MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 16, NUMBER 1, SPRING 1988

Predicting Statistics Achievement: A Prototypical Regression Analysis

Rodney J. Presley and Carl Huberty
University of Georgia

The purposes of the current study are: (a) to demonstrate a viable approach to the conduct of a multiple regression/correlation analysis; and (b) to illustrate the approach in the context of predicting achievement in an introductory statistical methods course. The analysis is proposed as being appropriate if the basic intent of a study is that of prediction as opposed to that of explanation. That is, the intent is to arrive at a model for predicting a criterion in as efficient a manner as the data on hand will allow. No model, causal or otherwise, is being posited or verified.

There are five dimensions of the suggested approach: 1) designing the study; 2) examining the data; 3) searching for an efficient prediction model; 4) using regression diagnostics; and 5) assessing the model(s). Each dimension of the study is presented in sections below, each of which includes an application in the context of predicting statistics achievement. [This list does not necessarily imply a sequential step-by-step analysis.]

An effective model for predicting statistics achievement may be useful in addressing three questions related to instruction and curriculum: 1) Can a fairly accurate rule be determined for predicting achievement in introductory statistics courses? 2) How effective are easily obtained graduate-level student test scores in predicting "high-achievers"? 3) In predicting "low-achievers"? Having some knowledge of predicted achievement

---

A special thanks is extended to Stephen Olejnik, David Payne, and John Stauffer (at The University of Georgia) for their cooperation in this study.

may be helpful in an obvious way to instructors. Furthermore, having rules for accurately predicting high and low achievers would possibly suggest either a special "advanced" section or some remedial pre-course experience.

Previous studies predicting achievement in introductory statistics courses have varied in predictor models used and in subject sample characteristics. Predictor variable domains employed in previous studies include computation skills, mathematics symbolism, previous mathematical experience, logical thinking, attitudes, anxiety, self appraisal, impulsiveness, arithmetic/mathematics achievement, and other biographical characteristics (e.g. gender, age, college major). Such predictor domains and others may be found in the studies by Bending and Hughes (1954), Bledsoe and Perkins (1976), Elmore and Vasu (1980), Feij (1976), Feinberg and Halperin (1978), Harvey, Plake, and Wise (1985), and Pruzek (1964). The size of the sample studied and the academic level of the students in the sample varied somewhat in these studies. For example, Bending and Hughes employed 71 undergraduate level students, while Elmore and Vasu (N=188) and Pruzek (N=112) employed graduate students; Feinberg and Halperin employed undergraduate (209) as well as graduate (94) level students, while Harvey et al. (1985) employed 47 and 41 undergraduate and graduate level students, respectively.

As might be expected most of the studies reviewed used a multiple regression/correlation analysis. Typically, squared multiple correlation coefficients were reported (along with some type of "variable selection" results and some kind of regression weights). The percent of variance shared between statistics achievement and one or more variables (from predictor variable domains as listed above) has generally been in the range of 30 to 45 (based on <u>unadjusted</u> squared multiple correlation coefficients).

Designing the Study

In conducting a multiple regression/correlation study one must clearly define the population for which the prediction model is intended, select a meaningful criterion, and select a useful set of predictors.

Subjects

The target population of interest in this study is graduate students enrolled in the introductory statistical methods course. Students in eight sections of an introductory statistical methods course offered in The University of Georgia College of Education served as the experimental units. The first class enrolled in Summer Quarter 1984 and the last in Fall Quarter 1986. Most of the students were in College of Education graduate degree programs. [It is the opinion of the junior author, who has taught this course for several years, that these classes are representative of previous and subsequent classes in the same course.] Students in six of the classes (five of which were taught by the junior author) were administered equivalent tests and examinations. Students from these classes constituted the design sample. Students from the two remaining classes constituted the "model assessment" sample.

Some descriptive information on all students who completed the course in the eight classes is given in Table 1. Only those students who had taken the Graduate Record Examinations prior to enrollment were considered in the final analysis. There were 122 students in the design sample (classes 1-6) and 51 students in the model assessment sample (classes 7 & 8).

Criterion

Since it is difficult to maintain contact with students after they complete the course, we decided to focus on an immediate criterion as opposed to an intermediate or ultimate criterion (Crocker & Algina, 1986, p. 225). The immediate criterion is end-of-course achievement in the introductory statistics class. Specifically the criterion variable, SCORE, is defined as a linear composite of Z transformations of the student scores on the in-class midterm and final examinations. The weights for midterm and final examination are 1.0 and 1.5, respectively: SCORE = 1.0 * ZMIDTERM + 1.5 * ZFINALEXAM. The raw-to-standard score transformation employed the mean and standard deviation based on classes 1-6.

Although four different textbooks (Glass & Hopkins, 1984; Hinkle, Wiersma, & Jurs, 1979; Iman & Conover, 1983; Wright, 1976) were used with the eight classes, the material covered in the course on introductory statistical methods was quite comparable across the classes. In classes 1-6 the midterm test (35 multiple-choice items) covered graphical and numerical descriptors for data distributions. In the same six classes, the final examination (45 multiple-choice items) covered probability, probability distributions, estimation, and introduction to statistical testing. (Some test and examination items pertained to computation; however, the focus was on concepts and higher-level cognitive performance.) It may be argued that instructional performance was fairly constant, and that the six midterm and final examinations had comparable difficulty and internal consistency levels. For one administration of the midterm, the mean number of correct responses (total score of 35) was 21.8 and the Cronbach alpha value was .84; the respective values for one administration of the final examination (total score of 45) were 27.7 and .83. In essence it is assumed that a common scale of measurement was used for all six midterm examinations and for all six final examinations.

Predictors

In selecting predictor variables, Pedhazur (1982, p. 138) suggests attending to theoretical considerations and previous research evidence. There is some empirical evidence (e.g., Bledsoe & Perkins, 1976; Brown, 1933(I); Woelke & Leitner, 1980) that basic mathematical abilities can contribute to the prediction of introductory statistics achievement. Educators generally believe that previous relevant knowledge and skill will affect student achievement in new learning situations. Elmore and Vasu (1980) conducted a study examining the relationship between several affective variables and achievement in statistics. In their review of previous studies they noted that the correlation between statistics achievement and affective variables was generally low. Elmore and Vasu did not consider measures of specific arithmetic and algebra skills in their study but did report significant correlations between two attitudinal variables and statistics achievement. Some type of specific arithmetic/algebra skill measures were included in most of the studies reviewed by these authors which reported low correlation between affective measures and statistics achievement. The present authors interpret this as indicating that affective variables contribute little to the prediction of statistics achievement when measures of specific arithmetic/algebra skills are also included as predictors. Based on previous research and instructional considerations, the current authors decided to consider predictor variables designed to measure mathematics/ algebra achievement or skill level in preference to affective predictors.

Various algebra and arithmetic achievement skills were sampled by a locally developed pre-statistics inventory. The seven scales of this inventory , the abbreviation as used throughout this paper, the content areas, and maximum number of points are listed below:

1) S1. Operations with integers, common fractions, and decimal fractions (25 points maximum),
2) S2. Proportions and percents (8 points),
3) S3. Squaring and extracting square roots (6 points),
4) S4. Operations with signed numbers (8 points),
5) S5. Operations with simple formulas and construction of simple formulas (8 points),

6) S6. Linear graphs (6 points), and

7) S7. Miscellaneous -- terms, inequalities, symbolism, etc. (13 points).

The sum of these seven scale scores, labeled TOTAL (74 points), was also considered as a predictor measure.

In addition to the seven scale scores and TOTAL score, three predictor measures were obtained from the Graduate Record Examinations; the Verbal score (GREV), Quantitative score (GREQ), and the product of the Verbal and Quantitative scores (GREVQ). Cohen (1978) has suggested the use of product scores in regression models to represent nonadditive or interaction effects between two variables. Because many statistics problems are presented in narrative form, the present authors believe that verbal and quantitative achievement may interact to effect achievement in statistics. It is interesting to note that in ten studies reviewed, the Graduate Record Examinations scores were used as predictor measures only by Elmore & Vasu (1986) and by Noble (1986). These scores are readily available for most students, being an admission requirement in many programs, and seem a natural choice for predictors with statistics achievement as the criterion. The GRE scores were selected because of their availability and their apparent relevance.

