

Using Geometry to Visualize Abstract Aspects of Statistical Formulae Relevant to Correlation and Regression

Kyle Nickodem¹, Ernest C. Davenport, Ph.D.¹, Qinjun Wang¹
Steven Andrew Culpepper, Ph.D.²

¹Univ. of Minnesota-Twin Cities, 56 E. River Rd, 250 EdSciB, Minneapolis, MN 55414

²University of Illinois at Urbana-Champaign

Abstract

In undergraduate and graduate statistics courses, concepts are often taught using algebraic or computational explanations. These may not, however, be optimal teaching methods for all students. Thus, the use of geometry can provide an alternative pedagogy that satisfies different learning styles without a loss of accuracy and rigor. This paper demonstrates an approach using geometric diagrams to introduce students to the concepts of bivariate, partial, and part correlations, multiple regression, R^2 , and R^2 change in a concrete manner. In addition to supplementing current teaching methods, a geometric approach provides a visual representation of concepts that enables students to *see* statistics while requiring only a high school level of geometrical background knowledge.

Key Words: geometry, statistics pedagogy, visual representation, regression, correlation

1. Objectives and Framework

Statistics courses are often challenging for students and some may view statistics as a stumbling block to degree attainment. Given that the M.A. and Ph.D. are both research degrees, students must develop some level of statistical competency. Thus, statistics instructors may need alternative pedagogies to make concepts accessible to students with different learning styles without sacrificing accuracy or rigor. Geometry can provide one such alternative. Rather than supersede the more common algebraic or computational approaches, geometrical representation of statistical concepts can be used as a supplement, providing students multiple opportunities to learn material and leading to better success for all students. This paper demonstrates how geometry can be used to introduce students to bivariate, partial, and part correlations, multiple regression, R^2 , and R^2 change in a very concrete manner allowing students to *see* statistics. It is hoped that a picture is truly worth a thousand words, or at least several algebraic formulas.

Current methods of teaching statistics typically fall into two broad categories: the algebraic approach and the computational approach (Wickens, 2014). The algebraic approach explains the mathematical underpinnings and formulas for computation. While these concepts are necessary for aspiring statisticians, for students in non-math fields the presence of complex formulae and mathematical notation elicits anxiety and negative attitudes toward the class (Cobb & Moore, 1997). Onwuegbuzie and Wilson (2003) found upwards of 80% of graduate students experience uncomfortable levels of statistics anxiety, particularly those in the social and biological sciences. Furthermore, the use of matrix algebra or formulas to teach concepts does little to develop students' statistical thinking and reasoning by failing to promote consideration of context, experimental design, assumptions, and limitations (Cobb & Moore, 1997). For novice students this makes statistics seem abstract, disjointed, and a skill only learned through rote memorization and cookbook application. Another criticism is the algebraic approach often neglects how statistical concepts relate to real world situations (Rumsey, 2002;

Yilmaz, 1996). Students might be able to calculate a standard deviation, and even define it using statistical language, but be unable to explain its relevance to on-the-job problems, such as properly interpreting a student's test score. In the absence of application to practical situations, formulas and definitions are quickly forgotten after an exam. Alternatively, it has been consistently argued that statistics instructors should deliberately minimize the use of complex formulas and mathematical notation, but if necessary, graphical and verbal methods should be used to first introduce the concept and care should be taken to connect concepts to the broader picture of why they are needed in a given situation (e.g. Bradstreet, 1996; Garfield & Ben-Zvi, 2007; Moore, 1997).

The computational approach is quite opposite of the algebraic approach. It focuses on the use of statistical software to transform data into interpretable results with minimal exposure to the underlying statistical theory and mathematics. In some ways this approach heeds the advice of many authors to use data exploration and analysis to introduce concepts because it encourages students to be active learners and illustrates various purposes and uses of statistics (Garfield & Ben-Zvi, 2007; Moore, 1997). However, the relative absence of underlying mathematics risks presenting statistics as magic (Cobb & Moore, 1997). Data can be inputted to software and with a few clicks produce results before any consideration is given to the quality of the data, appropriateness of the analysis, or validity of the interpretations. Thus, through the effort to make statistics more accessible to non-mathematicians there is a loss of rigor that can produce students who know just enough to be dangerous, as the idiom goes (Spirer, Spirer, & Jaffe, 1998).

Both the algebraic and computation approaches provide valuable information to aspiring researchers and data analysts, especially when they are used in tandem. Nonetheless, there lacks a framework connecting the formulas and techniques of how statistics are estimated to their use by statistical software in a manner that is still accessible to students interested in applying statistical methods to a wide variety of fields. In some ways students' learning of statistics can be likened to the challenged faced by the blind men in John Godfrey Saxe's poem *The Blind Men and the Elephant* (Saxe, J.G.). In Saxe's poem six inquisitive blind men seek to establish truth about the nature of an elephant by having each man explore a different portion of its body. The examinations lead each man to develop a unique understanding of the animal that ranged from a spear to a tree. All six men made "blind" conjectures based on the sense of touch and all six settled on a reality far from the truth. In some sense, the use of touch represents a pedagogical tool for learning—it is with touch that the men were able to derive some meaning of truth. In the case of the blind men, the pedagogy they employed produced a disjointed and incomplete understanding of the animal. To complete their intellectual conquest, the men require a supplemental tool that unifies each man's partial truth into a comprehensive picture. For students, the algebraic and computational approaches are each a tool in the process of learning and using statistics, but fail to provide a complete understanding. Furthermore, unlike the blind men who all had similar strengths and weaknesses in their ability to learn, students in undergraduate and graduate programs are rather diverse in their methods for optimal learning. Therefore, how can we help algebraically challenged individuals visualize statistical concepts in an accessible, yet rigorous manner to produce students well-versed in both the computation and application of statistics? If the old cliché that a picture is worth a thousand words holds any merit then maybe the answer lies with a visual or geometrical presentation of statistical concepts.

