

# Evaluating Univariate, Bivariate, and Multivariate Normality Using Graphical and Statistical Procedures

Tom Burdenski, Texas A & M University

This paper reviews graphical and statistical procedures for evaluating multivariate normality by guiding the reader through univariate and bivariate procedures that are necessary, but insufficient, indications of a multivariate normal distribution. A data set utilizing three dependent variables for two groups provided by George and Mallery (1999) is used to analyze kurtosis and skewness coefficients, Q-Q plots, the Shapiro-Wilk or Kolmogorov-Smirnov statistic, and bivariate scatterplots. A procedure programmed by Thompson (1990) is used to explore multivariate normality by plotting Mahalanobis distances against derived chi-square values in a scatterplot.

Reality is complex. Over time, researchers in the social sciences have become increasingly aware that simple univariate methods comparing an experimental group with a control group on a single dependent variable are inadequate to meet the needs of the complex phenomena that dominate educational and psychological research. In the majority of social science research, two or more dependent variables are necessary, because nearly every effect has multiple causes and nearly every cause has multiple effects. Even when studying a single construct, such as self-concept, it is often helpful to use multiple tools to measure elusive constructs (called "multi-operationalizing").

In a methodological shift that increasingly emphasizes honoring the complexity of reality, Grimm and Yarnold (1995) reported that the use of multivariate statistics in research has accelerated in the last 20 years and that it is difficult to find empirically based research articles that do not employ one or more multivariate analyses. In a comparison of the 1976 and 1992 volumes of the *Journal of Consulting and Clinical Psychology* (JCCP) Grimm and Yarnold found that the use of multivariate statistics in JCCP increased from 9% to 67% in that 16 year period.

Daniel (1990) noted that multivariate methods usually best honor the reality about which the researcher wishes to generalize. McMillan and Schumacher (1984) compellingly argued against the limitations of viewing the world through an overly-simplified univariate lens:

Social scientists have realized for many years that human behavior can be understood only by examining many variables at the same time, not by dealing with one variable in one study, another variable in a second study, and so forth. These [univariate] procedures have failed to reflect our current emphasis on the multiplicity of factors in human behavior. In the reality of complex social situations the researcher needs to examine many variables simultaneously. (pp. 269-270)

Thompson (1986, p. 9), stated that the reality about which most researchers strive to generalize is usually one "in which the researcher cares about multiple outcomes, in which most outcomes have multiple

causes, and in which most causes have multiple effects." Given this conception of reality, only multivariate methods honor the full constellation of inter-relating variables *simultaneously*.

## Experimentwise Error Rates

Whereas "testwise" error rates refer to the probability of making a Type I error for a given hypothesis test, "experimentwise" error rates refer to the probability of having made a Type I error *anywhere* within the study. Inflation of "experimentwise" error rates can be attributed to two factors: (a) the number of dependent variables in the study; and (b) the amount of correlation between the factors--if two factors are perfectly correlated there is no inflation. On the other extreme, very low correlations produce highly inflated "experimentwise" error rates. The Bonferroni inequality can be used to calculate the "experimentwise" error rate when the hypotheses or variables tested using a single sample are perfectly uncorrelated:

$$\alpha_{EW} = 1 - (1 - \alpha_{TW})^K$$

As noted by Thompson (1994):

... if three perfectly uncorrelated hypotheses (or dependent variables) are tested using a single sample, each at the  $\alpha_{TW} = .05$  level of statistical significance, the "experimentwise" Type I error rate will be:

$$\begin{aligned}\alpha_{EW} &= 1 - (1 - \alpha_{TW})^K \\ &= 1 - (1 - .05)^3 \\ &= 1 - (.95)^3 \\ &= 1 - (.95)(.95)(.95) \\ &= 1 - (.9025)(.95) \\ &= 1 - .857375\end{aligned}$$

$$\alpha_{EW} = 0.142625$$

Thus, for a study testing three perfectly uncorrelated dependent variables, each at the  $\alpha_{TW} = .05$  level of statistical significance, the probability is .142625 (or 14.265%) that one or more null hypotheses will be incorrectly rejected within the study. Most unfortunately, knowing this will not inform the researcher as to which one or more of the statistically significant hypotheses is, in fact, a Type I error.

As illustrated by Fish (1988) and Maxwell (1992) using heuristic examples, invoking multiple univariate tests instead of multivariate tests can also lead unwary researchers to fail to identify statistically significant results. The wrong-headed use of the so-called "Bonferroni correction" coupled with use of univariate tests is also inappropriate, because the application (a) severely attenuates power and (b) still does not honor a multivariate reality. Multivariate analyses can detect interaction effects between independent variables that would go undetected if multiple univariate measures were used in place of multivariate measures. Independent variables may have small, but noteworthy effects on multiple dependent variables that add up to an important pattern when examined as a composite, but otherwise appear meaningless in a univariate test (or series of tests) of a single dependent variable.

### Assumptions of Multivariate Statistics

Because use of multivariate statistics has become commonplace, it is imperative that researchers understand and honor the central assumptions that guide their use. The first assumption of most multivariate statistics is that the variance/covariance matrices across the  $k$  groups must be homogeneous (equal); and the second assumption, which is the focus of this paper, is that the interval response variables across the  $k$  groups must be multivariate normally distributed. The test for homogeneity of variance in multivariate statistics is Box's  $M$  (Box, 1949; 1954), which is a statistically powerful test of bivariate correlations (unstandardized  $r$ ) that is analogous to the Levene test in univariate analyses. If Box's  $M$  is favorable, you *do not reject* the homogeneity of variance assumption, which means that you have met the first assumption of multivariate analyses. Box's  $M$  tests the first assumption, but it is also sensitive to the second assumption of multivariate normality. In other words, if you don't reject the homogeneity of variance assumption, you may have a problem with multivariate normality (see Tabachnick & Fidell, 1983; 1989; 1996 for a detailed elaboration of the homogeneity of covariance assumption).

### Univariate Normality

Determining univariate normality is helpful when assessing multivariate normality, because one can do so even with a small sample size ( $n < 25$ ) and because univariate normality is a necessary precondition for multivariate normality (Gnanadesikan, 1977; Johnson & Wichern, 1992). The advantage of proceeding from a univariate to bivariate to multivariate examination of the data is that such a procedure provides useful information on which dependent variables to use before conducting a multivariate analysis. In order to build a foundation for a complete understanding of multivariate normality, a brief review of univariate normality is in order.

Parametric tests require that the sample data be drawn from a population with a known form, most typically the normal distribution, so that at least one population parameter can be estimated from the sample (Munro & Page, 1993). As noted by Bump (1991), the normal curve is determined by a mathematical equation that uses the mean and standard deviation values to determine two additional statistics--skewness and kurtosis. Both statistics are used to assess the normality of a univariate distribution. Skewness refers to the degree of symmetry of the distribution. Kurtosis refers to the shape of the distribution against the normal distribution, by comparing relative height to width. The mean and standard deviation are used to convert the measured scores to z-scores, which are then used to compute the skewness, as explained by Glass and Stanley (1970, p. 91):  $K_x = ((\sum Z_i^4)/n)$ , most researchers and statistical packages, however, apply an additive constant of (-3) so that the skewness will be equal to 0 in a univariate normal distribution."