A matrix of correlations (see Table 2) among the predictors and between the predictors and the criterion may be useful in screening initially chosen measures. Predictors having near zero correlation with the criterion would be suspect as useful predictors. For the current study correlations of the predictors ith the criterion range from a minimum of r=.20 for GREV to a aximum of r=.50 for GREQ. Therefore no potential predictors ere eliminated at this point because of low correlation with the riterion. Predictors which correlate highly with one another ay indicate redundancy of information. If two such variables re detected one may be eliminated from the analysis or when ogically appropriate the items used to measure the two variables ay be combined. For the current study the highest predictor ntercorrelation was between GREV and GREVQ (r=.79). This is not surprisingly strong correlation considering that GREVQ is a unction of GREV. No other predictor intercorrelation approached his magnitude. Therefore no variables were eliminated at this tage because of redundancy.

Pedhazur (1982, pp. 32-36) discusses the assumptions nderlying multiple regression analysis. He describes this nalysis technique as robust. Stevens (1984, p. 335) has uggested plotting the criterion values as a visual means of as-essing approximate normalcy. Such a plot of the criterion easures in this study suggest approximate normalcy (see Figure). In addition, Stevens suggests plotting the predictor ariables, not to check for normalcy, but as a visual aid in etecting outliers in the predictor space.

Examining the Data

Errors in the data may seriously distort efforts at rediction. Recording of data, transposing the data, and ntering the data into the computer are all opportunities for rrors. We used the computer to list the data as they were entered and compared this listing with the original data. Also, we find the use of frequency histograms and stem-and-leaf plots of predictor and criterion measures useful in detecting extreme values which may be errors. In addition, these plots help to identify segments of the predictor range which are sparsely represented by the data sampled. If the data set is quite large and variables can only assume restricted values, then one may write computer statements to isolate all observations with variable values out of the allowed range of values. This approach may still allow errors into the data set. The best approach, though time consuming, is to list the data and make comparisons to the original observation records.

Searching for an Efficient Model

Two questions must be answered before the parameters of a linear regression model are estimated. First, what is the optimum number of the available predictors that should be retained in the model? Secondly, what is the best combination of predictors for a subset of chosen size? [This brings up a related question: How is one model deemed better than another? Cross-validation results may be the ultimate test of the appropriateness of a prediction model. The use of a validation or assessment sample in the current study is discussed later.] Three indices of model effectiveness will be examined at this time. A better model will account for more of the variability in the criterion variable and reduce the error in the predicted scores. Since the adjusted R-squared value reflects the proportion of variance in the criterion accounted for by the model, one index of a good model is the adjusted R-squared value. The higher the adjusted R-squared value the better the model fits the sample data. The RSQUARE procedure in SAS (SAS Institute Inc., 1985) was used to calculate the adjusted R-squared values for all possible combinations of the predictor variables in all possible size subsets of the predictor variables. The adjustment formula used by SAS is

$$\text{adjusted R-squared} = 1-(1-\text{R-squared})(n-1)/(n-p)$$

where n is the number of units sampled and p is the number of parameters in the model including the intercept. The highest adjusted R-squared value for each predictor subset size may be plotted against the subset size (see Figure 2).

A second index is the Mean-Square Error which is equal to Sum-of-Squares Error)/(n-p). The model with the lowest Mean-quare Error value has minimized the error and reflects a good it of the model to the sample data. The lowest Mean-Square rror for each subset size may be plotted against the subset size see Figure 3). A third index, Mallows' Cp statistic, is a easure of bias in estimating the parameters of the regression odel (Chatterjee & Price, 1977, pp. 198-199). A model that is o simple (omits important predictors) may result in biased egression weights and biased prediction, while an overly omplicated model (including predictors that add little or othing in addition to the predictors already in the model) may esult in large variance both in the regression weights and the edicted values (Myers, 1986, pp. 112-114). As Cp exceeds p the bias in estimation of model parameters becomes more severe. Especially in the use of regression for prediction, one wishes to minimize the bias of estimating the model parameters. The values of Cp against p may also be plotted (see Figure 4). A good model will have a "low" value of Cp and one that is "close" to p.

These three indices, adjusted R-squared value, Mean Square Error, and Mallows' Cp, may be examined simultaneously to determine a good subset size. The three indices may not point to exactly the same subset size. After simultaneously considering the three indices one may decide to retain two or more predictor subset sizes. Examination of Figure 2 reveals that a model with three predictors will achieve the largest adjusted R-squared value. The smallest Mean-Square Error value is associated with a model of three predictors as can be seen in Figure 3. Examination of Figure 4 suggest that a model with more than three predictors may be desirable. As the predictor subset size is increased the value of Cp approaches p. But, at the same time the value of adjusted R-square begins to fall and the value of Mean-Square Error increases. It should be noted, as often happens, that neither of the three statistics indicates a predictor subset size that is greatly superior to others. Accordingly, we considered models of five and six predictors. [One additional model was considered; TOTAL score along with GREV and GREQ constituted the predictors of a third model. This model is simple and may reveal the advantages or disadvantages of summing the scale scores of the pre-statistics inventory into one score.]

Now that we have decided to look at models of five and six predictors, we must decide which particular subset of variables to use in our model. In the SAS computer printout (see Table 3 for subset of six predictors) the combinations of variables in each subset size are ordered in accordance with the adjusted R-squared value. One might feel compelled to select the best combination of variables as indicated by the highest adjusted R-squared value (lowest Mean-Square Error, or Cp value closest to p). Examination of the actual values will reveal negligible difference in the adjusted R-squared value for the best and second best combination of variables in each subset size. Since the regression procedure capitalizes on sample specific relationships one need not feel bound to select the subset of variables with the highest adjusted R-squared value realizing that when the difference between the adjusted R-squared value for the best and second best subsets is negligible, the order of the best and second best set of variables of a given subset size may very well be reversed when a different sample is examined. With his in mind the present authors chose the models retaining the ollowing variables for the five and six predictor variables odels, respectively; S4, S5, S6, GREV, GREVQ and S1, S4, S5, S6, REV, GREVQ. It was desirable from a substantive viewpoint to etain a variable subset with the GREV and GREVQ variables.

Using Regression Diagnostics

Regression diagnostic methodology is relatively new and the ury is still out on the relative usefulness of indices to detect nfluential data points and outliers. We restricted our diagnostics to examination of the influence of single data points; the study of the influence of groups of data points is in its infancy, with very little practical guidance having been offered--see discussion by Atkinson and by Hoaglin and Kempthorne in Chatterjee and Hadi (1986). Also, little guidance has been suggested for the simultaneous consideration of predictor variable selection and outlier detection. [We selected predictors first and diagnosed second with an admission of potentially misleading results.]

In this section we will discuss the practical application of some of these techniques. After selecting the variables for models of five and six predictors the SAS PROC REG (regression procedure) was used to estimate a linear model relating the predictors to the criterion. Options were selected to print the actual criterion value and the predicted criterion value for each observation. The difference between the predicted value and the observed value is the simple residual value. These values were examined en masse and individually.

Assumptions Check

A plot of the residuals against the predicted score may reveal model underspecification (omission of important predictor variables), violation of the assumption of homogeneity of variance, departure from normalcy in the model errors, and extreme or suspect data points (Draper & Smith, 1981, pp. 141-147; Myers, 1986, p. 138). Consider the hypothetical plots in Figure 5. With an appropriately fitted linear regression model, the plot of the residual values against the predicted scores should look similar to plot 1 in Figure 5. A graph such as plot 2 in Figure

5 indicates that the variances are not constant suggesting a need for a weighted least squares analysis or a transformation of the criterion variable. A graph such as plot 3 in Figure 5 indicates an error in analysis; the departure from the fitted equation is systematic. This effect can also be caused by incorrectly omitting an intercept term in the model. A graph such as plot 4 in Figure 5 indicates an inadequate model--need for extra terms in the model (e.g. squares or crossproducts) or need for a transformation on the criterion values before analysis. After visually inspecting Figure 6, the graph of residuals against predicted scores for the five variable model, concerns of the type just discussed were set aside.

Outliers

An outlier is defined as an individual observation with a relatively large absolute value of residual score. We proceed to examine outliers individually. Since any model is an approximation of the data, outliers are not uncommon. Outlier observations may represent data error or they may be units that for some reason represent a population different than the majority of units in the sample. Outliers may have some characteristic in common that determines a different functional relationship between the predictor and criterion variables for them than for the majority of the sample. If this is so then one can search for the characteristic and determine if it is an important variable that should be included in future predictor models. Outliers may have an excessively strong influence on the estimation of regression weights compared to the influence of other data points. If this is the case the outlier is also an influential observation point. Stevens (1984) (and others; e.g. Draper & Smith, 1981, p. 169, Weisberg, 1985, pp. 114-125, Chaterjee & Hadi, 1986, p. 380) point out that an outlier may or may not be an influential observation in determining estimates o regression parameters. Conversely, an observation may be influential and not be an outlier. We will identify outlier observations mindful of their impact on fit of the model to the sample data and their influence on estimation of the regression parameters. Also, observations which are not outliers but which are influential will be identified and examined. This will be discussed below. For a more technical discussion of regression diagnostics pertaining to outliers and influential data points see Cook and Weisberg (1982).

