On the History of the Use of Geometry in the General Linear Model

DAVID G. HERR*

The question of why a geometric or coordinate-free approach to linear models has been subordinated to an algebraic approach is considered by reviewing selected papers having a geometric slant. These begin with R.A. Fisher's 1915 paper on the distribution of the correlation coefficient and continue through William Kruskal's elegant 1975 paper on the geometry of generalized inverses. The thesis is put forward that the relative unpopularity of the geometric approach is not due to an inherent inferiority but rather to a combination of inertia, poor exposition, and a resistance to abstraction.

KEY WORDS: General linear model; Geometry; Coordinate free; Least squares; History of linear models.

1. INTRODUCTION

Although there is a good deal more involved in the general linear model than least squares estimation, the fundamental ideas in least squares estimation are the fundamental ideas in the general linear model. In both cases, one considers a space (set) in which data must lie, a subspace (subset) of this space that corresponds to some assumptions on the data, and the relationship of the observed data to this subspace (subset).

There are two general points of view taken with respect to linear models—an algebraic one and a geometric one. Consideration of least squares estimation from each point of view will illustrate these different perspectives.

We suppose we have data $y = (y_1, \ldots, y_n)'$, which lie in Euclidian n-space, R^n. We further assume that there is a parameter vector $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$, which lies in R^k, $k \leq n$, and $n \times k$ matrix X, so that $y = X\beta$ + error. We wish to estimate β.

2. LEAST SQUARES ESTIMATION—ALGEBRAIC VIEWPOINT

Following (Gauss 1857) and (Legendre 1806), we choose as estimates of the β_i's those values $\hat{\beta}_i$ that minimize

$$Q = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{k} x_{ij}\beta_j)^2, \tag{2.1}$$

where x_{ij} is the (i,j) element of X. To minimize Q, we consider the necessary conditions for a local minimum:

$$\partial Q(\beta_1, \ldots, \beta_k)/\partial\beta_p = \sum_{i=1}^{n} 2(y_i - \sum_{j=1}^{k} x_{ij}\beta_j)(-x_{ip}) = 0 \tag{2.2}$$

for $p = 1, \ldots, k$. We are thus led to the normal equations

$$\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ip}x_{ij}\beta_j = \sum_{i=1}^{n} y_i x_{ip}, \quad p = 1, \ldots, k, \tag{2.3}$$

or in matrix form

$$X'X\beta = X'y. \tag{2.4}$$

If $\hat{\beta}$ is the solution of these equations, $\hat{\beta}$ is called a least squares estimate of β. The essential facts that a solution exists and that it minimizes Q are not inherent in the algebraic derivation of the normal equations. It is, of course, possible to prove both facts. That, however, is not the point. The point is that setting derivatives equal to zero does not, in and of itself, guarantee the existence of solutions or that the solutions, when they exist, will yield the desired extreme of the function. This point is made by Cox and Hinkley (1974, p. 284), Mood, Graybill, and Boes (1974, p. 283), and C.R. Rao (1973, pp. 222–223). In Rao's book, the normal equations, obtained algebraically, are followed by what amounts to the geometric argument for existence and minimization.

3. LEAST SQUARES ESTIMATION—GEOMETRIC VIEWPOINT

To find values $\hat{\beta}_i$ that minimize Q, we rewrite Q as

$$Q = \|y - X\beta\|^2 \tag{3.1}$$

and notice that Q is the squared distance of y from $[X]$, the subspace of R^n spanned by the columns of X. Minimizing Q corresponds, then, to finding the point in $[X]$ closest to y. The answer is readily visualized as the "point in $[X]$ directly below y," that is, the perpendicular projection of y on $[X]$. If $X\hat{\beta}$ denotes the perpendicular projection of y on $[X]$, then $X\hat{\beta}$ is unique and satisfies

$$y = X\hat{\beta} + Z, \quad Z \text{ perpendicular to } [X]. \tag{3.2}$$

Multiplying by X' we have

$$X'y = X'X\hat{\beta}, \tag{3.3}$$

since $X'Z = 0$. Thus $\hat{\beta}$ must satisfy the normal equations. We appear to have arrived at the same place as in the algebraic viewpoint. Appearances are deceiving! At this point we know there is a solution, $\hat{\beta}$, that $X\hat{\beta}$ is unique and that for any solution, $\hat{\beta}$,