2. Theoretical Background

The field of statistics has a rich history of using geometry to illustrate, derive, and explain statistical concepts. Some of the most prominent statisticians of the 20th century have presented complex statistical ideas with geometry (Bartlett, 1934; Durbin & Kendall, 1951; Fisher, 1915; Kendall, 1961; Kruskal, 1961, 1968, 1975; Watson, 1967; Zyskind, 1967). Kruskal (1961) notes one benefit of using a geometric approach is that, "...it permits a simpler, more general, more elegant, and more direct treatment of the general theory of linear estimation than do its notational competitors." More recently, Saville & Wood (2012) add, "use of geometry clarifies and unifies our understanding of the basic statistical techniques."

For instance, the geometric approach played a prominent role in popularizing the least squares method. DeLaubenfels (2006) explains that it was the geometrical features of least squares that ultimately led to its preference over earlier, more intuitive approaches such as the sum of absolute deviations¹. While a squared lost function is not as intuitive, its application along with concepts of orthogonality allows the use of Pythagorean's Theorem for distance to be related to the vector decomposition of the total sums of squares; i.e., $SS_T = SS_E + SS_M$ where SS_T is the total sums of squares, SS_E is the sums of squares error, SS_M is the sums of square for the model. The formula for this decomposition is given in summation notation below (equation 1). The decomposition, however, is more readily understood when represented geometrically (Figure 1) where one can more clearly see that error is minimized by the orthogonal projection (a) from the vector representing SS_M (b) to the vector representing SS_T (c).

$$\sum(Y - \bar{Y})^2 = \sum(Y - \hat{Y})^2 + \sum(\hat{Y} - \bar{Y})^2 \quad (1)$$

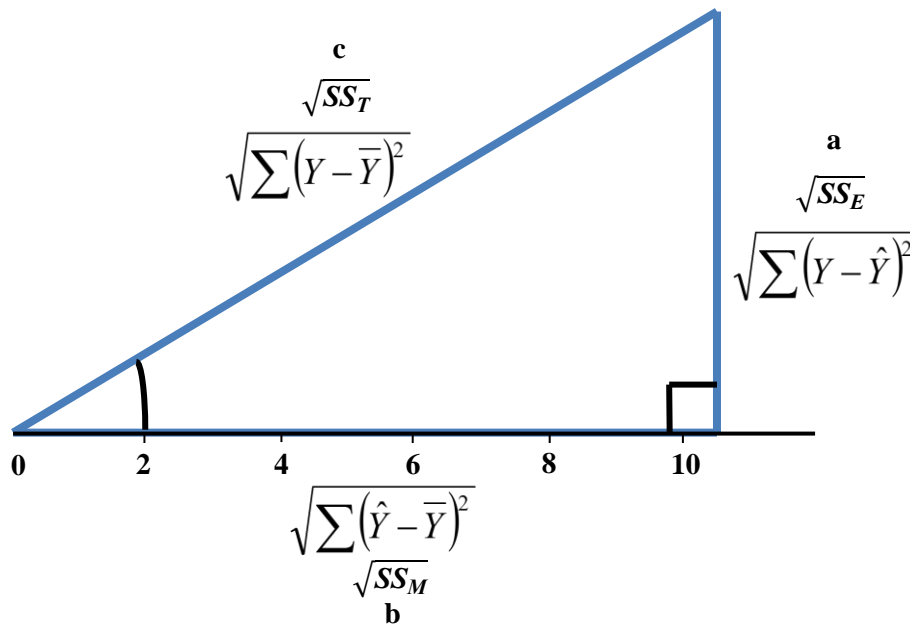


Figure 1: Decomposition of the total sums of squares.

¹ Under conditions of non-redundancy, Least Squares is also preferred to the sum of absolute deviations, since it will always guarantee a unique solution for the estimates.

There has also been a plethora of prior research developing geometric demonstrations of various aspects of the general linear model. Herr (1980) chronicles the early contributions of prominent statisticians while much of the research during the last four decades has focused on the geometric presentation of regression and correlations. These include Leung and Lam's (1975) short discussion about the relationship between correlations and the angle between two n -dimensional vectors; Schulman's (1979) illustration of a model of rank correlations (e.g. Spearman's rho and Kendall's tau); Thomas and O'Quigley's (1993) extension of the traditional geometrical display of correlations with the use of spherical triangles; and Bring's (1996) excellent treatment of the geometric representation of regression to compare the relative importance of variables in a model.

Another benefit of geometrical presentation of statistical concepts is it can sometimes provide an intuitive understanding to initially perplexing situations. Leung and Lam (1975) link algebraic facts about correlations to a geometric display and note that, "...many facts concerning correlation coefficients can be seen in terms of the geometry of representation" (p. 129). This idea has been enumerated in other articles, as well. For instance, Saville and Wood (1986) provide a thorough presentation of a geometric display of randomized block designs and simple linear regression and Schey (1993) employs geometric techniques to display the conditions for suppressor variables. Additionally, Hamilton (1987) used geometry to inform an analytical discussion while presenting a case where two predictors are essentially uncorrelated with a criterion, yet, the coefficient of determination equaled one—which is perhaps one example that would not be immediately accessible in an algebraic discussion. Also, Davenport, Kuang, Davison, Nickodem, and Wang (2015) use geometry to show the equivalence of several models that at first glance would appear discrepant.

If geometry can be a useful vehicle for those who do not intuitively understand statistics or have different learning styles, why is geometry not employed more frequently in the social and biological sciences to teach statistics? Two possible reasons have been posited: (1) While the 19th century was the golden age of geometry, the rise of statistics in the 20th century coincided with algebraic methods coming into vogue. Computation was also easier with algebraic expressions than geometric calculations, although a century of advances in computers has rendered this point moot. Nonetheless, a tradition of teaching statistic algebraically was established and had not until recently been challenged. (2) Some of the early uses of geometry during the reign of the algebraic tradition were reserved solely for advanced statistics students rather than novices and beginning students. This precedent led to the misconception that only those with a deep knowledge of mathematics could connect with a geometric approach (Herr, 1980; Saville & Wood, 2012). Neither tradition nor misconception are justifiable reasons for ignoring geometry as a valuable tool for improving novice students' understanding of statistics.

The goal of this study is to present statistical concepts, namely multiple regression, bivariate correlations, partial correlations, part correlations, R^2 , and R^2 change with geometric diagrams in order to promote their use when teaching statistics to undergraduate and graduate students in the social and biological sciences. As a supplement to algebraic and computational approaches, a geometric approach to teaching statistic provides a mechanism for integrating the calculation and application of statistics into a unified framework that is both rigorous and accessible to students from a variety of backgrounds and learning styles. This approach also uniquely provides a visual representation of concepts that enables students to *see* statistics while requiring only a high school level of geometrical background knowledge.

3. Data

This study used data collected by the Office for Institutional Research at the University of Minnesota from the college application process for 2,975 high schools in the state of Minnesota who enrolled at the University for Fall 2000. The relationships among three variables were examined: 1) high school percentile rank (HSR), 2) percentage of non-free/reduced priced lunch (%NFR) students in each entering students' high school, and 3) ACT composite score. This paper employed four statistical analyses: bivariate correlations, partial correlations, part correlations, and linear regression. Table 1 presents the Pearson Product Moment, partial, and part correlations among the three variables.

Table 1: Bivariate, Partial, and Part Correlations

	ACT	HSR	%NFR	\hat{ACT}
ACT	1.00	0.36	0.24	0.45
HSR		1.00	-0.10	0.79
%NFR			1.00	0.53
\hat{ACT}				1.00
Partial Correlation		0.40	0.30	
Part Correlation		0.39	0.28	

Note. Partial and part correlations were calculated with ACT Composite as the criterion, holding the other predictor constant. ACT = ACT Composite, HSR = High school percentile rank, %NFR = % non-free/reduced priced lunch students in the high school, \hat{ACT} = the predicted ACT score with HSR and %NFR as independent variables. All correlations were significant at $p < 0.001$.

The first three columns contain the correlations among HSR, %NFR, and ACT, while the last column contains the predicted ACT composite score, \hat{ACT} , with HSR and %NFR as the independent variables. The correlations between ACT and HSR and ACT and %NFR were 0.36 and 0.24, respectively. The correlation between HSR and %NFR was -0.10. The bivariate correlations were all statistically significant at the 0.001 rejection level. Interestingly, the positive relationship between %NFR and ACT suggests that as the student body becomes less impoverished or more affluent, individual students' composite ACT score tend to be higher.

4. An Algebraic Overview of Bivariate, Part and Partial Correlations

The purpose of this section is to provide an overview of the traditional presentation of bivariate, part, and partial correlations with analytic formulae. A bivariate correlation expresses the strength of the relationship between two variables in a single summary statistic that ranges from -1 to 1. For variables X and Y , the correlation is calculated with the formula:

$$r_{xy} = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_i [(x_i - \bar{x})^2][\sum_i (y_i - \bar{y})^2]}} \quad (2)$$

Partial correlation is then the relationship between two variables while controlling for the effect of a third variable, Z , on both X and Y . The formula for partial correlation is presented below:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (3)$$

Part correlation is similar to partial correlation, but only accounts for the link between Z and either X or Y , but not both. In the formula below we control for the effect of Z on X :

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)}} \quad (4)$$

The formulas do not offer an intuitive understanding of how they characterize the relationship between the variables, X , Y , and Z , even with accompanying explanations.

During this discussion consider the three variables to be standardized to have a mean of zero and a standard deviation of one. In the case of the bivariate correlation, the equation (2) shows that a correlation is the ratio of the covariance of X and Y and the product of the standard deviations for X and Y . Since we are assuming the variables are standardized, the denominator will be 1 and the covariance and bivariate correlation will be equal. The partial correlation between Y and X controlling for Z ($r_{xy.z}$) is simply the correlation between the error terms from when Z is regressed separately on both X and Y . So the partial correlation can also be represented as $r_{e_X e_Y}$, where $e_X = X - r_{XZ}Z$ and $e_Y = Y - r_{YZ}Z$ with r_{XZ} and r_{YZ} being the correlations between X and Z and Y and Z , respectively. Note that $r_{XZ}Z$ is the predicted value of X using Z and $r_{YZ}Z$ is the predicted value of Y using Z . Incorporating these new variables controlled for Z in the formula for correlation yields equation (3). This becomes a ratio of the covariance between e_X and e_Y over the product of the standard deviations of e_X and e_Y . A simple algebraic derivation shows that $r_{XY} - r_{XY}r_{YZ}$ is the covariance between e_X and e_Y while $\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}$ is the product of their respective standard deviations.