However, Glass and Stanley (1970) noted that in a univariate distribution, skewness has a very minor effect on alpha or power in ANOVA if the design is balanced (i.e. there are an equal number of observations in each cell) and kurtosis also has a very slight effect on alpha levels and only effects the power of a test when the distribution is platykurtic (flattened as compared to the normal distribution). The severity of the effect of kurtosis on power increases proportionately with the presence of kurtosis in more than one variable.

### Graphical and Statistical Tests of Univariate Normality

According to Stevens (1996), one of the most popular graphical methods for testing univariate normality is the normal probability plot or Q-Q Plot (quantile-versus-quantile) in which observations are ordered in increasing degree of magnitude and then plotted against expected normal distribution values. Three additional graphical tests are the box-and-whisker plot, stem-and-leaf plot, and a histogram of the dependent variables. These tests allow a quick and simple means of evaluating the shape of the univariate distribution for each dependent variable. Stevens (1996) recommends that with samples of less than 50 cases, prudent researchers use non-graphical tests such as the chi-square goodness of fit, Kolmogorov-Smirnov test, the Shapiro-Wilk test, *and* an evaluation of the skewness and kurtosis of the distribution to make an evaluation about univariate normality. The Shapiro-Wilk test (Wilk, Shapiro, & Chen, 1968) was developed to detect a wide variety of variations from a normal univariate distribution. The smaller the  $W$  value, the greater the departure from normality. As a guideline, Gnanadesikan (1977) stated that for  $p_{\text{calculated}}$  values of 0.1 or higher, normality is a reasonable assumption.

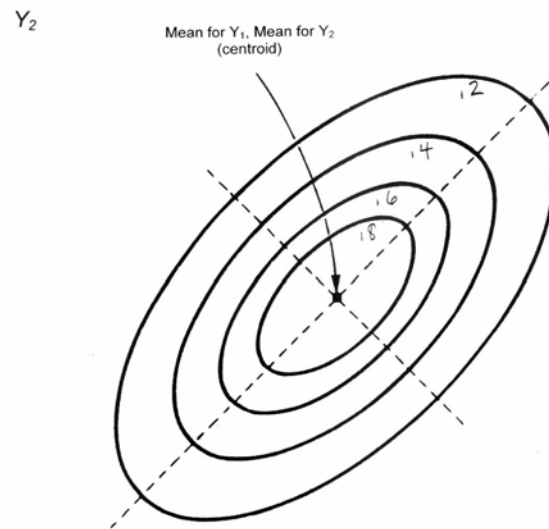
Wilk, Shapiro, and Chen (1968) concluded that for sample sizes under 20, the combination of the skewness and kurtosis coefficients or the Shapiro-Wilk method were most sensitive to detecting extreme non-normality. Stevens (1996) recommended that researchers evaluate univariate normality by examining the Shapiro-Wilk statistic and examining the kurtosis and skewness coefficients (along with their standard errors) because Shapiro-Wilk has the most power and a review of the skewness and kurtosis can help determine the cause of non-normality whenever it is present. The Shapiro-Wilk test is recommended for samples of less than 25 and the Kolmogorov-Smirnov test is recommended for samples greater than 25. Both the Shapiro-Wilk and the Kolmogorov-Smirnov tests perform an aggregate test of skewness and kurtosis in the univariate case. You do not want to find statistical significance because the null says the distribution is normal and you do not want to reject the assumption of normality.

### Bivariate Normality

As noted by Stevens (1996), in addition to establishing univariate normality, two additional characteristics of a normal multivariate distribution are that the linear relationship of any combination of variables is distributed normally, and that all possible subsets of the sets of variables are normally distributed. The relationship between bivariate and multivariate normality is complex. Statistical significance tests like those used in MANOVA require that the distribution of each dependent variable are normally distributed about each of the other dependent variables in any " $X_1$  and  $X_2$ " comparison.

Two distributions that are univariate normal might also be bivariate normal, but just because two distributions are univariate normal does not mean that they will be bivariate normal. In a bivariate comparison, we compare each person's score on two measures, so we are thinking in three dimensions--the X-axis, Y-axis and a third axis to demonstrate frequency of scores. This requirement means that a circular or elliptical pattern will emerge in a scatterplot when examining the correlation of any two dependent variables in a bivariate normal distribution. The narrower the ellipse in the bivariate scatterplot, the greater the correlation between the dependent variables, and subsequently, the greater the likelihood that the assumption of multivariate normality will hold.

Figure 1 is a graphical representation of a bivariate frequency distribution in two-dimensional form. In this drawing, the viewer is looking down at the distribution from above. The largest concentric circle is the footprint or floor of the bell or mound. The footprint of the bell is not a circle in this example, because the standard deviation for each person on the X-axis is roughly twice as large as the standard deviation on the Y-axis. A series of *contour lines* is used to demonstrate a series of ellipses with varying amounts of distance from the common center, called the *centroid*.

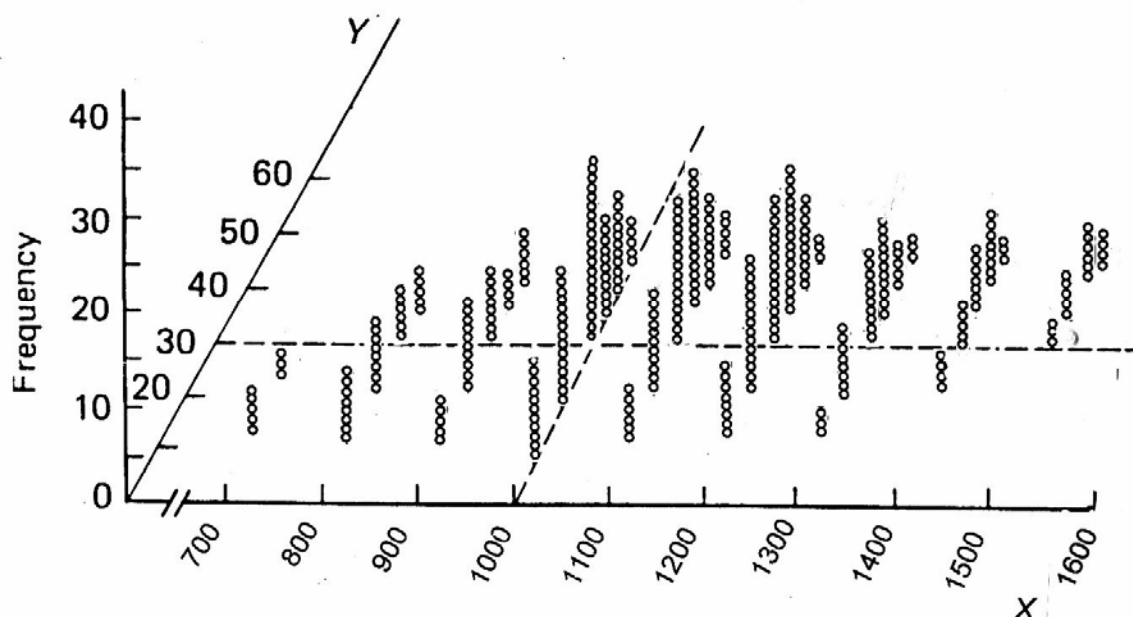


**Figure 1.** Contour Diagram for a Bivariate Normal Surface

The advantage of drawing the centroid with contour lines is that you can graphically demonstrate the probability that a random bivariate observation (plotted on the  $X_1X_2$  plane) will lie within the elliptical region, which is equivalent to the area under a portion of the normal curve in a univariate distribution of scores (Neter, Kutner, Nachtsheim, & Wasserman, 1996).

