final answer of how to assess variable importance is given, it is still useful to illustrate the different measures geometrically to gain a better understanding of their properties.

KEY WORDS: Coefficient of determination; Perpendicular projection; Relative importance; Standardized regression coefficients; t values.

1. INTRODUCTION

Geometry is a very powerful and illustrative tool to describe regression analysis. Despite its merits, geometry is seldom used in regression analysis, except for the use of scatterplots. By reviewing the history of the geometric approach, Herr (1980) tries to answer the question, "Why is the geometric approach so seldom used?" His answer can be summarized in three points;

1. Tradition of an algebraic approach is so strong that it will take a lot of effort and time to make a change.
2. The use by Fisher (1915) and Durbin and Kendall (1951) of the pure geometric approach convinced two generations of statisticians that geometry might be all right for the gifted few, but it would never do for the masses.
3. To fully appreciate the analytic geometric approach and to be able to use it effectively in research, teaching, and consulting requires that the statistician have an affinity for an talent in abstract thought. Dealing with abstractions is essentially a mathematical endeavor, and some statisticians eschew mathematics whenever possible (Herr 1980, p. 46).

In many regression studies one purpose is to compare the importance of different explanatory variables. Despite the fact that regression is an old and well-known technique, there is still debate about how to assess the importance of the explanatory variables. Several measures have been suggested and used, such as t values, standardized regression coefficient, elasticities, hierarchical partitioning, commonality analysis, increment in R^2, semi-partial correlations, etc. Thorough discussions of some of these measures can be found in Darlington (1968, 1990), Pedhazur (1982), and Bring (1994d).

The aim of this article is twofold: first, to show the power of the geometric approach in studying regression, and second, to use this technique to shed some light on the problem of how to compare variables in a regression model.

In Section 2 a short summary of the geometric approach is given. Good introductions on how to use the geometric approach in statistics are given by Margolis (1979), Bryant (1984), Saville and Wood (1986), Saville and Wood (1991), and Draper and Smith (1981, chap. 10.5). In Sections 3–5 some measures used for comparing variables will be given a geometric interpretation, and this is followed by a summary in Section 6.

2. A GEOMETRIC PRESENTATION OF MULTIPLE REGRESSION

In most introductory texts on simple linear regression geometry is used to gain a better understanding of the least squares method. The observations are plotted with the independent variable (x) on one axis and the dependent variable (y) on the other. This presentation will be called the *variable-axes presentation.* In more advanced textbooks the calculations are usually solved by using matrix algebra. To geometrically illustrate the matrix calculations the variable-axes presentation is not suitable. A better way is to use what I will call the *observation-axes presentation.* Consider the following data matrix:

$$\begin{bmatrix} y_1 & x_{11} & x_{12} & \dots & x_{1k} \\ y_2 & x_{21} & x_{32} & \dots & x_{2k} \\ y_3 & x_{31} & x_{32} & \dots & x_{3k} \\ \dots & \dots & \dots & \dots & \dots \\ y_n & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}.$$

It contains n observations on $k+1$ variables, one dependent (y) and k independent ($x_1 \dots x_k$). One way of thinking about such a data matrix is that we have $k+1$ variables spanning a $k+1$-dimensional space. In this space we have n observations. In simple linear regression, where $k=1$, it is easy to illustrate the data geometrically.

In matrix calculations each variable is considered as a vector in an n-dimensional space. Instead of thinking of n observations in a $k+1$-dimensional space we could think

Table 1. Hypothetical Data for Adam and Eve

	Height (cm)	Weight (kg)
Adam	182	85
Eve	164	52

Johan Bring is a Statistician with the Regional Oncology Center, University Hospital, S-751 85 Uppsala, Sweden. This article was written as a portion of the author's doctoral dissertation for the Department of Statistics at Uppsala University. The author thanks Axel Bring and the referees for their helpful comments.

