Emily Hadley

RTI Application

Re: Exploring Factors Related to Income Status Above $50K in the 1996 Census

This analysis explores data from the 1996 Census to understand what factors might predict whether an individual makes more than $50,000 per year. The investigation began by exploring the variables in the dataset and looking for patterns. No variable had excessive missing observations, but some variables had to be adjusted. One example is capital gains. The graphic below shows the relationship between capital gains and income over $50,000 per year. It is technically a boxplot, but the boxes have disappeared into the ‘0’ line because over 90% of respondents did not report capital gains. Yet when respondents do report gains, there does seem to be difference where, as one might expect, those reporting higher gains tend to make over $50k per year. Individuals were grouped together by capital gain amount and these groupings were included in the prediction model.

The initial prediction model was selected from variables likely to be important to prediction based on prior knowledge and variable exploration (namely age, sex, education level, and marital status). New variables were tested and added to the model if they were significant. The final hand-built model was compared to one built by a statistical algorithm. The hand-built one performed better and included the following variables: age, sex, education level, marital status, race, capital gain, capital loss, and hours worked per week.

The strongest predictor was the capital gains category, where someone reporting capital gains above $10,000 was nearly 35,000% more likely to make $50,000 per year than someone without capital gains. This extraordinary result is because high capital gains are nearly a perfect predictor; in the figure below, 244 people account for the point at $99,9999 and all made over $50,000. Other results for making more than $50,000: males were more likely than females, individuals with postsecondary experience (especially a doctorate or professional degree) were more likely than high school graduates, and each year increase in age increased the likelihood of making more. The results that predicted making less than $50,000: individuals who were separated, divorced, widowed, or never married were far less likely as compared to individuals in civil marriages, individuals who identify as non-white as compared to individuals who identify as white, and those with less than a high school education as compared to those with a high school degree.

