Interpreting Regression Analysis Results: An Example

Carl J Huberty
J. Brady Allen
University of Georgia

An example of the application of multiple regression is presented in some detail. Predictor variable scores are based on the three parts of the Graduate Record Examination (GRE). Criterion variable scores are based on the performance of graduate students in an introductory statistical methods course. Even though the general predictive power of the GRE measures is assessed, the focus of the interpretation is on the prediction of the criterion for specified profiles of predictor measures.

The Graduate Record Examinations (GRE) are in widespread use across graduate-level universities in the United States. Typically, performances on the verbal test (GREV), the quantitative test (GREQ), and the analytical test (GREA) are used for admission purposes; sometimes score sums, such as GREV + GREQ, are utilized. Not only is the use widespread, so is the questioning of the use of GRE performance for admission purposes (Morrison & Morrison, 1995). The questioning position often espoused is usually based on the low relationship between performance on the GRE and performance in graduate school (as assessed by graduate grade point average). That is, what is being questioned is the predictive validity of GRE performance relative to performance in graduate school.

It might be argued that there are two related difficulties in assessing GRE performance predictive validity. One difficulty is that variability in the GRE scores is necessarily restricted because only those students with higher GRE scores are typically admitted to graduate school. The other difficulty is that the variability of *overall* performance in graduate school, as typically assessed by A-B-C grading, is quite restricted because of typical current grading practices (see, e.g., Cole, 1993).

But, how about using GRE scores in predicting the performances in a particular area of study in graduate school? Some restriction was considered by Kluever and Green (1992) in their prediction study for students in a college of education. An intent of the current study is to determine how useful performance on the GRE is in predicting performance in introductory graduate level statistical methods courses. The overriding purpose of this article is to illustrate how specific prediction information might be obtained; that is, information more specific than an overall index of relationship.

Analysis Units

A graduate statistical methods course taught at The University of Georgia in the College of Education might be titled Educational Statistics I. Topics covered in ESI include data description, correlation, and inference regarding a mean, proportion, and correlation.

The analysis units used in this study are students from a collection of six ESI classes. There was a total of 135 ESI students. Eleven of these students had not taken any of the GRE and, therefore, were not considered analysis units. [Class performance data for these 11 students were, however, used in calculating standard scores for the remaining 124 students.] Of the 124 ESI students, 48 were master level, 16 specialist level, and 60 were doctoral level. About 92 percent of the 124 ESI students were enrolled in Education graduate programs. Two textbooks were used with the ESI classes. For the first five classes, Moore and McCabe (1989, chaps. 1-8) was used; for the sixth class, Moore (1995, chaps. 1-7) was used.

Criterion Measure

Student performance was based on three types of assessment, four quizzes, a test, and an examination. The sequence of assessments used is Q_1, Q_2, T, Q_3, Q_4, E. The final examination covered material in the second half of each course. Three scores were obtained for each student in each class: (highest) sum of three 10-point quizzes, score on the 35-item test, and score on the 45-item examination. [All items on the quizzes, test, and examination were of the multiple-choice variety, focusing mostly on concepts. Typical score ranges were approximately 27-15 for the quiz sum, 33-15 for the test, and 35-15 for the examination.] Each of these three scores was transformed to z scores using the mean and standard deviation based on all six classes for ESI. [In "real life" the three scores are transformed using data on a current class plus the three most recent classes.] A composite of the three z scores,

$$Z = 0.5\, z_Q + 1.0 z_T + 1.5 z_E\,,$$

served as the criterion variable score for this study. [The composite Z is the basis used in course grading.]

Table 1. Descriptors for GRE and Class Scores for ESI students

Variable	MIN	C_{25}	C_{50}	C_{75}	MAX
V	320	450	510	580	800
Q	350	350	560	620	800
A	250	500	550	610	760
V + Q	700	990	1045	1185	1510
V+Q+A	1070	1493	1605	1768	2160
Z	-8.58	-1.47	0.27	1.65	5.26

Note: V = GREV, Q = GREQ, and A + GREA.

Measurement characteristics of the quizzes, the midterms tests, and the examinations are judged to be acceptable. Specifically, content validity of the three types of scores is judged to be very respectable. Values of the Kuder-Richardson 20 index (of internal consistency) for the five midterm tests ranged from about .65 to about .85. It is to be expected that $K\text{-}R_{20}$ values for the quizzes would be lower; recall that the sum of the three highest quiz scores was used for each student. The internal consistency of the scores on the final examination was somewhat higher than that for the midterm test scores. It is also assumed that a common scale of measurement is used across classes for the quizzes, for the tests, and for the examinations.

Predictor Measures

Three parts of the Graduate Record Examinations were utilized in this study to serve as bases of predictor measures; Verbal (GREV), Quantitative (GREQ), and Analytical (GREA). Thus, the three predictor variables considered are verbal aptitude, quantitative aptitude, and analytical aptitude. For ESI, the data matrix has 124 rows and four columns (three predictors and one criterion). Completion of the basic data matrix will now be briefly discussed.