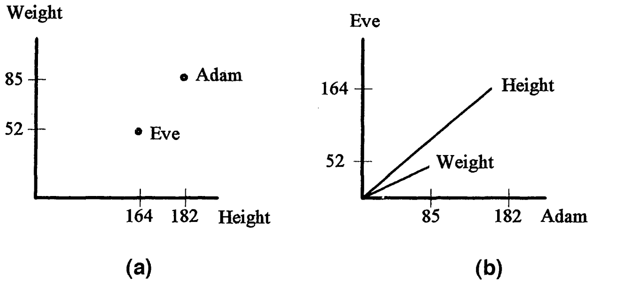

Figure 1. Two Geometrical Presentations of the Data in Table 1: (a) Variable-Axes and (b) Observation-Axes.

of $k+1$ vectors in an n-dimensional space. In other words, instead of having one axis for each variable we have one axis for each observation. To clarify the difference, let us consider a simple example, Table 1 and Figure 1.

In the variable-axes presentation observations are represented as points in a variable space, whereas in the observation-axes presentation variables are represented as vectors in observation space.

If we add a third person, we would, in the observation-axes presentation, get a third axis for that person, and the two vectors would now be in the three-dimensional space. Hence, independent of the number of observations, the data for height and weight will always be represented by two vectors. It is difficult to draw the vectors when there are more than three observations. However, as long as there are only two vectors, they span a two-dimensional subspace, and can therefore always be compared in a two-dimensional space.

From now on only the observation-axes presentation will be used. However, the base axes will be omitted and only the variable vectors will be displayed. To simplify the presentation we standardize the variables, $x_i = (x_i^+ - \bar{x}_i^+)/s_{x_i^+}\sqrt{(n-1)}$ $y = (y^+ - \bar{y}^+)/s_{y^+}\sqrt{(n-1)}$, where x^+ and y^+ are the original variables. By this standardization the vectors will have length 1, and no intercept is needed in the regression model.

When fitting a regression equation we get the estimated vector $\hat{\mathbf{y}} = \mathbf{XB}$, where $\mathbf{X}$ is the matrix with explanatory variables and $\mathbf{B}$ contains the standardized regression coefficients. In Figure 2 a model with one explanatory variable is illustrated.

The model vector, $\hat{\mathbf{y}}$, is the perpendicular projection of $\mathbf{y}$ on $\mathbf{x}$. To understand why this is the best estimate note that the squared length of $\mathbf{y}$, $\|\mathbf{y}\|^2 = \sum y_i^2$, is the total sum

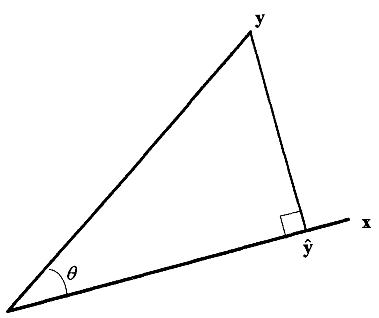

Figure 2. Linear Regression of y on x.

of squares, SS_{Tot}. The squared length of $\hat{\mathbf{y}}$, $\|\hat{\mathbf{y}}\|^2 = \sum \hat{y}_i^2 = SS_{Reg}$ and the squared length of the vector $\mathbf{y} - \hat{\mathbf{y}}$, $\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \sum (y_i - \hat{y}_i)^2 = SS_{Res}$. In other words, the observation vector is decomposed into a model vector and an error vector. The least squares method estimates $\mathbf{B}$ by minimizing SS_{Res}, and the shortest possible length of $\mathbf{y} - \hat{\mathbf{y}}$ is found by projecting $\mathbf{y}$ perpendicular on $\mathbf{x}$.

