A Method for Teaching Statistics Using
N-Dimensional Geometry

D. J. SAVILLE and G. R. WOOD*

A concrete geometric method is described that enables the basic techniques of statistics (analysis of variance and regression) to be presented to students in a rigorous but elementary fashion. The few necessary ideas from geometry are summarized, and their use is illustrated by means of examples. The method works always with two objects (the observation vector and the model space) and two processes (projection of the observation vector onto the model space and application of Pythagoras’s theorem).

KEY WORDS: Elementary statistics; Analysis of variance; Regression; Vector; Subspace; Projection; Pythagoras’s theorem.

1. INTRODUCTION

The bulk of commonly used contemporary statistical methods is based on a relatively simple application of the mathematics of Euclidean $N$-dimensional space. Unfortunately this fact is rarely acknowledged in elementary or medium-level statistics courses. Such courses are therefore forced into a “cookbook approach,” and at best provide only semiplausible explanations of concepts such as degrees of freedom, orthogonality in experimental design, orthogonal contrasts, and orthogonal polynomials. Higher-level statistics courses, however, tend to go to the other extreme. In these courses the mathematics is so sophisticated that a nonmathematician quickly becomes lost. In fact, there appears to be a real need for a teaching method that bridges the gap between the two extremes: a method that conveys an understanding of the underlying mathematical principles at an elementary level.

The approach that seems most likely to achieve this goal is the geometric approach. Within the last few years this has been increasingly recognized by statisticians, as evidenced by three papers published recently in this journal. Margolis (1979) presented several examples of the use of vector geometry in the derivation of elementary statistical results and advocated geometry for presenting basic statistical concepts to serious beginning students. Herr (1980) reviewed selected statistical papers with a geometric slant, from R. A. Fisher’s 1915 paper to William Kruskal’s 1975 paper, and concluded 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” (p. 43). Bryant (1984) stated that “geometry seems to be the natural way to emphasize the unity of the fundamental ideas” (p. 38) and gave an outline of a method of teaching statistics using elementary vector geometry.

Convinced of the value of the geometric approach, we have been independently developing ideas on the teaching of statistics by geometric methods. The first major expression of this thinking was the introduction in 1984 of a course teaching statistics entirely along geometric lines, for second-year undergraduate statistics students at the University of Canterbury in Christchurch, New Zealand. The second major expression was in 1985 when a graduate-level applied statistics course was taught using geometric methods at the University of California in Davis. On each occasion the course was well received, and in our view constituted a substantial improvement over the more traditional cookbook methods. The aim was to introduce students to the theory and methods of analysis of variance and regression in a rigorous but elementary geometric setting, at the same time highlighting the unity of the area. To do this every effort was made to use only the minimal set of vector geometric tools. It is this minimal tool kit and its use that we present in this article.

2. ELEMENTARY VECTOR GEOMETRY

(THE TOOL KIT)

The handful of concepts and results from vector geometry that are essential to our development will now be summarized (see Fig. 1). In any course these would need to be expanded with justifications or proofs at an appropriate level. Each idea is presented both algebraically and geometrically, highlighting the duality of the two approaches. The interplay

---

*D. J. Saville is Biometrician, Agricultural Research Division, Ministry of Agriculture and Fisheries, Lincoln, New Zealand. G. R. Wood is Senior Lecturer, Department of Mathematics, University of Canterbury, Christchurch, New Zealand.
1. A vector \( \mathbf{y} \) in \( N \)-dimensional space is an array of the form 
\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_n
\end{bmatrix}
\]

2. Vector addition is defined componentwise by 
\[
\begin{bmatrix} 
x_1 \\
\vdots \\
x_n
\end{bmatrix} + 
\begin{bmatrix} 
y_1 \\
\vdots \\
y_n
\end{bmatrix} = 
\begin{bmatrix} 
x_1 + y_1 \\
\vdots \\
x_n + y_n
\end{bmatrix}
\]

3. Multiplication by a scalar, \( c \), is defined componentwise by 
\[
c \begin{bmatrix} 
x_1 \\
\vdots \\
x_n
\end{bmatrix} = 
\begin{bmatrix} 
cx_1 \\
\vdots \\
cx_n
\end{bmatrix}
\]

