Some Geometrical Considerations

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Some Geometrical Considerations

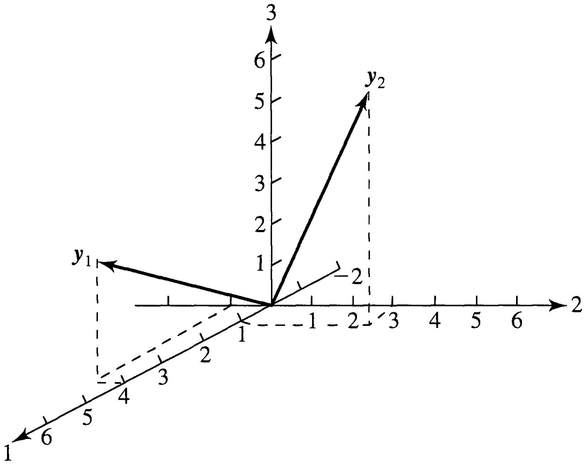
Introduction

- In our treatment of linear and multiple regression algebra, we have, so far, relied on the most traditional algebraic approach.
- This began, in the case of simple bivariate linear regression, by presenting the data for n observations on two variables X and Y as points plotted in a plane.
- This approach is of course quite useful, but another quite different approach has also proven extremely useful.
- In the sample, this approach involves presenting variables as vectors plotted in the n-dimensional "data space."

A Variable as a Vector

- For example, suppose $n = 3$ and the variable y_1 has the values $y_1' = (4, -1, 3)$. The variable y_2 has values $y^{W}{\prime}_2 = (1, 3, 5)$.
- We can plot them in 3-dimensional space as shown on the next slide, taken from Johnson and Wichern (2002).

A Variable as a Vector



A Variable as a Vector

A Vectorspace and its Basis

- Recall the operations of scalar multiplication and vector addition as already defined.
- Recall also that a set of vectors is linearly independent if and only if no vector is a linear combination of the others.
- Now consider a set of linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots _k$. They are said to be basis vectors that span a k-dimensional vectorspace.
- The vectorspace itself is defined as the set of all linear combinations of its basis vectors.

A Variable as a Vector

Length of a Vector

- As an extension of the Pythagorean Theorem, the Euclidean length of a vector, denoted $||\mathbf{x}||$, is the square root of the sum of squares of its elements, i.e.,

$$||\mathbf{x}|| = \sqrt{\mathbf{x}'\mathbf{x}} \tag{1}$$

A Variable as a Vector

Angle Between Two Vectors

- The cosine of the angle θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ satisfies the equation

$$\cos(\theta_{\mathbf{x},\mathbf{y}}) = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}} \tag{2}$$

- Conversely, the scalar product of two vectors can be computed as

$$\mathbf{x}'\mathbf{y} = ||\mathbf{x}||||\mathbf{y}|| \cos(\theta_{\mathbf{x},\mathbf{y}}) \tag{3}$$

A Variable as a Vector

Relationship between Correlation and Angle

- Equation 2 on the preceding slide shows some intimate connections between statistics and geometry.
- Suppose that both $\mathbf{x}$ and $\mathbf{y}$ are in deviation score form. Since the variance of X is then $\mathbf{x}'\mathbf{x}/(n-1)$ and the covariance between $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{x}'\mathbf{y}/(n-1)$, the following facts immediately follow:
  1. The lengths of a group of deviation score vectors in $n-1$ dimensional space are directly proportional to their standard deviations.
  2. The correlation between any two deviation score vectors in $n-1$ dimensional space is equal to the cosine of the angle between them.

Projection and Least Squares Estimation

Properties of Projectors

- Projection is a key concept in geometry.
- The projection or shadow of a vector $\mathbf{y}$ on another vector $\mathbf{x}$ is defined as

$$\frac{\mathbf{x}\mathbf{x}'}{\mathbf{x}'\mathbf{x}}\mathbf{y} = \mathbf{P}_{\mathbf{x}\mathbf{y}} \tag{4}$$

- As we prove in Homework 01, for a vector $\mathbf{x}$, the orthogonal projector $\mathbf{P}_\mathbf{x} = \mathbf{x}(\mathbf{x}'\mathbf{x}^{-1})\mathbf{x}'$ and its complementary projector $\mathbf{Q}_\mathbf{x} = \mathbf{I} - \mathbf{P}_\mathbf{x}$ have a number of key properties, most of which trace back to the following:

$$\mathbf{P}_\mathbf{x} = \mathbf{P}_\mathbf{x}{}' = \mathbf{P}_\mathbf{x}^2$$

$$\mathbf{Q}_\mathbf{x} = \mathbf{Q}_\mathbf{x}{}' = \mathbf{Q}_\mathbf{x}^2$$

$$\mathbf{P}_\mathbf{x}\mathbf{Q}_\mathbf{x} = \mathbf{0}$$

$$\mathbf{P}_\mathbf{x}\mathbf{x} = \mathbf{x}, \;\; \mathbf{Q}_\mathbf{x}\mathbf{x} = \mathbf{0}$$