Other important measures in regression analysis are the correlation coefficients and the coefficient of determination, which can also be given geometrical representations.

$$R^2 = \frac{SS_{Reg}}{SS_{Tot}} = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} = \|\hat{\mathbf{y}}\|^2. \tag{1}$$

(The y vector is standardized to have length 1.)

$$r_{xy} = \begin{cases} \|\hat{\mathbf{y}}\| & \text{if } \theta \le 90^\circ \text{ or } \theta \ge 270^\circ \\ -\|\hat{\mathbf{y}}\| & \text{if } 90^\circ \le \theta \le 270^\circ \end{cases} \tag{2}$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. Another way of calculating the correlation is $r_{xy} = \cos(\theta)$.

If there are two independent variables, the estimated vector $\hat{\mathbf{y}}$ is found by projecting $\mathbf{y}$ perpendicular on the plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$; see Figure 3. The fit could now be measured either by the angle (θ) between $\mathbf{y}$ and the plane or by the length of $\hat{\mathbf{y}}$, $R^2 = \|\hat{\mathbf{y}}\|^2 = \cos^2(\theta)$.

In some of the figures below we will omit the y vector to make the figure easier to grasp. In these cases it is important to note that even without the y vector it is still possible to find the correlation between y and the x variables. The perpendicular projection of $\mathbf{y}$ on $\mathbf{x}_1$ is the same as the perpendicular projection of $\hat{\mathbf{y}}$ on $\mathbf{x}_1$; see Figure 3.

When there are two or more independent variables in the model it is often of interest to compare the relative importance of the independent variables. What is meant by importance varies from study to study, and therefore there are several measures available to measure relative importance. In the following sections some of the suggested measures will be given a geometrical interpretation.

3. STANDARDIZED REGRESSION COEFFICIENTS B_i

A common way to compare explanatory variables in the medical and social sciences is to use standardized regression coefficients (beta coefficients). These can be calculated in two different ways, yielding the same result: (1) standardize all variables to have mean zero and standard deviation 1, and then calculate ordinary regression coefficients; (2) use the unstandardized variables, and then multiply the regression coefficients by the ratio between the standard deviation of the respective independent variable and the standard deviation of the dependent variable.

How can these coefficients be illustrated geometrically? With two independent variables the regression equation is $\hat{\mathbf{y}} = B_1\mathbf{x}_1 + B_2\mathbf{x}_2$. The vector $\hat{\mathbf{y}}$ is a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$. The estimated coefficients are uniquely determined as long as the x variables span a two-dimensional space. Figure 4 illustrates the standardized regression coefficients for models with two and three explanatory vari-

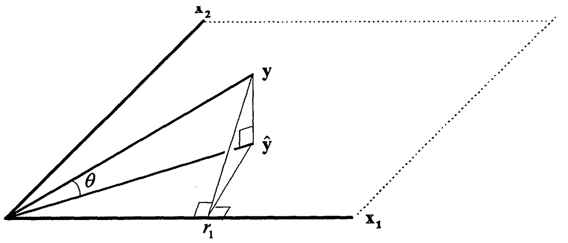

Figure 3. Regression With Two Explanatory Variables and Two Ways to Estimate the Correlation (r_1) Between x_1 and y.

ables. The standardized coefficient B_i is the signed distance traveled parallel with $\mathbf{x}_i$. Note that B_i can be negative.

Whether these coefficients are good indicators of variable importance is still under debate; see Pedhazur (1982), Greenland, Schlesselman, and Criqui (1986), Darlington (1990), and Bring (1994a, 1994b). Afifi and Clarke (1990), positive toward the use of standardized coefficients, give the following interpretation:

> The standardized coefficients of the various X variables can be directly compared in order to determine the relative contribution of each to the regression plane. The larger the magnitude of the standardized B_i the more X_i contributes to the prediction of Y (Afifi and Clarke 1990, p. 155).

When the explanatory variables are uncorrelated then the B_i's could be used as indicators of contribution to the prediction of y. In this case the x vectors are orthogonal, and by using the Phythagoras theorem we get the following partitioning of the model vector:

$$\|\hat{\mathbf{y}}\|^2 = B_1^2 + B_2^2 + \cdots + B_k^2. \qquad (3)$$

Note that in this case the standardized coefficients coincide with the correlation coefficients between y and x_i. If the explanatory variables are correlated the partitioning in (3) does not hold, as can easily be seen from Figure 4a. Then there seems to be no justification why the B_i's should represent relative contribution to the regression plane. Moreover, removing the variable with the smallest B_i does not necessarily cause the smallest reduction in R^2; see Bring (1994a).

4. *t VALUES*

When estimating a regression equation, a t value is usually calculated for each independent variable. These t values

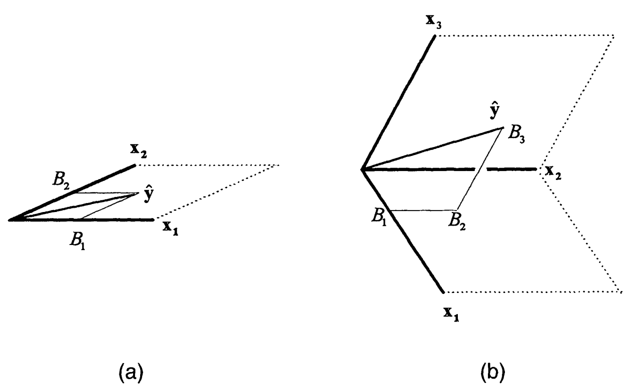

Figure 4. Standardized Regression Coefficients for Models With Two and Three Explanatory Variables (y Not in the Figure).

are often used to decide if a variable is significant or not. If significant variables are included in the regression model and the nonsignificant excluded, then the t values are in a sense used as indicators of importance. The significant variables are considered important and the nonsignificant unimportant. Why not use the t values to rank the variables regarding their relative importance, that is, the variable with the largest absolute t value is the most important?

The t values are related to R^2. For example, the squared t value for β_1 is

$$t_1^2 = \frac{(R^2_{1,2,3,\ldots k} - R^2_{2,3,\ldots k})}{(1 - R^2_{1,2,3,\ldots k})/(n-1-1)}, \qquad (4)$$

where $R^2_{1,2,3,\ldots k}$ is the coefficient of determination with all k variables included in the equation and $R^2_{2,3\ldots k}$ is the coefficient of determination with all variables except x_1 in the model. If we compute t values for two variables within the same regression model, by computing the ratio between them, the denominators are the same for both variables. Therefore this comparison is equivalent to comparing the reduction in R^2 caused by eliminating each variable and retaining all the other variables in the model:

$$\frac{t_1^2}{t_2^2} = \frac{(R^2_{1,2,3,\ldots k} - R^2_{2,3,\ldots k})}{(R^2_{1,2,3,\ldots k} - R^2_{1,3,\ldots k})}. \qquad (5)$$

This comparison can be given a geometric representation. Consider the three estimated vectors in Figure 5; $\hat{\mathbf{y}}$ is y regressed on x_1, x_2, and $x_3, \ldots \tilde{\mathbf{y}}$ is y regressed on x_1 and x_2, and $\ddot{\mathbf{y}}$ is y regressed on x_2 and x_3.

In the full model with x_1, x_2, and x_3, R^2 equals $\|\hat{\mathbf{y}}\|^2$. Without x_3 in the model $R^2 = \|\tilde{\mathbf{y}}\|^2$. The reduction in R^2 caused by excluding x_3 could be measured either by comparing the lengths of $\hat{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ or evaluating the length of the vector $\hat{\mathbf{y}} - \tilde{\mathbf{y}}$. The squared length of this vector represents the increase in SS_{Res} caused by excluding x_3. If x_1 is removed instead of x_3 the new R^2 will be $\| \ddot{\mathbf{y}} \|^2$. Comparing the t values for x_1 and x_3,

$$\frac{t_1^2}{t_3^2} = \frac{R^2_{123} - R^2_{23}}{R^2_{123} - R^2_{12}} = \frac{\|\hat{\mathbf{y}}\|^2 - \| \ddot{\mathbf{y}} \|^2}{\|\hat{\mathbf{y}}\|^2 - \|\tilde{\mathbf{y}}\|^2}$$
$$= \frac{\|\hat{\mathbf{y}} - \ddot{\mathbf{y}}\|^2}{\|\hat{\mathbf{y}} - \tilde{\mathbf{y}}\|^2} = \frac{\sin^2\varphi}{\sin^2\theta}. \qquad (6)$$

