The Effects of Inheritance, Industry, and Age on Billionaire's Net Worth

Jayden Lewis

Introduction and Data

On April 2nd, 2024, Forbes reported that there are 2,781 billionaires in the world. Amongst this small and unique population, there are a variety of net worths, making some billionaires wealthier than others. This project aims to analyze some of the variables that attribute to a billionaire's net worth and draw conclusions on how we may be able to predict who is the biggest bread winner.

To Explore this topic, we will be using a csv file from the "CORGIS Data Set Project," created by Ryan Whitcomb on May 17th, 2016 titled "billionaires." Scholars at Peterson Institute for International Economics have built off the Forbes World's Billionaires list from 1996-2014 and added more variables. The data set has 2614 observations and 22 columnns. Each observation represents 1 of the 2,614 billionares from Forbes's list and the 22 columns include numeric variables such as the billionaire's age, the year that that the company was founded, the billionaire's net worth, and the "Gross Domestic Product" of the country where the billionaire has citizenship. Some categorical variables include company name, the billionaire's relationship to the company,the country that the billionaire is a citizen of, the industry the billionaire is in, and the way that the money was inherited.

The general research question we will be answering is how does inheritance status, company industry, and current age (as of May 17th, 2016) affect a billionaire's net worth? To help us answer this question, we have created three sub questions that are relevant to the variables: 1. Do billionaires who inherited their companies tend to be worth more than billionaires who did not? 2. Do billionaires in certain industries tend to be worth more than others, and within these industries, does the same effect of inheritance remain constant? 3. Does a billionaire's age tend to affect how much they are worth?

Our response variable (numeric) is the number of billion of dollars that the billionaire is worth (3.5 for example), represented in the data as "wealth.worth.in.billions". The first explanatory variable is the inheritance status, originally represented in the data as wealth.how.inherited. Within this variable, the are the categories "not inherited," "spouse/widow," "father," 3rd

generation," "4th generation," and "5th generation or longer." However, because our research questions is specifically looking at weather the billionaire inherited their wealth or not, we have created a new variable called "inheritance" with categories "Inherited" and "Not inherited." The second explanatory variable is the industry that the billionaire is in, originally represented in the data as wealth.how.industry. Within this variable, there are 18 different industries, some of which overlap. We also found that 16 observations had a "0" under this variable, representing that the industry was not reported/was unknown. To help efficiently answer our research question, we created a new variable called "industry" which grouped the industries into "Technology," "Finance," "Consumer/Retail," and "Other." We also renamed "Other" to "Aother" so that when analyzing the industries within our model, it is in reference to Other industries. The final explanatory variable is the current age of the billionaire as of May 17th, 2016 (when the data set was created). This variable is represented as demographics.age in the data set. Because 383 billionaire's ages were not available when the data was collected, and were represented with a "0", we filtered the variable to include all numbers greater than 0. We put these new and filtered variables into a new data set titled "project_variables," which we will use to create all plots, models, and summary statistics. The data set we will be working with now has 2229 observations.