An understanding of equation (4) for part correlations can be obtained using the same steps as presented for the partial correlation. The difference between part and partial correlations is that part correlations measure the relationship between Y and X with the influence of Z partialled only from X , but not from Y . Hence, another name for part correlations is semi-partials. Equation (4) is then obtained by computing the ratio of the covariance between Y and e_X over the standard deviation of Y and e_X . In this case the denominator of equation (4) contains only the standard deviation for e_X since Y is standardized with a mean of zero and a standard deviation of one and one times the standard deviation for e_X is e_X .

While the formulas and explanations shed some light on how to conceptualize the relationship between three variables, such as for ACT, HSR, and %NFR, many students will still be in the dark and feeling a fair amount of statistical anxiety. Conversely, the computational approach would bypass the use of mathematical notation in the above discussion entirely and rely on the output in Table 1 to explain the concepts. Doing so certainly aids in comprehending and interpreting the strength of relationships between variables without the overwhelming nature of the formulas. Yet, even in the computational approach, how variables relate to each other is still presented in an abstract manner. Some students might be comfortable with abstract thinking and be able to mentally represent the variables in a comprehensible fashion, but others may still struggle to grasp the meaning of these correlations beyond the single statistic used to describe them. The use of geometry, however, can transform these abstractions into concrete visuals that are more intuitive and accessible to a wider audience.

5. Geometry of Bivariate, Part, and Partial Correlations and Regression

5.1 Bivariate Correlation

Previous research notes the inherent geometric interpretation of bivariate correlations, specifically how bivariate correlations relate to the cosine function. Using ACT and %NFR as exemplars, Draper and Smith (1966) as well as Leung and Lam (1975) show that the following expression holds:

$$r_{ACT, \%NFR} = \cos(\theta) \tag{5}$$

where $\cos(\theta)$ is the cosine of the angle between the vectors of ACT and %NFR on a coordinate plane and $r_{ACT, \%NFR}$ is the bivariate relationship between ACT and %NFR. This relationship is expressed visually in Figure 2a:

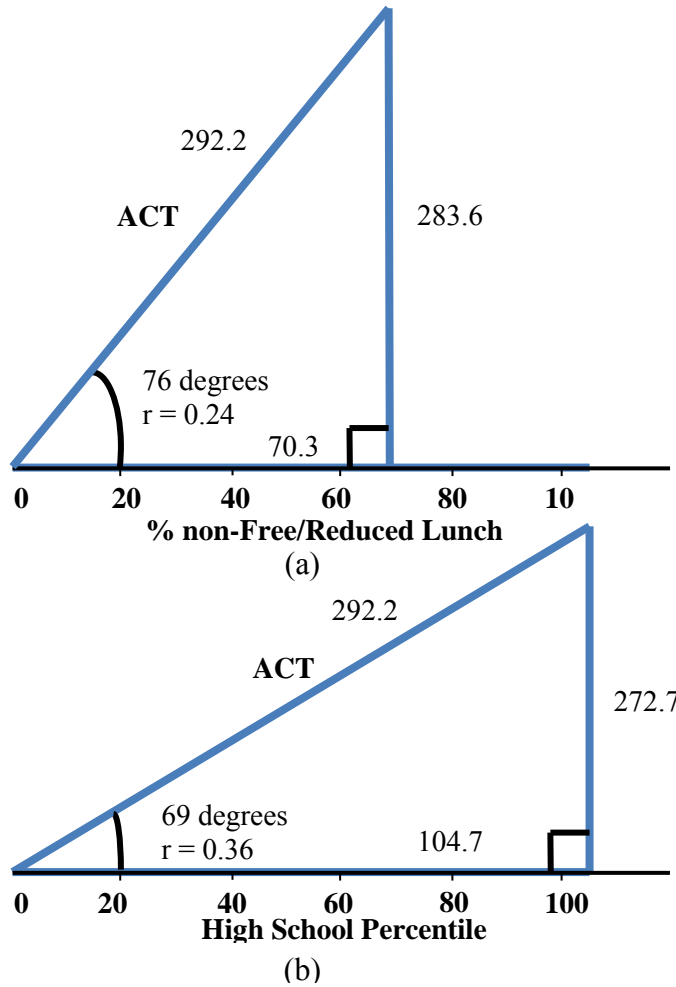


Figure 2: Geometric representation of bivariate correlations

In Figure 2a, the angle between ACT and %NFR is 76° corresponding to a correlation of 0.24, which is the same correlation we see in Table 1 and would obtain through equation (2). It is important to note that a correlation of one has a corresponding angle of 0° , meaning the variable vectors lie in the same space. The idea that two variables with a perfect positive relationship would share the same space is rather intuitive and provides a more concrete representation of the relationship between the variables than the statistic $r = 1$. Additionally, a correlation of zero suggests that the variables are orthogonal,

meaning they will not share any space on the plane. Unsurprisingly, a correlation of zero corresponds to a 90° angle making the vectors perpendicular. From these principles we can show that the more positively correlated two variables are, the more acute the angle between them will be. Therefore, even without any numeric information a student could identify HSR (Figure 2b) as being more closely related to ACT than %NFR (Figure 2a) simply by looking at the geometric visual (assuming the visuals are drawn proportionally) and noting that Figure 2b has a more acute angle. Such an assertion can then be confirmed geometrically with $\cos(\theta)$, algebraically, or with software.