Hence comparing the t values of x_1 and x_3 is basically equivalent to comparing the distance from $\hat{\mathbf{y}}$ to the plane spanned by $\mathbf{x}_2$ and $\mathbf{x}_3$ to the distance between $\hat{\mathbf{y}}$ and the plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$. The geometrical presentation illustrates nicely that removing x_3 does not cause much loss in predictive power, while the contrary is true for x_1.

5. PARTITIONING R^2

The ability of the independent variables to explain the variation in y can be measured by the length of $\hat{\mathbf{y}}$ ($\|\hat{\mathbf{y}}\|^2 = R^2$). If it is possible to decide how much of the length of $\hat{\mathbf{y}}$ that is due to each individual independent variable, these proportions could be used as indicators of the variables'

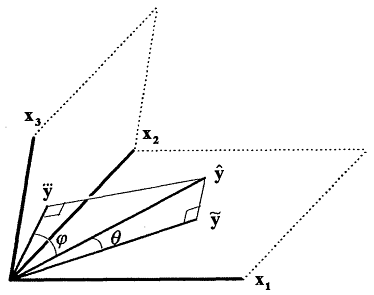

Figure 5. Three Regression Models ($\hat{\mathbf{y}}$, $\tilde{\mathbf{y}}$, and $\ddot{\mathbf{y}}$) Based on Different Sets of Independent Variables (y Not in the Figure).

relative importance. In other words how much of R^2 should be contributed to each of the independent variables?

Several measures of variable importance are calculated by partitioning R^2 between the independent variables. If the independent variables are uncorrelated this partitioning is straightforward:

$$R^2 = r_1^2 + r_2^2 + \cdots + r_k^2. \tag{7}$$

However, when the variables are correlated we get different partitionings depending on the choice of method. In the next subsections three methods for partitioning R^2 will be illustrated geometrically.

5.1 Stepwise

One way of selecting variables to be included in the regression equation is to use stepwise regression (forward); see Draper and Smith (1981). When this procedure is used most computer packages report the increment in R^2 for each successive variable included. Unfortunately, these increments are sometimes used as indicators of the independent variables' relative importance. By using geometry it can be demonstrated that this a rather arbitrary approach for the assessment of relative importance.

Assume that there are only two variables to choose between, x_1 and x_2. The stepwise procedure first selects the variable with the largest correlation with y. In geometric terms this means that the procedure compares the lengths of the vectors $\mathbf{a}$ and $\mathbf{c}$; Figure 6. If $\mathbf{a}$ is longer than $\mathbf{c}$, x_1 is selected first; if $\mathbf{c}$ is longer than $\mathbf{a}$, x_2 is selected first.

If x_1 is selected first, R^2 increases from 0 to $||\mathbf{a}||^2$. When x_2 is included as a second variable R^2 increases with $||\mathbf{b}||^2$. If x_2 were selected first, its contribution to R^2 would be $||\mathbf{c}||^2$, and the increase from x_1 would be $||\mathbf{d}||^2$ instead of $||\mathbf{a}||^2$. Only if $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal is $||\mathbf{a}||^2 = ||\mathbf{d}||^2$. Hence the more correlated the independent variables are, the more their assessed importance depends on the selection order. There is, with very few exceptions, no justification for selecting the variables in the order selected by the stepwise procedure. The stepwise method is therefore an unwise method for assessing relative importance; see, for example, Leigh (1988), Bring (1994c). Note that $||\mathbf{b}||^2$ and $||\mathbf{d}||^2$ correspond to the increase in SS_{Res} caused by excluding each variable from the full model as discussed with the t values. These quantities are also the squared semipartial correlations between x_2 and y and between x_1 and y, respectively.

