



Modeling Travel Behavior – Madison Fitzpatrick and Aimee Karagiannis

Purpose

This lesson is intended to introduce students the process of computational model construction and interpretation by allowing them to create a model of their own design and estimate its parameters using real data.

Overview

Part 1. The students begin by completing a sample travel survey, so that they get a sense of where the data set came from and what it contains. Then the teacher will give a brief lecture about travel behavior modeling so that the students understand what these models are used for and what kinds of variables they include. The teacher will also review the concept of linear regression and extend it conceptually to the case of 3 or more dimensions. From this, the students should understand which variable is the outcome and which are the inputs, as well as what the parameters are and how they can be interpreted.

Part 2. In the next part of the lesson, the students choose their outcome and input variables and justify these choices. They also make predictions about how the inputs will affect the outcome and why. This is a good point at which to stop, because the teacher will need time to collect the data after the students finish specifying their variables.

Part 3. The teacher must make an Excel spreadsheet for each group that includes data for their chosen variables. Each group should have a computer and access to their data. The teacher will walk the class through some basic tasks in Excel, like viewing and sorting data, then explain how to use the LINEST formula to estimate a multiple linear regression model from their data.

Part 4. Once all the students have estimated their parameters, the next task is to interpret them. The students must decide in what direction each of their input variables actually affects the outcome and offer a plausible explanation for why this might be true.

Student Outcomes

Students should be able to...

Part 1

- Understand where travel behavior data come from and why travel behavior models are important.
- Understand how to interpret the coefficient of an input variable in a multiple linear regression model.

Part 2

- Define the outcome of interest in a travel behavior model.
- Choose variables to include in a travel behavior model, and justify those choices.
- Predict the impact that a variable will have on the outcome of interest.



Part 3

- Use Excel to display and sort data.
- Estimate a multiple linear regression model using Excel.

Part 4

- Interpret and communicate the results of the model estimation.

Time

About 4 hours. This lesson was completed with an Advanced Placement class over the course of two double-periods, with each part taking about one period. Less advanced classes may need more time.

Level

High School AP Environmental Science

Materials and Tools

- [Travel Survey](#)
- [Lecture slides](#) (optional)
- [Travel Behavior Modeling worksheets](#)
- Computers (1 for every 4 students) with Excel installed
- Travel behavior data

Prerequisites

An understanding of the equation for a line ($y = mx + b$) is essential. It is helpful if students have seen linear regression before.

Background

In order to succeed in this lesson, students should be able to:

- Understand a data point written in the form $(Y, X_1, X_2, X_3, X_4\dots)$.
- Look at a linear equation ($y = mx + b$) and explain how y would change (increase, decrease, or stay the same) if x were increased.

Teaching Notes

Part 1: Travel Survey and Lecture

The purpose of the travel survey is to help the students understand where travel data come from and what are some of the issues that may cause the data to be imperfect. For the travel diary part, ask students to recall the previous weekday (if it's Monday, they should think about the previous Friday; otherwise, think about yesterday). Students are likely to have questions about exactly how they should write things down, and they are likely to moan and groan about how hard it is to remember and how annoying it is to have to write every single trip.

When they are done, ask them, "How many of you thought that survey was a pain to fill out? How many of you needed me to clarify something for you? If you received this survey in the mail at home, with a letter asking you to fill it out for research purposes, would you do it?" The answers to these questions serve to illustrate some of the problems that researchers experience with this kind of data. Many people fail to respond to the survey at all, so the final sample is usually not a perfect reflection of the whole population. In addition, many people find the survey confusing, and so they may enter responses that are unclear or incorrect. This makes the data imperfect, and researchers have to figure out how to deal with that.



The lecture should serve to answer the following questions.

What are travel behavior models?

Travel behavior models describe the relationship between an individual person's characteristics and their travel choices. First, think about one person. This person has an age, gender, ethnicity, preferences/attitudes, and lots of other things that describe them personally. He or she also lives in a household, which has characteristics like household income, the number of people in the household, the number of cars in the household. These can be considered to apply to our person, to be included in their characteristics. The person also has a neighborhood, which has characteristics like the density of roadways, the number of transit stops, and the population density. These neighborhood characteristics can also apply to our person, because he or she lives in the neighborhood. Our person also makes choice about how to travel, and we can describe these with quantitative measures, like the distance they travel in a day. Travel behavior models take the persons characteristics (the input variables) and relate them to their travel choices (the outcome variables). If we have a lot of people, all with these same input and outcome variables, we can estimate one model for the whole group of people.

Why do we use them?

Models can tell us two kinds of interesting things. The first thing, which we will focus on in this lesson, is how different input variables affect the outcome. Let's say the outcome is total time spent traveling in one day. Some of the questions that a model could answer would be: do women tend to travel more than men? Does being older make you likely to travel more or less than younger people? Do people who live in the city spend more or less time traveling than people in the suburbs? There are many other similar questions we could investigate with a travel behavior model.

The second thing models can do is answer hypothetical questions. Let's say we created a travel behavior model for a city. We did this using real survey data, but now we want to know about the future. How will our outcome be different in 30 years, when the population has grown and changed? What if we build a new train system? What if gas prices were twice as high as they are now? We can make up hypothetical data, feed it into our model, and answer these kinds of questions.