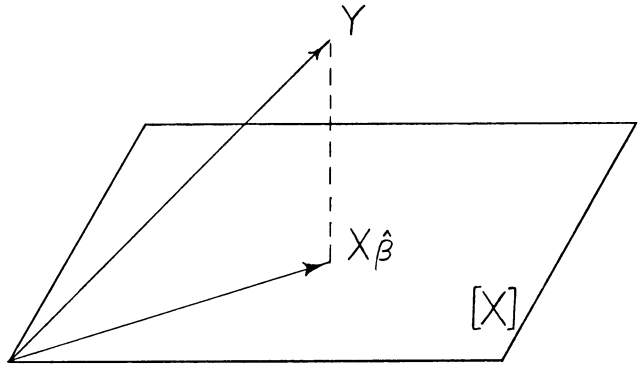
$$\|y - X\hat{\beta}\|^2 \leq \|y - X\beta\|^2, \tag{3.4}$$

for all β in R^k. Thus $\hat{\beta}$ yields the global minimum of Q.

A few remarks on the previous sketches are in order. In both, considerable detail has been omitted.

---

* David G. Herr is Associate Professor of Mathematics, University of North Carolina at Greensboro, Greensboro, NC 27412. He wishes to thank the editor, an associate editor, and referees for their thought-provoking comments and a referee for pointing out the reference Box, Hunter, and Hunter (1978).



The Least Squares Estimate

The derivation of partial derivatives and the theorem, which says that among the solutions of the equations formed by setting the $\partial Q/\partial\beta_p$ equal to zero are to be found the local extrema of Q, have been omitted from the algebraic viewpoint. Similarly, the theorem that says that the closest point in a subspace to a point not in that subspace is the perpendicular projection of that point onto the subspace has been omitted from the geometric viewpoint. The derivation of partial derivatives involves limits and some geometric ideas, however. For a unified treatment see Apostol (1969). On the other hand, the ideas of finite dimensional vector spaces used in the geometric viewpoint are finite and do not involve the idea of limits. For an excellent treatment see Halmos (1958). Thus it is the author's opinion that in the particular problem of deriving the normal equations, as well as in numerous other related problems, the geometric viewpoint is conceptually simpler than the algebraic viewpoint.

As an additional elementary example of the fact that the geometric approach is simpler and more complete than the algebraic one, consider the problem of finding the distribution of $s^2 = (n - 1)^{-1} \sum(X_i - \bar{X})^2$, for $X_1, \ldots, X_n$ iid $N(\mu,\sigma^2)$. Let Y be the vector of X_i's. Then $Y = \mu j + e$, where j is the $n \times 1$ vector of 1's and e is a vector of iid $N(0,\sigma^2)$ rv. Let $PY = \bar{x}j$, the perpendicular projection of Y on $[j]$. Then $s^2 = (n - 1)^{-1}\|(I - P)Y\|^2$, where I is the $n \times n$ identity matrix. Note that there exists an orthonormal basis $\{g_1, \ldots, g_n\}$ of R^n, such that $\{g_1 \ldots, g_{n-1}\}$ is an orthonormal basis of the orthogonal complement of $[j]$. Then

$$\|(I - P)Y\|^2 = (g'_1Y)^2 + (g'_2Y)^2 + \ldots + (g'_{n-1}Y)^2.$$

Because g'_iY is a linear combination of independent normal rv, it follows that g'_iY is distributed as $N(\mu g'_ij,\sigma^2)$. Because g_i is orthogonal to j, $g'_ij = 0$ and the g'_iY are distributed as $N(0,\sigma^2)$, $i = 1, 2, \ldots, n - 1$. Because g_i is orthogonal to g_j, $i \neq j$, the g'_iY are uncorrelated and thus independent. It then follows immediately that $(n - 1)s^2/\sigma^2$ is a sum of $n - 1$ independent chi-squared rv, each with one degree of freedom.

Another, less elementary, example of the advantage of the geometric approach is in the analysis of unbalanced, two-way designs. Suppose it is desired to test the hypothesis that there is no difference in the simple averages of cell means averaged over columns (along each row). If the hypothesis of no interaction is accepted, but the interaction sum of squares is not pooled with error, there are two sums of squares for rows that could be used—the sum of squares as if there were interaction and the sum of squares using the assumption of no interaction. Intuitively, the latter sum of squares should give a more powerful test than the former if the error sum of squares without the interaction sum of squares is used in each case. But at least one group of researchers found this impossible to prove using an algebraic approach. It is very easy to prove using the geometric approach. The basic idea is to show that the noncentrality parameter for the test assuming no interaction is larger than that for the other test. This is accomplished by showing that the noncentrality parameter for the test assuming no interaction is the length of the hypotenuse of a right triangle, while the other noncentrality parameter is the length of one of the legs of the same triangle.

Because the geometric approach to least squares estimation has considerable merit and usefulness, why is it not more widely used? This question has bothered the author for some time and has led to a consideration of the history of the use of geometry in the study of linear models. Before proceeding, let it be said that, regrettably, no claim to comprehensiveness can be made for this study. In fact, the author earnestly solicits any information on additional references from readers of this paper.

The easiest and perhaps best answer to the question may be tradition. It appears that neither Gauss nor Legendre could have thought of least squares from the geometric viewpoint described here. They did not have the requisite vector space ideas (May 1977). Thus momentum may be keeping this tradition alive.

4. PAST EXAMPLES OF THE GEOMETRIC APPROACH

R.A. Fisher (1915), however, was not dissuaded from thinking of statistical questions geometrically. This may have been due to his oft-referred-to insight or, as H.O. Hartley (1978) surmised recently, his poor eyesight. Whatever the reason, his eyes or his mind, he definitely thought geometrically from time to time. The clarity with which Fisher saw concepts was not easily transferred to the readers of his papers. Therein lies the germ of another possible answer to the question. To explore this embryonic explanation, let us consider papers of seven prominent statisticians who have written about linear models using a geometric approach. They are Fisher (1915), Bartlett (1933–34), Durbin and Kendall (1951), Kruskal (1961, 1968, 1975), Zyskind (1967), and Watson (1967). A short comment on the use of geometry or geometric thinking in each of the eight papers follows.

4.1. Fisher 1915

In this paper, Fisher is concerned with the distribution of the correlation coefficient of a sample of n pairs from a bivariate normal distribution. His approach might be termed *pure geometric* in that there is very little of an analytic nature between the verbal description of the geometry and the conclusion stated as a formula. A reasonably typical example is provided by a part of his discussion of transforming the $2n$-variate normal density of the sample of pairs (x_i, y_i) to a density in terms of the quantities $\bar{x}$, $\mu_1 = [n^{-1} \sum (x - \bar{x})^2]$, $\bar{y}$, $\mu_2 = [n^{-1} \sum (y - \bar{y})^2]$, and the sample correlation coefficient r.

> An element of volume in this n dimensional space may now without difficulty be specified in terms of $\bar{x}$ and μ_1; for given $\bar{x}$ and μ_1, P *(the point* $(x_1, x_2, \ldots, x_n)$*)* must be on a sphere in $n - 1$ dimensions, lying at right angles to the line *OM (the line through the origin in the direction* $(1, 1, \ldots, 1)$*)*, and the element of volume is
>
> $$C\mu_1^{n-2} d\mu_1 d\bar{x},$$
>
> where C is some constant, which need not be determined.

If you see it, it's beautifully elegant; if you don't, there is very little there to help improve your vision. It seems the kind of discussion that inspires the reader to honor the genius that produced it, but does not inspire him to try to emulate the approach.

4.2. Bartlett 1933–34

Bartlett is concerned with the implications and advantages of thinking of a sample of size n as an n dimensional vector. His approach might be termed *analytic geometric* in that each geometric idea is represented by analytic formulas. This enables the reader to educate whatever geometric insight he has with the analytic formulas in an iterative way, so that the net result is a deeper understanding of both the geometric view and the analytic one. An example is provided by Bartlett's discussion of the analysis of the row vector S of observations from a Latin square design.