With more than two variables the stepwise procedure increases the model space by one dimension at a time in the direction that increases the length of $\hat{\mathbf{y}}$ the most. An alternative method of assessing the independent variables' importance is to use average stepwise which will be discussed in Section 5.2.

5.2 Average Stepwise

Instead of using only the order selected by the stepwise procedure it is possible to consider all possible orderings. If there are k independent variables, there are $k!$ possible orderings. By considering all possible orderings, a variable's contribution to R^2 can be calculated as the average increment in R^2. Chevan and Sutherland (1991) give a good overview of this approach.

With the average stepwise approach the importance of x_1 in Figure 6 is equal to the average of $||\mathbf{a}||^2$ and $||\mathbf{d}||^2$, and the importance of x_2 is the average of $||\mathbf{b}||^2$ and $||\mathbf{c}||^2$. In terms of R^2 the contributions of x_1 and x_2 would be

average contribution:

$$\begin{aligned} x_1 &= (R_1^2 + (R_{1,2}^2 - R_2^2))/2 \\ x_2 &= (R_2^2 + (R_{1,2}^2 - R_1^2))/2. \end{aligned}$$

By using this approach the problem of multicollinearity is reduced due to the fact that all possible orderings of the variables are considered.

5.3 The Product Measure $B_i r_i$

The fact that $R^2 = B_1 r_1 + B_2 r_2 + \cdots + B_k r_k$ (proof in the appendix), where B_i is the standardized regression coefficient and r_i is the correlation between x_i and y, has led some researchers to suggest that each variable's con-

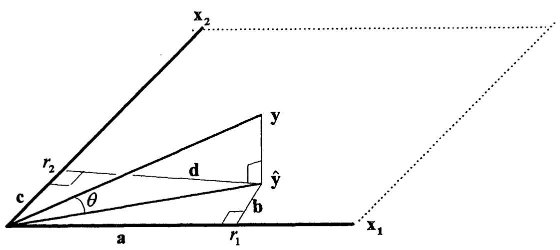

Figure 6. Selection of the First Variable by the Stepwise Regression Procedure.

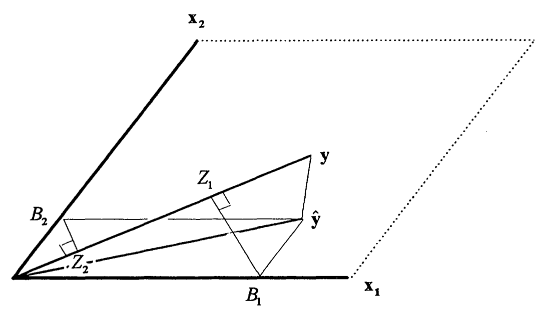

Figure 7. Geometrical Representation of the Product Measure, $Z_i = B_i r_i$.

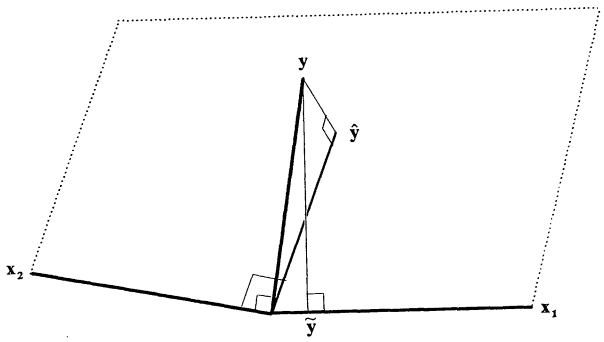

Figure 8. Regression Model With Two Independent Variables.

tribution to R^2 is equal to $B_i r_i$. Pratt (1987) gives a solid justification for the use of this measure.

