Abstract

The purpose of this research is to provide insight into student understanding about graphs of quantitative, univariate data. Specifically, students' understanding of the variability of data displayed graphically is investigated. This research also utilizes an ensemble of machine learning algorithms to further investigate this knowledge and expedite the investigation process for large data sets.

A total of nine constructed-response items are disseminated to students through an online homework platform. The responses to these questions are examined to determine the prevalence of particular misconceptions about variability in graphs and to investigate the relationships between these misconceptions. In addition to the nine online homework items, this research includes face-to-face, task-based interviews with 19 students from the same introductory statistics course at the University of Georgia. Students are asked a series of questions that are isomorphic to their online counterparts. The differences in completeness and correctness are analyzed between responses given online and those given during face-to-face interviews.

A rubric is constructed for each of the nine constructed-response items and used to categorize student responses. Each rubric has multiple bins, and student responses may be assigned to one or more of the rubric bins. Multiple statistics PhD students read a small subset of student responses for each item and categorize these responses into the appropriate bin. Next, a series of eight machine learning algorithms is constructed using the previously categorized responses as training data. These algorithms are then tuned to make accurate predictions about the uncategorized responses.

Finally, an ensemble of the eight machine learning algorithms is constructed to combine the votes of each of the algorithms. The results of these

ensemble categorizations show that students struggle to compare the variability between two graphs, even when students have a correct understanding of the statistical definition of variability. Students often assess the variability in a graph by the variation in the heights of the bars or dots. This research provides valuable information about the different ways students view variability in graphs and demonstrates a method in which constructed responses can be categorized in an automated fashion.