Statistical significance testing applies to the bivariate case in terms of the distance from the centroid or Cartesian coordinate for each person on the X and Y axes. The closer the scores aggregate toward the centroid, the greater the chance of being included in the sample because of nearness to the Cartesian coordinate. The first contoured line shows a value of .8 meaning there is an 80% chance of being included in the sample. The last contoured line has a value of .2 meaning that there is only a 20% chance of being included in the sample.

If a group of 400 people is measured in two ways--for example, each person's composite (Verbal + Quantitative) GRE score (X) and self-esteem (Y)--the data can be represented in a bivariate frequency relationship as shown in Figure 2. If we had bivariate normality, the circles would be concentric in a sense. We are comparing two variables, but have three axes. The third axis is height, which graphically shows the frequency of the bivariate scores. In this example, height is a measure of frequency and not a third variable. For each person, there is a pair of scores, a score on X and a score on Y. A bivariate frequency distribution is a picture of the frequency with which different pairs of X and Y scores occur in a group of persons. In Figure 2, a bivariate frequency distribution is displayed for about 400 people on GRE Composite (Verbal + Quantitative) Score (X-axis) and self-esteem (Y-axis). In this example, the highest frequency of scores is a GRE Composite Score of 1000 and a self-



**Figure 2.** Bivariate Frequency distribution for Persons Measured on Total GRE (X) and Self-Esteem (Y).

esteem score of 30. This point is the Cartesian coordinate for the two sets of scores and also forms the highest point of the distribution of scores. When the height of the line is compared to the vertical scale of frequency, we can determine that approximately 20 persons had a composite GRE score of 1000 and a self-esteem score of 30.

A surface or "roof" drawn on the top of a large number of scores in a bivariate frequency distribution takes the shape of a three-dimensional bell or hat as demonstrated in Figure 3. The shape is formed by conceptualizing the one-dimensional bell-shaped normal distribution and stretching it in the X and Y directions and rotating it around its center (i.e. the Cartesian coordinate) in the XY plane. All bivariate normal distributions have the following characteristics:

- (a) For each value of X, the distribution of its associated Y value is a normal distribution and vice-versa.
- (b) The Y means for each value of X are linear (i.e., they fall on a straight line) and the same is true for the X means for each value of Y.
- (c) The scatterplots demonstrate homoscedasticity--the variance in the Y values is uniform across all values of X and the variance in X values is constant for all values of Y.

If you were to multiply all of the z-scores on the X axis by 2 in Figure 3 and place those scores on the Y axis, the base of the three-dimensional bell will be an ellipse instead of a circle because the Y scores will be twice as spread out as the X scores. However, a non-circular base can still be normal because a

multiplicative constant of two will not change the skewness, kurtosis, or mean of zero.

The shape of the mound or hat is determined by the amount of correlation between the two variables. If both dependent variables are expressed in standard deviation units, the more correlated the variables, the narrower the mound or hat because correlation causes the probability to concentrate along a line (see Figure 4;  $r = .8$ ). In the extreme case that dependent variable  $X_1$  is completely correlated with dependent variable  $X_2$ , all points would be exactly on the regression equation, the standard deviation for  $X_1$  and  $X_2$  would be equal to zero and the "contour" would all be straight lines with no areas.

Furthermore, if the distribution is bivariate normal, any plane perpendicular to the  $X_1X_2$  plane will cut the surface into a normal curve and a plane parallel to the  $X_1X_2$  plane will cut in an ellipse. The bivariate normal distribution has the property that the regression of  $X_1$  on  $X_2$  is linear. Therefore, if we have a bivariate normal distribution, we know that if we trace the means of  $X_2$  for each  $X_1$ , the result will be a straight line. It does not necessarily follow, however, that if the regression is linear, the joint distribution is bivariate normal.

### Multivariate Normality

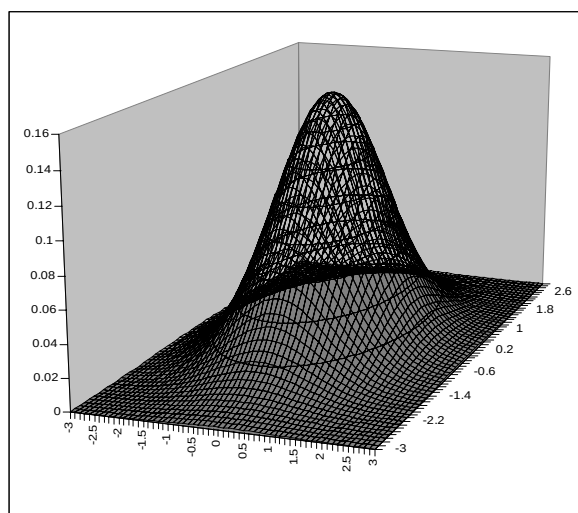
For a data set of two or more dependent variables, all of the variables must be univariate normal and all possible pairs of the variables must also be normal as

necessary but insufficient conditions for multivariate normality. The mathematical model that serves as the basis for MANOVA and other multivariate techniques is based on the multivariate normal distribution. This means that both the sampling distributions of the means of dependent variables in each cell are normally distributed as are the linear combinations of dependent variables. The central limit theorem states that for large samples, the sampling distribution of means in the univariate case will approach normality. Mardia (1971) demonstrated that MANOVA is robust to modest violation of normality if the violation is caused by skewness rather than outliers.

In some instances, researchers can examine multivariate outliers by simply examining z-scores and looking for extreme scores on each dependent variable. However, this technique does not identify a set of scores for a person that are slightly deviant on several variables. Fortunately, a statistic called Mahalanobis distance ( $D^2$ ) can be used to detect scores that deviate from the mean (above or below) for a group of dependent variables as a set. Detecting multivariate outliers from a set of dependent variables is a much subtler process than detecting univariate or bivariate outliers.

The Mahalanobis distance is the distance of a case from the centroid where the centroid is the point defined by the means of all the variables taken as a whole. The Mahalanobis distance demonstrates how far an individual case is from the centroid of all the cases for the predictor variables. When the distance is great, the observation is an outlier. According to Krzanowski (1988) and Stevens (1996), the Mahalanobis distance is accepted by researchers as the measure of distance between two multivariate populations and it is independent of sample size. The Mahalanobis distance can be written in terms of the covariance matrix  $S$  as:

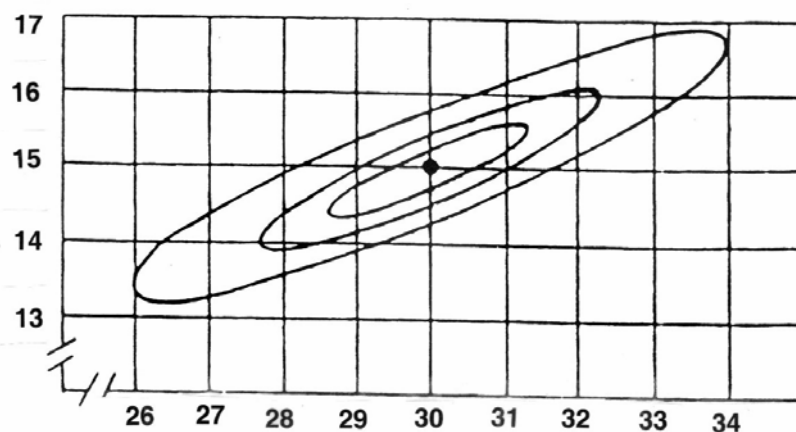
$$D_i^2 = (X_i - \bar{x})' S^{-1} (X_i - \bar{x}),$$