4. The span of a set of vectors, 
\[
x_1 = 
\begin{bmatrix} 
x_{11} \\
\vdots \\
x_{N1}
\end{bmatrix}, \ldots, x_k = 
\begin{bmatrix} 
x_{1k} \\
\vdots \\
x_{Nk}
\end{bmatrix}
\]
is the set of all vectors of the form 
\[
c_1 x_1 + \ldots + c_k x_k
\]
(that is, all linear combinations of \( x_1, \ldots, x_k \)).
Any such span is termed a subspace of \( N \)-dimensional space.

5. The squared length of a vector \( \mathbf{y} \) is 
\[
\|\mathbf{y}\|^2 = y_1^2 + y_2^2 + \ldots + y_N^2.
\]

A unit vector, \( \mathbf{U} \), is a vector of length 1; that is, \( \|\mathbf{U}\| = 1 \).

6. The angle, \( \theta \), between two vectors \( \mathbf{x} \) and \( \mathbf{y} \) is given by 
\[
\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{x.y}{\|x\| \|y\|}
\]
\((x.y = x_1 y_1 + \ldots + x_N y_N) \) is termed the dot product of \( \mathbf{x} \) and \( \mathbf{y} \).

Vectors \( \mathbf{x} \) and \( \mathbf{y} \) are orthogonal if \( \theta = 90^\circ \). This occurs iff \( x.y = 0 \).
7. An orthogonal coordinate system for $N$-space is a set of $N$ orthogonal unit vectors $U_1, \ldots, U_N$. 

8. The orthogonal decomposition of an arbitrary vector $y$ in terms of such a coordinate system is 

$$y = (y.U_1)U_1 + \cdots + (y.U_N)U_N.$$ 

Orthogonal decomposition of $y$ 

9. The nearest point to $y$ in $M$, the span of $\{U_1, \ldots, U_k\}, k \leq N$, is 

$$(y.U_1)U_1 + \cdots + (y.U_k)U_k.$$ 

This vector is called the projection of $y$ onto the subspace $M$. 

Projection onto the subspace spanned by $U_1$ and $U_2$. 

10. Pythagoras's theorem in $N$-space says that 

$$\|y\|^2 = (y.U_1)^2 + \cdots + (y.U_N)^2.$$ 

That is, 

$$(\text{length})^2 = \text{sum of squared lengths of projections.}$$ 

Pythagoras's theorem in $N$-space, showing squared lengths of projections. 

Figure 1. The Mathematical Framework (the ten ideas necessary for the geometric development of analysis of variance and regression).
of the algebra and geometry will be exploited in the method proposed in the next section.

3. ANALYZING DATA (USING THE TOOL KIT)

In any analysis of variance or regression problem a model is assumed and data \( y_1, \ldots, y_N \) are collected. This provides us with two objects; the data provide an observation vector

\[
y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}
\]

in \( N \)-space, and the model determines a subspace of \( N \)-space \( M = \text{span} \{ U_1, \ldots, U_k \} \) (see Fig. 2). Examples will show how \( M \) is determined.

The data are used to fit the model and to test hypotheses about the model parameters (see Fig. 3). This is routinely performed by means of two geometric processes. To fit the model we project \( y \) onto \( M \), forming \( (y.U_1)U_1 + \cdots + (y.U_k)U_k \), the fitted model. To test hypotheses about the model parameters we use Pythagoras’s theorem to express \( \|y\|^2 \) as \( (y.U_1)^2 + \cdots + (y.U_N)^2 \), where \( \{ U_1, \ldots, U_N \} \) is a full orthogonal coordinate system. Test statistics are produced by comparing averages of appropriate sets of these squared lengths.

In summary, two objects, \( y \) from the data and \( M \) from the assumed model, together with two processes, projection and application of Pythagoras’s theorem, suffice for the analysis. We now illustrate the method in a sequence of four graded examples.

