Understanding Consumer Decisions on Alternative Fuel Vehicles Using Discrete Choice Models

Sayeeda Tuz Jahin Aishee, 2018 CTECH REU Intern

Engineering School, Cornell University

REU Program: 2018 Center for Transportation Environment and Community Health Research Experience for Undergraduates (CTECH REU) Program

CTECH REU Principal Investigator: Prof. Linda Nozick, Civil and Environmental Engineering, Cornell University

CTECH REU Mentor: Dr. Esther Chiew, Civil and Environmental Engineering, Cornell University

Primary Sources of CTECH REU Funding: Center for Transportation Engineering and Community Health, Cornell University, and the National Science Foundation

Contact: sta48@cornell.edu, lkn3@cornell.edu, wc437@cornell.edu

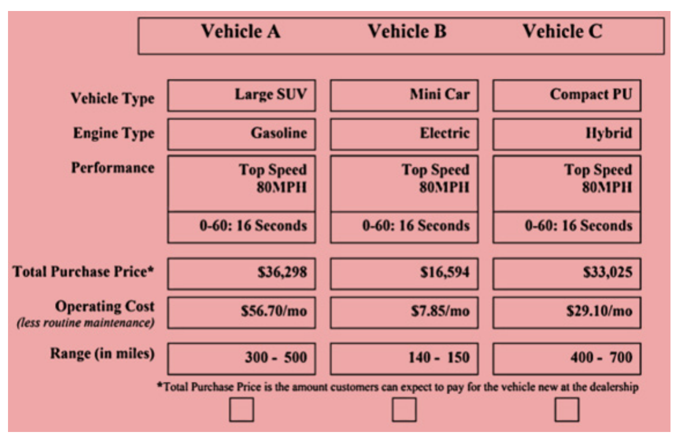
Website: http://www.cnf.cornell.edu/cnf_2018reu.html

Abstract:

In a rapidly transforming technological world, air pollution levels are highly dependent on people's decision to buy less polluting cars. In this study, we used Discrete Choice Modeling to analyze data of vehicle choice to recommend the most attractive attributes of alternative fuel vehicles to businesses. With this information, businesses could cater to customers' preferences, resulting in higher sales of alternative fuel vehicles and, ultimately, reducing toxic air pollution. Using R, a computer modeling software, we found that consumers highly value cost efficiency, high-performance, and having a variety of fueling options. These attributes could increase the market share of alternative fuel vehicles so that six out of seven people would buy an alternative fuel vehicle.

Summary of Research:

In 2016, 6.1 million deaths were caused by air pollution related health risks [1]. Additionally, 28% of US global warming emissions are caused by the transportation sector; that's more than any other sector in the nation. We wanted to combat this by having more consumers opt for alternative fuel vehicles instead of gas vehicles. To figure out what people want the most, we analyzed a data set from a survey of around 1000 individuals from California, shown in Figure 1. Individuals were required to choose a vehicle from different alternatives while considering their attributes. We used discrete choice modeling as the data included continuous variables such as purchase price, operational cost, and range along with dummy variables such as engine type, vehicle type, and performance. Discrete models specifically analyze choices that customers make between products or services. They produce utility, or the willingness to buy a product, to predict the probability of choosing an alternative. Utility is usually directly related to probability, so the higher the utility, the higher the probability.



	Vehicle A	Vehicle B	Vehicle C
Vehicle Type	Large SUV	Mini Car	Compact PU
Engine Type	Gasoline	Electric	Hybrid
Performance	Top Speed 80MPH	Top Speed 80MPH	Top Speed 80MPH
	0-60: 16 Seconds	0-60: 16 Seconds	0-60: 16 Seconds
Total Purchase Price*	$36,298	$16,594	$33,025
Operating Cost (less routine maintenance)	$56.70/mo	$7.85/mo	$29.10/mo
Range (in miles)	300 - 500	140 - 150	400 - 700

*Total Purchase Price is the amount customers can expect to pay for the vehicle new at the dealership

Figure 1: One question from the survey. Consumers had to choose one vehicle while considering the attributes.

We used Multinomial Logit Models from the realm of Discrete Choice Models to portray our data set. The key inaccuracy of this model comes from only being able to detect systematic variation and suggesting proportional substitution. Systematic variation doesn't account for inherent biases of individuals or any random choices they make. Proportional substitution implies that if the probability for choosing one alternative increases the probability for choosing the other alternatives decrease proportionally. We derived a function that includes all the alternative specific attributes and produces generic coefficients. We tested that this was the most accurate function using Log-Likelihood Ratio Test. Since this function produced the highest log likelihood ratio compared to the other functions, which excluded some attributes, it was the most representative of the data set.

