Income and Commute Time Relationship for Minnesota Workers

**Introduction**

The Minnesota Department of Transportation is interested in adding more means of public transport into Minnesota cities. They are considering whether their public transport will impact the income of citizens within the city. MDOT have contacted us to analyze whether they should add new lines to public transport based on income, therefore, we are analyzing and modeling the relationship income has with vehicle accessibility and travel time to work—variables measured in the US census.

After analyzing the model, we recommend that MDOT should not consider income when adding or removing public transport from MN cities as the variables we chose to roughly describe public transport have little to no effect on total personal income.

**The Data**

Chart, histogram

Description automatically generated To reach these conclusions, data from the 2021 United States census in the state of Minnesota was used to analyze public transportation and personal annual income. Of the thousands of variables gathered from US households, we selected total person’s income as our response variable. As seen in Figure 1, the spread is quite right-skewed, so we log-transformed the income variable to control this.

The explanatory variables selected were age, sex, race, vehicles owned per household, average hours worked per week, daily commute time, employment status, class of worker, and Medicaid status. Basic demographic variables were included to account for capturing the diversity of the population with respect to age, sex, and race. For example, we would expect the annual income of a younger individual out of college to be lower than an individual who has been in the work force for several decades. This also goes for discrepancies in the gender wage gap and racial discrimination in wages or opportunity. Average hours worked per week, employment status, and class of worker capture information about an individual’s work status. Vehicles owned per household and daily commute time are the variables of greatest interest that best represent the transportation habits of individuals. If these variables are highly correlated positively with predicted income in the model, then increasing access to transportation should increase individuals’ income and increase quality of life for the population.

Chart, histogram

Description automatically generated In selecting the variables to use in the model, it was apparent that a transformation on commute time was necessary as there was right-skewed behavior similarly to total person’s income. This transformation helps normalize the data, as can be seen in Figure 2. We also observed that the race and Medicaid status variables were not going to be as helpful as there was very little correlation between commute time, race or Medicaid status, and income, as seen in Figures 3 and 4.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

**The Model**

Several candidate models were created using a combination of the aforementioned variables. Using an ANOVA test to minimize residual deviance and minimizing AIC, the best model used number of vehicles, sex, age, average hours worked per week, and the log transform of travel time to work to predict the log transform of total person annual income, as seen in the following equation:

or

Using the coefficient output from our model, we can generalize the trend found to be the best fit. The model predicts a female of average age with a 0-minute commute, 0 vehicles, and works 40 hours a week makes $32,418.01 annually, and a male under the same conditions makes $36,618.23. For commute time, our model predicts a one percent increase in commute time corresponds to an increase of 0.09% in annual income. For vehicles owned, our model predicts a difference in one vehicle owned corresponds to a decrease in annual income by 3.12%. For age, our model predicts a difference of one year corresponds to an increase in annual income by 1.95%. For average hours worked per week, our model predicts a difference of one hour corresponds to an increase in annual income by 3.73%.

From these coefficient interpretations, we can see that having more vehicles, according to the model, lowers the predicted annual income which is counterintuitive. The variables that seem to have the most sway are age and average hours worked per week, and this will be displayed visually in the upcoming analysis.

Scatter chart

Description automatically generated**Analyzing the Model**

Chart, scatter chart

Description automatically generatedStatistically, the model created predicts annual income quite well, however it is being influenced most heavily by the demographic variables due to their strength in the model. When looking at the residual plot in Figure 5, we can generally say that 95% of the data is within two standard deviations of a difference of 0 between predicted annual income and actual annual income. In Figure 6, it can also be observed that the predicted versus actual annual incomes are quite similar as you can nearly fit a straight line between them plotted against each other. The selected model also has a good AIC and the best residual deviance, characteristic of the best model, but these statistical Chart, scatter chart

Description automatically generatedfindings are misleading as the variables leading us to these positive conclusions are not the ones desired in the original proposal.

Chart

Description automatically generatedIn an analysis of the trend lines of each variable included in the model and the predicted annual incomes holding all other variables constant at reasonable values for each sex, it can be seen that the travel variables do not have a very strong, or even positive, impact on the predicted annual income. In Figure 7, we can see that as travel time increases, predicted income increases only slightly, which follows from the coefficient interpretation. Similarly in Figure 8, the straight line slope over number of vehicles indicates that there was not a very large correlation between an increase in vehicles and income. In-fact, the slope was actually negative, meaning as one owns more cars, the predicted income lowers, which is not real-world reasonable.

Chart, scatter chart

Description automatically generated

Figure 9 shows the trend lines for average hours worked per week and age versus predicted income. The steeper slopes indicate a more substantial impact on predicted income than commute time or number of vehicles owned. This indicates that 95% of the residuals being within two standard deviations was most likely due to the high correlation of these demographic variables and not necessarily the travel variables.

**Conclusion**

The Minnesota Department of Transportation is worried that adding or removing public transportation in Minnesota cities may impact citizen’s income. Using data from the 2021 census, we found that this was not the case. Two of our explanatory variables, number of vehicles available and travel time to work, were used as a stand-in for public transport use. We found that these variables have little to no impact on income. Therefore, the department of transportation should not consider income as a reason for increasing the amount of public transportation in MN cities.