Example 1: A Single Population

We begin with a random sample of size \( n = 3 \), \( y_1, y_2, \) and \( y_3 \), assumed to come from a normal distribution with unknown mean \( \mu \) and unknown variance \( \sigma^2 \) (for example, grain yield responses of 1.1, 1.4, and 1.4 tonnes/ha for the treatment of wheat seed with fungicide on three farms). Our aim is to estimate \( \mu \) and \( \sigma^2 \) and to test the null hypothesis \( H_0 : \mu = 0 \) against the alternative hypothesis \( H_1 : \mu \neq 0 \).

Our observation vector (in 3-space) is

\[
y = \begin{bmatrix} 1.1 \\ 1.4 \\ 1.4 \end{bmatrix}
\]

The very simple model assumed here is that \( y_i = \mu + e_i \), where \( e_i (i = 1, 2, 3) \) are independent \( N[0, \sigma^2] \) values. So if \( \mu \) were known, we could express \( y \) in the (vector) form

\[
\begin{bmatrix} 1.1 \\ 1.4 \\ 1.4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix},
\]

where \( e_i = y_i - \mu \) for each \( i \). We have arrived at the model space, \( M \), for this problem, namely the span of the vector \( \{ 1, 1, 1 \}' \). This completes the collection of the raw material.

The model demands that we express \( y \) as a multiple of \( \{ 1, 1, 1 \}' \) together with an error vector. How can we best approximate this decomposition of \( y \), or in the usual phrasing, how do we fit the model? We use the method of least squares and choose \( \mu \), the estimate of \( \mu \), so that \( \hat{\mu} \{ 1, 1, 1 \}' \) is closest to \( y \). That is, we must project \( y \) onto \( M \) in order to find \( \hat{\mu} \). This means that our least squares approximation to \( \mu \{ 1, 1, 1 \}' \) is \( (y.U_1)U_1 \), where \( U_1 = \{ 1, 1, 1 \}' \sqrt{3} \) is the unit vector in the \( \{ 1, 1, 1 \}' \) direction. Now

\[
(y.U_1)U_1 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \sqrt{3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},
\]

or \( \bar{y} \) is the least squares approximation to \( \mu \). We have arrived at the fitted model

\[
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \bar{y} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \bar{y} \\ 1 \bar{y} \\ 1 \bar{y} \end{bmatrix},
\]

or more succinctly, \( y = \bar{y} + (y - \bar{y}) \), which for our example is

\[
\begin{bmatrix} 1.1 \\ 1.4 \\ 1.4 \end{bmatrix} = 1.3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -.2 \\ .1 \\ .1 \end{bmatrix}.
\]

To summarize, we have broken \( y \) into two orthogonal components, a model vector and an error vector, as shown in Figure 4.

Suppose now that we wish to test \( H_0 : \mu = 0 \) against \( H_1 : \mu \neq 0 \). We take any orthogonal coordinate system for 3-space that includes \( U_1 \). A convenient one is

\[
U_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad U_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad U_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}
\]

The space spanned by \( U_2 \) and \( U_3 \) we term the error space, since the error vector always lies in this space. Pythagoras’s theorem now tells us that

\[
\|y\|^2 = (y.U_1)^2 + (y.U_2)^2 + (y.U_3)^2
\]

or

\[
5.13 = 5.07 + .045 + .015.
\]
Observation vector
\[ y = \begin{bmatrix} 1.1 \\ 1.4 \\ 1.4 \end{bmatrix} \]

Error vector
\[ (y - \bar{y}) = \begin{bmatrix} -0.2 \\ 0.1 \\ 0.1 \end{bmatrix} \]

Mean vector (model vector)
\[ \bar{y} = \begin{bmatrix} 1.3 \\ 1.3 \\ 1.3 \end{bmatrix} \]

Figure 4. Orthogonal Decomposition of the Observation Vector (single population).