The simple residual, the standardized residual, and the studentized residual all are indicators of outliers in the crit rion space. We accept the argument of Stevens (1984, p. 336) that the studentized residual is a more sensitive detector of outliers. For more discussion on this and alternate names for these statistics, see Chatterjee and Hadi (1986). A studentiz residual is referenced to the Student t distribution with N-p- degrees of freedom (Chatterjee & Hadi, 1986 p. 380). As the choice of alpha level in hypothesis testing is arbitrary, so i the choice of a critical value for studentized residuals. A stem-and-leaf plot of residuals may be constructed to identify data points which are outliers relative to other data points i the sample.

Observations may be outliers in the predictor space

(Stevens, 1984, p. 337) because of extreme values on one or more predictor measures or because they represent a rare combination of predictor values. Such observations will have a relatively large diagonal element in the so-called HAT matrix, h sub ii. These observations are also called high leverage points. High leverage points may or may not be influential. How large is a relatively large HAT diagonal element? A critical value of 2p/n has been suggested (Chatterjee & Hadi, 1986). For a discussion of critical values for influence indicators in general see Belsley, Kuh, and Welsh (1980). We prefer to consider the h sub ii values in context with the values for all observations by constructing a stem-and-leaf plot. An example will follow in the subsection, Illustration.

Influence Indicators

Several indicators of influence are reviewed by Chatterjee and Hadi (1986). Seven excellent comment reviews follow that article. There is some confusion about just what is being influenced in the influence measure. In addition there are only rule-of-thumb guidelines for the analyst to use in deciding when an influence measure is large enough to warrant concern. In regard to the latter, instead of adopting a rule-of-thumb critical value a stem-and-leaf plot may be constructed for each influence indicator. A visual inspection of those plots will reveal observations with influence indicator values that are large relative to others in the sample. This approach may be criticized as being arbitrary, as are the rule-of-thumb approaches. It is believed that these graphical approaches will give the researcher a better feel for his/her data than employing rule-of-thumb values. The influence indicators considered here reflect influence on the b vector of regression weight estimates, the variance/covariance of the b vector, or a combination of both, and the influence on a single b value estimating a single model predictor parameter.

Cook's D or Cook's distance, sometimes abbreviated D sub i and C sub i (Chatterjee & Hadi, 1986, p. 383) measures the change in distance between the b vector as estimated with the ith observation in the model and the b vector as estimated with the ith observation removed from the model. It therefore indicates the influence of the ith observation on the parameter estimates of all the predictor weights (see comments by Hoaglin in Chatterjee and Hadi, 1986). The same information is also provided by Welsh's distance, and a modified Cook's distance. Different rule-of-thumb critical values are suggested for these influence indicators (Chatterjee & Hadi, 1986). Each of these indicators should identify influential observations in the same rank order.

The covariance ratio (CVR) and the Cook-Weisberg statistic provide information on the influence of the ith observation on the variability of the parameter estimates of the b vector elements. An index called DFFITS indicates influence on both the estimates of the b vector and the variance/covariance of the predictor parameter estimates.

Finally an observation may have strong influence on only one of the b values. This is indicated by an index called DFBETA. Plots of DFBETA against observation number are also referred to as partial regression leverage plots.

The numerous plots referred to above are not all reproduced herein. They are easily obtained from popular computer software packages such as SAS and SPSS. Regression diagnostics were conducted for the three models considered in this paper. For economy of space, only the diagnostics for the five variable model are discussed in detail. At the end of this discussion the reader is appraised of which observations we decided to eliminate from each model. Other researchers examining the exact same data and indicators of influence and outliers may reach slightly different decisions about eliminating observations. Finally it should be noted that observations which are outliers in the predictor space but, which are not excessively influential, may represent areas in which the sample data are sparse. Such observations may prompt the researcher to collect more data.

Illustration

We turn now to the predictor models studied in the context of predicting statistics achievement. Outliers and influential data points will be identified for one model (Model 2) and the decision to delete or not delete the associated observation will be addressed. The three models and their adjusted R-squared values are listed below;

Model 1	SCORE=GREV GREQ TOTAL	adj R**2=.2983
Model 2	SCORE=S4 S5 S6 GREV GREVQ	adj R**2=.3138
Model 3	SCORE=S1 S4 S5 S6 GREV GREVQ	adj R**2=.3093

The stem-and-leaf plot of the studentized residual (RSTUDENT) for Model 2 is given in Figure 7 (each stem-and-leaf plot is accompanied with a tabular listing of extreme observations and their values). It is apparent that observation 215 and 176 have high studentized residual values relative to the sample . Observations 88 and 148 have relatively low studentized residual values. A small studentized residual value implies that the predicted criterion value for that observation is lower than the actual criterion value. Of these four observations only 215 is a relative outlier in the predictor space as indicated by the stem-and-leaf plot of h sub ii in Figure 8. At this point one may wonder if observation 215 is representative of the population from which it is believed the sample was drawn. In this study specifically, is there something about observation 215 that makes this person not representative of students enrolled in introductory statistics courses? This question is not addressed in this paper. Merely the point is made that regression diagnostics may lead the researcher to identify data points which have some characteristic different from the majority of the sample.

We now examine the influence indicators to identify observations which have an unusually strong influence on the paramaterization of the model. Examination of the stem-and-leaf plot of Cook's D (Figure 9) reveals that observation 215 and 176 are relatively influential in determining the estimates in the $\underline{b}$ vector. The stem-and-leaf plot for the DFFITS indicator is given in Figure 10. This suggests that observation 215 and 176 are influential in determining the $\underline{b}$ vector and/or the variance of the estimates in the $\underline{b}$ vector. Examination of the stem-and-leaf plot of COVRATIO (see Figure 11) reveals observation 215 but not

176 to be influential in increasing the variance of the b vector. In essence observation 215 receives a double indictment for its influential role in determining the b vector and its relatively strong contribution to lack of fit of the model to the sample data. Elimination of these two observation points and recalculation of the regression equation should improve the predictive accuracy of the model. In addition, the removal of observation 215 and to a lesser extent 176 should increase the fit of the model to the sample data.

In examining Figure 9 and Figure 10 the reader may have noticed that observation 144 is relatively influential in determining the b vector and/or the variance of the b vector. However, this observation is not a relative outlier in the criterion space or the predictor space. Examination of stem-and-leaf plots and frequency histograms of all the model variables does not indicate that observation 144 came from a sparse region of the data. No further consideration is given to deleting this observation at this time.

Plotting DFBETA for each predictor against observation number, the so-called partial regression leverage plot, did not indicate observations which were excessively influential in estimating the b value for one predictor.

Observation 215 and 176 were removed from the sample data and the regression equation for Model 2 was recalculated. The adjusted R-squared value rose from .3138 to .3759, an increase of over 6% explained variance.

After examining stem-and-leaf plots of the outlier measures and influence indicators for the other two models we decided to drop observation 215 and 176 from Model 1 and observation 215, 176, and 144 from Model 3. The change in adjusted R-squared for Model 1 was from .2983 to .3761 and for Model 3 from .3093 to .4047.

Assessing the Model(s)

Information was gathered from classes 7 and 8 (N=29 and 22, respectively) in order to assess the usefulness of the models. Because the same criterion was not available for these two classes, this assessment differs from the traditional "cross validation" study. The instructors in these two classes were asked to rank-order their students based on performance. The regression models were applied to the predictor values for each student in these classes to obtain a predicted criterion score. These predicted criterion scores were rank-ordered and correlated with rankings assigned by each instructor. Using Model 2, the one discussed most extensively in this paper, the correlation for class 7 was $r=.524$ and for class 8 $r=.607$. Using Model 1 and Model 3 the respective correlations were all at least .60.

Finally we examined the use of Model 2 to predict high achievers who might benefit from accelerated instruction and low achievers who might benefit form remedial instruction. The junior author (five classes) plus the instructor of one other class identified those students who were judged to have been capable to benefit from an accelerated instructional experience in statistical methods. The judgments were based on such things as completed work, perceived maturity in quantitative methods, work habits, persistence, etc., as well as on test performance. The judgments were made not knowing the predicted or actual SCORE value for each student.