Of the 124 ESI students, six had not taken the Analytical part of the GRE. Two ways of imputing these six scores were considered. One way was simply to use the mean GREA based on the remaining 118 students. The second imputation method used was to regress GREA on GREQ and GREV using the complete data on the 118 students. To determine the way of choice, two 4x4 correlation matrices were determined using the three GRE scores and the composite, *Z*; one matrix was based on the GREA mean and the other was based on regressed GREA scores. The benchmark correlation matrix is the "available case" matrix where all but six correlations are based on 124 students; the remaining six correlations are based on 118 students. The three correlation matrices were visually compared; for the purposes of this article, the regressed GREA value was used to replace the six missing scores. So, a full 124 x 4 data matrix was used in the analyses.

Results

Three sets of predictor measures were considered: (1) GREV (denoted V), GREQ (Q), GREA (A); (2) V + Q, A; and (3) V + Q + A. So then, the composite *Z* was regressed on V, Q, and A, on V + Q and A, and on V + Q + A. Table 1 shows descriptive information for each predictor and the criterion measured on the 124 ESI students.

The correlations among the five predictors and between each predictor and *Z* are reported in Table 2. [All three scatterplots (not reported herein) revealed reasonable linearity; normal probability-plots indicated no aberrations, as did a plot of *Z* versus .]

The predictability of performance in an introductory level statistical methods course as measured by *Z*) using the GRE scores as predictor scores may be *broadly* assessed via a multiple correlation coefficient value. The broad results for the three regression analyses are given in Table 3. The adjustment used to get R^2_{adj} is that proposed by M. Ezekiel in 1930 presented by Huberty (1994) wherein the *F*-test involving R^2_{adj} is also discussed.

Very often in prediction studies, the researcher is interested in determining a relative ordering of the predictor variables. That is, it may be of interest to determine the most and least important predictors. In a multiple regression context, we view the most important predictor as the one which when deleted from the total set of predictors will decrease the error mean square value (or, equivalently, the R^2_{adj} value)) the most -- focus is on overall predictive accuracy. This approach to assessing predictor importance is discussed by Huberty (1989) and Huberty and Petoskey (1999).

For the first prediction model that involved three predictors (V, Q, A), an ordering of the importance of the predictors may be obtained by deleting, in turn, each predictor. The *adjusted* two-predictor R^2 values were $R^2_{(Q)} = .250$, $R^2_{(A)} = .370$, and $R^2_{(V)} = .386$, where $R^2_{(Q)}$ denotes the adjusted obtained by deleting Q. Thus, Q (i.e., GREQ) is judged to be the most important predictor, with V and A of about equal importance (or unimportance). For the second prediction model (using V + Q and A), it was found

Table 2. Correlations among GRE and Class Scores

Variable	Q	A	V+Q	V+Q+A	Z
V	$.05^{ns}$	.35	.73	.66	.34
Q		.30	.72	.63	.53
A			.45	.80	.47
V + Q				.90	.60
V+Q+A					.64

Note: ^{ns} indicates not significant ($p > .05$). All other correlations were significant ($p < .001$).

Table 3. Overall Prediction Results for ESI data.

Predictor(s)	R^2	R^2_{adj}	F	df
V, Q, A	.436	.422	15.27	3.97, 120
V+Q, A	.412	.402	21.02	2.65, 121
V+Q+A	.409	.404	42.04	1.33, 122

Note: All models were significant ($p < .001$).

that the R^2_{adj} values obtained by deleting each of the predictors were $R^2_{(V+Q)}$ = .226 and $R^2_{(A)}$ = .354. Clearly, V + Q is more important than A when it comes to broad, overall prediction of Z.

More specific prediction information may be obtained by examining some particular predictor profiles. To do this, five clusters of profiles of ESI students were identified for the three-predictor model, one for the two-predictor model, and one for the one-predictor model (see Table 4). Our rationale for the cluster definitions is based on the various prediction models used at different universities; some use only predictors V and Q, some include A along with V and Q, and others use V + Q and/or V + Q + A. Also, we were interested in determining prediction quality for those who are generally high test scorers, low test scorers, and those who were high on some predictors and low on others. The question then becomes: How well can the composite Z score for ESI students be predicted for each the various profiles? The goodness of prediction was based on the magnitude of the standardized residual (see Montgomery & Peck, 1992, p. 68). If $|Z - \hat{Z}| <$ .80, it was judged that we had a "good" prediction. [The composite Z scores typically ranged from about 5.00 to about -8.00. Other cut-offs may be more appropriate in other prediction situations.]

Table 5 summarizes how well Z scores can be predicted for students with each of the 12 profiles.