> We have a classification in rows, columns, and treatments. We write
>
> $$S = R + C + T + E + M, \tag{4.1}$$
>
> where $R = (\bar{x}_r - \bar{x})$ is the vector representing the differences of row means from the general mean, $C = (\bar{x}_c - \bar{x})$ similarly for columns, $T = (\bar{x}_t - \bar{x})$ for treatments, $M = (\bar{x})$ as before, and
>
> $$E = (\bar{x} - \bar{x}_r - \bar{x}_c - \bar{x}_t + 2\bar{x}) \tag{4.2}$$
>
> is the residual error term. From the algebraic relations
>
> $$RC' = RT' = \ldots = EM' = 0, \tag{4.3}$$
>
> we have, analogously to (4.1)
>
> $$S^2 = R^2 + C^2 + T^2 + E^2 + M^2. \tag{4.4}$$

He had previously mentioned that the *algebraic relations* meant the corresponding vectors were perpendicular and that S^2 would be used for the squared length of S. Although the notation is rather lean, the ideas and approach have the fullness of the modern approach of Kruskal, for example.

4.3. Durbin and Kendall 1951

In their study of the geometry of estimation, Durbin and Kendall seem to this author to revert to the pure geometric approach of Fisher. As an example consider their discussion of finding the minimum variance, unbiased estimator of the common mean of a sample $x_1, \ldots, x_n$ of independent, identically distributed random variables with common variance σ^2. The estimator is of the form $\sum \lambda_i x_i$ with the unbiasedness restriction $\sum \lambda_i = 1$.

> Consider now a Euclidian $[n]$ space with co-ordinates $\lambda_1, \ldots, \lambda_n$, which we call the estimator space. The hyperplane (2)($\sum \lambda_i = 1$) corresponds to the range of values of λ giving unbiased estimators and any point P in it determines just one estimator. Now the variance of the estimator is $\sigma^2 \sum \lambda^2_j$ and hence is $\sigma^2 OP^2$ where O is the origin. It follows that this is a minimum when P is the foot of the perpendicular from O on to the hyperplane. Symmetry alone is enough to show that the values of the λ's are then all equal.

Again if you see it, it is elegant; if you don't, it is a little hard to follow.

4.4. Kruskal 1961

One of the two stated purposes of this paper is "to describe the coordinate-free approach to Gauss-Markov (linear least squares) estimation." This approach is, like that of Bartlett (1933–34), an analytic geometric one. Kruskal notes in this paper that "it is curious the coordinate-free approach to Gauss-Markov estimation, although known to many statisticians, has infrequently been discussed in the literature on least squares and analysis of variance." He further indicates that there are two major motivations for emphasizing the coordinate-free approach.

> First, it permits a simpler, more general, more elegant, and more direct treatment of the general theory of linear estimation than do its notational competitors, the matrix and scalar approaches. Second, it is useful as an introduction to infinite-dimensional spaces, which are important, for example, in the consideration of stochastic processes.

Kruskal credits L.J. Savage with introducing him to the coordinate-free approach. The second section of this paper provides a succinct primer on the coordinate-free approach. It seems that Kruskal hoped his paper would encourage more statisticians to adopt this approach to linear models. It does not appear that this hope was realized during the next 10 years or so.

4.5. Zyskind 1967

In stating conditions under which sample least squares estimators are also best linear unbiased estimators, Zyskind refers to "r orthogonal eigenvectors" of the variance-covariance matrix forming "a basis for the column space" of the design matrix.

Although this work should probably be regarded as an example of the analytic geometric approach, there is a great deal more emphasis on the analytic than on the geometric. In fact, its inclusion in this list is due in large measure to this author's view of the paper as the precursor of the more geometric work that Zyskind did with his student Justus Seely (1970a, b). It is interesting to note also that (Kruskal 1961) is *not* referenced in this paper.

4.6. Watson 1967

Following close on the heels of Zyskind (1967), this paper illustrates Kruskal's contention that coordinate-free linear models are closely related to stochastic processes, for Watson makes extensive and effective use of the spectral decomposition of the variance-covariance matrix to study the error vectors in least squares regression. This use of a convenient basis rather than a basis fixed at the outset is an excellent illustration of the fact that coordinate free does not mean freedom from coordinates so much as it means freedom to choose the appropriate coordinates for the task at hand. That Watson is thinking geometrically is illustrated by the following solution to least squares regression.