Of the 122 design-sample students, 11 were judged to have been capable of succeeding in an accelerated course. [The junior author had taught two such course sequences prior to 1984.] Of these 11, nine obtained a predicted SCORE value (via Model 2) above +1.75. [The use of a cut-off value of +1.75 was judged reasonable, based on the junior author's use of SCORE with many other classes.] There was one false-positive, i.e., one student was empirically predicted to have been capable but was not judged capable by the instructor. And there were two false-negatives. [See Table 4.] With a false-positive error judged as being more serious, the resulting "hit-rate" was .82 (9/11). On the other hand, the hit-rate for predicting those students who might benefit from some remedial experience was extremely low (less than chance). It appears that Model 2, at least, has reasonable predictive validity in the sense that it is potentially useful for identifying those students who would be capable of benefiting from an accelerated course experience, whereas model validity is lacking for predicting remedial-instruction student candidates.

Discussion

In general one may question the representativeness of students enrolled in introductory statistical methods courses offered by the College of Education at The University of Georgia. The mean scores on the Graduate Record Examinations for these students were near the national average. The variability in end-of-course achievement scores not accounted for by the models is typical of, if not lower than, that found in other studies with a similar purpose. One might hypothesize various factors that could account for this remaining variance--e.g., motivation, study habits, test taking skills, academic persistance, academic maturity, and research experience. It was assumed in this study that a serious effort was put forth in completing the pre-statistics inventory, and that the reported GRE scores were correct.

Predictive measures used in the models are readily obtainable and all contributed significantly to the obtained predictive acuracy. The effectiveness of each model was assesed in three ways: (1) an adjusted R-squared value; (2) correlation of instructor-judged rank orderings of two assesment classes against rank orderings of predicted SCORE; and (3) prediction of those students who might be advised to enroll in an accelerated course. The three assessment measures were considered "respectable": (1) adjusted R-squared values (after deletion of observations identified as outliers and/or influential) of .376, .376, and .405 for Models 1 through 3, respectively; (2) rank correlations of about .6; (3) and a ratio of 9 out of 11 students judged by instructors as capable of benefiting from an accelerated instructional experience correctly identified. Thus of the three questions posed at the outset of the paper concerning regression and statistics achievement, the first two may be answered in the affirmitive and the lattter negatively for this study.

References

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley.

Bendig, A. W., & Hughes, J. B. (1954). Student attitude and achievement in a course in introductory statistics. *Journal of Educational Psychology*, *45*, 268-276.

Bledsoe, J. C., & Perkins, M. L. (1976). Prediction of success in elementary statistics: Three replications. *Psychological Reports*, *38*, 723-726.

Brown, R. (1933). *Mathematical difficulties of students of educational statistics*. Contributions to Education, No. 569. New York: Teachers College, Columbia University.

Chatterjee, S., & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression (with comments). *Statistical Science, 1,* 379-416.

Chatterjee, S., & Price, B. (1977). *Regression analysis by example*. New York: Wiley.

Cohen, J. (1978). Partialed products *are* interactions; Partialed powers *are* curve components. *Psychological Bulletin, 85,* 858-866.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression.* New York: Chapman and Hall.

Crocker, L, & Algina, J. (1986). *Introduction to classical & modern test theory.* New York: Holt, Rinehart & Winston.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: Wiley.

Elmore, P. B., & Vasu, E. S. (1980). Relationship between selected variables and statistics achievement: Building a theoretical model. *Journal of Educational Psychology, 72,* 457-467.

Elmore, P. 8., & Vasu, E. S. (1986). A model of statistics achievement using spatial ability, feminist attitudes and mathematics-related variables as predictors. *Educational and Psychological Measurement, 46,* 215-222.

Feij, J. A. (1976). Field independence, impulsiveness, high school training, and academic achievement. *Journal of Educational Psychology, 68,* 793-799.

Feinberg, L. B., & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. Journal of Experimental Education, 46, 11-18.

Glass, G. V., & Hopkins, K. D. (1984). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall.

Harvey, A. L., Plake, B. S., & Wise, S. L. (1985, April). The validity of six beliefs about factors related to statistics achievement. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1979). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin.

Iman, R. L., & Conover, W. J. (1983). A modern approach to statistics. New York: Wiley.

Myers, R. H. (1986). Classical and modern regression with applications. Boston: Duxbury.

Noble, R. F. (1986, April). Multiple regression analysis of six predictor variables of academic achievement in the course introduction to research. Paper presented at the annual meeting of the Americal Educational Research Association, San Francisco.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Pruzek, R. M. (1964). Prediction of success in elementary statistics. Journal of Educational Measurement, 1, 165-167.

SAS Institute Inc. (1985). SAS user's guide: Statistics, version 5 edition. Cary, NC: SAS Institute Inc.

Stevens, J. P. (1984). Outliers and influential data points in regression analysis. Psychological Bulletin, 95, 334-344.

Weisberg, S. (1985). Applied linear regression (2nd ed.). New York: Wiley.

Zoelke, P. L., & Leitner, D. W. (1980). Gender differences in performance on variables related to achievement in graduate-level statistics. Psychological Reports, 47, 1119-1125.

Wright, R. L. D. (1976). Understanding statistics. New York: Harcourt Brace Jovanovich.

Table 1

Gender and Degree Program for Subjects

	Design Sample	Assessment	Sample
Class(es)	1-6	7	8
Gender			
F	87	13	20
M	35	9	9
Degree			
Master	87	15	18
Specialist	7	1	0
Doctorate	28	6	11

le 2

dictor/Criterion Correlations, Means, and Standard Deviations

S1	S2	S3	S4	S5	S6	S7	GREV	GREQ	GREVQ	Mean	SD
1.000										20.7	3.45
.387	1.000									5.8	2.65
.569	.335	1.000								3.6	1.95
.422	.287	.423	1.000							6.7	1.43
.289	.268	.222	.339	1.000						6.8	1.51
.364	.204	.343	.474	.293	1.000					3.3	1.90
.536	.279	.430	.594	.521	.576	1.000				9.8	2.55
.115	.048	.142	-.019	-.008	-.086	.027	1.000			516.0	98.80
.527	.307	.538	.448	.267	.520	.541	.003	1.000		535.2	84.10
.488	.233	.427	.259	.168	.263	.356	.791	.598	1.000	276200.8	72115.17
.355	.211	.330	.328	.228	.417	.378	.204	.497	.472	0.0	2.08