**Figure 3.** Bivariate Normal Distribution

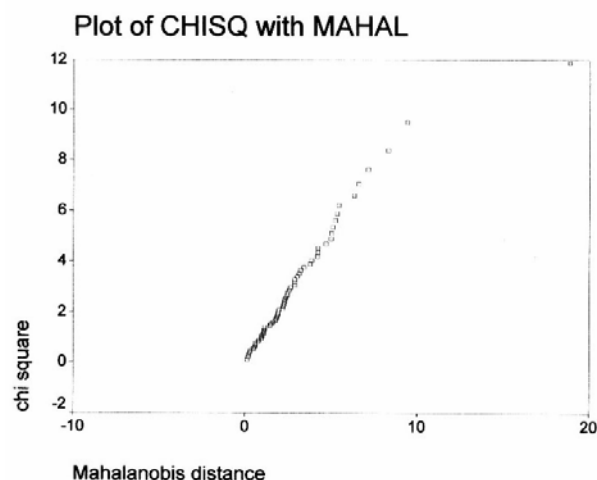
where  $D_i^2$  is the Mahalanobis distance for a given individual,  $S$  is covariance matrix with variances on the diagonal and covariances off the diagonal. The rank for  $S$  is the number of rows and columns for the covariance matrix, which is  $3 \times 3$ , if there are three dependent variables.

The assumption of MANOVA, for example, is that in each group, multivariate normality holds regarding the dependent variables, so if there are a total of 105 cases (as in the heuristic example below) with 64 cases in the female group and 41 cases in the male, both have to have multivariate normality. In group 1, there are three interval variables and the rank of the correlation matrix is  $3 \times 3$ .  $X_i$  is the composite of three scores of a given individual with a rank of  $3 \times 1$ . Person #1 has three scores with one column. The matrix of means also has a rank of  $3 \times 1$  (three means with one column) which yields a product of  $3 \times 1$  and is not conformable to  $3 \times 3$ . The transpose ( $'$ ) notation means you flip the



Mean for  $X_1 = 30$ ,  $SD_1 = 2$ . Mean for  $X_2 = 15$ ,  $SD_2 = 1$ .  $r = 0.8$

**Figure 4.** Elliptical Bivariate Normal Distribution for 2 Variables with Dissimilar Standard Deviations and Means



**Figure 5.** Scatterplot of Chi-Square with Mahalanobis Distance for 64 females without transforming or deleting scores.

3 x 1 and it becomes 1 x 3. The right most part of the matrix is also a 3 x 1 but it does not have a transpose symbol, so it is not flipped on its side.

From the formula, the Mahalanobis distance is descriptive of how far each case's set of scores is from the group means adjusting for correlation of the variables (in the example, a measure of the distance of the each person from the group means adjusted by how correlated the three variables are). In Figure 5, the smallest Mahalanobis distance is for participant #32 because each of the three scores (3.0, 6.1, and 9.8, respectively) is closest to the mean for each variable (2.89; 6.23; and 10.3, respectively).

Having correlated dependent variables is commonplace in social science research. The correlation of dependent variables must be taken into account when calculating the Mahalanobis distance because deviations from the means of two highly correlated dependent variables are partially redundant whereas the deviations from the mean for two highly uncorrelated dependent variables are not redundant. More concretely, say in a set of three dependent variables all with a standard deviation of 5, that the mean of  $X_1$  is 10, the mean of  $X_2$  is 11 and the mean of  $X_3$  is 2,  $X_1$  is highly correlated with  $X_2$  but  $X_1$  is highly uncorrelated with  $X_3$  and  $X_2$  is highly uncorrelated with  $X_3$ . If person #1 has a score one standard deviation above the mean on  $X_1$  ( $X_1=15$ ) and  $X_2$  ( $X_2=16$ ) and scores at the mean of  $X_3$  ( $X_3=7$ ), that person will have a smaller  $D^2$  than person #2 who scores at the mean on  $X_1$  ( $X_1=7$ ) and one standard deviation above the mean on  $X_2$  ( $X_2=16$ ) and  $X_3$  ( $X_3=7$ ). The  $D^2$  for person #1 includes redundant distance from the means because the scores on  $X_1$  of 15 and  $X_2$  of 16 are very similar. In a sense,  $X_1$  and  $X_2$  are measuring the same thing, so the deviation from the means is due in part to similarity in the variables. Person #1 will have a lower  $D^2$  because the deviation from the means is redundant whereas the  $D^2$  for person #2 will be much greater because the

Mahalanobis distance is not due to distance from similar means of the variables but rather to substantial distance from *dissimilar* means ( $X_1=10$ ;  $X_2=16$ ;  $X_3=7$ ).

There are two evaluations to be done when examining the Mahalanobis distance by chi-square scatterplot--the first is whether or not the points form a straight line or not. If the points on the scatterplot form a straight line, you have multivariate normality. The second consideration is whether or not there are anomalous persons with scores on the scatterplot that are a noteworthy distance from the centroids. You can have a perfectly straight line and still have outliers in the data set, but it is rare to have a person whose scores are outlying on all of the dependent variables in a data set. Before eliminating outliers, a prudent researcher will examine whether or not the extreme score on the multivariate scatterplot is due to an anomalous score on one dependent variable by examining each univariate distribution before eliminating the person from the data set. If only one score is anomalous, it is more prudent to transform the score on that variable rather than eliminate valuable information from the analysis, or to eliminate that variable from the data set.

#### Evaluating Univariate Normality: A Heuristic Example

To make the discussion about testing bivariate and multivariate normality more concrete, a data set developed by George and Mallery (1999) will be analyzed using SPSS version 8.0 to test the distribution of scores for 64 female and 41 male students taught by the same professor in three sections of a course. The three dependent variables in this analysis are each student's GPA previous to taking the course (PREVGPA), final exam grade (FINAL) and total points for the course (TOTAL). In such a data set, it might be interesting to examine the differences between males and females (an independent variable with two levels) on all three dependent variables--previous GPA, final exam grade, and total points in the course. The SPSS syntax for the female group ( $n = 64$ ) appears in Appendix A and the syntax for the male group ( $n = 41$ ) appears in Appendix B. For the sake of brevity and clarity, univariate normality will be assumed and only the bivariate and multivariate output from the female group will be analyzed in detail in this paper.

As noted by Marascuilo and Levin (1983), multivariate normality is a requirement for utilizing the statistical inference procedure that is the basis of all "OVA" designs. The test for univariate normality for the grades data for the female group was done by using the MULTINOR program developed by Thompson (1990) on SPSS 8.0 (Appendix A). The MULTINOR program generates graphical and non-graphical analyses of the distribution of each dependent variable separately.