It is very easy to show that \( y.U_1 = (y_1 + y_2 + y_3) / 3 \) comes from an \( N[\sqrt{3}, \mu, \sigma^2] \) distribution, whereas \( y.U_2 \) and \( y.U_3 \) come from an \( N[0, \sigma^2] \) distribution. If \( H_0 \) holds, the squares of these projection coefficients will average to \( \sigma^2 \) over many repeats of the experiment. If \( H_0 \) is false, just the first of these quantities, \( (y.U_1)^2 \), will be inflated. To test \( H_0 \), we use

\[
\frac{(y.U_1)^2}{[(y.U_2)^2 + (y.U_3)^2]/2},
\]

which comes from an \( F_{1,2} \) distribution if \( H_0 \) holds. For our data this is

\[
\frac{5.07}{(.045 + .015)/2} = 169,
\]

from which we would reject \( H_0 \) at the 1% level.

Note that we can rewrite our test statistic as

\[
\frac{||\bar{y}||^2}{||y - \bar{y}||^2/(n - 1)} = \frac{n \bar{y}^2}{s^2},
\]

the usual expression (where \( s^2 \) = sample variance).

Figures 5 and 6 sum up the relevant facts.

**Example 2: Two Populations (Completely Randomized Design)**

We now consider two populations having \( N[\mu_1, \sigma^2] \) and \( N[\mu_2, \sigma^2] \) distributions and samples of size two from each. We will want to estimate the unknown parameters \( \mu_1, \mu_2, \) and \( \sigma^2 \) and test \( H_0 : \mu_1 = \mu_2 \) against \( H_1 : \mu_1 \neq \mu_2 \). Our numerical example will comprise the sugar beet yields in tonnes/ha on a fresh weight basis for control plots (39.2 and 40.4) and nitrogen-fertilized plots (45.3 and 46.3).

Our observation vector, in 4-space, is \([39.2, 40.4, 45.3, 46.3]'\), and the model assumed is that \( y_{ij} = \mu_i + e_{ij} \), for \( i, j = 1, 2 \), where \( y_{ij} \) is the \( j \)th observation from the \( i \)th population. As before the \( e_{ij} \)'s are independent \( N[0, \sigma^2] \) values. If \( \mu_1 \) and \( \mu_2 \) were known, we could express \( y \) as

\[
\begin{bmatrix} 39.2 \\ 40.4 \\ 45.3 \\ 46.3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \mu_1 + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \mu_2 + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \end{bmatrix},
\]

so the model space \( M \) for this problem is a plane, the span of

\[
U_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}/\sqrt{2} \quad \text{and} \quad U_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}/\sqrt{2}.
\]

To fit the model we form

\[
(y.U_1)U_1 + (y.U_2)U_2 = \bar{y}_1, \quad \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \bar{y}_2, = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_2 \end{bmatrix}.
\]

Figure 5. Pythagoras's Theorem Applied to \( y = (y.U_1)U_1 + (y.U_2)U_2 + (y.U_3)U_3 \).
Distribution of $y.U_2$ \(N[0, \sigma^2]\)

Distribution of $y.U_1$ \(N[\sqrt{3}\mu, \sigma^2]\)

Figure 6. Distributions of the Projection Coefficients, $y.U_i$.

where $\bar{y}_i$ is the mean of the $i$th sample, $i = 1, 2$. So the fitted model is

\[
\begin{bmatrix}
y_{11} \\
y_{12} \\
y_{21} \\
y_{22}
\end{bmatrix} = \begin{bmatrix}
\bar{y}_1 \\
\bar{y}_2 \\
\bar{y}_2 \\
\bar{y}_2
\end{bmatrix} + \begin{bmatrix}
y_{11} - \bar{y}_1 \\
y_{12} - \bar{y}_1 \\
y_{21} - \bar{y}_2 \\
y_{22} - \bar{y}_2
\end{bmatrix}
\]

or more succinctly, $y = \bar{y}_i + (y - \bar{y}_i)$. For our example this decomposition is

\[
\begin{bmatrix}
39.2 \\
40.4 \\
45.3 \\
46.3
\end{bmatrix} = \begin{bmatrix}
39.8 \\
39.8 \\
45.8 \\
45.8
\end{bmatrix} + \begin{bmatrix}
-.6 \\
.6 \\
-.5 \\
.5
\end{bmatrix}
\]