Table 4. Clusters of GRE Profiles for ESI students

Cluster	Size	Part		Centile	Score
1	8	Q	≥	70	650
		A	≥	60	590
		V	≥	50	480
2	17	A	≥	80	650
3	12	Q	≥	80	710
4	9	Q	≤	30	480
		A	≤	40	510
		V	≤	50	480
5	23	Q	≥	50	570
		V	≤	50	480
6	25	A	≥	60	
		V+Q	≥		1100
7	20	V+Q+A	≥		1800
		V+Q+A	≤		2000

Note: V = GREV, Q = GREQ, and A + GREA.

Some summary statements are given below:

1. Of the five ESI GRE score profiles for the three-predictor model, clusters 1 and 2 had over 80% small ($|Z - \hat{Z}| < .80$) prediction errors. That is, for students with high Q, A, and V scores or students with very high A scores, it was judged that the percent of good prediction of class performance was respectable.
2. For students performing poorly on all GRE parts (cluster 4), prediction was not considered very respectable (only 44% good prediction).
3. For those students who score above the median on GREQ and below the median on GREV (cluster 5), prediction was not very respectable.
4. For the two-predictor model (V+Q and A), respectable prediction resulted (80% small prediction errors) for students with "high" scores on both predictors (cluster 6).
5. Groups of students for whom respectable prediction resulted (clusters 1, 2, 6) were not dominated by high-performing (i.e., "A") students, except, possibly, for cluster 1 students with three "high" GRE scores.
6. Residuals for students in cluster 1 (with high GRE scores) indicated a dominance of over-prediction, while for cluster 6 (also with high GRE scores) there was a dominance of under-prediction; it was for these two clusters that residuals were judged respectable.

Conclusions

The results of this study would indicate that prediction of performance in an *introductory* graduate-level statistical methods course (ESI) using scores on the three GRE parts can be accomplished in a fairly successful manner. First of all, accounting for about 42% of the variability in overall course performance is judged to be fairly high, especially in relation to that found by Elmore, Lewis, and Bay (1993) and Goldberg and Alliger (1992) where *unadjusted* R^2 values ranged from .09 to .29. Secondly, accuracy of prediction in the current study was judged to be respectable for some subgroups of students. Success resulted in predicting overall course performance (as assessed by a composite Z) for subgroups/clusters of students with "high" GRE scores. Also, prediction of Z for students with "low" GRE scores was judged to be poor.

The above conclusions suggest to us that restricting the view of regression analysis results to looking at an adjusted R^2 value may very well result in placing unnecessary limitations on interpretation possibilities. For a real practical research situation it may very well be informative to learn about the predictive accuracy for analysis units with particular predictor variable score profiles.

References

Cole, W. (1993, January 6). By rewarding mediocrity we discourage excellence. *Chronicle*, pp. B1-B2.

Elmore, P. B., Lewis, E. L., & Bay, M. L. G. (1993, April). *Statistics achievement: A function of attitudes and related experiences*. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

Goldberg, E. L., & Alliger, G. M. (1992). Assessing the validity of the GRE for students in Psychology: A validity generalization approach. *Educational & Psychological Measurement, 52*, 1019-1027.

Huberty, C. J (1989). Problems with stepwise methods -- Better alternatives. In B. Thompson (Ed.), *Advances in social science methodology* (*pp*. 43-70). Greenwich, CT: JAI Press.

Huberty, C. J (1994). A note on interpreting an R^2 value. *Journal of Educational and Behavioral Statistics, 19*, 351-356.

Huberty, C. J, & Petoskey, M. D. (1999). Use of multiple correlation analysis and multiple regression analysis. *Journal of Vocational Educational Research, 24*, 15-43.

Kluever, R. C., & Green, K. E. (1992). Prediction of achievement of doctoral students in education. *Perceptual and Motor Skills, 74*, 419-423.

Montgomery, D. C., & Peck, E. A. (1992). Introduction to linear regression analysis. New York: Wiley.

Moore, D. S. (1995). *The basic practice of statistics*. New York: Freeman.

Moore, D. S., & McCabe, G. P. (1989). *Introduction to the practice of statistics*. New York: Freeman.

Morrison, T., & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the Graduate Record Examination with graduate grade point average representing the criterion of graduate success. *Educational & Psychological Measurement, 55*, 309-316.

Table 5. Results for Predicting Composite Z for Clusters of ESI students

Cluster[1]	Size	No.[2]	%	A	B	C-D	Pos.[3]	Neg.[4]
1	8	7	88	5	2	0	1	6
2	17	14	82	8	5	1	7	7
3	12	8	67	6	2	0	2	6
4	9	4	44	0	2	2	2	2
5	23	11	48	2	8	1	6	5
6	25	20	80	11	9	1	15	5
7	20	14	70	8	6	0	12	2

Note: [1]For Cluster definitions, see Table 4. [2]Number of students with a standardized residual magnitude < 0.8. [3] "Pos." indicates under-prediction. [4] "Neg." indicates over-prediction.