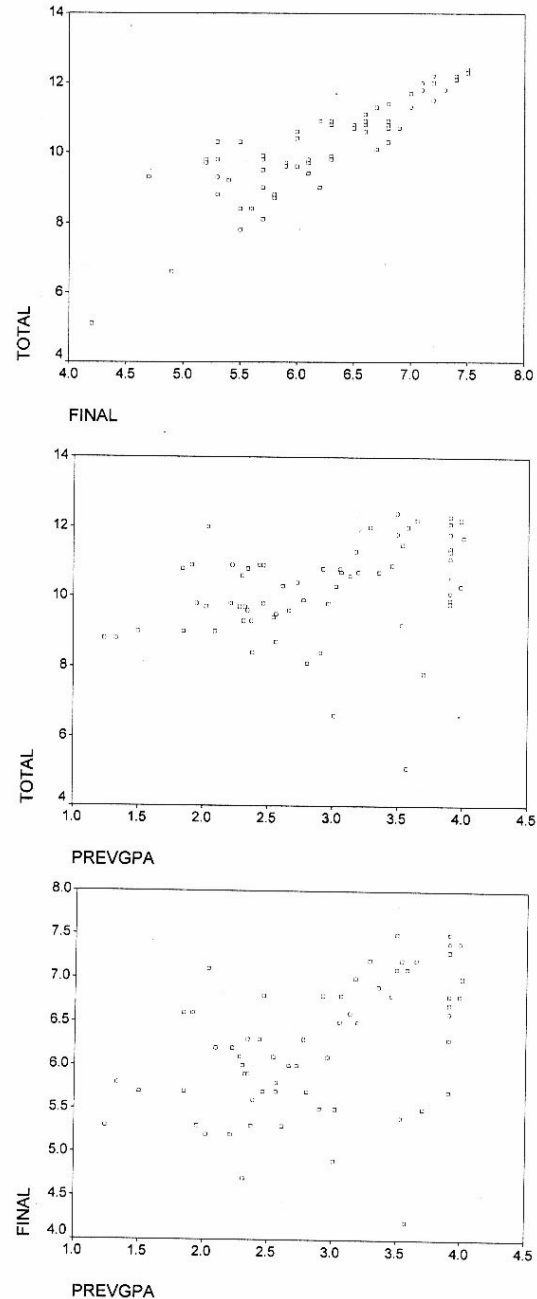
### Bivariate Normality

If the three dependent variables displayed univariate normality (bearing in mind that univariate normality is a necessary, but insufficient foundation for multivariate normality), the next step would be to examine the bivariate correlations between each of the dependent variables. You can attain univariate normality, but fail to demonstrate bivariate normality, which examines each pair of variables--PREVGPA with FINAL, PREVGPA with TOTAL and FINAL with TOTAL. This was done in this example by using the MULTINOR program (Appendix B) by requesting scatterplots and noting elliptical patterns for the three possible combinations of variables. In Figure 6, the scatterplot for each possible pair reveals a clear elliptical pattern between FINAL and TOTAL, but the scores in the scatterplots for PREVGPA with FINAL and PREVGPA with TOTAL are widely scattered and are thus *not* bivariate normal. When the pattern of the scores in a bivariate plot are less clear, researchers can examine the percentage of scores that converge around the centroid (e.g., 80%, 60%, 40%, 20%, 10%) as a guide to deciding whether or not an elliptical pattern is displayed.

At this stage of the analysis, a prudent researcher might stop and consider replacing PREVGPA with another dependent variable or go back and transform the scores in each of the univariate distributions to make them more normal. As noted earlier, Tabachnick and Fidell (1996) recommended that researchers start by taking the square root of the scores, but the scores can also be squared, or the natural log or log-ten (LG10) can be used:

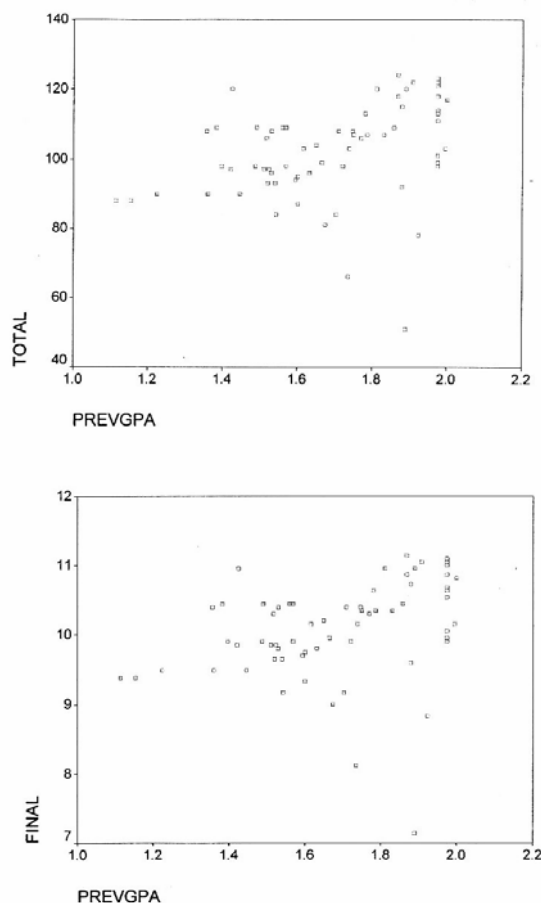
...transformations may improve the analysis, and may have the further advantage of reducing the impact of outliers. Our recommendation, then, is to consider transformation of variables in all situations unless there is some reason not to. If you decide to transform, it is important to check that the variable is normally or near-normally distributed after transformation. Often you need to try one transformation and then another until you find the transformation that produces the skewness and kurtosis values nearest zero, the prettiest picture, and/or the fewest outliers. (p. 82)

After transforming the univariate distributions, the bivariate distributions could be examined again to determine if the three pairs of variables have become bivariate normal due to the univariate transformation of scores. For this set of scores, four data transformations were conducted: (a) square root of scores (Figure 7), (b) squared scores (Figure 8) (c) natural log (Figure 9), and (d) log 10 (Figure 10). In none of these transformations did the bivariate relationships between PREVGPA and TOTAL or PREVGPA and FINAL become bivariate normal. Because PREVGPA appear-



**Figure 6.** Bivariate Scatterplots of PREVGPA, TOTAL, and FINAL.

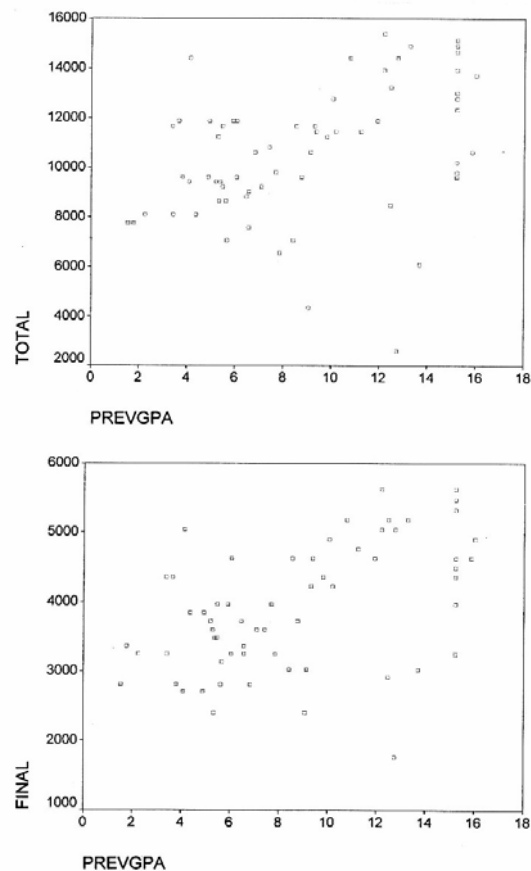
ed to be the problematic DV, a decision was made to create a new DV that was comprised of the sum of the quiz grades in the course. This new DV was named QUIZTOT and a new evaluation of univariate, bivariate, and multivariate normality was conducted as before. The syntax commands for the new variable are shown in Appendix D. Figure 11 shows that the variable QUIZTOT has a bivariate normal relationship with both FINAL and TOTAL and is a big improvement over the variable PREVGPA.



