

SDS 328M Project

Zach Ford

May 12, 2018

MW: 11-12:15; 6:00-7:00



TEXAS

The University of Texas at Austin



Introduction

Objectives: Does the amount of miles someone in Austin, TX or surrounding area, drive per week have any affect given what political party they affiliate with and number of children they have?

- The variables that I am investigating are the number of children someone has and what political party that person affiliates with.
- The reason that I chose this topic was because these two explanatory variables of political party and number of children, at first glance do not appear to have any affect on the response variable of the amount of miles driven, but I was curious to see if the explanatory variables have any impact on the response variable.

Research Hypotheses: For political party affiliation, I do not expect any kind of effect with the amount of miles someone drives. For number of children someone has, I expect the more children that they have they will drive more miles. For an interaction between the two variables, I do not expect this to be a significant interaction effect.

Methods

Sample: My sample data includes 60 subjects who live in Austin, TX or surrounding area and who have a car.

- The subjects that are included in this sample are people who live in Austin, TX or surrounding area and who have a car.
- The amount of samples that are used in my data are 60 samples and I did not have to remove any outliers.
- For political party affiliation the categories used were Republican Party, Democratic Party, and other. For number of children someone has the categories used are 0, 1, 2+. For the amount of miles someone drove I used miles and this variable was the response variable therefore it was fill in the blank.

Analysis Method: All of my graphs and tables are going to be made with R-Studio version 3.4 and Apple numbers, my data was collected using Google forms.

- The way that I collected my data was using a Google form questionnaire. Once all the data was collected I transferred the data into Apple numbers. All the graphs came from R-Studio with version 3.4.

Descriptives

Response Variable:

	Median	Five Number Summary
Amount of Miles Driven	100	0, 50, 100, 300, 600

Explanatory Variables:

Political Party:

Republican Party	Democratic Party	Other
47.62%	20.95%	31.43%

Number of Children:

0	1	2+
8.57%	10.48%	80.95%

Results

Results Table:

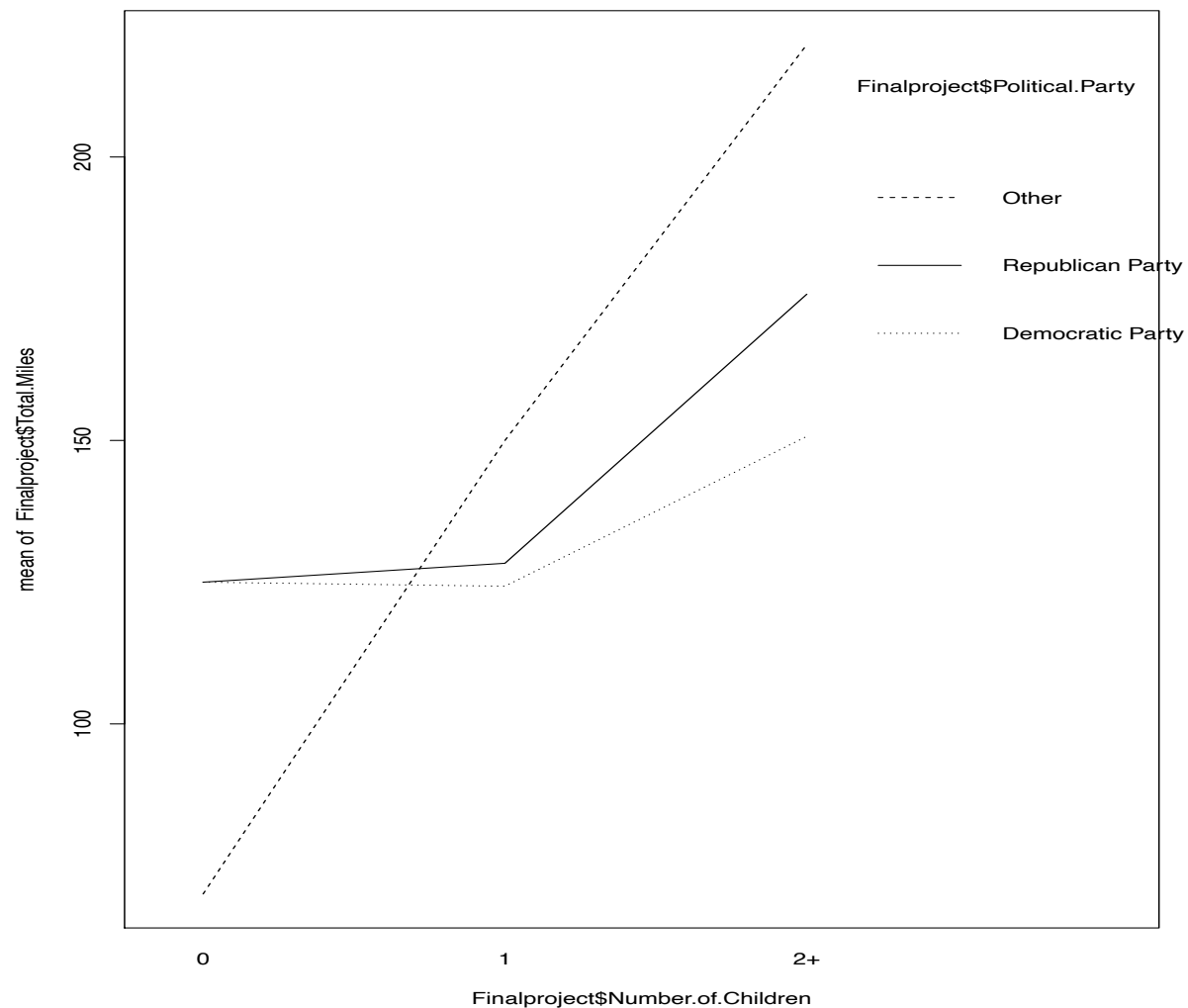
Variable	Estimate/SS	Test Statistic	P-value
(Intercept)	20753	0.9596	0.3297
Political Party	36628	0.8468	0.4318
Number of Children	89520	2.0696	0.1316

Intercept:

Variable	Estimate/SS	Test Statistic	P-Value
(Intercept)	15625	0.7012	0.4045
Political Party	6722	0.1508	0.8602
Number of Children	3532	0.0793	0.9239
Political Party*Number of Children	23549	0.2642	0.9003

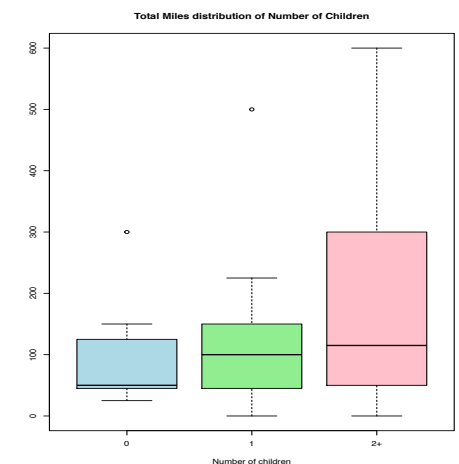
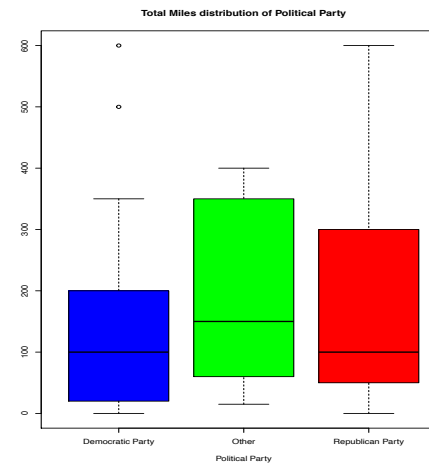
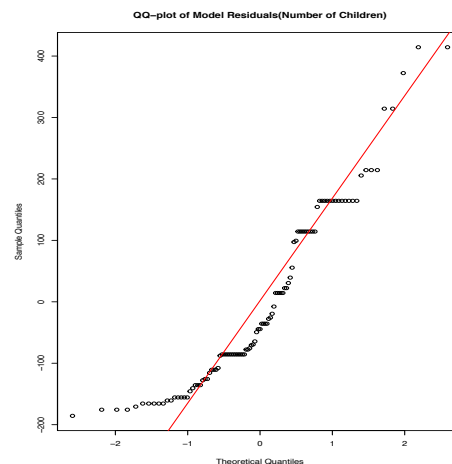
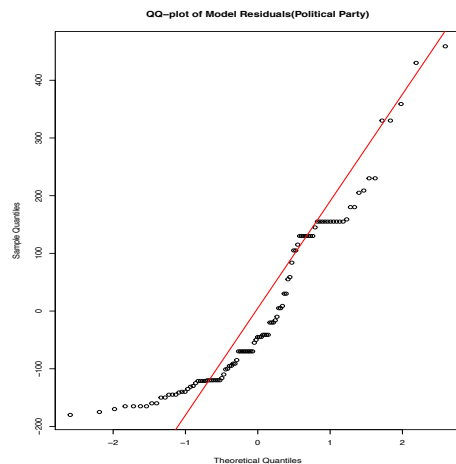
Interaction Graph:

