

Capstone Project Proposal

Anthony Stachowitz

July 11, 2018

Project Idea: Census income data set

Central Question

Predict the if someone makes over \$50,000 a year based on different conditions using Logistic Regression.

To help answer this central question, I would investigate the following more specific questions:

1. What is the relation between working class and making over \$50,000 a year?
2. What is the relation between education and making over \$50,000 a year?
3. What is the relation between marital status and making over \$50,000 a year?
4. What is the relation between occupation and making over \$50,000 a year?
5. which variable is the best predictor of making over \$50,000 a year?

Client and Motivation

I am looking at this issue from the perspective of an insurance company looking to write policies based on the appropriate amount coverage given their income.

Data

The data can be found in the GitHub repository at <https://github.com/stackdata/insurance> (<https://github.com/stackdata/insurance>). It's in an easily downloadable xls format.

Outline of Approach

1. Clean up data and fill any missing data with appropriate information and rename columns appropriately.
2. Start to explore data through statistical summaries and plots. look at relation between working class and making over \$50,000 a year. look at the relation between education and making over \$50,000 a year. look at the relation between marital status and making over \$50,000 a year. look at the relation between occupation and making over \$50,000 a year.
3. Identify relationship connections and use statistical tests to quantify the strength of the relationship.

Deliverables

Deliver a report with my findings and a conclusion.