Results and Conclusions:

Using the function (Figure 2), we derived coefficients for all the attributes shown in Figure 3. The negative coefficients of the continuous variables signify that as the actual value of the attributes increase, people would be less inclined to choose an alternative with the increased amount of attribute and vice versa. For example, since the coefficient for purchase price (pprice) is negative, it implies that as pprice increases the willingness of a customer to choose an alternative with an increased pprice decreases. It's the opposite with range, since the coefficient is positive. People are more likely to choose an alternative with an increased range. When it comes to the discrete variables, if a coefficient for an attribute is negative people are less likely to choose an alternative with that attribute. According to Figure 3, people are less willing to buy a vehicle that is electric (negative coefficient) and more likely to buy a vehicle that is hybrid (positive coefficient). However, overall $1/5^{th}$ of the market was willing to buy electric vehicles and that market share increased when a vehicle was a small or a compact vehicle. When the pprice of a vehicle was lower than the median pprice of the market, the operational cost (ocost) of a vehicle was lower than the median ocost of the market, and the range of the vehicle was greater than the median range of the market, 84.41% of the market chose to buy hybrid or electric vehicles. Consumers would also pay $16,000 more for a hybrid vehicle instead of a gas vehicle, $83 more for one extra mile in an electric vehicle, and $4,000 more for a high-performance vehicle over a mid-performance vehicle.

```
mlogit(formula = choice ~ pprice + ocost + range + electric +
    hybrid + hperf + lperf + mini + small + large + sSUV + mSUV +
    lSUV + compPU + fullPU + minivan + 0, data = data1, method = "nr",
    print.level = 0)
```

Figure 2: The R command that lists the function used to analyze the data set.

```
Frequencies of alternatives:
      1       2       3
0.33172 0.35337 0.31491

nr method
5 iterations, 0h:0m:0s
g'(-H)^-1g = 0.000247
successive function values within tolerance limits

Coefficients :
          Estimate Std. Error  z-value  Pr(>|z|)
pprice   -0.0532178  0.0017621 -30.2019 < 2.2e-16 ***
ocost    -0.0265221  0.0017894 -14.8219 < 2.2e-16 ***
range     0.4458927  0.0769181   5.7970 6.752e-09 ***
electric -1.0975419  0.0768052 -14.2899 < 2.2e-16 ***
hybrid    0.7546017  0.0428463  17.6118 < 2.2e-16 ***
hperf     0.1884808  0.0217873   8.6510 < 2.2e-16 ***
lperf    -0.2556816  0.0222666 -11.4827 < 2.2e-16 ***
mini     -1.3091796  0.0796035 -16.4462 < 2.2e-16 ***
small    -0.3351288  0.0687550  -4.8742 1.092e-06 ***
large     0.1452564  0.0690467   2.1037 0.0354010 *
sSUV     -0.0055082  0.0660140  -0.0834 0.9335016
mSUV      0.6510700  0.0713574   9.1241 < 2.2e-16 ***
lSUV      0.4632372  0.0834508   5.5510 2.840e-08 ***
compPU   -0.3520306  0.0687153  -5.1230 3.007e-07 ***
fullPU    0.0283859  0.0718255   0.3952 0.6926905
minivan   0.2633300  0.0708870   3.7148 0.0002034 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -6939.9
```

Figure 3: The summary of the data set with coefficients and significance levels.

Future Work:

Analyzing updated data based on real market consumer decisions from around the nation would provide a more accurate prediction of the market shares.

Acknowledgements:

I would like to thank my professor Dr. Linda Nozick, my mentor Dr. Esther Chiew, the program coordinators Celia Szczepura-McLean, Joe Rowe, and Melanie-Claire Mallison. This project was funded with the help of Center for Transportation Engineering and Community Health, Cornell University, and the National Science Foundation.

References:

[1] "Air Pollution Contributed To More Than 6 Million Deaths In 2016 [Infographic]." McCarthy, Niall. Forbes, Forbes Magazine, 18 Apr. 2018, www.forbes.com/sites/niallmccarthy/2018/04/18/air-pollution-contributed-to-more-than-6-million-deaths-in-2016-infographic/#4a28acaf13b4.

[2] "Sources of Greenhouse Gas Emissions." EPA, Environmental Protection Agency, 11 Apr. 2018, www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions.