

MEMORANDUM

To: Carole Voulgaris, Course Instructor, SES 5215
From: Jane Doe, Student, SES 5215
Date: October 28, 2022
Subject: Assignment 4, Regression with one predictor

The purpose of this memo is to present the results of a set of linear regression models that are relevant to the question:

What is the effect of a person's income on their typical travel time to work, after accounting for the effects of transportation mode and population density?

This analysis draws on census microdata from the 2015-2019 American Community Survey, which I accessed through IPUMS (Ruggles et al. 2022). My dataset includes survey all Survey respondents from California who indicated a usual commute time of more than zero minutes (804,800 people). The variables in my data set are:

- Commute time (outcome): The self-reported typical time (in minutes) it took to get from home to work in the week prior to the survey.
- Commute mode (predictor): The most common mode of transportation used to reach work in the week prior to the survey.
- Income (predictor): The person's pre-tax income in the year prior to the survey.
- Population density (predictor): The number of people per square mile of land area in the public-use microdata area (PUMA) in which the person lives.

Relationship between commute mode and commute time

Table 1 shows the results of the model that predicts commute time based on commute mode. Commute time was log-transformed. The R-squared value for this model was 0.08, suggesting that about 8 percent of the variation on commute time can be explained by differences in commute mode. The coefficients for bike, transit, and walking all had significant coefficients, indicating that commutes by each of these modes had travel times that were significantly different from the travel time for those who commute by car. The "other" mode category had a coefficient with a p-value of 0.49, suggesting that there is not a significant difference between commute times by modes in this category and car commute times. The coefficients for bike and walk are both negative, indicating that commuters by these modes spend less time commuting than do those who commute by car. The coefficient for transit is positive, indicating that transit commuters spend more time commuting than those who commute by car.

Table 1: Results of regression model predicting commute time based in travel mode

Coefficient	Estimated value	p-value
Bike	-0.293	<0.01
Transit	0.687	<0.01
Walk	-0.972	<0.01
Other	-0.005	0.49
Intercept	3.106	<0.01

Relationship between commute time and continuous variables

Table 2 shows the results of two different regression models: one predicting commute times based on income, and the other predicting commute times based on population density. All variables are log-transformed in both models. Each of the two models separately predicts just under two percent of the total variation in commute time, with R-squared values of 0.017 for the income model and 0.016 for the population density model. The coefficient in each model is significant and positive, indicating that, without controlling for other factors, higher incomes and higher population densities are both associated with longer commute times.

Table 2: Results of regression models predicting commute time based on income and population density.

Predictor	Estimated coefficient (in regression with a single predictor)	p-value	Model R ²
Income	0.09	<0.01	0.017
Population density	0.06	<0.01	0.016

References

Steven Ruggles, Sarah Flood, Ronald Goeken, Megan Schouweiler and Matthew Sobek. *IPUMS USA: Version 12.0* [dataset]. Minneapolis, MN: IPUMS, 2022.
<https://doi.org/10.18128/D010.V12.0>