5.2 Single Predictor Regression

The triangle in Figure 1 can be linked to those in Figure 2 along with the sum of squares in Table 2 to demonstrate a geometric representation of regression. More specifically, a generalization of Pythagorean's Theorem to Euclidian distance shows that the length of a vector equals the square root of the sum of the squared projections onto each dimension. In other words, equation (1) is equivalent to $SS_T = SS_E + SS_M$ which is also equivalent to $c^2 = a^2 + b^2$. Regardless of which of the three equations are used, taking the square root of each term produces the length of each of the three corresponding vectors in the triangle in Figure 1. With the assistance of the geometrical diagram, it is easy to show that the hypotenuse (side c) is equal to the square root of the total sums of squares while the side b corresponds to the model sums of squares which is necessarily orthogonal to side a -- the error sums of squares. While the mathematical notation in equation (1) might be intimidating to some students, all undergraduate and graduate students should be familiar with Pythagorean's Theorem. Therefore, by relating the decomposition of the sums of squares to a common geometric principle, students are provided with a picture to more easily comprehend the least squares regression orthogonal projection.

Table 2: Summary of Model Sum of Squares for Three Regression Models

	Sum of Squares	df	F	P-value	R ²
$ACT_i = \beta_{00} + \beta_{10}\%NFR + e_{i,1}$					
SS _{Model}	4,935.8	1	182.5	<0.001	.058
SS _{Residual}	80,417.9	2,973			
SS _{Total}	85,353.7	2,974			
$ACT_i = \beta_{01} + \beta_{11}HSR + e_{i,2}$					
SS _{Model}	10,965.6	1	438.3	<0.001	.128
SS _{Residual}	74,388.1	2,973			
SS _{Total}	85,353.7	2,974			
$ACT_i = \beta_{02} + \beta_{12}\%NFR + \beta_{22}HSR + e_{i,1}$					
SS _{Model}	17,639.5	2	387.1	<0.001	.207
SS _{Residual}	67,714.2	2,972			
SS _{Total}	85,353.7	2,974			
SS _(HSR %NFR)	12,703.7				
SS _(%NFR HSR)	6,673.9				

Consider a one-predictor model with %NFR predicting ACT, shown below as well as at the top of Table 2:

$$ACT_i = \beta_{00} + \beta_{10}\%NFR + e_{i,1} \quad (6)$$

As with the output from most statistical software, students can use the sums of squares from Table 2 to calculate the lengths of the corresponding vectors in Figure 2a. For instance, the total sums of squares for the model is 85,353.7, so we can find the length of the hypotenuse of Figure 2a by calculating $\sqrt{85,353.7} = 292.2$. The line segment adjacent to the 74° angle is the SS_M (side b). Given that %NFR is the only predictor, then the estimate of ACT as represented by the model must lie in the space of %NFR or along this vector. The length of which is determined by where the perpendicular projection of ACT hits the vector of %NFR which is represented by SS_M . Thus, its length is equal to the square root of the model sum of squares, $\sqrt{4,935.8}$ which equals approximately 70.3. Finally, the error that is left over from predicting ACT with %NFR must necessarily be orthogonal to the vector %NFR -- which houses the model used to predict ACT -- since whatever relationship exists between %NFR and ACT is in the model with the remainder in error. This final vector, SS_E , equals $\sqrt{80,417.9}$ which is approximately 283.6. As anticipated: $85,353.7 = 4,935.8 + 80,417.9$ (re: Pythagorean's $c^2 = a^2 + b^2$). The extent to which SS_M equals SS_T is the extent to which the model is helpful. A geometric description of the relationship between ACT and HSR can be developed in a similar manner from the regression equation below and its accompanying information in Table 2:

$$ACT_i = \beta_{01} + \beta_{11}HSR + e_{i,2} \quad (7)$$

5.3 Multiple Regression

There are only two variables in a single predictor model: the predictor and the criterion. As such, the model resides in a two dimensional space, like we see in Figure 2. Each additional predictor added to the model also adds a dimension to the geometric space in which the regression is represented. Therefore, the number of dimensions is equal to $k + 1$ predictors. While it is more difficult to visually depict models beyond the third dimension the principle demonstrates to students how the abstract notions of the interrelationship between variables can be transformed into a concrete illustration. A two predictor model with ACT regressed on both %NFR and HSR follows the equation:

$$ACT_i = \beta_{02} + \beta_{12}\%NFR + \beta_{22}HSR + e_{i,1} \quad (8)$$

Figure 3a shows the relationship of %NFR (vector D) and HSR (vector E) with the predicted ACT Composite score, \hat{ACT} (vector A), which lies in the plane defined by the two predictors and can be represented two-dimensionally. The length of D in the multiple regression model is the same as the square root of the SS_M for the %NFR as the sole predictor model (equation 6) and represented on the x-axis in Figure 2a. Likewise, the length of E in the multiple regression model is equivalent to the square root of the SS_M for the HSR as the sole predictor model (equation 7) and represented on the x-axis in Figure 2b. Whereas D and E are the SS_M s for the single predictor models, A is the square root of the SS_M for the multiple regression model. Vector B represents the contribution of HSR to the SS_M for the two-predictor model above and beyond the contribution of %NFR, which is represented in Table 2 as $SS_{(HSR | \%NFR)}$. C is the contribution of %NFR ($SS_{(\%NFR | HSR)}$). These are called incremental, or Type III, sums of squares. The length for each of these vectors is calculated from the sums of squares in Table 2:

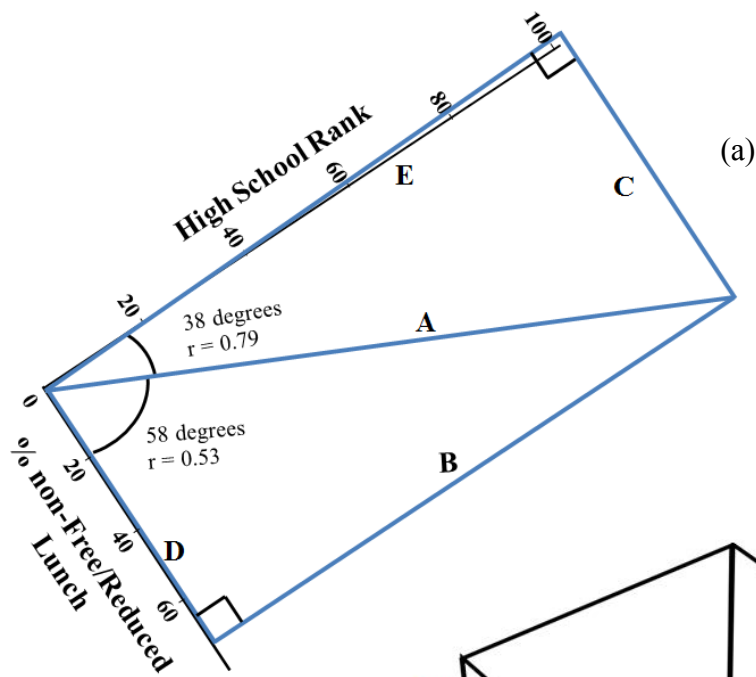
$$D = \sqrt{SS_M^{\%NFR}} = \sqrt{4,935.8} = 70.3$$

$$E = \sqrt{SS_M^{HSR}} = \sqrt{10,965.6} = 104.7$$

$$A = \sqrt{SS_M^{\%NFR, HSR}} = \sqrt{17,639.5} = 132.8$$

$$B = \sqrt{SS_M^{\%NFR, HSR} - SS_M^{\%NFR}} = \sqrt{17,639.5 - 4,935.8} = \sqrt{12,703.7} = 112.7$$

$$C = \sqrt{SS_M^{\%NFR, HSR} - SS_M^{HSR}} = \sqrt{17,639.5 - 10,965.6} = \sqrt{6,673.9} = 81.7$$



(b)

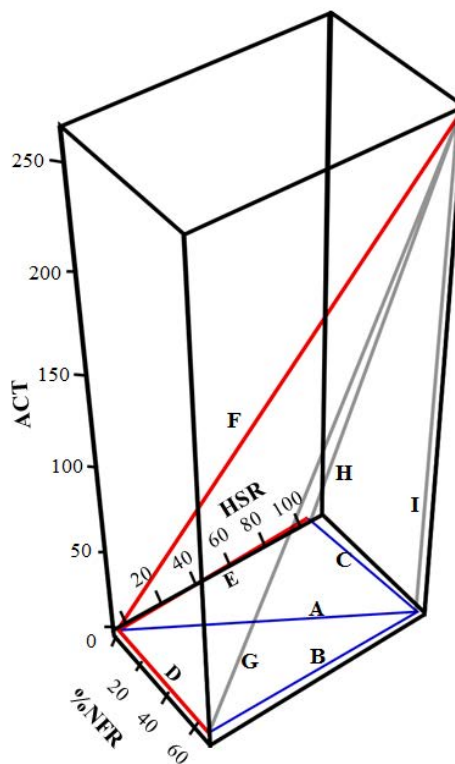


Figure 3. Two-dimensional (a) and three-dimensional (b) representations of the regression model with %NFR and HSR as predictors and ACT as the criterion.

The idea that C represents the amount of variation in \hat{ACT} that is unaccounted for by HSR, but accounted for by %NFR is further illustrated by noting that C is orthogonal to E, which is the square root of the SS_M for the regression of HSR onto \hat{ACT} —i.e. C is the variation of \hat{ACT} that is independent of HSR. Likewise, B is the amount of variation in \hat{ACT} that is unaccounted for by %NFR. Also, since D represents the square root of the SS_M with %NFR as the independent variable and \hat{ACT} as the dependent variable, B represents the portion in \hat{ACT} that is accounted for by HSR and that is orthogonal to %NFR.

The three dimensional interpretation of the multiple regression model is displayed in Figure 3b. Of particular note is the bivariate relationships depicted in Figure 2 are maintained in the multiple regression relationship shown in Figure 3b. The linchpin is the SS_T which is the same for all three models (equations 6, 7 and 8) as shown in Table 2. In Figure 3b the square root of the SS_T is signified by vector F, which extends into the third dimension. Using vector F as our common thread we can see that model with %NFR as the only predictor (equation 6) is represented by triangle FDG where D is the SS_M for the model and G is the SS_E . Similarly, triangle FEH depicts the model with HSR as the single predictor (equation 7) where E corresponds with the SS_M and H with the SS_E . Note that the SS_E for both of these models project down from the third dimension onto each of the SS_M vectors for the single predictor models. Similarly, the SS_E vector for the multiple regression model (vector I) projects down from the third dimension onto the two-dimensional plane defined by the predictor variables creating triangle FAI. Once again, the length of each of these vectors is calculated with the information from Table 2:

$$F = \sqrt{SS_T} = \sqrt{85,353.7} = 292.2$$

$$G = \sqrt{SS_T - SS_M^{\%NFR}} = \sqrt{85,353.7 - 4,935.8} = 283.6$$

$$H = \sqrt{SS_T - SS_M^{HSR}} = \sqrt{85,353.7 - 19,965.5} = 272.2$$

$$I = \sqrt{SS_T - SS_M^{\%NFR, HSR}} = \sqrt{85,353.7 - 17,639.5} = 260.2$$

Another way to view Figure 3b is from an architectural standpoint. Suppose I wish to build a structure and have an entity arising from the origin as is SS_T (remember that SS_T is in a third dimension). The best unidimensional support for SS_T and the one that will take less material is a support that has a perpendicular projection down from SS_T to the vector represented by %NFR or the vector represented by HSR -- the SS_M for each single predictor model. Any other projection will use more material and provide less support due to its angle. These supports are represented by the endpoints of vectors D and E, respectively. In like manner, if one wishes to use the plane defined by HSR and %NFR simultaneously, the support using the least materials and providing maximal support is provided by a perpendicular projection into the plane. That support is represented here by the endpoint of vector A.