**Figure 7.** Bivariate Scatterplots of PREV GPA with TOTAL and FINAL (square-root transformation).

### Multivariate Normality

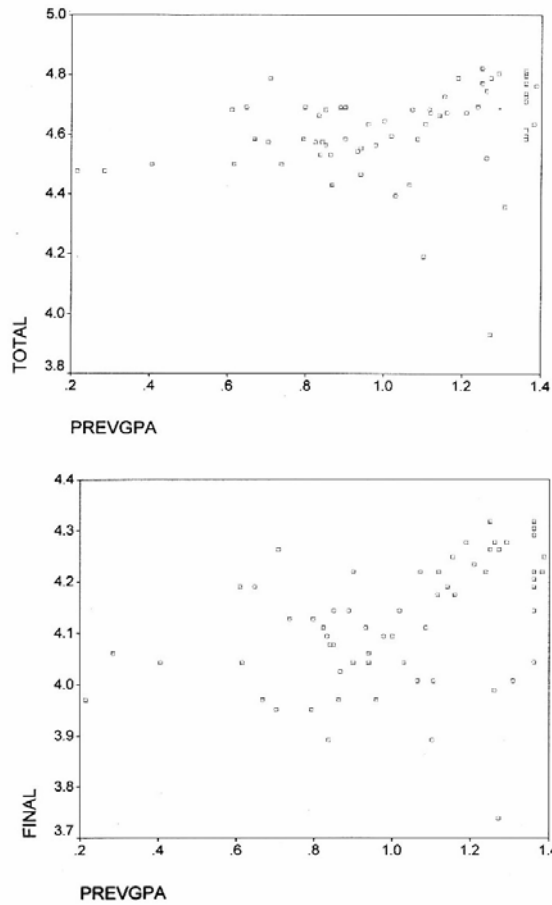
Assuming that both univariate and bivariate normality are attained after transforming the univariate scores or replacing a dependent variable (as done in this example), the third level of assessment is to examine the Mahalanobis distance by chi-square scatterplot to assess multivariate normality. As noted earlier, the Mahalanobis distance is accepted by researchers as the measure of distance between two multivariate populations and it is independent of sample size (Krzanowski, 1988; Stevens, 1996). If we examine the scatterplot of Mahalanobis distance ( $D^2$ ) values with chi-squares (Thompson, 1990) for this data set in Figure 12 we can see that we have a fairly straight line, which suggests multivariate normality. The second issue is the presence of outliers. This scatterplot has one extreme score in the upper right hand corner that is well off the line with an approximate  $D^2$  score of 62 and a chi-square score of about 12. If we look at the listing of Mahalanobis distances which are ranked from lowest to highest in Figure 16, we can determine that the outlier is case #61 and the  $D^2$  is more than four times larger than the next



**Figure 8.** Bivariate Scatterplots of PREV GPA with TOTAL and FINAL (squared-score transformation).

largest  $D^2$  (case #36). Because case #36 in turn is twice as large as the next largest  $D^2$  (case #45), both case #61 and #36 can be considered outliers. Again, assuming univariate and bivariate normality has been demonstrated, because we have multivariate normality except for two outliers, we can remove or transform the outliers and then look at the univariate and bivariate relationships again because removal of the extreme scores will change the means for both variable X and variable Y, which means that the Mahalanobis distance for each variable will change. If after examining the raw data for case #36 and #42, we discover that they both had very high quiz scores (QUIZTOT) and very low scores on the FINAL, we might call these two students and ask why they did so poorly on the final exam. If we learn that they both had the flu the day of the exam, but took the exam anyway, we might delete their scores from the data set because their illness likely produced "fluky" or abnormal scores (i.e. high quiz scores and low final exam scores). Figure 13 shows the Mahalanobis distance and chi-square values for this data set after the outliers are re-

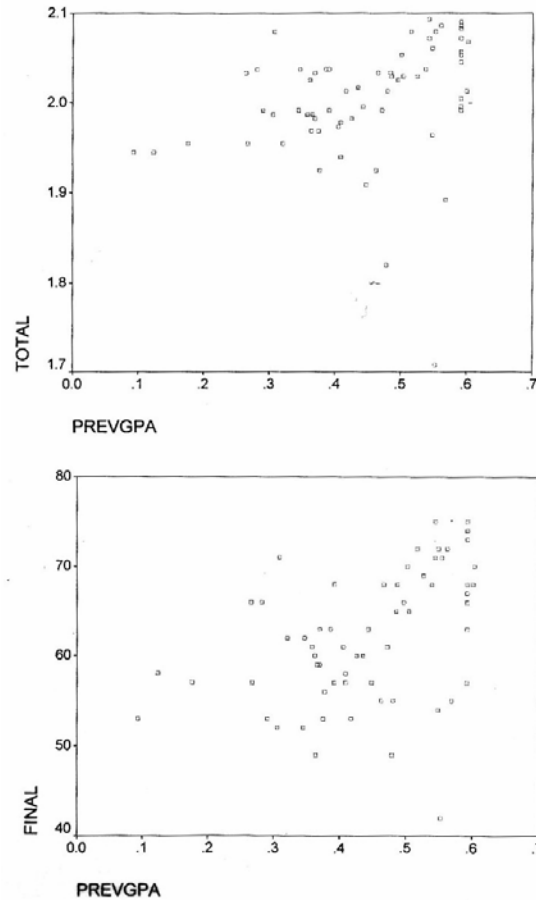




**Figure 9.** Bivariate Scatterplots of PREV GPA with TOTAL and FINAL (natural log transformation).

moved. Note that while the line appears to become less straight, in actuality the scale for the Mahalanobis distance is being reduced from 70 units to 12 units, thus showing more precisely the linear relationships between the two variables.

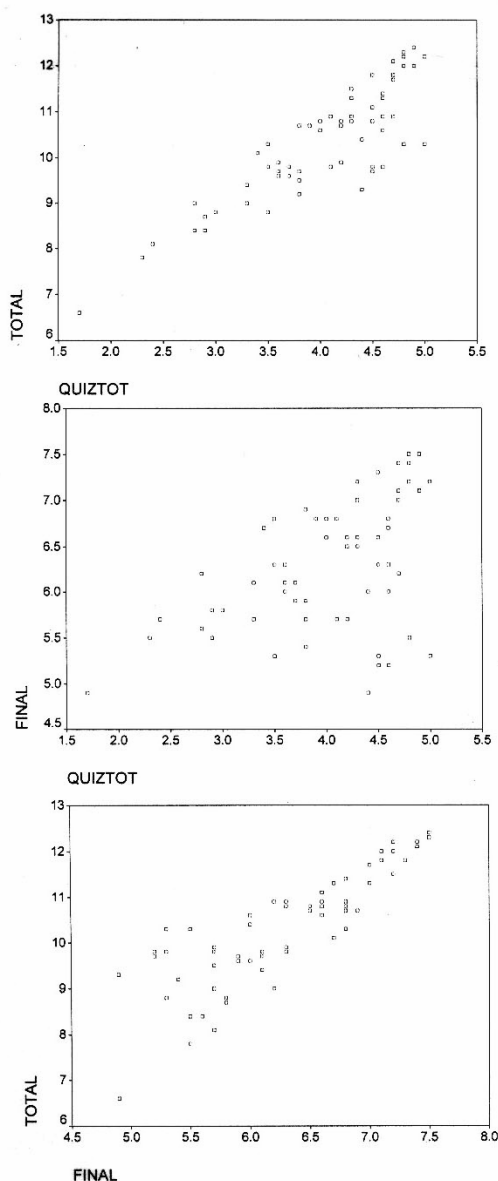
An alternative to the stair-step approach of examining the univariate, bivariate, and multivariate normality of the proposed dependent variables in sequence for the multivariate analysis is to plot the Mahalanobis distance against the chi-square values straight away--if you get a straight line, you can stop there because multivariate normality subsumes univariate and bivariate normality. However, plotting Mahalanobis distance against chi-square is only useful with samples greater than 25. If you fail to obtain a straight line, you can remove scores when you can justify doing so, or transform an individual's scores or a set of scores.



**Figure 10.** Bivariate Scatterplots of PREV GPA with TOTAL and FINAL (log-10 transformation).

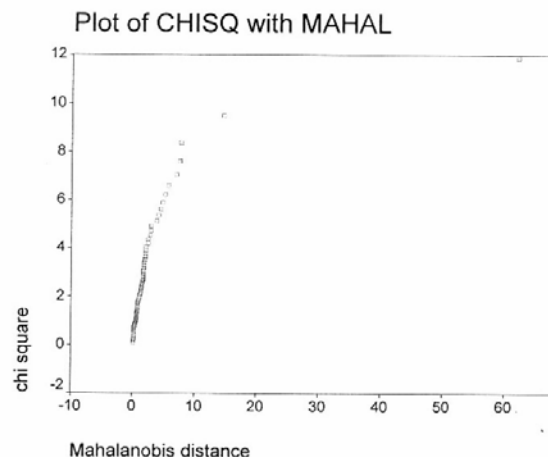
### References

- Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria, *Biometrika*, 36, 317-346.
- Box, G.E.P. (1954). Some theorems on quadratic forms applied in analysis of variance problems: II. Effect of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, 25, 484-498.
- Bump, W. (1991, January). *The normal curve takes many forms: A review of skewness and kurtosis*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio. (ERIC Document Reproduction Service No. ED 342 790)
- Daniel, L.G. (1990, January). *Use of structure coefficients in multivariate educational research: A heuristic example*. Annual Meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 315 451)

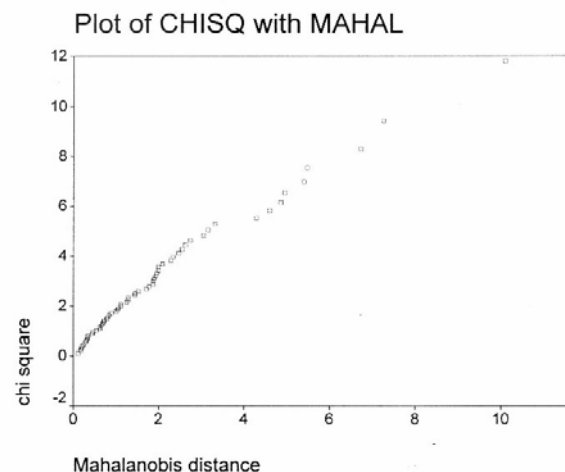


**Figure 11.** Bivariate Scatterplots of QUIZTOT with TOTAL and FINAL.

- Fish, L. (1988). Why multivariate methods are usually vital. *Measurement in Evaluation and Counseling and Development*, 21, 130-137.
- George, D., & Mallery, P. (1999). *SPSS for WINDOWS step by step*. Boston: Allyn & Bacon.
- Glass, G.V., & Stanley, J.C. (1970). *Statistical methods for education and psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Gnandesikan, R. (1977). *Methods for statistical analysis of multivariate observations*. New York: Wiley.
- Grimm, L.G., & Yarnold, P.R.. (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.



**Figure 12.** Scatterplot of Chi-Square with Mahalanobis Distance for 64 females after replacing PREVGPA with QUIZTOT.



**Figure 13.** Scatterplot of Chi-Square with Mahalanobis Distance for 64 females after replacing PREVGPA with QUIZTOT and deleting two outliers.

- Johnson, N., & Wichern, D. (1988). *Applied multivariate statistical analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Krzanowski, W.J. (1995). *Recent advances in descriptive multivariate analysis*. Oxford: Clarendon
- Marascuilo, L.A., & Levin, J.R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Monterey, CA: Brooks/Cole.
- Mardia, K.V. (1971). The effect of non-normality on some multivariate tests and robustness to non-normality in the linear model. *Biometrika*, 58, 105-121.
- Maxwell, S. (1992). Recent developments in MANOVA applications. In B. Thompson (ed.), *Advances in social science methodology* (Vol. 2, pp. 137-168). Greenwich, CT: JAI Press.

- McMillan, J.H., & Schumacher, S. (1984). *Research in education: A conceptual approach*. Boston: Little, Brown.
- Munro, B., & Page, E. (1993). *Statistical methods for health care research* (2nd ed.). Philadelphia: J. B. Lippincott.
- Neter, J., Kunter, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear statistical models* (4<sup>th</sup> ed.). Chicago: Irwin.
- Stevens, R. (1991). *Applied multivariate statistics for the social sciences* (2<sup>rd</sup> ed.). Mahwah, NJ: Erlbaum.
- Stevens, R. (1996). *Applied multivariate statistics for the social sciences* (3<sup>rd</sup> ed.). Mahwah, NJ: Erlbaum.
- Tabachnick, B.G., & Fidell, L.S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper & Row.
- Tabachnick, B.G., & Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.
- Thompson, B. (1986, November). *Two reasons why multivariate methods are usually vital*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis.
- Thompson, B. (1990). MULTINOR: A FORTRAN program that assists in evaluating multivariate normality. *Educational and Psychological Measurement*, 50, 845-848.
- Thompson, B. (1994, February). Why multivariate methods are usually vital in research: Some basic concepts. In *Paper presented as a Featured Speaker at the biennial meeting of the Southwestern Association for Research in Human Development, Austin, T. (ERIC Document Reproduction Service No. ED 367 687)*