This measure is not as easy to understand as the two previous measures. The main reason for the difficulty is that the importance of each variable is calculated as a product of two factors. However, by using inner products it is easy to find a geometrical representation of this measure.

$$\begin{aligned} B_i r_i &= \pm\|B_i \mathbf{x}_i\| \cos\theta = \pm\|B_i \mathbf{x}_i\| \cos\theta \|\mathbf{y}\| \\ &= \pm\langle B_{i1}\mathbf{x}_i, \mathbf{y}\rangle = Z_i \end{aligned} \tag{8}$$

where θ is the angle between $\mathbf{x}_i$ and $\mathbf{y}$ and $\langle\quad\rangle$ represents an inner product. Z_i is the component of $B_i\mathbf{x}_i$ in the direction of $\mathbf{y}$; see Figure 7.

Despite the ease in finding a geometrical interpretation, it is still difficult to comprehend the meaning of this measure. However, in Figure 8 an example is given where this measure gives a counterintuitive result.

The angle between $\mathbf{x}_2$ and $\mathbf{y}$ is 90°, which means that the correlation between these variables is zero. Hence the importance of x_2 in explaining R^2 is zero according to the product measure. Therefore x_1 is responsible for all the length of $\hat{\mathbf{y}}$. However, if we exclude x_2, x_1 would not explain much at all of the variation in y, only $\|\tilde{\mathbf{y}}\|^2$. Hence with the product measure, x_2's contribution to R^2 is zero and x_1's contribution to R^2 equals R^2. However, the inclusion of x_2 greatly increases R^2. Therefore it is not satisfactory that x_2's contribution should be zero.

Figure 8 also illustrates the interesting situation where $R^2 > r_1^2 + r_2^2$. For a discussion of when this occurs, see Bertrand and Holder (1988), Hamilton (1987), and Schey (1993). (Schey uses geometry to illustrate when this occurs.)

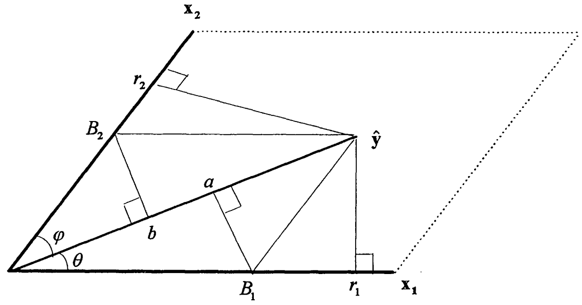

Figure A.1. Relationship Between R^2 and $B_1r_1 + B_2r_2$.

6. SUMMARY

Geometry is a very powerful tool for illustrating many statistical techniques. However, it requires an ability for abstract thought. To attain this ability requires training, and it is therefore important that there are illustrations available at a rather elementary level.

This article has given a brief introduction to the geometric approach to regression analysis, and I hope the presentation has been at a suitable level even for the reader unfamiliar with geometrical thinking. The geometric approach was also used to illustrate some measures that could be used for comparing the "importance" of the explanatory variables in multiple regression models. There is no unique definition of the concept importance, and no clear-cut answer can be given as to which measure to use. However, displaying these measures geometrically can increase the understanding of what they are measuring, and especially indicate situations when the measures are not suitable.

APPENDIX

Proof that $R_{12}^2 = B_1 r_1 + B_2 r_2$.

$$R^2 = \|\hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}\|(a+b)$$

where a is the part of $B_1\mathbf{x}_1$ parallel to $\hat{\mathbf{y}}$ and b is the part of $B_2\mathbf{x}_2$ parallel to $\hat{\mathbf{y}}$ (see Fig. A.1).

$$\begin{aligned} \|\hat{\mathbf{y}}\|(a+b) &= \|\hat{\mathbf{y}}\|(B_1 \cos\theta + B_2\cos\varphi) \\ &= \|\hat{\mathbf{y}}\|\cos\theta(B_1) + \|\hat{\mathbf{y}}\|\cos\varphi(B_2) \\ &= r_1 B_1 + r_2 B_2. \end{aligned}$$

The proof could easily be extended to more than two variables.