A decomposition of $||y||^2$ with respect to an orthogonal coordinate system including $U_1$ and $U_2$ would permit us to test $H_0: \mu_1 = 0$ or $H_0: \mu_2 = 0$, since the associated projection coefficients $y.U_1$ and $y.U_2$ reflect the size of $\mu_1$ and $\mu_2$. But these are not of interest, so we choose a second, appropriate, orthogonal coordinate system for $M$, \{[$U'_1$, $U'_2$]\}, as follows:

\[
U'_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \sqrt{4}, \quad U'_2 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \sqrt{4}
\]

\[
U_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \sqrt{2}, \quad U_4 = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \sqrt{2}
\]

Here $U_3$ and $U_4$ complete the system and span the error space, the space in which the error vector lies. It is easy to show that $y.U'_1$ comes from an $N[\sqrt{4}\mu, \sigma^2]$ distribution [where $\mu = (\mu_1 + \mu_2)/2$ is the overall mean], $y.U'_2$ comes from an $N[\mu_1 - \mu_2, \sigma^2]$ distribution, whereas $y.U_3$ and $y.U_4$ come from an $N[0, \sigma^2]$ distribution. Notice that we are now in a position to test two hypotheses: $H_0: \mu = 0$ and $H_0: \mu_1 = \mu_2$. We are only interested, however, in the second of these hypotheses. Pythagoras’s theorem tells us that

\[
||y||^2 = (y.U'_1)^2 + (y.U'_2)^2 + (y.U_3)^2 + (y.U_4)^2
\]

or

\[
7364.58 = 7327.36 + 36 + (.72 + .5)
\]

for our data. If $H_0: \mu_1 = \mu_2$ is true, then the final three
squared lengths will average to $\sigma^2$ over many repeats of the experiment, whereas if $H_0$ is false just $(y.U_i)^2$ will be inflated. We calculate

$$\frac{(y.U_i)^2}{(y.U_3)^2 + (y.U_4)^2}/2,$$

which comes from an $F_{1,2}$ distribution if $H_0$ holds. In our example this is

$$\frac{36}{(.72 + .5)/2} = 59.02.$$  

On this basis we reject $H_0: \mu_1 = \mu_2$ at the 5% level of significance.

Fitting the model with respect to this second coordinate system produces the decomposition

$$y = (y.U_1)U_1' + (y.U_2)U_2' + [(y.U_3)U_3 + (y.U_4)U_4]$$

$$= \bar{y}.. + (\bar{y}_i - \bar{y}..) + (y - \bar{y}_i),$$

where $\bar{y}..$ is the overall sample mean. We have decomposed the model vector, $\bar{y}_i$, into the overall mean vector, $\bar{y}..$, plus the treatment vector, $(\bar{y}_i - \bar{y}..$. This is illustrated in Figure 8. Evidently our test statistic is

$$\frac{\|\bar{y}_i - \bar{y}..\|^2}{\|y - \bar{y}..\|^2/2},$$

which leads to the usual algebraic form.

**Example 3: Two Populations (Randomized Block Design)**

Suppose now that the four plots of Example 2 are paired and the two treatments are allocated at random within each pair. Such a design is termed a randomized block design with two blocks and two treatments. The observation vector, together with the origin of the data, is as follows:

$$y = \begin{bmatrix} 39.2 \\ 40.4 \\ 45.3 \\ 46.3 \end{bmatrix} \text{ Control, Block 1} \quad \text{Control, Block 2} \quad \text{Fertilized, Block 1} \quad \text{Fertilized, Block 2}.$$  

The model assumed is a refinement of that in Example 2, namely

$$y_i = \mu_i + \beta_j + \epsilon_{ij},$$

where $\beta_j = \gamma_j - \mu$ is the $j$th block effect, $\gamma_j$ being the $j$th block mean. It follows that $\beta_1 + \beta_2 = 0$. To determine the model space, we write

$$\begin{bmatrix} 39.2 \\ 40.4 \\ 45.3 \\ 46.3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{bmatrix}.$$

