

Ordinal vs. Multinomial – “Behavioral Risk” prediction modeling

Ordinal vs. Multinomial. Does model selection make a difference?

Corey Leadbeater

National University

04/27/2019

### **Abstract**

Regression modeling is one of the most highly sought-after skills by employers seeking new data scientists.<sup>1</sup> While most data science curricula tend to include regression modeling techniques, nuances between theoretical and practical applications remain difficult concepts for students to grasp. In this paper we explore the impact of model selection on a fictitious, but common real-world scenario. Specifically, how modeling output might be used to determine resource allocation for a behavioral intervention program and how improper model selection might negatively impact program adoption and efficacy. Here we provide a case study for properly evaluating multinomial vs. ordinal model selection as well as the theoretical and practical applications for each.

A cross-sectional secondary analysis using Pearson’s Chi-Square test statistic for independence was conducted on the 2016 Monitoring Futures Study dataset to determine the association between “behavior risk” (v7335), “parental communication” (v7254), “time spent alone after school” (loner) and grades (v7221). Statistically significant associations were identified, and adjusted odds ratios were produced using both multinomial and ordinal regression models. Output from both models were evaluated and compared to demonstrate the utility of ordinal modeling as a “higher-view” utility whereas multinomial models are more appropriate for more specific group-level comparisons.