Projection and Least Squares Estimation

Properties of Projectors

- The key point of the homework assignment is that $\mathbf{P_x}$ and $\mathbf{Q_x}$ can be used to decompose a vector $\mathbf{y}$ into two component vectors that are orthogonal to each other, with one component collinear with $\mathbf{x}$ and the other orthogonal to it.
- Specifically, for any $\mathbf{y}$, define

$$\hat{\mathbf{y}} = \mathbf{P_x y}, \ \ \mathbf{e} = \mathbf{Q_x y} \tag{5}$$

- Clearly $\hat{\mathbf{y}}$ is collinear with $\mathbf{x}$, since

$$\mathbf{P_x y} = \mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{y} = \mathbf{x}b \tag{6}$$

with

$$b = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} \tag{7}$$

Projection and Least Squares Estimation

Properties of Projectors

- It also follows that

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e} \tag{8}$$

since

$$\begin{aligned} \hat{\mathbf{y}} + \mathbf{e} &= \mathbf{P_x y} + \mathbf{Q_x y} \\ &= \mathbf{P_x y} + (\mathbf{I} - \mathbf{P_x y}) \\ &= (\mathbf{P_x} + \mathbf{I} - \mathbf{P_x})\mathbf{y} \\ &= \mathbf{Iy} = \mathbf{y} \end{aligned} \tag{9}$$

and that

$$\mathbf{e}'\hat{\mathbf{y}} = 0 \tag{10}$$

Projection and Least Squares Estimation

Column Space Projectors

- Now consider an $\mathbf{X}$ of full column rank with more than one column. Similar results to the preceding ones can be established, as follows:
- We define the column space of $\mathbf{X}$, $\mathrm{Sp}(\mathbf{X})$, as the set of all linear combinations of the columns of $\mathbf{X}$, that is, a vectorspace with the columns of $\mathbf{X}$ as its basis.
- The column space orthogonal projector $\mathbf{P_x}$ and its complementary projector $\mathbf{Q_x}$ are defined essentially the same as before, i.e.

$$\mathbf{P_X} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$$

and

$$\mathbf{Q_X} = \mathbf{I} - \mathbf{P_X}$$

Projection and Least Squares Estimation

Column Space Projectors

- Now for any *matrix* $\mathbf{Y}$, the columns of

$$\hat{\mathbf{Y}} = \mathbf{P_X Y}$$

are in the column space of $\mathbf{X}$, since

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{X}\left\{\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y}\right\} && (11) \\ &= \mathbf{XB} && (12) \end{aligned}$$

Moreover, as before, we can define $\mathbf{E} = \mathbf{Q_x Y}$ and establish results analogous to those in Equations 8–10.

- Just as we say that $\mathbf{P_x}$ projects any vector into $\mathrm{Sp}(\mathbf{X})$, $\mathbf{Q_x}$ projects any vector into $\mathrm{Sp}(\mathbf{X})^{\perp}$, the orthogonal complement to $\mathrm{Sp}(\mathbf{X})$.
- These results are central in linear regression.

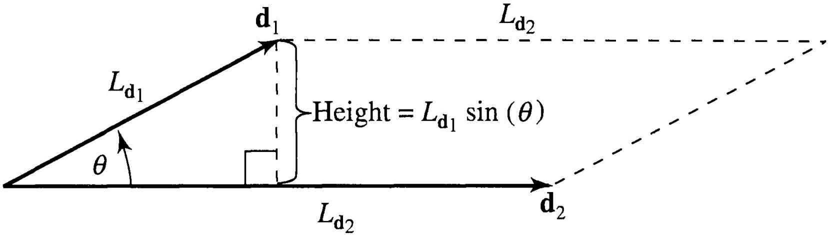
Demos in 3D

- Let's digress and examine the geometry of statistics with an active demonstration in $n = 3$ dimensions.
- Although being stuck in 3 dimensions constrains our ability to visualize, many of the concepts become clearer.
- Create a working directory. Download the files *GeometrySupport.R* and *GeometryDemos.R* to it from the website. startup R, and make sure that the `rgl` and `geometry` packages are installed.
- If they are not, please download them and install them.
- Then, open the file *GeometryDemos.R* in RStudio, and set the working directory to where this file is located.

The Determinant as Generalized Variance

- In our 3D demo, we saw how two vectors can be thought of as defining a parallelogram.
- We have also pointed out that the length of a vector of deviation scores is equal to $\sqrt{n-1}$ times its standard deviation, so that the length of a deviation score vector is directly proportional to the standard deviation of the variable it represents.
- It turns out that, just as the square root of the variance of a single variable is proportional to its length, the square root of the determinant of the covariance matrix of a pair of variables is directly proportional to the area of the parallelogram they "carve out" in deviation score space.
- Here is a picture from Johnson and Wichern.

The Determinant as Generalized Variance



The Determinant as Generalized Variance

- If $\mathbf{S}$ is a 2×2 matrix, it is well known that

$$|\mathbf{S}| = s_{11}s_{22} - s_{21}s12 = s_{11}s_{22} - s^2 12$$

.

- But since

$$s_{12} = r_{12}\sqrt{s_1 1 s_2 2}$$

we have

$$|\mathbf{S}| = s_{11}s_{22}(1 - r_{12}^2)$$

The Determinant as Generalized Variance

- But since the area of the parallelogram is $L_{d_1} \times Height$, and (recalling that $\sin^2\theta + \cos^2\theta = 1$)

$$Height = L_{d_1} \sin\theta = L_{d_1}\sqrt{1 - r^2}$$

  we have

$$Area = L_{d_1} L_{d_2}\sqrt{1 - r^2} = (n-1)\sqrt{s_{11}s_{22}(1 - r^2)}$$

- Consequently,

$$Area^2 = (n-1)^2|\mathbf{S}|$$

  and

$$Area = (n-1)|\mathbf{S}|^{1/2}$$

The Determinant as Generalized Variance

- More generally, as proven by T.W. Anderson in his classic textbook on multivariate analysis, with p variables the relationship is

$$Volume^2 = (n-1)^p |\mathbf{S}|$$

- So $|\mathbf{S}|^{1/2}$ is the multivariate analog of the standard deviation, and the determinant is a multivariate analog of variance.