We have found the model space, the span of the first four vectors on the right side. The four vectors, however, do not form an orthogonal coordinate system for $M$. So we rewrite the model in an equivalent, but appropriate, way as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \end{bmatrix} + [((\mu_2 - \mu_1)/2] \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + [(\beta_2 - \beta_1)/2] \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{bmatrix}.$$

Now the model vectors are orthogonal. Fitting the model by projecting yields

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \bar{y}.. + [(\bar{y}_2 - \bar{y}_1)/2] \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + [(\bar{y}_2 - \bar{y}_1)/2] \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{bmatrix},$$

where $\bar{y}_j$ is the mean for the $j$th block. In brief, this is readily shown to be

$$y = \bar{y}.. + (\bar{y}_i - \bar{y}..) + (\bar{y}_j - \bar{y}..) + (y - \bar{y}_i - \bar{y}_j + \bar{y}..),$$

Figure 8. A Refined Orthogonal Decomposition of the Observation Vector (two populations, completely randomized design). The squared length of the treatment vector, $(y.U_2)^2$, divided by the mean squared length of the error vector, $((y.U_2)^2 + (y.U_4)^2)/2$, produces the $F$ ratio for treatments.
which for our example is

\[
\begin{bmatrix}
39.2 \\
40.4 \\
45.3 \\
46.2
\end{bmatrix} =
\begin{bmatrix}
42.8 \\
42.8 \\
42.8 \\
3.0
\end{bmatrix} +
\begin{bmatrix}
-3 \\
3 \\
-3 \\
1.0
\end{bmatrix} +
\begin{bmatrix}
-55 \\
-55 \\
-55 \\
-55
\end{bmatrix} +
\begin{bmatrix}
-0.5 \\
-0.5 \\
-0.5 \\
-0.5
\end{bmatrix}
\]

<table>
<thead>
<tr>
<th>Observation vector</th>
<th>Overall Mean vector</th>
<th>Treatment vector</th>
<th>Block vector</th>
<th>Error vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment vector</td>
<td>Overall Mean vector</td>
<td>Treatment vector</td>
<td>Block vector</td>
<td>Error vector</td>
</tr>
</tbody>
</table>

This decomposition is illustrated in Figure 9.

To test \( H_0 : \mu_1 = \mu_2 \) against \( H_1 : \mu_1 \neq \mu_2 \), the appropriate coordinate system is

\[
U_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{4}, \quad U_2 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} / \sqrt{4},
\]

\[
U_3 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} / \sqrt{4}, \quad U_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{4}.
\]

It is readily shown that \( y.U_2 \) comes from an \( N[\mu_2 - \mu_1, \sigma^2] \) distribution, whereas now only \( y.U_4 \) comes from an \( N[0, \sigma^2] \) distribution. By Pythagoras's theorem,

\[
\| y \|^2 = (y.U_1)^2 + (y.U_2)^2 + (y.U_3)^2 + (y.U_4)^2
\]

or

\[
7364.58 = 7327.36 + 36 + 1.21 + .01.
\]

If \( H_0 \) is true, then \( (y.U_2)^2/(y.U_4)^2 \) will come from an \( F_{1,1} \) distribution. In our example this is 36/.01 = 3,600, so we reject \( H_0 \) at the 5% level of significance.

**Example 4: Simple Linear Regression**

An experiment is conducted on five cars to investigate the relationship between a petrol additive and nitrogen oxides in the exhaust. The results are as follows:

- **Amount of additive, \( x \):** 1 2 3 4 5
- **Reduction in nitrogen oxides, \( y \):** 2.1 3.1 3.0 3.8 4.3

A linear relationship between \( x \) and \( y \) is assumed, so we may write our model in the form \( y_i = \beta_0 + \beta_1(x_i - \bar{x}) + e_i \), where \( \bar{x} \) is the mean of the five \( x \) values. Our aim is to estimate \( \beta_0 \) and \( \beta_1 \) and test \( H_0: \beta_1 = 0 \) against \( H_1: \beta_1 \neq 0 \).

\[
\beta_1 \neq 0.
\]

The observation vector and its decomposition are

\[
\begin{bmatrix}
2.1 \\
3.1 \\
3.0 \\
3.8 \\
4.3
\end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix},
\]

revealing a two-dimensional model space and an appropriate orthogonal coordinate system for this model space, namely

\[
U_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{5} \quad \text{and} \quad U_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{10}.
\]

To fit the model we calculate

\[
(y.U_1)U_1 + (y.U_2)U_2 = 3.26 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + .51 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},
\]

so the least squares estimates of \( \beta_0 \) and \( \beta_1 \) are 3.26 and .51, respectively. Our decomposition of \( y \) is pictured in Figure 10.

