**PUTTING IT ALL TOGETHER: A STUDENT PROJECT REPRINTED**

The following is a project submitted by a student that uses the *NELS* data set to examine the relationship between mathematics achievement in twelfth grade and several independent variables. While more than two independent variables are studied, this project shows to what extent the two independent variable case is easily generalized to more than two variables. The project is a good example of the kind of analysis and writing required by this type of investigation.

Responding to mounting pressure from federal and state agencies to monitor the success of schools in the United States, the National Center of Education Statistics (NCES) of the US Department of Education conducted a survey in the Spring of 1988 on a nationally representative sample of approximately 25,000 eighth graders to measure achievement outcomes in English, history, mathematics, and science, in addition to personal, familial, social, institutional, and cultural factors that might relate to these outcomes. A follow-up of these students was conducted during tenth grade in the spring of 1990; a second follow-up was conducted during the twelfth grade in the spring of 1992; and, finally, a third follow-up was conducted in the spring of 1994.

Approximately 5,000 students who responded to all four administrations of the survey were always at grade level (neither repeated nor skipped a grade) and pursued some form of post-secondary education. Using a random sample of 500 of these 5,000 students, this project seeks to model the relationship between mathematics achievement in twelfth grade, as measured by a standardized test of mathematics achievement, and a number of variables purported to influence such achievement, including student and home background variables (gender, socio-economic status), self-concept, educational aspirations, enrollment in advanced math in eighth grade, and computer ownership.

### Measuring the variables

According to the NELS:88 handbook, the development of the survey “was guided by the research objectives of this longitudinal study. Items were chosen based on their utility in predicting or explaining future outcomes as measured in later survey waves. Questions were framed to provide consistency and continuity with earlier education longitudinal studies, as well as to address new areas of policy concern and to reflect recent advances in theory.”

The variables used in this study were measured as follows:

*Mathematics achievement* in twelfth grade was assessed using a standardized test of mathematics achievement. The actual range of scores is 39.28 through 71.12, from low to high achievement.

*Socioeconomic status (SES)* is a composite of father’s education level, mother’s education level, father’s occupation, mother’s occupation, and family income. Each component was standardized to a mean of 0 and a standard deviation of 1 and then averaged to yield the socioeconomic composite. After linear transformation to avoid negative values, the actual range for SES is 2 through 32, from low to high SES.

*Sex* is a dichotomous variable coded 0 for males and 1 for females.

*In advanced math* is a dichotomous variable that measures whether the student was enrolled in an advanced, enriched, or accelerated math class in eighth grade. The variable was coded 1 if the student was enrolled and 2 if the student was not enrolled.

*Does family have a computer* is a dichotomous variable coded 1 if the family owns a computer and 2 otherwise.

*Self-esteem* is a composite score based on responses to the following four items: I feel good about myself, I feel I am a person of worth, the equal of other people, I am able to do things as well as most other people, On the whole, I am satisfied with myself. Each of these four items was standardized separately to a mean of 0 and standard deviation of 1. After linear transformation to avoid negative scores, the actual range is 0 through 43, from low to high esteem.

*Educational aspiration* measures how far in school the student expects to go. Coding is as follows: 1= Some high school; 2 = finish/earn high school equivalency; 3 = vocational/trade-less than 2 yrs; 4 = vocational/trade-2+ yrs; 5 = college-less than 2 yrs; 6 = college/2+yrs-Assoc; 7 = college/Bachelor’s; 8 = College-Master’s degree; 9 = College program - PhD; 10 = MD, LLB,JD,DDS, or equivalent.

All variables, other than the three described above as measured dichotomously, are treated in this study as if they were continuous.

**Examining the variables individually and in pairs**

The sample of 500 cases consists of 227 males and 273 females. Numeric summaries of each of the variables under study are presented in Table 1 for the total group of 500.

Table 1. Numeric summaries of all variables for total group of 500 cases.



According to Table 1, none of the variables appear to be sufficiently skewed to warrant the use of symmetrizing transformations. To verify this impression, the distribution of the worst case scenario, that of highest level of education expected (skewness = -.901), is graphed in Figure 1. Figure 1 visually corroborates the notion that symmetrizing transformations are not necessary in this case, because the normal curve describes reasonably well the shape of the histogram.



Figure 1. Histogram of highest level of education expected.

Table 2 presents numeric summaries of all variables by gender, and Table 3 presents the matrix of intercorrelations between all pairs of variables. According to Table 2, males and females have similar profiles in terms of these variables.

Table 2. Numeric summaries of selected variables by gender



Table 3. Matrix of intercorrelations between all pairs of variables.



According to Table 3, each of the six regressors, owning a computer, educational aspirations, self-esteem, SES, sex, and enrolled in advanced math in eighth grade, is statistically significantly related to the criterion, mathematics achievement in twelfth grade. Moreover, both highest level of education expected and SES account for between 10 and 14 percent of the variance of the criterion. Yet, the intercorrelations between regressors are sufficiently low to suggest that redundancy of regressors is not a problem.