Fifteen "Best" Subsets of Size Five, Six, and Seven

IN	R-SQUARE	ADJUSTED	MSE	C(P)	VARIABLES IN MODEL
4	0.338203	0.315577	2.96871	-0.03907	S1 S6 GREQ GREVQ
4	0.338835	0.316282	2.96565	-.154462	S5 S6 GREQ GREVQ
4	0.338963	0.316363	2.96530	-.167647	S4 S6 GREQ GREVQ
4	0.339483	0.316902	2.96296	-.255735	S1 S6 GREV GREVQ
4	0.339583	0.317005	2.96252	-.272582	S5 S6 GREV GREVQ
4	0.340093	0.317532	2.96023	-.358947	S4 S6 GREV GREVQ
5	0.339834	0.311379	2.98692	1.68491	S5 S6 S7 GREV GREVQ
5	0.339940	0.311490	2.98644	1.66694	S1 S6 S7 GREV GREVQ
5	0.340046	0.311600	2.98596	1.649	S1 S2 S6 GREV GREVQ
5	0.340055	0.311609	2.98592	1.64758	S3 S5 S6 GREV GREVQ
5	0.340077	0.311632	2.98582	1.6438	S2 S5 S6 GREV GREVQ
5	0.340154	0.311712	2.98547	1.63086	S4 S6 GREV GREQ GREVQ
5	0.340236	0.311798	2.98510	1.6169	S3 S4 S6 GREV GREVQ
5	0.340344	0.311910	2.98461	1.59868	S4 S6 S7 GREV GREVQ
5	0.340616	0.312194	2.98338	1.55267	S2 S4 S6 GREV GREVQ
5	0.340668	0.312249	2.98315	1.5438	S1 S4 S6 GREQ GREVQ
5	0.340869	0.312459	2.98224	1.50978	S1 S5 S6 GREQ GREVQ
5	0.341267	0.312873	2.98044	1.44252	S4 S5 S6 GREQ GREVQ
5	0.341824	0.313454	2.97792	1.34834	S1 S5 S6 GREV GREVQ
5	0.341938	0.313573	2.97740	1.32907	S1 S4 S6 GREV GREVQ
5	0.342108	0.313751	2.97663	1.30017	S4 S5 S6 GREV GREVQ
6	0.341591	0.307239	3.00488	3.38775	S2 S4 S5 S6 GREQ GREVQ
6	0.341828	0.307488	3.00379	3.34767	S1 S5 S6 S7 GREV GREVQ
6	0.341838	0.307499	3.00375	3.34591	S1 S3 S5 S6 GREV GREVQ
6	0.341889	0.307553	3.00351	3.33728	S1 S5 S6 GREV GREQ GREVQ
6	0.341947	0.307613	3.00325	3.32752	S1 S4 S6 GREV GREQ GREVQ
6	0.341951	0.307618	3.00323	3.32681	S1 S3 S4 S6 GREV GREVQ
6	0.341958	0.307625	3.00320	3.32568	S1 S4 S6 S7 GREV GREVQ
6	0.342072	0.307745	3.00268	3.30638	S1 S2 S5 S6 GREV GREVQ
6	0.342111	0.307786	3.00250	3.29973	S4 S5 S6 S7 GREV GREVQ
6	0.342215	0.307896	3.00203	3.28212	S4 S5 S6 GREV GREQ GREVQ
6	0.342221	0.307902	3.00200	3.28118	S3 S4 S5 S6 GREV GREVQ
6	0.342236	0.307918	3.00193	3.27855	S1 S2 S4 S6 GREV GREVQ
6	0.342357	0.308045	3.00138	3.25814	S2 S4 S5 S6 GREV GREVQ
6	0.342550	0.308248	3.00050	3.22552	S1 S4 S5 S6 GREQ GREVQ
6	0.343555	0.309306	2.99591	3.05546	S1 S4 S5 S6 GREV GREVQ
7	0.342227	0.301838	3.02830	5.28006	S3 S4 S5 S6 S7 GREV GREVQ
7	0.342251	0.301863	3.02820	5.27609	S1 S2 S4 S6 GREV GREQ GREVQ
7	0.342252	0.301864	3.02819	5.27593	S1 S2 S4 S6 S7 GREV GREVQ
7	0.342271	0.301884	3.02810	5.27263	S1 S2 S3 S4 S6 GREV GREVQ
7	0.342292	0.301907	3.02800	5.26901	S3 S4 S5 S6 GREV GREQ GREVQ
7	0.342359	0.301978	3.02770	5.25768	S2 S4 S5 S6 S7 GREV GREVQ
7	0.342422	0.302045	3.02741	5.24706	S2 S3 S4 S5 S6 GREV GREVQ
7	0.342471	0.302097	3.02718	5.23875	S2 S4 S5 S6 GREV GREQ GREVQ
7	0.342603	0.302237	3.02657	5.2164	S1 S3 S4 S5 S6 GREQ GREVQ
7	0.342681	0.302319	3.02622	5.20333	S1 S4 S5 S6 S7 GREQ GREVQ
7	0.342742	0.302384	3.02593	5.193	S1 S2 S4 S5 S6 GREQ GREVQ
7	0.343565	0.303257	3.02215	5.05378	S1 S3 S4 S5 S6 GREV GREVQ
7	0.343589	0.303283	3.02204	5.0497	S1 S4 S5 S6 GREV GREQ GREVQ
7	0.343682	0.303382	3.02161	5.03391	S1 S4 S5 S6 S7 GREV GREVQ
7	0.343684	0.303384	3.02160	5.03354	S1 S2 S4 S5 S6 GREV GREVQ

Table 4

Number of Students Predicted to Benefit from Accelerated Course

		Model 2 Prediction: Yes	No	
Instructor	Yes	9	2	11
Judgment	No	1	110	111
		10	112	122

Note. Judgments/predictions are for the six design-sample classes.

Figure 1. Frequency histogram of SCORE.

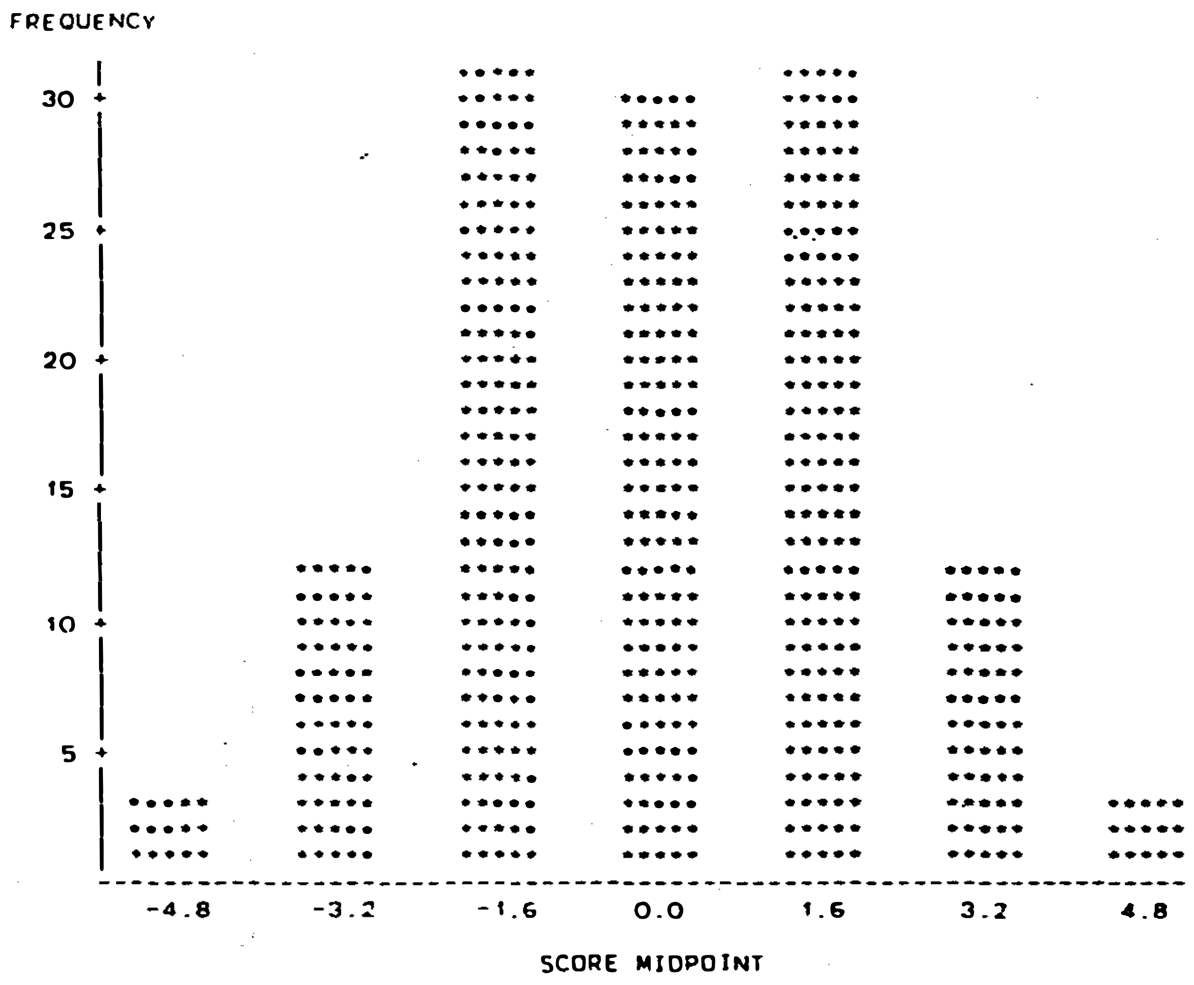


Figure 2. Plot of adjusted R^2 against sub set size.

