Verran, J.A. & Ferketich, S.L. (1987). Testing linear model assumptions: Residual analysis. Nursing Research, 36, 127-130.

Testing Linear Model Assumptions: Residual Analysis

JOYCE A. VERRAN • SANDRA L. FERKETICH

Analysis of data involves a variety of methods. Classical statistical approaches include procedures that confirm hypothesized relationships. Recent advances in the analysis of research data emphasize the search for possible models that describe the universe of interest. This orientation views the empirical world as offering significant information to the researcher willing to explore the data.

Exploratory data analysis techniques are newer and less widely used by nurse scientists. These techniques are designed to assist in detecting different types of underlying data structures. Exploratory analytic methods are used with flexibility and the unexpected in mind but without restrictive model assumptions as prerequisites (Hartwig & Dearing, 1979). One such exploratory method is residual analysis. Although there are a variety of uses for this technique, the purpose of this article is to describe residual analysis as a method for building a more stable mathematical linear model.

In the context of the linear model, residual analysis may be viewed as a technique of exploratory data analysis initiated to examine the efficacy of the specified model. Inherent in the equation that defines the model is a set of mathematical assumptions. This set of assumptions forms the critical underpinnings of the model and the analysis procedures. If these assumptions are violated, the statistical techniques used to confirm the model may be so inappropriate that results are uninterpretable or are interpreted incorrectly.

Assumptions of the Linear Model: In the strictest sense, the assumptions of a linear regression model are stated about the error terms of the equation. In the early 1960s a body of literature developed that proposed the observable residuals from the sample regression could and should be used to examine for assumption violations (Anscombe & Tukey, 1963).

For the linear model, the usual set of underlying assumptions, as specified by Graybill (1976) and Neter and Wasserman (1974), are:

1. The residual mean is zero.
2. The residual variance is equal at all points of the predicted dependent variable.
3. The residuals are normally distributed.
4. The residuals indicate the independent variables have a fixed distribution.
5. The residuals show no evidence of departure from linearity.
6. The residuals are independent.

In addition to these assumptions, the independent variables in an equation are assumed to be measured without error.

A model may depart from its mathematical assumptions in a variety of ways. Although regression analysis is a robust procedure, each of these departures, or combination of violations, has an impact on the stability of the final model. For example, regression coefficients (parameter estimates) may be unstable or the explained variance (R^2) may be biased (Belsley, Kuh, & Welsch, 1980; Hey, 1974). In addition, tests of significance, such as F tests, may be inaccurate. Thus, the researcher cannot have confidence in the stability of the results of the analysis.

Types of Residual Analysis: Residuals may be analyzed in two ways to assess for model violations. First, statistical tests that examine for specific assumption violations have been developed. Second, graphs may be used to examine for unusual or unexpected patterns. A judicious combination of these two methods is perhaps the best approach to analysis of residuals. It may seem that the two methods are mutually exclusive; however, they complement each other. For example, a graphic approach may demonstrate a curvilinear relationship that would not have been apparent if a linear summary measure of association, such as a Pearson correlation coefficient, had been used. Alternatively, the subjective interpretation of a graph can be confirmed by the use of the appropriate summary statistic. In other words, the tandem use of graphs and statistics provides a base for the researcher not easily obtained when there is exclusive use of one approach.

Residual Analysis Procedure: This article suggests an approach to residual analysis which uses both graphs and summary statistics.

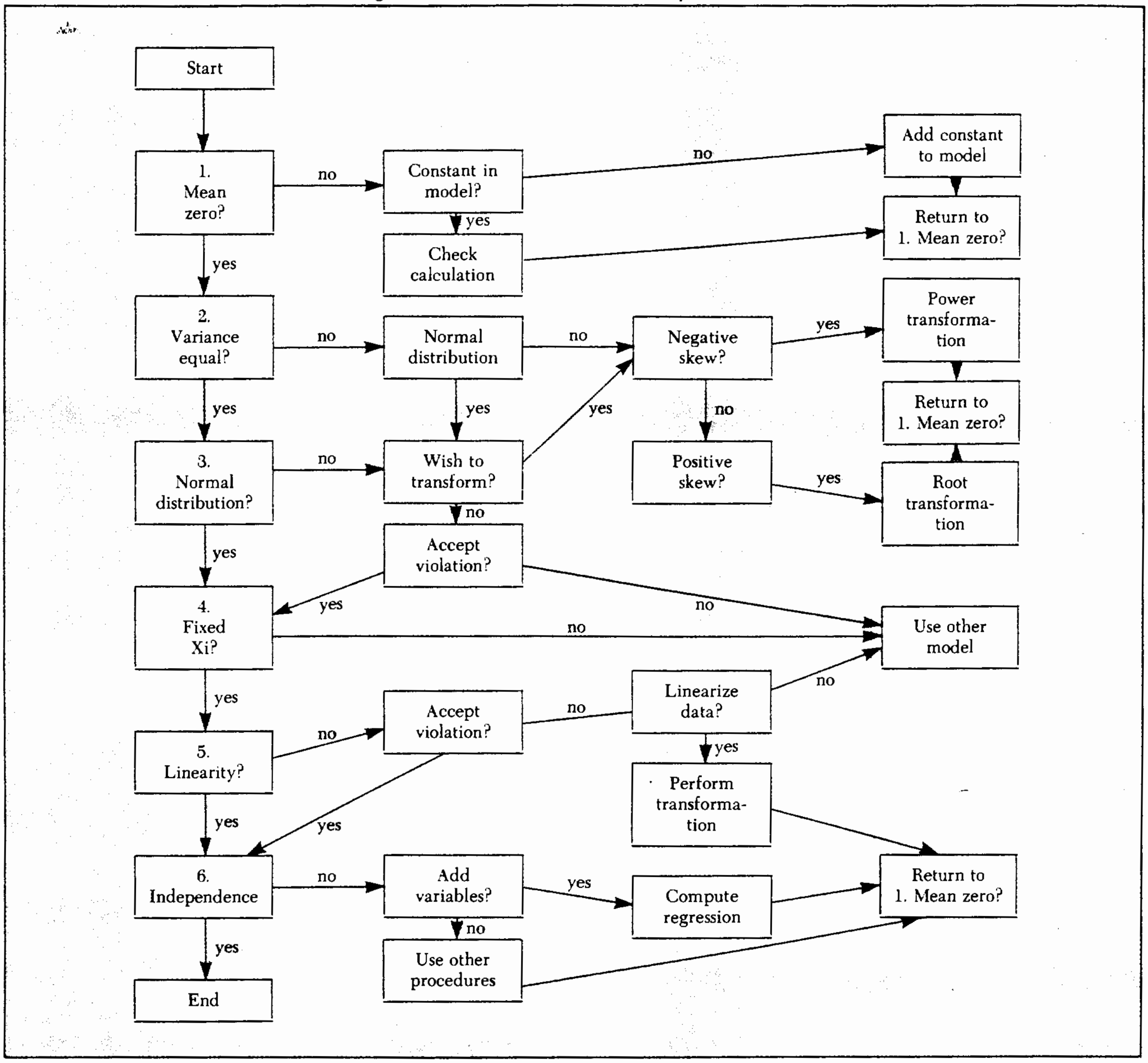
The flow chart in Figure 1 summarizes a pattern of examination for model assumptions. The order of examination is based on a natural flow of the analysis procedures and not on

JOYCE A. VERRAN, PHD, RN, is an assistant professor and division coordinator, Medical-Surgical Nursing, College of Nursing, University of Arizona, Tucson, AZ.

SANDRA L. FERKETICH, PHD, RN, is an assistant professor and coordinator of the Parent-Child Program, Department of Physiological Nursing, School of Nursing, University of California, San Francisco, CA.

Accepted for publication January 15, 1986.

Figure 1. Flow Chart for Residual Analysis Procedure



any ranking of importance. The *yes* or *no* beside the arrowed lines indicates the answer to the question posed in each of the boxes. Table 1 summarizes graphic and mathematic approaches useful in obtaining an answer to the posed questions in boxes 1 through 6.

Although the flow chart is self-explanatory, two portions require particular attention. The questions regarding zero mean and independence are crucial, and no acceptance of violations of these assumptions can be tolerated. The flow chart shows closed loops that permit no progress in the analysis of the model until these violations have been corrected.

Zero Mean: Examination for zero mean is accomplished by obtaining the summary statistics of the mean and its 95% confidence interval for the unstandardized residuals. If the confidence interval crosses zero, one can assume with reasonable confidence that the assumption of zero mean has not been violated. When an intercept or constant is fit as part of the linear estimation process, violation of this assumption is infrequent.

Normality: When investigating whether residuals fit a normal distribution, a variety of sources of information should be used. The manifestations of a nonnormal distribution can occur in a multitude of ways; therefore, data should be examined both graphically and statistically.

Graphic techniques for the test of the normality assumption include histograms, frequency polygons, and normal probability plots. For the researcher unfamiliar with the interpretation of probability plots, Daniel and Wood (1980) provide an excellent review of the variations that can occur within such plots.