"If a perpendicular is dropped from the point in n-space with position y onto the regression space, the foot of the perpendicular is Xb, where b is the least squares estimate of β." Watson's work is more nearly an example of the analytic geometric approach than Zyskind's (1967), but it is still rather heavy on the analytic and, except as noted, uses the usual coordinate system. Watson is aware of Kruskal's work in as much as he foretells the existence of Kruskal (1968), but he still does not reference Kruskal's 1961 article.

4.7. Kruskal 1968, 1975

These two papers are elegant examples of the analytic geometric approach to linear models. In Kruskal (1968), the question of equality of simple least squares and best linear unbiased estimates, which was considered in Zyskind (1967) and Watson (1967), is treated using a coordinate-free approach. The comparison of the parts of the three papers dealing with this question is very instructive. The simplicity and beauty of the coordinate-free approach is clearly demonstrated by such a comparison.

In Kruskal (1975), an analytic geometric approach is used with such skill and grace that the paper ought to be required reading for anyone who might be tempted to deal with generalized inverses.

We have singled out eight papers to discuss in some detail, but they do not tell the whole story of the use of geometry in linear models. As we have mentioned, L.J. Savage evidently was instrumental in getting Kruskal interested. Professor R.C. Bose, whose notes on linear models were used for years by graduate students in statistics at Chapel Hill, has, through these notes, acquainted a large segment of the statistical profession with the comprehensiveness of the analytic geometric approach. Yet he did not stress the geometric ideas when teaching this course. G.A.F. Seber (1966, 1977) is another author who obviously appreciates the geometric ideas inherent in linear models. The book by Scheffé (1959) is a classic in which the geometric ideas appear as asides. It is as though Scheffé appreciated the elegance of the geometry but didn't believe the book would be accepted if it were all done geometrically. The dust jacket features "the" picture for illustrating the geometry of hypothesis testing in a linear model. Other papers using an analytic geometric approach that have appeared in the last 10 years include: Seely (1970a, 1970b), Seely and Zyskind (1971), Cleveland (1971), Burdick et al. (1974), Haberman (1975), and Herr (1976). Except for Kruskal and possibly Bartlett, no one seems to have made an attempt to promote the coordinate-free or analytic geometric approach to linear models in print.[1]

5. CONCLUSION

There still appears to be a great reluctance on the part of many to adopt this approach. Why?

Theory 1: The tradition of an algebraic approach is so strong that it will take a lot of effort and time to make a change.

Theory 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 a gifted few, but it would never do for the masses.

Theory 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 and talent in abstract thought. Dealing with abstractions is essentially a mathematical endeavor, and some statisticians eschew mathematics whenever possible.

Theory 4: The analytic geometric approach is inherently inferior to the more common matrix algebra approach.

An Opinion: It is the present author's opinion that Theory 4 is untenable. The papers by Kruskal (1961, 1968, 1975) are sufficient evidence to refute this theory. As for the other three, they are probably all true to some extent. Theory 3 represents a most dangerous state of affairs. The paper (Box 1976) notwithstanding, it seems a mistake to give any encouragement to pro-

[1] Several references not explicitly mentioned in the text are nevertheless of interest. These are Box, Hunter, and Hunter (1978), Corstan (1958), Draper and Smith (1966), Kolmogorov (1946), Zyskind and Martin (1969), Maes (1967), Eaton (1970), Bose (1944, 1961) and Magness and McGuire (1962).

grams that tend to produce statisticians of the kind described in Theory 3.

[*Received May 1978. Revised June 1979.*]

REFERENCES

Apostol, Tom M. (1969), *Calculus* (Vol. 2) (2nd ed.), Waltham, Mass.: Blaisdell Publishing Co.

Bartlett, M.S. (1933–34), "The Vector Representation of a Sample," *Proceedings of the Cambridge Philosophical Society*, 30, 327–340.

Bose, R.C. (1944), "The Fundamental Theorem of Linear Estimation," Abstract, *Proceedings of the 31st Indian Scientific Congress*, 2–3.

——— (1961), unpublished lecture notes for *Statistics 150* at the University of North Carolina, Chapel Hill.

Box, George E.P. (1976), "Science and Statistics," *Journal of the American Statistical Association*, 71, 791–799.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York: John Wiley & Sons.

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.