We now wish to test \( H_0: \beta_1 = 0 \). Elementary methods can be used to show that \( y.U_2 \) comes from an \( N[\sqrt{10}\beta_1, \sigma^2] \) distribution, whereas \( y.U_i \) comes from an \( N[0, \sigma^2] \) distribution for any unit vector \( U_i \) in our three-dimensional error space. So if \( U_3, U_4, \) and \( U_5 \) complete our orthogonal coordinate system,

\[
\frac{(y.U_2)^2}{(y.U_3)^2 + (y.U_4)^2 + (y.U_5)^2/3}
\]

comes from a \( F_{1,3} \) distribution, under \( H_0 \). For our data this ratio is \( 2.601/.0703 = 37.00 \), allowing us to reject \( H_0: \beta_1 = 0 \) at the 1% significance level.

## 4. DISCUSSION

The simple examples presented in this article were chosen to convey some of the flavor of the approach. The method

---

**Figure 9.** Orthogonal Decomposition of the Observation Vector (randomized block design).
clearly extends, however, to much more sophisticated situations. The courses we have taught have included the class comparison, factorial, and rates categories of orthogonal contrast in multiple, single, and fractional replicate studies; polynomial regression using orthogonal polynomials; multiple regression; randomized block, latin square, and split plot designs; regression through the origin; and analysis of covariance. Examples of such applications are currently being prepared for a textbook that we expect to publish during 1987.

For us the geometric approach provides so much more insight than a purely algebraic approach, that we are surprised to find geometric ideas not widely used at an early stage in the teaching of statistical methods. In speculating on the reasons for the dominance of algebraic methods in statistical teaching the following passage from Yaglom (1981) is illuminating.

It is well known that the history of science exhibits ebbs and flows; if the 19th century was the golden age of geometry, then our times are distinguished by the preeminence of algebra, by the distinctive "algebraization" of all branches of mathematics reflected in the acceptance of Nicolas Bourbaki's mathematical structures, converting even geometry virtually into a part of algebra. (p. 258)

There is no doubt that the influence of the Bourbaki school, at its most extreme epitomized in Dieudonné's slogans of "Down with Euclid!" and "Death to triangles!," has been far-reaching. This spirit of formalism, arriving at a time when statistical methods were being developed, has assuredly had a marked influence on the presentation of the subject to two or more generations of statisticians.

But times are changing. Disillusionment with "the new math" and the rise of radically different mathematical philosophies [such as those of Kitcher (1983) and Lakatos (1976)] presage a new era in which geometry may be restored to its rightful place in the development of mathematics and the mathematical mind. To quote from Davis and Hersh (1981):

In recent years, a reaction against formalism has been growing. In recent mathematical research, there is a turn toward the concrete and the applicable. In texts and treatises, there is more respect for examples, less strictness in formal exposition. The formalist philosophy of mathematics is the intellectual source of the formalist style of mathematical work. The signs seem to indicate that the formalist philosophy may soon lose its privileged status. (p. 344)

The fashion of formalism, together with the difficulty of understanding Fisher's often telegraphic geometric style [commented on in Herr (1980)], would appear to have been the major influences in pushing statistics into an algebraic mold over the past 50 years.

For us, geometry appears close to the genesis of intuition and offers perspectives that enrich the formal linear algebra. It is interesting to note in this regard that there appears to be some evidence that the hemispheres of our brains specialize in somewhat distinct functions: whereas the left hemisphere is responsible for "discrete," symbolic (hence algebraic) processing, the right hemisphere is responsible for "spatial," visual (hence geometric) processing. An excellent summary of such findings over the past century is found in Corballis and Beale (1976, pp. 101–106). In the light of this evidence it would seem unwise for us to forge ahead using only half of our faculties!