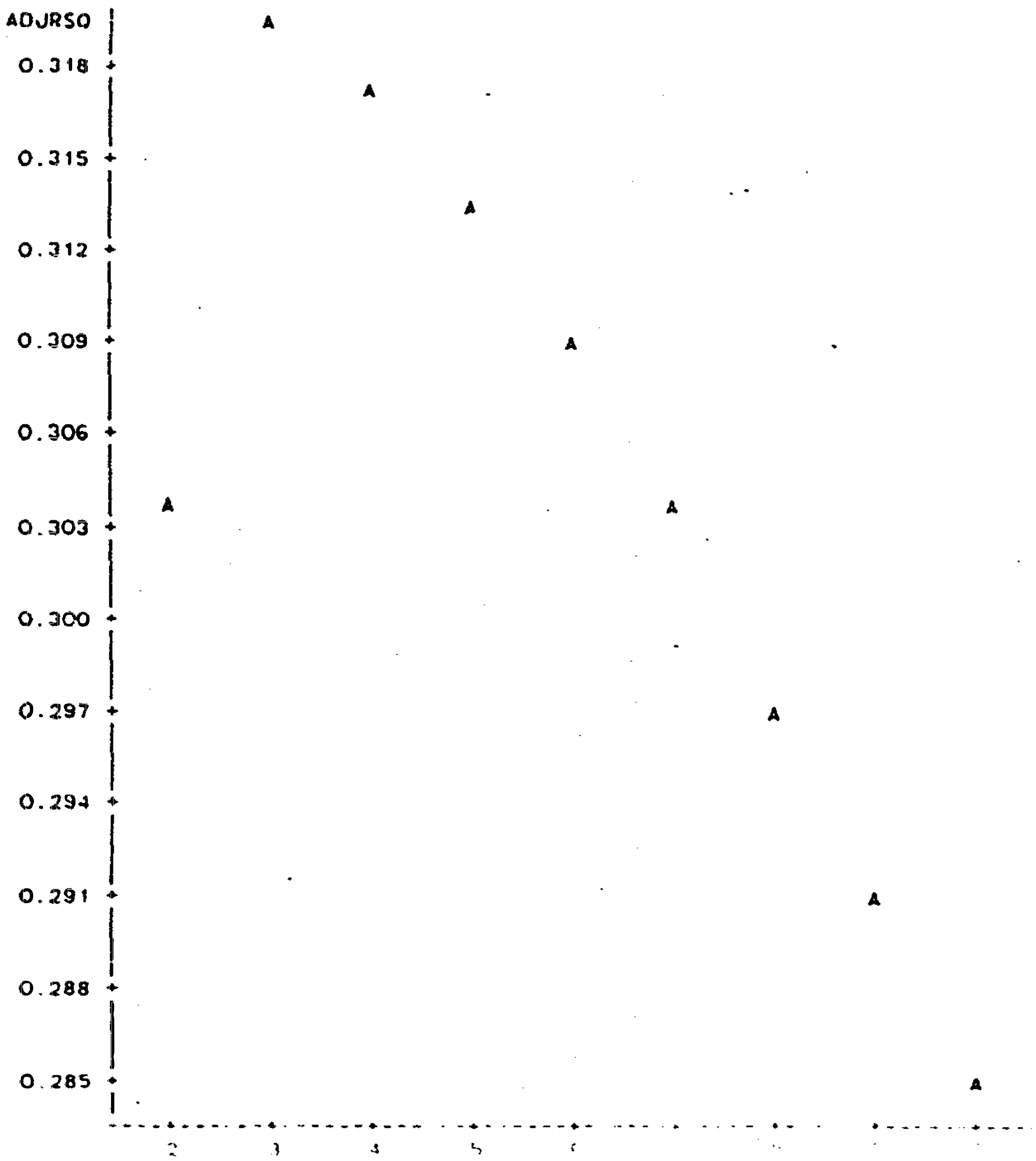


Figure 3. Plot of mean square error against sub set size.

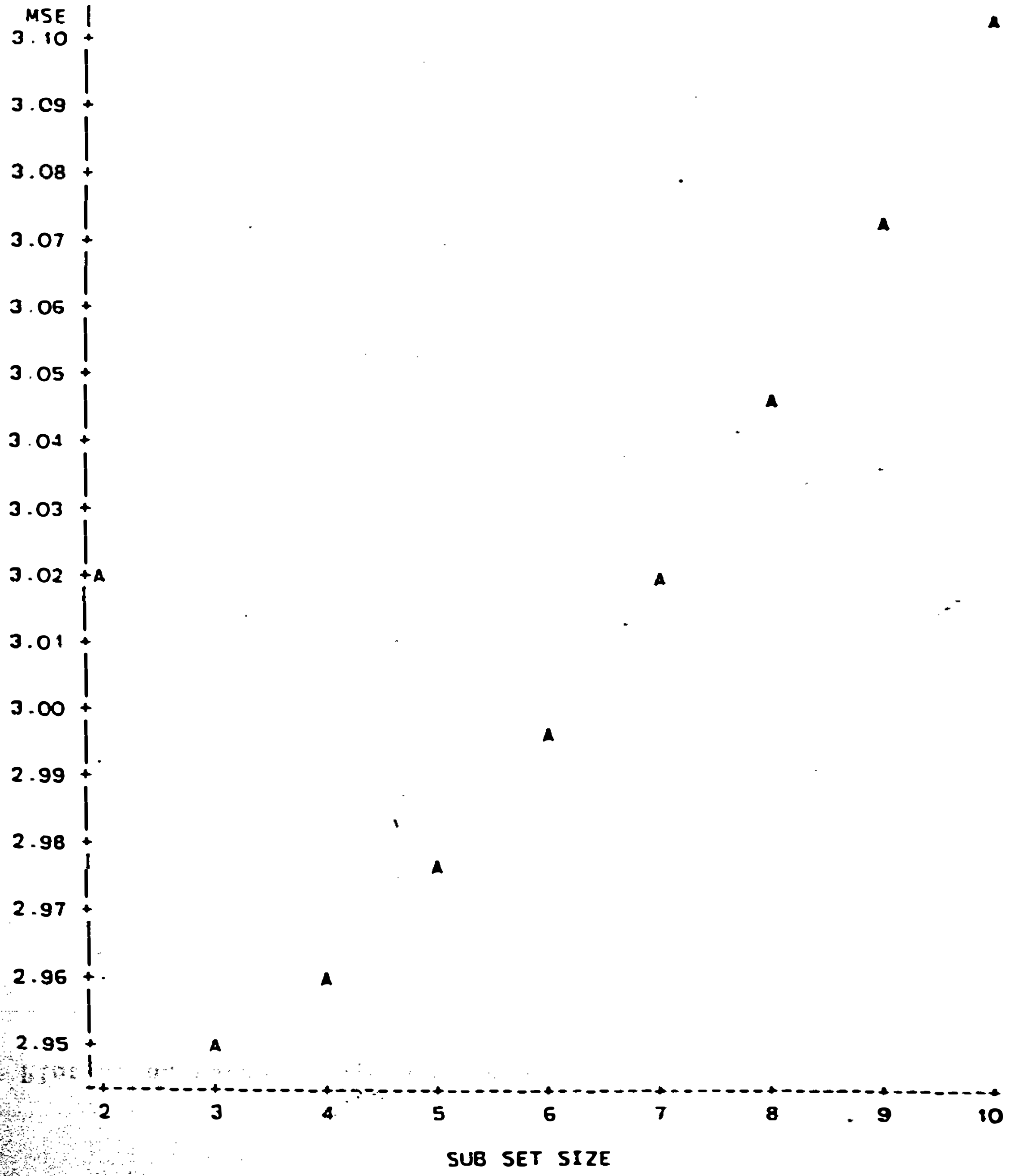


Figure 4. Plot of Cp against p.