What kinds of variables do they include?

There are three main categories of variables that we will focus on using for our models. These are socioeconomic/demographic variables, local geographic variables, and personal attitudes/preferences. (See the "More Information" section for partial lists of these types of variables.)

Part 2: Outcome and Input Variable Selection

Students should get into groups of 4 to create their models. The first step is to choose an outcome of interest. The worksheet will ask them to justify their choice and explain what they want to learn about their outcome. Once the teacher has approved their choice, they can select their input variables. They should choose variables that they think would strongly affect their outcome. On the worksheet, they will justify why they chose each variable, predict in which direction it will affect the outcome, and explain why they think so. The students will need a lot of hands-on guidance throughout this process. It is helpful to require the students to check in with the teacher at certain points so that the teacher can ensure they are on the right track before they move on.

Once students have chosen their variables, the teacher will have to extract those variables from the raw data set to create one spreadsheet for each group. The spreadsheet should contain only that group's chosen variables and no others, to minimize confusion during model estimation.

Part 3: Model Estimation

Once the students have their data spreadsheets open in Excel, it may be helpful to show them some easy ways to view and sort data in Excel, if they are not very familiar with the program. Tips I like to share include freeze panes, which keeps the first row visible while you scroll down, and filters, which allow the user to easily sort and filter data. These tips are nice but not essential for the core of the lesson. The students will use the LINEST function to estimate their models. See the Excel help files for a full explanation of how to use and interpret the results of this function. The teacher will have to walk them through the process slowly, step by step. Even then, many groups may get confused or make a mistake. It is important to check that every group has done it right before moving on. As a final step, the students should copy their estimated parameters onto their modeling worksheet in equation form, with the names of their variables written out in place of the X and Y symbols, and their numeric parameters in place of the M and B symbols.

Part 4: Model Interpretation

For each variable, the students will explain how it affects the outcome variable based on the estimated coefficient. They will state whether the result matched their prediction and offer a plausible explanation for why the result might be true. Once again, the students will need a lot of hands-on guidance with this. If students misinterpret a result or offer an explanation that does not make sense, challenge their thinking and ask them to correct their answers before turning in their final report.

Assessment

The students will be assessed informally for their understanding of concepts introduced in part 1. Outcomes for parts 2-4 will be assessed based on their written justifications and explanations of their variable choices as well as their model results and written interpretations.

Additional Information

There are some important restrictions on the types of variables that students may choose to include in their models. Only variables that can be measured on a continuous scale (or a near enough approximation of continuous) can be used. Categorical variables, such as the mode of transportation used for commuting (car, bus, train, walk, bike, etc.) cannot be used because they cannot be measured on a meaningful numeric scale. Binary variables (e.g. gender, where male = 0 and female = 1) can be used because if a coefficient were estimated for this variable, one would be able to say, for instance, that being female makes the outcome more or less likely. Also, variables that are measured on a numeric scale that is not continuous (e.g. an importance rating from 1-7; or employment status, where 0 = unemployed, 1 = employed part time, 2 = employed full time) are also usable, because a higher number represents more of something (more important, or more time spent working). These restrictions hold true for both outcome and input variables.

Some students may be interested in outcomes that only involve a portion of the general sample. For instance, if students choose their outcome to be distance to work, or time spent traveling to school, the



only survey participants of interest are those who have jobs or go to school. In this case, it is important to restrict the data to these sub-samples. Otherwise, their data would include a lot of people for whom the outcome is zero because they don't go to work or school at all. Including these people would bias the results of the model. If you do not want to deal with the hassle of creating these sub-samples, you can limit the options that students have for their outcome variable.

The Excel function that estimates a multiple linear regression model is LINEST. This is an array function, so it is important to enter the function properly so that all the results are visible. Check the Excel help files for specific instructions on how to use the function correctly. It is helpful to provide students with the following suggestions for their variables.

OUTCOMES:

- Time spent traveling OR distance traveled in a day
 - Total
 - For a particular purpose (work, school, shopping, etc.)
 - On a particular mode (car, transit, walk/bike, etc.)

INPUTS:

- Socioeconomic and demographic variables
 - Age
 - Gender
 - Race/ethnicity
 - Income
 - Household size
 - Number of children
 - Possession of driver's license
 - Possession of transit pass
 - Number of vehicles per person (in the household)
- Local geographic variables
 - Residential population density
 - Job density
 - Transit stop density
 - Roadway density
 - Neighborhood median income
 - Neighborhood median home price
 - Neighborhood racial/ethnic composition
- Attitude variables (questions from the survey → rate the importance of...)
 - Ability to arrive on time
 - Ability to travel without changing vehicles
 - Not having to deal with traffic congestion
 - Short travel time
 - Day-to-day costs like gas, parking, and bus fare/pass
 - Protection from weather
 - Having a seat
 - Short wait time
 - Ability to read while traveling
 - Ability to travel when desired
 - Flexibility to change plans
 - Making few stops

- Parking availability
- Safety of vehicle from accidents
- Freedom from threats to personal safety
- Avoiding stress
- Minimizing pollution