---

### Appendix A

#### SPSS Commands for Female Group (n=64)

```

SET BLANKS=SYSMIS UNDEFINED=WARN printback=list.
TITLE 'MULTINOR.SPS  tests multivar normality graphically*****'.
COMMENT The original MULTINOR computer program was presented, with examples, in:
COMMENT Thompson, B. (1990). MULTINOR: A FORTRAN program that assists in
COMMENT evaluating multivariate normality. Educational and Psychological Measurement, 50, 845-848.
COMMENT The data source for the example are from: George, D. J., and Mallery, P. (1999). SPSS for
COMMENT Windows step by step. Boston: Allyn & Bacon.
COMMENT Here there are 3 variables for which multivariate normality is being confirmed.
DATA LIST
  FILE='a:normgrad.dat' FIXED RECORDS=1 TABLE
  /1 gender 1 ethnict 3 year 5 lowup 7 section 9 prevgpa 11-14 (1) final 16-17 (1) total 19-21 (1) .
list variables=all/cases=9999/format=numbered .
COMMENT 'y' is a variable automatically created by the program, and does not have to modified for different data sets.
select if (gender eq 1) .
compute y=$casenum .
print formats y(F5) .
regression variables=y prevgpa to total/
  descriptive=mean stddev corr/
  dependent=y/enter prevgpa to total/
  save=mahal(mahal) .
sort cases by mahal(a) .
execute .
list variables=y prevgpa to total mahal/cases=9999/format=numbered .
COMMENT In the next TWO lines, for a given data put the actual n in place of the number '64' used for the example data set.
loop #i=1 to 64 .
  compute p=($casenum - .5) / 64.
COMMENT In the next line, change '3' to whatever is the number of variables. The p critical value of
COMMENT chi square for a given case is set as [the case number (after sorting) - .5] / the sample size].
  if (gender eq 1) chisq=idf.chisq(p,3) .
end loop .
print formats p chisq (F8.5) .
list variables=y p mahal chisq/cases=9999/format=numbered .
plot
  vertical='chi square'/
  horizontal='Mahalanobis distance'/
  plot=chisq with mahal .

```

## Appendix B

### SPSS Commands for Male Group

```

SET BLANKS=SYSMIS UNDEFINED=WARN printback=list.
TITLE 'MULTINOR.SPS  tests multivar normality graphically*****'.
COMMENT *****.
COMMENT The original MULTINOR computer program was presented, with examples, in:
COMMENT   Thompson, B. (1990). MULTINOR: A FORTRAN program that assists
COMMENT   in evaluating multivariate normality. Educational and Psychological Measurement, 50, 845-848.
COMMENT   The data source for the example are from:
COMMENT   George, D. J., and Mallery, P. (1999). SPSS for Windows step by step. Boston: Allyn & Bacon.
COMMENT   Here there are 3 variables for which multivariate normality is being confirmed.
DATA LIST
  FILE='a:normgrad.dat' FIXED RECORDS=1 TABLE
    /1 gender 1 ethnicit 3 year 5 lowup 7 section 9 prevgpa 11-14 (1)  final 16-17 (1)
  total 19-21 (1) .
list variables=all/cases=9999/format=numbered .
COMMENT 'y' is a variable automatically created by the program,
COMMENT and does not have to be modified for different data sets.
select if (gender eq 2) .
compute y=$casenum .
print formats y(F5) .
regression variables=y prevgpa to total/
  descriptive=mean stddev corr/
  dependent=y/enter prevgpa to total/
  save=mahal(mahal) .
sort cases by mahal(a) .
execute .
list variables=y prevgpa to total mahal/cases=9999/format=numbered .
COMMENT In the next TWO lines, for a given data set put the
COMMENT actual n in place of the number '41' used for the
COMMENT example data set.
loop #i=1 to 41 .
compute p=($casenum - .5) / 41.
COMMENT In the next line, change '3' to whatever is the number
COMMENT of variables.
COMMENT The p critical value of chi square for a given case
COMMENT is set as [the case number (after sorting) - .5] / the
COMMENT sample size].
if (gender eq 2) chisq=idf.chisq(p,3) .
end loop .
print formats p chisq (F8.5) .
list variables=y p mahal chisq/cases=9999/format=numbered .
plot
  vertical='chi square'/
  horizontal='Mahalanobis distance'/
plot=chisq with mahal .

```

## Appendix C

### SPSS Syntax for Evaluating Univariate and Bivariate Normality

```

PLOT
/VARIABLES=prevgpa
/NOLOG
/NOSTANDARDIZE
/TYPE=Q-Q
/TIES=MEAN
/DIST=NORMAL .
GRAPH
/HISTOGRAM=prevgpa .
EXAMINE
VARIABLES=prevgpa final total
/PLOT BOXPLOT STEMLEAF HISTOGRAM NPLOT
/COMPARE GROUP
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL .
GRAPH
/SCATTERPLOT (BIVAR)=prevgpa WITH total
/MISSING=LISTWISE .
PLOT
/VERTICAL='prevgpa' REFERENCE (6,4)
/HORIZONTAL='total' REFERENCE (6,7)
/PLOT=prevgpa WITH total .
GRAPH
/SCATTERPLOT (BIVAR)=prevgpa with final
/MISSING=LISTWISE .
PLOT
/VERTICAL='prevgpa' REFERENCE (6,4)
/HORIZONTAL='final' REFERENCE (6,9)
/PLOT=prevgpa WITH final .
GRAPH
/SCATTERPLOT (BIVAR)=final with total
/MISSING=LISTWISE .
PLOT
/VERTICAL='final' REFERENCE (6,9)
/HORIZONTAL='total' REFERENCE (6,7)
/PLOT=final WITH total .
COMMENT is set as [the case number (after sorting) - .5] / the
COMMENT sample size].
compute p=($casenum - .5)/62 .
compute chisq=idf.chisq(p,3) .
end loop .
print formats p chisq (F8.5) .
list variables=y p mahal chisq/cases=9999/format=numbered .
plot
vertical='chi square'/
horizontal='Mahalanobis distance'/
plot=chisq with mahal .

```

## Appendix D

### SPSS Commands for New Dependent Variable

```

SET BLANKS=SYSMIS UNDEFINED=WARN printback=list.
TITLE 'MULTINOR.SPS  tests multivar normality graphically****'.
COMMENT *****
COMMENT  The original MULTINOR computer program was presented,
COMMENT  with examples, in: Thompson, B. (1990). MULTINOR: A FORTRAN program that assists
COMMENT  in evaluating multivariate normality. Educational and Psychological Measurement, 50, 845-848.
COMMENT  The data source for the example are from:
COMMENT  George, D. J., and Mallery, P. (1999). SPSS for Windows step by step. Boston: Allyn & Bacon.
COMMENT *****
COMMENT  Here there are 3 variables for which multivariate normality is being confirmed.
DATA LIST
  FILE='a:norgrades.txt' FIXED RECORDS=1 TABLE
  /1 quiztot 1-2 (1) final 4-5 (1) total 7-9 (1) .
list variables=all/cases=9999/format=numbered .
COMMENT  'y' is a variable automatically created by the program,
COMMENT  and does not have to be modified for different data sets .
compute y=$casenum .
execute .
print formats y(F5) .
regression variables=y quiztot to total/
  descriptive=mean stddev corr/
  dependent=y/enter quiztot to total/
  save=mahal (mahal) .
sort cases by mahal(a) .
execute .
list variables=y quiztot to total mahal/cases=9999/format=numbered .
COMMENT  In the next TWO lines, for a given data set put the
COMMENT  actual n in place of the number '62' used for the
COMMENT  example data set.
loop #i=1 to 62 .
COMMENT  In the next line, change '3' to whatever is the number of variables.
COMMENT  The p critical value of chi square for a given case

```