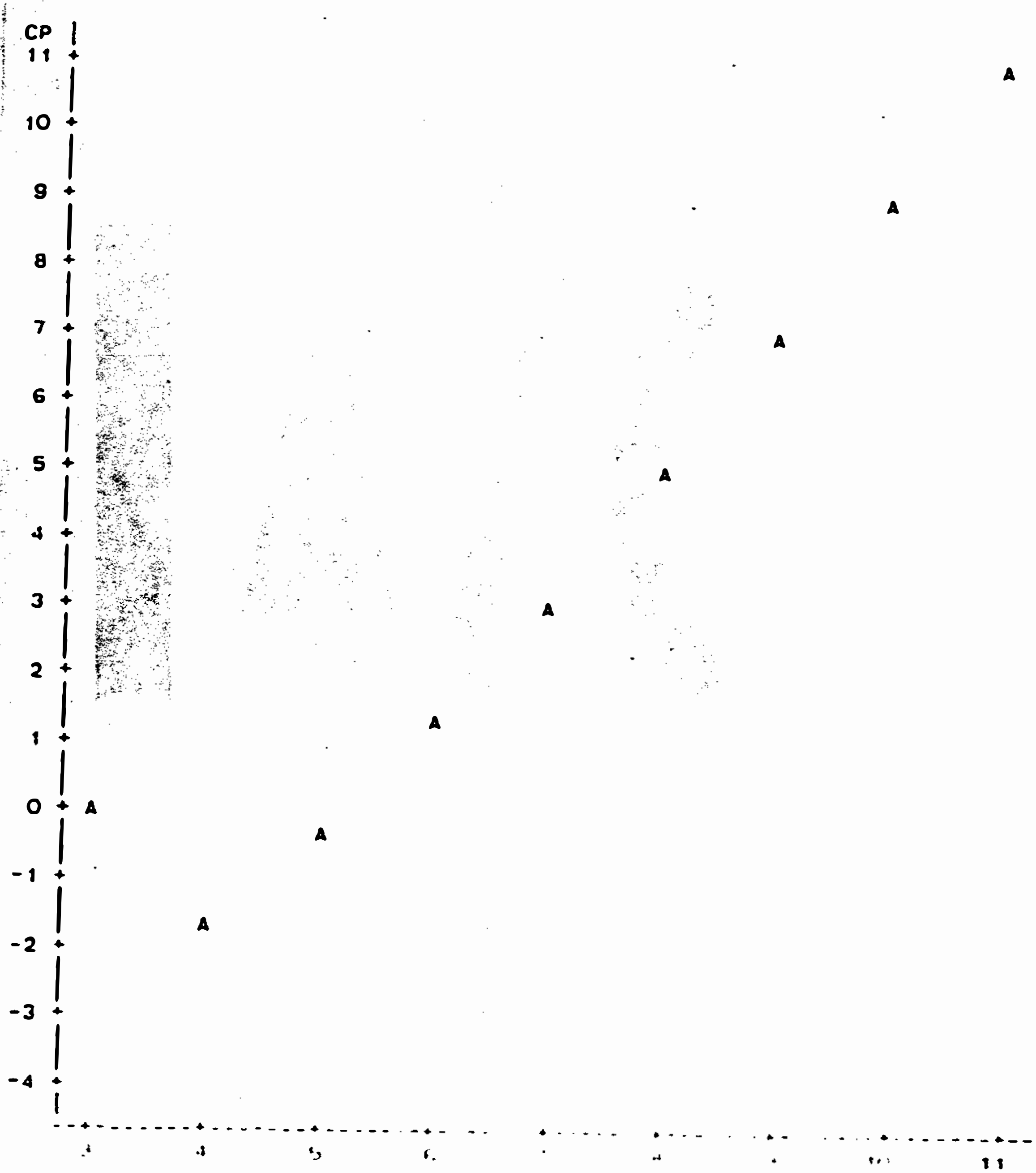


Figure 5. Hypothetical plots of residuals.

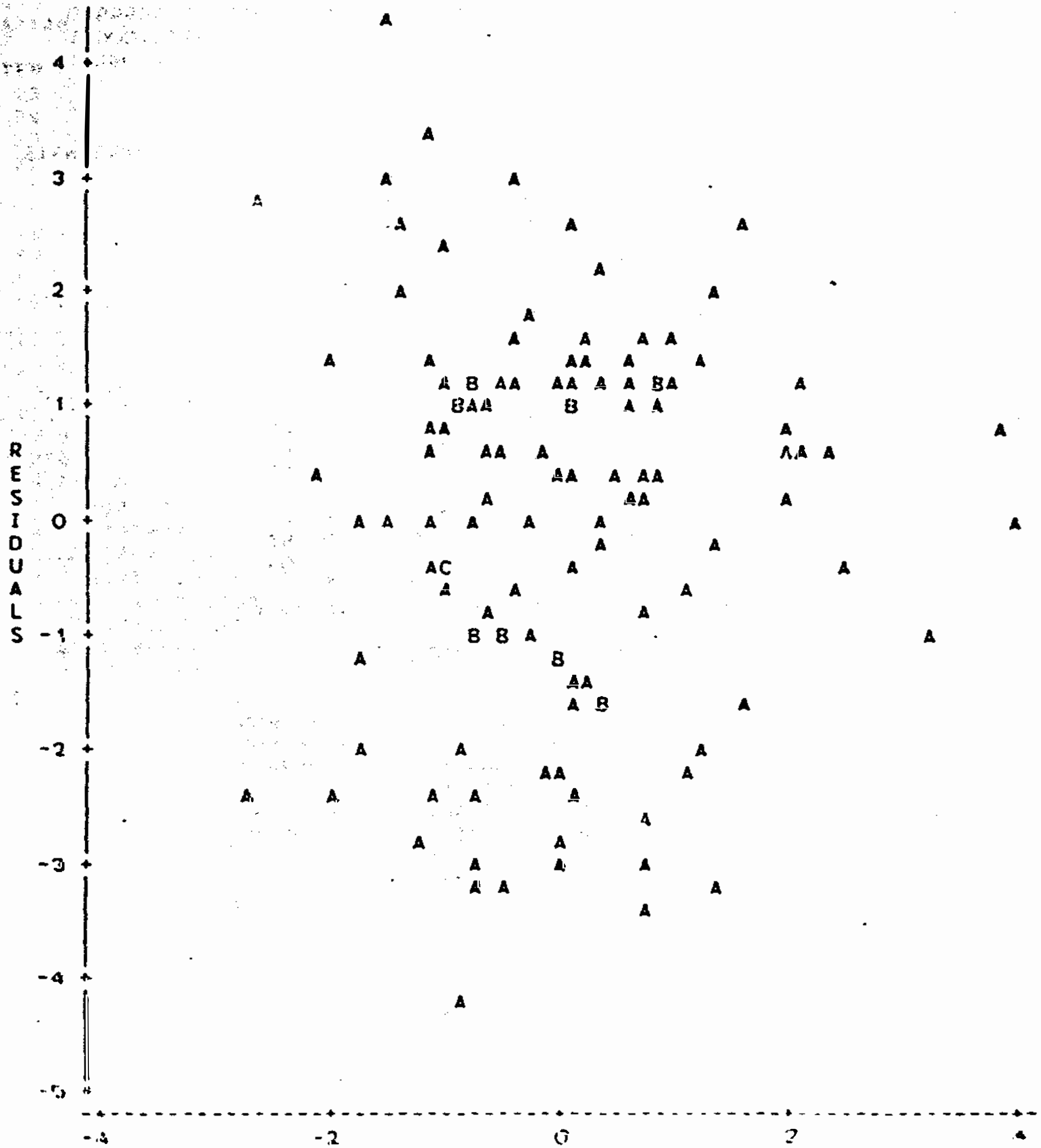
1

2

3

4

Figure 6. Plot of residuals against predicted value for Model 2.



```
VARIABLE=RSTUDENT          STUDENTIZED RESIDUAL

                                MOMENTS

              N                 122   SUM WGTS           122
              MEAN       0.00442784   SUM           0.540197
              STD DEV       1.01898   VARIANCE       1.03831
              SKEWNESS    -0.158568   KURTOSIS    -0.0695343
              USS           125.638   CSS            125.636
              CV            23012.9   STD MEAN     0.0922538
              T:MEAN=0    0.0479963   PROB>|T|      0.961798
              SGN RANK        162.5   PROB>|S|      0.678941
              NUM ¬= 0          122

                             QUANTILES(DEF=4)

               100% MAX     3.10708       99%       2.87259
                75% Q3     0.704903       95%       1.60965
                50% MED    0.179083       90%       1.16035
                25% Q1    -0.661022       10%      -1.47887
                 0% MIN     -2.5126        5%      -1.82563
                                           1%      -2.40553

               RANGE        5.61968
               Q3-Q1        1.36592
               MODE         -2.5126

                                EXTREMES

                  LOWEST       ID         HIGHEST      ID
                 -2.5126(      88)        1.736(      216)
                -2.04709(     148)      1.75613(       17)
                -1.92431(     212)      1.79609(      144)
                -1.92394(     151)      2.08756(      176)
                -1.88357(     207)      3.10708(      215)

STEM LEAF                                  #                    BOXPLOT
   3 1                                     1                       0
   2
   2 1                                     1                       |
   1 556788                                6                       |
   1 00012234                              8                       |
   0 55555666666666777777778888888899     32                    +-----+
   0 111122222333333444                   18                    *--+--*
  -0 43333222222000000                    17                    |     |
  -0 999877666665555                      15                    +-----+
  -1 433222100                             9                       |
  -1 9998877655555                        13                       |
  -2 0                                     1                       |
  -2 5                                     1                       |
     ----+----+----+----+----+----+--
```

Figure 8. Diagonal elements of the HAT matrix.

```
VARIABLE=H                H LEVERAGE

STEM LEAF                                 #
  25 2                                    1
  24
  23
  22
  21
  20
  19 4                                    1
  18
  17
  16
  15
  14 7                                    1
  13 0                                    1
  12
  11 16                                   2
  10 3                                    1
   9 8                                    1
   8 67789                                5
   7 0133345677                          10
   6 011137889                            9
   5 000222456899                        12
   4 1122222223345567788                 19
   3 00111222222233445667788899          26
   2 01111122334455556788899             23
   1 2356788999                          10
     ----+----+----+----+----+-
  MULTIPLY STEM.LEAF BY 10**-02
```