[Received June 1993. Revised September 1994.]

REFERENCES

Afifi, A. A., and Clarke, V. (1990), *Computer Aided Multivariate Analysis* (2nd ed.), New York: Van Nostrand Reinhold.

Bertrand, B., and Holder, R. (1988), "A Quirk in Multiple Regression: The Whole Can Be Greater Than Its Parts," *The Statistician*, 37, 371–374.

Bring, J. (1994a), "How to Standardize Regression Coefficients," *The American Statistician*, 48(3), 209–213.

——— (1994b), "Standardized Regression Coefficients and Relative Importance of Pain and Functional Disability to Patients with Rheumatoid Arthritis," (Letter to the Editor with Reply), *Journal of Rheumatology*, 21(9), 1774–1775.

——— (1994c), "Relative Importance of Factors Affecting Blood Pressure," (Letter to the Editor), *Journal of Human Hypertension*, 8, 297.

——— (1994d), "Variable Importance and Regression Modelling," Doctoral thesis, Department of Statistics, Uppsala University, Sweden.

Bryant, P. (1984), "Geometry, Statistics, Probability: Variation on a Common Theme," *The American Statistician*, 38, 38–48.

Chevan, A., and Sutherland, M. (1991), "Hierarchical Partitioning," *The American Statistician*, 45, 90–96.

Darlington, R. B. (1968), "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin*, 69, 161–182.

——— (1990), *Regression and Linear Models*, New York: McGraw-Hill.

Draper, N., and Smith, H. (1981), *Applied Regression Analysis* (2nd ed.), New York: John Wiley.

Durbin, J., and Kendall, M. G. (1951), "The Geometry of Estimation," *Biometrika*, 38, 150–158.

Fisher, R. A. (1915), "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," *Biometrika*, 10, 507–521.

Greenland, S., Schlesselman, J. J., and Criqui, M. H. (1986), "The Fallacy of Employing Standardized Regression Coefficients and Correlations as Measure of Effect," *American Journal of Epidemiology*, 123(2), 203–208.

Hamilton, D. (1987), "Sometimes $R^2 > r^2_{yx1} + r^2_{yx2}$, Correlated Variables Are Not Always Redundant," *The American Statistician*, 41, 129–132.

Herr, D. G. (1980), "On the History of the Use of Geometry in the General Linear Model," *The American Statistician*, 34, 131–135.

Leigh, J. P. (1988), "Assessing the Importance of an Independent Variable in Multiple Regression: Is Stepwise Unwise?," *Journal of Clinical Epidemiology*, 41(7), 669–677.

Margolis, M. S. (1979), "Perpendicular Projections and Elementary Statistics," *The American Statistician*, 33, 131–135.

Pedhazur, E. J. (1982), *Multiple Regression in Behavioral Research* (2nd ed.), New York: Holt, Rinehart & Winston.

Pratt, J. W. (1987), "Dividing the Indivisible: Using Simple Symmetry to Partition Variance Explained," in *Proceeding of the 2nd International Tampere Conference*, eds. T. Pukkila and S. Puntanen, University of Tampere, 245–260.

Saville, D. J., and Wood, G. R. (1986), "A Method for Teaching Statistics Using N-Dimensional Geometry," *The American Statistician*, 40, 205–214.

——— (1991), *Statistical Methods: The Geometric Approach*, New York: Springer-Verlag.

Schey, H. M. (1993), "The Relationship Between the Magnitudes of $SSR(x_2)$ and $SSR(x_2|x_1)$: A Geometric Description," *The American Statistician*, 47, 26–30.