# A MULTINOMIAL LOGIT ANALYSIS OF ACADEMIC PERFORMANCE AND CLASSROOM ATTENDANCE USING A DISCRETE CLASSIFICATION FOR THE ATTENDANCE VARIABLE 

Samuel L. Seaman, Graziadio School of Business \& Management, Pepperdine University, Malibu, CA 90263, 310.506.4401, samuel.seaman@pepperdine.edu<br>Charles Delaney, Hankamer School of Business, Baylor University, Waco, Texas 76798, 254.710.1011, charles_delaney@baylor.edu


#### Abstract

The relationship between classroom performance, measured on an ordinal scale with 5 levels, and a number of possible explanatory variables has been investigated using a multinomial logit model. Individuals in the study were subject to an attendance policy which mandates a grade of F for any student missing more than $25 \%$ of classes. Unique to this study, the attendance variable, measured in days absent, has been treated as an ordinal variable allowing the authors to demonstrate a significant, yet disproportional effect on performance. The results may be important to institutional policy makers considering the imposition of a mandatory attendance policy, and the model used in the study could serve as an archetype for future research involving categorical dependent and independent variables.


## INTRODUCTION

The occurrence of "categorical" variables in applied business research is a common phenomenon. The appropriate use of a "categorical technique" when such data emerge is not. Several authors have encouraged their readers to use an appropriate qualitative response analysis rather than the usual quantitative technique (Becker \& Walstad, 1987; and Ananth \& Kleinbaum, 1997). More recent exhortations have demonstrated the superiority of the qualitative approach when studying the relationship between academic performance and classroom attendance (Park \& Kerr, 1990; Glasure, 2002; and Ellis, Durden, \& Gaynor, 2003). Only one of the appropriately qualitative studies mentioned above (Park and Kerr), has treated the outcome variable as a multi-level, ordinal response variable. And, with a mandatory attendance policy, the absence variable should probably have been treated as a categorical variable in the logit model. As far as we can tell, none of the studies cited above, treated the attendance variable in this way

## THE DATA

The final data set consisted of the following recorded classroom variables: each student's final course grade ( $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}$, or F ), the student's gender, the student's mid-term examination score, and the number of absences for each student. The absence variable had been recorded as a numeric value between 0 and 10 (after 10, a student would not attend any additional classes and received an automatic F in the course). Recognizing that this variable is uniquely ordinal, rather than continuous, we converted the number of days absent to a categorical variable with five levels ( $0-2$ absences, 3-4 absences, 5-6 absences, 7-8 absences, and $9+$ absences).

## METHODOLOGY

When modeling a discrete outcome variable (course grade in this case) on a set of predictor variables, the usual "quantitative techniques" (OLS for example) may be inappropriate since basic assumptions underlying those techniques are likely to be violated. In this study, the situation is further complicated by the multi-level, ordinal nature of the discrete response. Ananth \& Kleinbaum (1997) have reviewed and evaluated the preferred methods for such a contingency and, given their recommended models, three variations of the multinomial logit model seem most appropriate for the application at hand: the cumulative logit model, the continuation-ratio model, or the adjacent-categories logit model. For reasons given later, the cumulative logit, or proportional-odds logit model, was used in our study.

## RESULTS AND DISCUSSION

In the original analysis of the classroom data, we modeled the ordinal outcome variable (grade in class) as a function of six independent variables: gender, mid-term exam grade, and four dummy variables for the ordinal, attendance variable. Results of this initial analysis yielded evidence of a statistically significant effect due to absences. No other independent variables were statistically significant in the model. A reduced model with ordinal outcome variable, grade, and just four "attendance" dummy variables, fit the data as well as our original model. Additionally, the hypothesis of "proportional odds" could not be rejected for either the original or final models, and so, we determined that the proportionalodds cumulative logit model would be most appropriate for our analysis. Since the intercept terms have little information in them relative to the issue being investigated, we have not reported their values here. We have, however, reported the estimated parameters (with p-values and estimated odds ratios) for each of the attendance dummy-variables in the table below

## Table of Estimated Coefficients for the Attendance Dummy Variables

| Attendance | Parameter | P-Value | Odds Ratio |
| :---: | :---: | :---: | :---: |
| Variable | Estimate |  | Estimate |
| 0-2 absences | 3.801 | <. 0001 | 44.746 |
| 3-4 absences | 3.580 | <. 0001 | 35.889 |
| 5-6 absences | 3.304 | <. 0001 | 27.245 |
| 7-8 absences | 1.858 | <. 0001 | 6.414 |

Since all attendance variables are dummy variables (where 9+ absences was the baseline or referent group), the parameter estimate associated with each one classification (say 0-2 absences), estimates the log-odds of being in a higher rather than a lower grade classification whenever the student missed 0-2 rather than $9+$ days of class. After converting these estimates to odds-ratios, the following observation for the first dummy variable ( $0-2$ days absent) can be generalized to all other absence classifications: The odds of being in the higher rather than lower grade classification are 44.746 times greater for a student that missed just 0-2 days, rather than $9+$ days. What is particularly interesting about the results given above is the trend that can be seen in the values of the odds ratios. Notice that the odds of being in a higher versus a lower grade classification are very similar for the first three attendance classifications, yet very different for the last attendance classification ( $7-8$ absences). The practical implication, here, is that students missing fewer than 6 classes tended to be significantly better off (in terms of their odds of being in a higher grade classification) than students missing 9 or more classes (these students generally failed the course). Students with 7 or 8 absences, whilst more likely to be in the higher grade classification, reduce their odds of doing so substantially

## COMMENTS

The study described above was not comprehensive. Whilst we included two covariates in the investigation, along with the attendance variables, many other pertinent factors have been identified in the literature. Unfortunately, we did not have access to those variables for this study. We have acknowledged the ordinal nature of the outcome variable and have used an appropriate multinomial logit model, a relatively unique approach in the applied business literature. Perhaps even more interesting, we have established that the effect of the absence variable on performance is not linear and, in fact, changes substantially once a student has missed more than 6 class periods. We hope that policy makers will use great care should they decide to use a mandatory attendance policy. Especially important is the "number of absences allowed". If that number is too large, it is quite possible that students will assume (falsely) that their performance will not suffer until they have approached the maximum number permitted. Lastly, we hope that this simple example of the use of a multinomial logit model (for an ordinal outcome variable and a variety of independent variables having different scales) will encourage other applied researchers to consider the use similar methods when analyzing similar data.

## SELECTED REFERENCES

[1] Avanth, C.V. \& Kleinbaum, D.G. (1997). Regression models for ordinal responses: a review of methods and applications. International Journal of Epidemiology, Vol. 26 (6), pp. 1323-1333.
[2] Becker, W.E. \& Walstad, W.B. (1987). Econometric modeling in economic education research. Boston: Kluwer-Nijhoff.
[3] Ellis, L.V., Durden, G.C., \& Gaynor, P.E. (2003). Evidence on factors that influence the probability of a good performance in the principles of economics course. A Journal of Applied Topics in Business and Economics, Spring, pp. 1-11.
[4] Glasure, Y.U. (2002). Does absenteeism matter in course performance? Journal of the Academy of Business Education, Fall, pp. 32-34.
[5] Park, K.H. \& Kerr, P.M. (1990). Determinants of academic performance: a multinomial logit approach. Research in Economic Education, Vol. 21 (Spring), pp. 101-111.