Since all of the slopes are different that means that each of the political party's have an interaction with each other.



Assumptions

Assumptions: There are four assumptions that I had to check before I ran the two-way ANOVA. The first two of random sampling and independent observations are confirmed based off of the experimental design. Each group being normally distributed is the next assumption and it can be checked by looking at a histogram and QQ-plot, or can be checked by running a formal test of the Shapiro-Wilk test, which both of the explanatory variables meet with the graphs below. The next assumption of equal variance among all groups with a visual check of a box plot and each box with equal spread. The formal test to check for equal variance is Levene's test and want the overall p-value to be greater than .05 to meet the assumptions, which political party's p-value is $p > .05$ and number of children's p-value is $p > .05$ as well as the graph shown below.



Discussion

Interpretation: When controlling for number of children, the amount of miles someone drove in Austin, TX did not differ based on which political party someone affiliated with (F -value=0.9596, Df =1/100, P -value=0.3297). When controlling for political party, the amount of miles someone drove Austin, TX did not differ based on the number of children someone had (F -value=0.8468, Df =2/100, P -value=0.4318). Political party affiliation and number of children someone had explain 1.8% of the amount of miles someone in Austin, TX drive. For the interaction, there was not a significant interaction between political party and the number of children had (F -value=0.2642, Df =4/96, P -value=0.9003). This data is not what I expected, I expected the number of children to have a significant difference in means when controlling for political party but that is not the case. I did not have any expectations for political party when controlling for number of children therefore the results were normal.

Limitations: Some of the limitations that I had with my study was that I was not able to send it out and get mass data due to the time table that was on the assignment. Another limitation that I had with my study was that since my response variable was more of an estimate than a plain answer there could be some error associated with some of the responses.

- There was selection bias in my study since everyone who had access to the study needed the internet therefore it was not completely random. Another type of bias is the that most of the people that participated in the study are from the same area since this is where most of my friends/family members live.
- The only assumption that could have been violated is the random sampling assumption since most of the subjects in my study are from the same area and they are not spread out over the Austin, TX area.
- One confounding variable in my study was that the questionnaire was an online survey which could have some bias to it because the only people that could take the questionnaire are people who have internet access.

Implications: One thing that I would change with my research would be to make the response variable categorical answer because I think if it more people instead of guessing how many miles they drove would nonetheless have a range of values they could choose from. I think this would make my results more confined or closer together because the subjects participating in the study would have a broad idea of how many miles they drive instead of trying to guess or estimate. The method that was used any way was sufficient because I was still able to get valuable data from the people who participated in the survey. This data may have some error or bias but overall the data that was collected still has some quality.

References: None