The results of Table 3 suggest that an individual with higher mathematics achievement in twelfth grade tends to be a male with higher self esteem, higher SES, and a higher level of expected education. In addition, the individual tends to have taken advanced math in eighth grade and owns a computer. Not surprisingly, individuals with higher SES tend also to be those who own a computer and expect to complete a higher level of education.

Scatterplots between all pairs of continuous variables were examined and confirm that the relationships depicted are linear and that the correlations reported in Table 3 are appropriate measures of relationship in this case.

### Examining the variables multivariately with mathematics achievement as the criterion

Multiple regression analysis was used to fit a full model including all main effects and two-way interactions. The statistical significance of the set of 15 interaction terms (six regressors taken pairwise) is tested by the significance of the F change statistic in Table 4. According to the results of Table 4, the F change statistic is not significant (F (15,469) = 1.487, p = .105), suggesting that the .033increase in R-square is not statistically significant and that the set of two-way interactions does not contribute to explaining mathematics achievement variance. Therefore, the set of two-way interactions was removed from the model.

Table 4. Test for the significance of the set of two-way interactions

The reduced model with only the set of six main effects is statistically significant according to Table 4 (F(6,484) = 30.234, p = .000). Furthermore, the proportion of mathematics achievement variance explained is .273. That is, more than one-fourth of the variance of mathematics achievement in twelfth grade is explained by these six variables taken collectively. An analysis of which variables contribute individually to explaining mathematics achievement may be found in Table 5.

Table 5. Regression coefficients and their significance in predicting math achievement



According to Table 5, when all variables are considered simultaneously, neither self-esteem nor owning a computer contribute significantly to the model. In the case of owning a computer, it is likely that the relationship between owning a computer and mathematics achievement became zero after controlling for SES. A further reduced model, without owning a computer and self-esteem, is presented in Table 6. Note that dropping both variables at once might have a surprising effect, especially if they are highly correlated.

Table 6. Model summary and regression coefficients with four regressors





According to Table 6, a negligible loss of variance explained resulted from eliminating the two non-significant variables, owning a computer and self-esteem. With the four regressors remaining, the reduced model accounts for 0.269 of mathematics achievement variance. This reduced model accounts for virtually the same amount of math achievement variance with two fewer variables. The interaction terms were not retested.

Based on a review of the regression coefficients in Table 6, controlling for whether an individual enrolled in advanced math in eighth grade, as well as for SES and sex, the greatest single predictor of twelfth grade mathematics achievement is an individual’s highest level of education expected. Having high educational aspirations, controlling for other relevant factors, appears to associate with achievement in mathematics. While one cannot say, based on this analysis, that strong mathematics achievement in twelfth grade results from having high educational aspirations, we can conclude that these two variables, after controlling for other relevant factors, are related significantly.

In addition, we may conclude that controlling for SES, sex, and educational aspirations, whether or not one enrolls in advanced math in eighth grade is a strong predictor of mathematics achievement in twelfth grade. Those who enroll tend to achieve better in math in twelfth grade than those who do not, suggesting that achievement in math for both males and females is a stable trait across the years.

Finally, SES and sex, controlling for the other two variables and each other, are both related to math achievement in twelfth grade. Controlling for all other variables in the equation, boys tend to achieve better in math than girls and individuals with higher SES tend to achieve better in math than those with lower SES. That boys tend to surpass girls in math is a well-known, yet unfortunate fact. That those who are more privileged economically tend to achieve better than those who are less privileged is also not a surprising finding.

A series of residual analyses was carried out to examine more closely the fit of the model to these data. In particular, partial plots between the residual scores and each of the two continuous regressors were examined for evidence of possible nonlinearity. These plots appear as Figures 2 and3 and suggest that allassociation between mathematics achievement in twelfth grade and each of the significant regressors is accounted for by the regression model in Table 6.

 

Figure 2. Partial residual plot: Figure 3. Partial residual plot:

Ed aspirations SES

In addition, boxplots of Cook’s distance and leverage values, as depicted in Figure 4, suggest that no point or set of points is unduly influencing the results of the analysis and distorting the results obtained as all distance and leverage values cluster near zero.



Figure 4. Boxplots of Cook’s distance and leverage values

Thus, according to these diagnostics, the regression model, fit to these 500 cases, is appropriate.

In conclusion, the results of this analysis based on 500 randomly selected cases from 5,000 nationally sampled students who continued to postsecondary education are interesting for at least two reasons. First, they suggest that a large portion (greater than 25 percent) of the variance due to mathematics achievement in twelfth grade can be attributable to four variables, educational aspirations, enrollment in advanced math in eighth grade, socioeconomic status, and sex, taken collectively. Second, they support the notion that each variable contributes uniquely to explaining twelfth grade math achievement over and above the contributions made by each of the other variables in the equation.

While the results of this study do not support causal inference, the relationships observed suggest a direction of study to examine ways to promote achievement in mathematics in twelfth grade and tolook forother variables that have the potential to explain the remaining 75 percent of mathematics achievement variance left unexplained by the variables in this study.