VARIABLE=H   H LEVERAGE

MOMENTS

N	122	SUM WGTS	122
MEAN	0.0491803	SUM	6
STD DEV	0.0339644	VARIANCE	0.00115358
SKEWNESS	2.85553	KURTOSIS	12.3066
USS	0.434665	CSS	0.139583
CV	69.0609	STD MEAN	0.00307499
T:MEAN=0	15.9937	PROB>|T|	0.0001
SGN RANK	3751.5	PROB>|S|	0.0001
NUM ¬= 0	122		

QUANTILES(DEF=4)

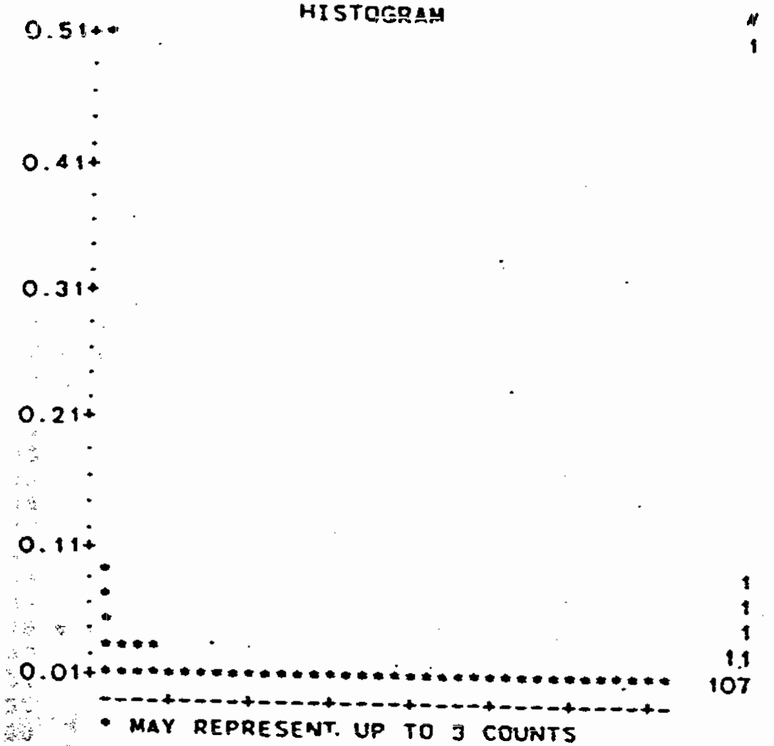
100% MAX	0.252143	99%	0.238775
75% Q3	0.0607935	95%	0.109586
50% MED	0.041395	90%	0.0866573
25% Q1	0.0283217	10%	0.0205924
0% MIN	0.0121976	5%	0.0178921
		1%	0.0123143
RANGE	0.239945		
Q3-Q1	0.0324718		
MODE	0.0121976		

EXTREMES

LOWEST	ID	HIGHEST	ID
0.0121976(	23)	0.115984(	172)
0.0127051(	15)	0.129566(	168)
0.0147533(	5)	0.14701(	144)
0.0162169(	178)	0.194024(	157)
0.0173934(	12)	0.252143(	215)

Figure 9. Cook's D.

VARIABLE=COOKD COOK'S D INFLUENCE STATISTIC



VARIABLE=COOKD COOK'S D INFLUENCE STATISTIC

MOMENTS

N	122	SUM WGTS	122
MEAN	0.0119143	SUM	1.45355
STD DEV	0.0467903	VARIANCE	0.00218933
SKEWNESS	9.87782	KURTOSIS	103.964
USS	0.282228	CSS	0.26491
CV	392.723	STD MEAN	0.0042362
T:MEAN=0	2.81251	PROB>|T|	0.00573749
SGN RANK	3751.5	PROB>|S|	0.0001
NUM ^= 0	122		

QUANTILES(DEF=4)

100% MAX	0.504815	99%	0.409618
75% Q3	0.00929867	95%	0.0341537
50% MED	0.00351305	90%	0.0229737
25% Q1	.000854951	10%	.000172662
0% MIN	5.709E-08	5%	.000016249
		1%	6.158E-08
RANGE	0.504815		
Q3-Q1	0.00844372		
MODE	5.709E-08		

EXTREMES

LOWEST	ID	HIGHEST	ID
5.709E-08(	139)	0.0387193(	41)
7.660E-08(	12)	0.0448415(	81)
7.434E-07(	169)	0.0680216(	176)
.000001004(	150)	0.0909179(	144)
9.217E-06(	137)	0.504815(	215)

Figure 10. DFFITS.

VARIABLE=DFFITS DIFFERENCE IN FIT INFLUENCE

```
STEM LEAF                          #
  18 0                             1
  17
  16
  15
  14
  13
  12
  11
  10
   9
   8
   7 5                             1
   6 5                             1
   5
   4 16                            2
   3 268                           3
   2 001244444667                 12
   1 11122233333445555677889      23
   0 1122333344455577778888899    25
  -0 9999777666554330000          19
  -1 99998876554320               14
  -2 9765443320                   10
  -3 65400                         5
  -4 96211                         5
  -5 2                             1
     ----+----+----+----+----+
 MULTIPLY STEM.LEAF BY 10**-01
```

VARIABLE=DFFITS DIFFERENCE IN FIT INFLUENCE

MOMENTS

N	122	SUM WGTS	122
MEAN	0.0210004	SUM	2.56205
STD DEV	0.272165	VARIANCE	0.0740736
SKEWNESS	2.32027	KURTOSIS	14.5353
USS	9.01671	CSS	8.96291
CV	1296	STD MEAN	0.0246406
T:MEAN=0	0.852268	PROB>|T|	0.395749
SGN RANK	228.5	PROB>|S|	0.560204
NUM ¬= 0	122		

QUANTILES(DEF=4)

100% MAX	1.80412	99%	1.56057
75% Q3	0.14309	95%	0.374837
50% MED	0.0324685	90%	0.240234
25% Q1	-0.147511	10%	-0.286331
0% MIN	-0.522705	5%	-0.400001
		1%	-0.514371
RANGE	2.32683		
Q3-Q1	0.290601		
MODE	-0.522705		

EXTREMES

LOWEST	ID	HIGHEST	ID
-0.522705(	81)	0.413054(	26)
-0.48647(	41)	0.457041(	10)
-0.456123(	88)	0.648031(	176)
-0.419952(	151)	0.745637(	144)
-0.409005(	162)	1.80412(	215)

Figure 11. Covariance ratio.

VARIABLE=COVRATIO   COVARIANCE RATIO INFLUENCE

```
STEM LEAF                       #
 130 2                          1
 128
 126
 124
 122
 120 0                          1
 118 0                          1
 116 0                          1
 114 086                        3
 112 567935                     6
 110 223770458999              12
 108 034567788802345666667     21
 106 012467778899922355558     21
 104 22245666822256678899      20
 102 5619                       4
 100 12351347                   8
  98 2522345                    7
  96 0279                       4
  94 0                          1
  92 2404                       4
  90 028                        3
  88 2                          1
  86 83                         2
  84
  82
  80
  78 0                          1
     ----+----+----+----+-
MULTIPLY STEM.LEAF BY 10**-02
```

VARIABLE=COVRATIO   COVARIANCE RATIO INFLUENCE

MOMENTS

N	122	SUM WGTS	122
MEAN	1.05385	SUM	128.569
STD DEV	0.0724998	VARIANCE	0.00525622
SKEWNESS	-0.602008	KURTOSIS	2.0435
USS	136.128	CSS	0.636003
CV	6.87955	STD MEAN	0.00656383
T:MEAN=0	160.554	PROB>|T|	0.0001
SGN RANK	3751.5	PROB>|S|	0.0001
NUM ^= 0	122		

QUANTILES(DEF=4)

100% MAX	1.30205	99%	1.28085
75% Q3	1.09568	95%	1.14665
50% MED	1.06704	90%	1.12601
25% Q1	1.01606	10%	0.946154
0% MIN	0.789554	5%	0.913275
		1%	0.807677
RANGE	0.5125		
Q3-Q1	0.0796142		
MODE	0.789554		

EXTREMES

LOWEST	ID	HIGHEST	ID
0.789554(	88)	1.15635(	219)
0.868347(	215)	1.17045(	155)
0.872611(	148)	1.19007(	172)
0.891833(	212)	1.20989(	168)
0.899888(	207)	1.30205(	157)