Summary statistics include omnibus tests for deviation from a normal distribution. The Shapiro–Wilk W is considered a uniformly powerful test for deviations from normality (Huang & Bolch, 1974; Shapiro, Wilk, & Chen, 1968). Although the

Table 1. Graphic and Statistical Means of Assessing Violations

VIOLATION	GRAPH	STATISTICAL TEST
1. Zero mean	Frequency distribution	95% confidence interval
2. Equal variance	Scatterplot	Split-half *F* test
3. Normality	Histogram Frequency polygon Probability plots	Shapiro–Wilk *W* Chi square
4. Fixed independent variables	Scatterplot	—
5. Linearity	Scatterplot	—
6. Independence	Scatterplot	Pearson correlation Durbin–Watson test Von Neuman ratio Runs test

W is an excellent test, its computation may be difficult and time-consuming if appropriate computer programs are not readily available. A chi-square goodness-of-fit test is hand computed easily and quickly provides a guide to normality which enhances the subjective information obtained from the graphic approach.

Variance Equality: Graphic analysis provides the best approach to the examination of residuals for the violation of equal variance. A scatterplot of standardized residuals, on the *y* axis, versus the predicted dependent variable, on the *x* axis, will demonstrate a random and equal scatter of points about the zero line of the residuals if there is no violation of this assumption. Alterations in scatter that indicate problems with variance are illustrated in Verran and Ferketich (1984).

A statistical test in support of the subjective interpretation of the scatterplot is a split-half *F* test (Younger, 1979). With this technique residuals may be split into two samples selected to best represent the observed unequal variance. An *F* test may then be performed on the variances of the residuals from the two samples.

Independence: Violation of the assumption of independence among the residuals may occur in a number of ways. Neter and Wasserman (1974) provide a helpful orientation to evaluating the data in regard to the independence assumption. They state that independence may be violated when one residual value depends on the value of another. This may occur when a variable is left out of the model, such as time, if the data were collected in a time sequence. Other variables, such as rater, care provider, or site, may cause relationships among the residuals that would not be discovered if only the more usual parameter of time were used for the residual plot. It is helpful to consider any other variable that may cause a correlation among the residuals and plot the standardized residuals against that particular variable or set of variables. If evidence of linear or nonlinear relationships is seen, follow-up statistical tests can be performed.

Summary statistical tests are selected on the basis of the evidence of the relationship observed in the graphic residual plot. For example, the statistical significance of a linear relationship can be supported by the use of a Pearson correlation coefficient. If residuals are sorted by the variable thought to cause dependency, the Durbin–Watson test, the Von Neuman ratio, or the nonparametric Runs test may also be used (Zar, 1974).

A further technique to test the impact of an external variable on the model is, quite simply, to add the variable to the linear equation and examine its impact on the percent of explained variance. The decision to add a variable is both statistical and theoretical in nature. Theoretical considerations involve the logical assessment of fit between the new variable and other model concepts. For example, the investigator may find that the time the data were collected and the individual raters who obtained the data both evidence patterns with residuals. However, the differences in how raters collected the data may indicate measurement error rather than theoretical significance. A logical decision would then be to add time to the model but to correct the rater error in other ways such as dropping the cases attributable to the rater or raters evincing the most error. Statistical considerations include the relationship among the number of variables added to the model, the sample size, and the consequent number of *F* tests performed in the regression sequence.

Linearity and Fixed Independent Variables: The statistical procedure used to fit the model assumes a linear relationship among variables. Whether the linear function is appropriate for the data may be evaluated by the study of scatterplots of residuals. As noted earlier, when residuals are plotted versus the predicted dependent variable, a random scatter about the residual zero line is expected. A scatter that curves across this line indicates that the linear function is inappropriate for the data under analysis. Similar plots of residuals versus the independent variables in the equation are also helpful in determining the specific variable or set of variables that may result in the curved function. These latter plots are also used to indirectly estimate whether the distributions of the independent variables are fixed.

Conclusions: As nurse researchers develop or begin to use sophisticated methods of analyzing data, it is critical to fully evaluate the data for underlying relationships and for possible violation of model assumptions. The field of exploratory data analysis can provide valuable techniques to achieve full analysis of data. Hoaglin, Mosteller, and Tukey (1983) noted that a major theme of exploratory data analysis is use of residuals. This view asserts that the analysis of a set of data is not complete until an examination of the residuals is performed.

Residual analysis can assist in supporting theory across studies. Replication studies would be more comparable if there were evidence that estimated parameters were stable. Then, differences or similarities in results would be less likely due to an artifact of the statistical analysis and more likely due to population differences.

