Emre Uzel
25 April 2021

INTL 550 Hw 2 Report

I explained most of my code with comments, hence I won't go into details of code here. I will just explain briefly what is my project about. My project idea is to calculate whether there is relationship development level of people in Istanbul. As going from center of the city to far places, we would expect development level decrease (Houses in centers are expensive therefore people with low income lives far away from city centers). From these ideas I decided to calculate it with family size. My assumption here; I say as family size increase, the level of development increase for the sake of project.

Hence, I pulled data from data.ibb.gov.tr which is a open source data platform managed by Istanbul Municipality. Distance variable is our variable of interest and gas consumption, waste size is our controlling variables and size of family is our target variable in this case.

To measure the distance between the districts, I used google maps api and use distance matrix api to calculate distance between districts. I decided Kadikoy to be the center of Asian part and Besiktas for the European part. I also give 1500 constant variables for Kadikoy and Besiktas. I calculate them and group by in one data. I named it as last_data.

I created 'linear_analysis' file and 'the_function' function to do multi-linear regression analyses. The resulting hypothesis (I am adding screenshot because I believe this explanation is very clear)and regression table as follows;

Hypothesis

We are curious about whether the distances of districts have significant effect on predicting Average Family Size

- H0 Hypothesis: Distances doesn't have effect determining Average Family Size
- H1 Hypothesis: Distances have effect determining Average Family Size.

To achieve this the abs(t0) value must be out of confidence intervals of t distribution for specified alpha = 0.05 and n-p-1.

Computationaly:

```
-IF abs(t0) > confidence_interval_value1 and  abs(t0) > confidence_interval_value2; then it has high effect rej
ect null hypothesis

-Otherwise; Reject Null Hypothesis
```

We can change District's column name and calculate for other variables as well.

I calculated t0 = -0.8. It is between the intervals. -2.04 < -0.8 <2.04. That's why we can't reject the null hypothesis. There is not enough evidence that distance from center have high effect on family size.

Regression table as follows:

	B coefficents	Standard Errors of Coefficents	Lower Bound of Credible Interval B	Upper Bound of Credible Interval B
B0	3.467671e+00	2.748991e-01	2.907011e+00	4.028331e+00
B1	-3.451880e-05	4.291097e-05	-1.220363e-04	5.299870e-05
B2	-1.030685e-08	4.907569e-09	-2.031591e-08	-2.978017e-10
B3	1.839594e-09	1.318947e-09	-8.504161e-10	4.529604e-09

As the 3 coefficents of variables are very close to 0. We can't say that they have a linear relationship with family size. The highest coefficient belongs to B1 and we found above that B1, there is not enough evidence to show that variable 1 has high effect on family size.