Cleveland, William S. (1971), "Projection With the Wrong Inner Product and Its Application to Regression With Correlated Errors and Linear Filtering of Time Series," *Annals of Mathematical Statistics*, 42, 616–620.

Corstan, L.C.A. (1958), "Vectors, a Tool in Statistical Regression Theory," *Mededelingen van de Landbouwhogeschool Wageningen*, 58, 1–92.

Cox, David R., and Hinkley, David V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Draper, Norman R., and Smith, Harry (1966), *Applied Regression Analysis*, New York: John Wiley & Sons.

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

Eaton, Morris L. (1970), "Gauss-Markov Estimation for Multivariate Linear Models: A Coordinate-free Approach," *Annals of Mathematical Statistics*, 41, 528–538.

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

Gauss, Carl F. (1857), *Theory of the Motion of the Heavenly Bodies. . .* , trans. C.H. Davis, Boston: Little, Brown.

Haberman, Sidney J. (1975), "How Much Do Gauss-Markov and Least Squares Estimates Differ? A Coordinate-free Approach," *Annals of Statistics*, 3, 982–990.

Halmos, Paul R. (1958), *Finite-Dimensional Vector Spaces* (2nd ed.), Princeton, N.J.: Van Nostrand Press.

Hartley, H.O. (1978), Personal communication.

Herr, David G. (1976), "A Geometric Characterization of Connectedness in a Two-Way Design," *Biometrika*, 63, 93–100.

Kolmogorov, A.N. (1946), "On the Proof of the Method of Least Squares," *Uspekhi Mathmaticeskih Nauk*, 1, 57–70 (in Russian).

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.

——— (1968), "When Are Gauss-Markov and Least Squares Estimators Identical? A Coordinate-free Approach," *Annals of Mathematical Statistics*, 39, 70–75.

——— (1975), "The Geometry of Generalized Inverses," *Journal of the Royal Statistical Society*, Ser. B, 37, 272–283.

Legendre, A.M. (1806), *Nouvelles Méthodes pour la determination des orbites des cometes*, Paris, France: Courcier.

Maes, J.M. (1967–68), "Partially Confounded and Non-orthogonal Effects in Analysis of Variance," *Cahiers Centre Eterdes Rich. Operat.*, 9, 134–155; 10, 32–34 (in French).

Magness, T.A., and McGuire, J.B. (1962), "Comparison of Least Squares and Minimum Variance Estimates of Regression Parameters," *Annals of Mathematical Statistics*, 33, 462–470.

May, K.O. (1977), Personal communication.

Mood, Alexander M., Graybill, Franklin A., and Boes, Duane C. (1974), *Introduction to the Theory of Statistics* (3rd ed.), New York: McGraw-Hill Book Co.

Rao, C.R. (1973), *Linear Statistical Inference and Its Application* (2nd ed.), New York: John Wiley & Sons.

Seber, G.A.F. (1966), *The Linear Hypothesis: A General Theory, Griffin's Statistical Monographs*, No. 19, London: Charles W. Griffin & Co.

——— (1977), *Linear Regression Analysis*, New York: John Wiley & Sons.

Seely, Justus (1970a), "Linear Spaces and Unbiased Estimation," *Annals of Mathematical Statistics*, 41, 1725–1734.

——— (1970b), "Linear Spaces and Unbiased Estimation—Application to Mixed Linear Model," *Annals of Mathematical Statistics*, 41, 1735–1748.

Seely, Justus, and Zyskind, George (1971), "Linear Spaces and Minimum Variance Unbiased Estimation," *Annals of Mathematical Statistics*, 42, 691–703.

Scheffé, Henry (1959), *The Analysis of Variance*, New York: John Wiley & Sons.

Watson, G.S. (1967), "Linear Least Squares Regression," *Annals of Mathematical Statistics*, 38, 1679–1699.

Zyskind, George (1967), "On Canonical Forms, Non-negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models," *Annals of Mathematical Statistics*, 38, 1092–1109.

Zyskind, George, and Martin, F.B. (1969), "On Best Linear Estimation and a General Gauss-Markov Theorem in Linear Models With Arbitrary Non-negative Covariance Structure," *SIAM Journal of Applied Mathematics*, 17, 1190–1202.