5.4 Part and Partial Correlation

Part correlations are the strength of the unique relationship of the predictor to the criterion above and beyond the relationship of the criterion with the other variables. Therefore, part correlations can be represented as what it takes to get from the prediction provided by the simple regression model for each predictor (equations 6 and 7) to the

prediction provided by the multiple regression model (expression 8) relative to the original total sums of squares. To get from the predictor of ACT provided by %NFR, D, to the predictor provided by both %NFR and HSR, A, requires an additional amount represented by B. Thus, the incremental increase in predictability, $B/F = 112.71 / 292.2 = 0.39$, is the part correlation for HSR with ACT when %NFR has been partialled from HSR only. Note that calculating the part correlation is as simple as finding the ratio of the square root of the incremental sums of square for HSR, $SS_{(HSR | \%NFR)}$ (vector B) and the square root of the SS_T (vector F). Not only does the geometric approach provide students with a visual representation of part correlation, but also offers a substantially easier method of computing the value than the formula in equation (4). In similar manner, to get from the predictor of ACT provided by HSR, E, to the predictor provided by both %NFR and HSR, A, requires an additional amount represented by C. Thus the incremental increase in predictability, $C/F = 81.7 / 292.2 = 0.28$ is the part correlation for %NFR when HSR is partialled only from %NFR.

G is the error from predicting ACT with %NFR, 283.6. Thus, G is the remainder of ACT after partialling out %NFR. Above we said that B was the contribution of HSR after partialling out %NFR. Thus, the partial correlation of HSR with ACT after partialling %NFR from each is $B/G = 112.7 / 283.6 = 0.40$. Likewise, H is the error from predicting ACT with HSR, 272.7. Thus, H is the remainder of ACT after partialling out HSR. Above we said that C was the contribution of %NFR after partialling out HSR. Thus, the partial correlation of %NFR with ACT after partialling HSR from each is $C/H = 81.7 / 272.7 = 0.30$. The part and partial correlations computed by taking the ratios of the appropriate vectors matches those found in Table 1 thereby demonstrating the accuracy of the geometric approach. In doing so, the geometric approach marries the simplicity of the computational approach by avoiding complex mathematics with the rigor of the algebraic approach while providing a more intuitive explanation of the processes underlying the calculation of correlations.

5.5 R^2 and R^2 Change

R^2 is the proportion of the total variation explained by the model. In other words, it is simply the ratio of the SS_M to the SS_T . For the single predictor model of ACT regressed on % NSF this produces an $R^2 = SS_M^{\%NFR} / SS_T = 4,935.8 / 85,353.7 = 0.057$. Expressed geometrically, this is the ratio of the squared vector lengths: D^2 / F^2 . Equivalently, equation (5) can be used to convert the angle between vectors D and F into their correlation by taking the inverse cosine. The R^2 value is then equal to the correlation squared. For multiple regression, the R^2 value can be derived from the ratio of the squared vector length for the criterion and the squared vector length for the predicted value. For the multiple regression model in equation (8), this is represented as A^2 / F^2 . Alternatively, the angle between ACT (F) and \hat{ACT} (A) is 63.3° , meaning the correlation between these variables is 0.45 and the $R^2 = .203$, which is what we see in Table 1.

R^2 change is the difference in the total variation explained when going from a model with one predictor to a multiple regression model. This concept is simply the part correlation of the predictor added to the model squared. We demonstrated above that the part correlation between HSR and ACT when controlling for %NFR was $B / F = 112.7 / 292.2 = 0.39$. Thus, the R^2 change from the model with %NFR as the sole predictor of ACT to the multiple regression model that included HSR is $B^2 / F^2 = 0.39^2 = .15$.

6. Closing Remarks

6.1 Educational Importance

The development of alternative pedagogies to address differing learning styles is a perennial concern for education. The current paper addresses the development of an alternative approach to teaching statistics with the incorporation of more annotated geometrical diagrams. It is possible that students who are more concrete learners may be better served by this approach than the current algebraic approach, which is more abstract. For most people, it is very difficult to see the results of an algebraic calculation. In contrast, geometrical diagrams present pictures from which many statistical concepts can be shown. Even if a picture is not worth a thousand words, it may be worth more than a couple algebraic equations.

The discussion above provides a link between the traditional algebraic and geometric approaches of explaining multiple regression, bivariate, part, and partial correlations, R^2 and R^2 change. The goal was to use geometric diagrams to help make abstract statistical concepts more concrete, and thus, more accessible to students in non-mathematics fields. That is not to say a geometric approach should replace algebraic and computational instruction, but rather be incorporated as a supplemental approach to provide students with a more complete picture of the calculation and application of statistical concepts. A key component of the geometric supplement is the basic geometric background knowledge necessary to understand the visuals and concepts does not exceed high school algebra and geometry. One need only understand how to use cosine to determine angles, Pythagorean's theorem for length when the vectors are orthogonal, as well as the generalization of Pythagorean's theorem to n-space to understand that each of the sums of squares in a variance decomposition is just a squared distance.