With the use of computers and packaged software programs, attention to examination of residuals becomes increasingly important. Researchers may tend to view the confirmation of relationships, evidenced by the computer printout, as the final step in analysis. But, further investigation, in an exploratory mode, may yield unexpected, yet theoretically valuable, information about relationships as well as provide evidence of the model's efficacy. NR

References

ANSCOMBE, S. J., & TUKEY, J. W. (1963). The examination and analysis of residuals. *Technometrics, 5,* 141–160.
BELSLEY, D. A., KUH, E., & WELSCH, R. E. (1980). *Regression diagnostics.* New York: John Wiley & Sons.
DANIEL, C., & WOOD, F. S. (1980). *Fitting equations to data.* New York: John Wiley & Sons.
GRAYBILL, F. (1976). *Theory and application of the linear model.* North Scituate, MA: Duxbury Press.
HARTWIG, F., & DEARING, B. E. (1979). *Exploratory data analysis.* Beverly Hills, CA: Sage Publications.
HEY, J. D. (1974). *Statistics in economics.* London: Martin Robertson & Company.
HOAGLIN, D. C., MOSTELLER, F., & TUKEY, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: John Wiley & Sons.
HUANG, C. J., & BOLCH, B. W. (1974). On testing of regression disturbances for normality. *Journal of the American Statistical Assoc*iaiton*, 69,* 330–335.
NETER, J., & WASSERMAN, W. (1974). *Applied linear statistical models.* Homewood, IL: Richard D. Irwin.
SHAPIRO, S. S., WILK, M. B., & CHEN, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association, 64,* 1343–1372.
VERRAN, J. A., & FERKETICH, S. L. (1984). Residual analysis for statistical model assumptions of regression. *Western Journal of Nursing Research, 6*(1), 27–40.
YOUNGER, M. S. (1979). *A handbook for linear regression.* North Scituate, MA: Duxbury Press.
ZAR, J. H. (1974). *Biostatistical analysis.* Englewood Cliffs, NJ: Prentice–Hall, Inc.

Multiple Regression Analysis With Small Samples: Cautions and Suggestions

PATRICIA A. PRESCOTT

Correlational techniques such as multiple regression have great appeal to nurse researchers, as they are appropriate for many multivariate research questions and can be used for both descriptive and inferential purposes. Despite the versatility of multiple regression, its limitations should be considered and its findings examined in light of these potential concerns. This article discusses limitations of multiple regression when used with small samples and illustrates use of a suggested procedure for dealing with problems of instability and inconsistency of findings. Inconsistency refers to the degree to which variables different from the original set of significant predictors appear as significant in repeated analyses. Instability, on the other hand, refers to the degree to which predictors that remain consistent from one analysis to another change in statistical magnitude and/or sign.

Limitations of Multiple Regression: Techniques based on correlational procedures such as multiple regression maximize chance associations in a data set and often produce findings that vary substantially across samples (Mosteller & Tukey, 1977). This problem is particularly acute in studies using many predictor variables and small samples. Clinical studies are often of this type, exploring the effect of many variables on a criterion variable using a small number of patients or clients.

There is no common definition of what constitutes a small sample, and the needed sample size depends on a number of factors, including the reliability of the estimate desired, the accuracy of the estimate, and the relative variance of the variable under consideration (Levy & Lemeshow, 1980). Despite these sources of variability, Thorndike's (1978) rule of thumb for a rough estimate of the minimal sample size for multivariate analysis (10 subjects per variable plus 50 additional subjects) indicates that a minimally adequate sample of 100 subjects would be needed in a study with five variables (p. 184). Using this criterion, many clinical studies examine multiple factors with small samples that maximize the likelihood of chance association among predictors.

Selected Nursing Studies: Twenty studies reporting findings based on multiple regression analysis and clearly identifying the number of predictor variables were located in a review of 1983 and 1984 issues of *Nursing Research, Western Journal of Nursing Research,* and *Research in Nursing and Health.* Of these studies, 12 samples maintained a ratio of at least 10 subjects per predictor variable. However, only 4 of the 12 studies used samples that met Thorndike's criterion for a minimally adequate multivariate sample. The remaining 8 studies neither met Thorndike's criterion nor the commonly used criterion of 10 subjects per predictor; they used samples ranging from a low of $N = 8$ with two pre-

Accepted for publication February 19, 1986.
This study was funded by the Robert Wood Johnson Foundation, 1980 to 1985, principal investigator, Dr. Patricia A. Prescott, School of Nursing, University of Maryland, Baltimore, MD.

PATRICIA A. PRESCOTT, PHD, RN, is a professor and chair, Graduate Department of Psycho Physiological Nursing, School of Nursing, University of Maryland, Baltimore, MD.