The method presented in this article relies on simple linear algebra for the formal proofs but derives its inspiration from geometric thinking. The result is a unified and comprehensible technique that enables analysis of variance, regression, and analysis of covariance to be taught rigorously at an elementary level.

[Received August 1985.]

REFERENCES


The American Statistician, August 1986, Vol. 40, No. 3 213
On Sums of Random Variables and Independence

BERNHARD K. FLURY*

Suppose that the density of the sum of two random variables $X$ and $Y$ is given by the convolution of the two marginal densities. Although this condition is stronger than uncorrelatedness of $X$ and $Y$, it does not imply stochastic independence, as is shown by three examples. A situation for which this fact may be relevant occurs in the construction of chi-squared tests for nested hypotheses.

KEY WORDS: Convolution; Correlation; Chi square.

1. INTRODUCTION

It is well known that uncorrelatedness of two random variables $X$ and $Y$ does not imply stochastic independence of $X$ and $Y$, except in special cases (e.g., when the joint distribution of $X$ and $Y$ is bivariate normal). Another well-known but frequently ignored fact is that univariate normality of both $X$ and $Y$ does not imply bivariate normality. In fact, many examples of nonnormal bivariate distributions with normal marginals have been published, and Kowalski (1973) gave a comprehensive overview of the relevant literature until 1972. Particularly appealing examples in which $X$ and $Y$ are uncorrelated and univariate normal, yet not bivariate normal, were published by Melnick and Tenenbein (1982).

The search for interesting counterexamples to popular beliefs has largely been restricted to the case of normal distributions. This note gives three counterexamples to another possible misspecification: Suppose that the univariate distributions of $X$ and $Y$ are given by the densities $f_X$ and $f_Y$ and that the density of $X + Y$ is the convolution of $f_X$ and $f_Y$. Does this imply stochastic independence of $X$ and $Y$? More formally, let $f(x, y)$ denote the joint density of $X$ and $Y$, then $f_X(x) = \int f(x, y) dy$ and $f_Y(y) = \int f(x, y) dx$ and the density of the sum is $f_Z(z) = \int f(x, z-x) dx$. The question is now: If

$$f_Z(z) = \int f_X(x)f_Y(z-x) dx, \tag{1}$$

does this imply that $f(x, y) = f_X(x)f_Y(y)$? It is not difficult to show that (1) implies $E[XY] = E[X]E[Y]$, that is, uncorrelatedness of $X$ and $Y$. Uncorrelatedness does not imply (1), however, as the first example in Melnick and Tenenbein (1982) shows, so (1) is a stronger condition. Yet the answer to the previous question is no, as I show by three examples.

2. EXAMPLES

Example 1

Start with a bivariate standard normal density,

$$g(x, y) = (2\pi)^{-1} \exp[-(x^2 + y^2)/2],$$

and divide the $(x, y)$ plane in eight areas bordered by the straight lines $x = 0$, $y = 0$, $x + y = 0$, and $x - y = 0$. This is shown graphically in Figure 1. Now set $f(x, y) = 2g(x, y)$ in the shaded areas and $f(x, y) = 0$ in the blank areas. The marginals are still both standard normal. Moreover, as integration along straight lines of constant $x + y$ shows, the distribution of $X + Y$ is normal with mean 0 and variance 2. More generally, for any positive power $k$, the distribution of $X^k + Y^k$ behaves as if $X$ and $Y$ were independent.

This example, which is the same as example 2 of Melnick

![Figure 1](image)

Figure 1. Construction of Example 1. The density is $\pi^{-1} \exp[-(x^2 + y^2)/2]$ in the four shaded areas and 0 in the unshaded areas. Both marginals are $N(0, 1)$, and the sum is $N(0, 2)$.

*Bernhard K. Flury is Oberassistent, Department of Statistics, University of Berne, Sidlerstrasse 5, CH-3012 Berne, Switzerland.