6.2 Future Research

There are opportunities for additional educational and psychological research in the use of geometry for explaining statistical concepts. First, there is a need to examine whether or not the geometric, algebraic, or computational approach is more effective in maximizing students' short- and long-term comprehension of statistical theory and methods. Of equal importance to comprehension is how effective the methods are in minimizing statistical anxiety while eliciting positive attitudes toward and motivation to learn statistics. In many ways, statistics courses act as gatekeepers of opportunity for many mathematically under prepared and anxious masters and doctoral students, since these courses are often requirements for degree completion. However, can geometric presentations of statistical concepts act as a remedy for those who cannot 'see' statistical concepts with traditional algebraic approaches or fail to grasp the statistical reasoning required for appropriate use of statistical software in the computational approach? It may be the case that some students benefit more from the use of geometry in the class room rather than the traditional algebraic presentation. In this case, maintaining educational fairness may require the adaptation of new or differentiated curriculum that caters to multiple learners. Additionally, research that examines the use of geometry for explaining statistics to K-12 and undergraduate students may be worthwhile. If a picture is worth a thousand words, then the use of geometry to present statistics may serve as a new tool for helping more students "see" the meaning of varied statistical concepts and thereby creating a larger statistically literate group of students.

A second area of research may provide additional insights into individual differences of learning. Are students from different cultural backgrounds or gender better served with a geometric approach to teaching statistics? This is analogous to asking, what are the factors that impact whether or not the blind men perceived the elephant as a wall or a tree?

Certainly, a goal of intellectual pursuit is to gain a clearer picture of reality. Sometimes our current understandings of questions and our theoretical constructs that we use to answer them are inherently biased. Other times, individuals' reality is blurred simply because the conception of a question is not appropriate for their understanding; much like that of Saxe's the *Six Blind Men and the Elephant*. In line with Howard Gardner's (1993) work on multiple intelligences, we know that individuals have different strengths and weaknesses. It is illogical and maybe even unethical to not provide individuals with differing pedagogies in light of that fact. The use of geometry with statistics may hold a key in focusing on an often blurred reality. It is the duty of educators to provide better ways to teach. Hopefully, this will lead to better educational researchers which will lead to better educational research – maybe in the area of teaching statistics.

References

- Bartlett, M.S. (1934). The vector representation of a sample. *Proceedings of the Cambridge Philosophical Society*, 30(3), 327-340.
- Bradstreet, T. E. (1996). Teaching introductory statistics courses so that non-statisticians experience statistical reasoning. *The American Statistician*, 50(1), 69-78.
- Bring, J. (1996). A geometric approach to compare variables in a regression model. *The American Statistician*, 50(1), 57-62.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104(9), 801-823.
- Davenport, Jr., E. C. Kuang, H., Davison, M. L., Nickodem, K., and Wang, Q. (2015). Making use of atypical regression models for theory building. A paper presentation to the Joint Statistical Meetings of the American Statistical Association at its 2015 annual meeting in Seattle, WA.
- DeLaubenfels, R. (2006). The victory of least squares and orthogonality in statistics. *The American Statistician*, 60(4), 315-321.
- Durbin, J., & 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.
- Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic Books.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396.
- Hamilton, D. (1987). Sometimes $R^2 > r_{yx_1}^2 + r_{yx_2}^2$: Correlated variables are not always redundant. *The American Statistician*, 41(2), 129-132.
- Herr, D. G. (1980). On the history of the use of geometry in the general linear model. *The American Statistician*, 34(1), 43-47.
- Kendall, M. G. (1961). *A course in the geometry of n dimensions*. London: Charles Griffin and Company.

- Kruskal, W.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.
- Kruskal, W.H. (1968). When are the Gauss-Markov and least squares estimators identical. *Annals of Mathematical Statistics*, 39(1), 70-75.
- Kruskal, W.H. (1975). The geometry of generalized inverses. *Journal of the Royal Statistical Society, Series B (Methodological)*, 37(2), 272-283.
- Leung, C., & Lam, K. (1975). A note on the geometric representation of the correlation coefficients. *The American Statistician*, 29(3), 128-130.
- Moore, D. S. (1997). *Statistics: Concepts and Controversies* (4th ed.), New York: W. H. Freeman and Company.
- Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: Nature, etiology, antecedents, effects, and treatments--a comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195-209.
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3), 6-13.
- Saville, D. J., & Wood, G. R. (1986). A method for teaching statistics using N-dimensional geometry. *The American Statistician*, 40(3), 205-214.
- Saville, D. J., & Wood, G. R. (2012). *Statistical methods: The geometric approach*. New York: Springer Science & Business Media.
- Saxe, J. G. *The Blind Men and the Elephant*. Retrieved September, 16, 2016 from <http://www.is.wayne.edu/mnissani/ElephantCorner/6blind.htm>.
- 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(1), 26-30.
- Schulman, R. S. (1979). A geometric model of rank correlation. *The American Statistician*, 33(2), 77-80.
- Spirer, H. F., Spirer, L., & Jaffe, A. J. (1998). *Misused statistics*. CRC Press.
- Thomas, G. & O'Quigley, J. (1993). A geometric interpretation of partial correlation using spherical triangles. *The American Statistician*, 47(1), 30-32.
- Watson, G. S. (1967). Linear least squares regression. *Annals of Mathematical Statistics*, 38(6), 1679-1699.
- Wickens, T. D. (2014). *The geometry of multivariate statistics*. Psychology Press.
- Yilmaz, M. R. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education*, 4(1), 1-9.
- Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Annals of Mathematical Statistics*, 38(4), 1092-1109.