CIS 9660 Project

Group 10

Quanming Feng, Yue Wang, Sky Song

Interaction terms in linear regression

Analysis on

2-3

Deep thoughts and answer question

Test causality

Discuss two issues for causality and spend time on this.

We choose Topic 2: Restaurants and Rodents as our project.

We will be using the two data set given and will also looking for extra data to support our findings if necessary.

In this project we will be prepared to answer the question if more restaurants will lead to more rodent inspections.

The question is to find whether there is a relationship between the number of restaurants and the number of rodent inspections. So, we should quantify these two targets. The first target can be calculated by number of restaurants in a region (area). As the dataset Open Restaurant Application contains a variable called NTA which indicate the region the restaurant belongs to. There is a total of 189 region provided by NTA. We show the top regions that have most restaurant in Fig. 1. The second target is to quantify the region have more rodent inspection or not. According to the dataset Rodent Inspection, we have the information of every rodent inspection in NYC. And we can link two datasets by the variable BBL which is contained in both datasets. Fig. 2 shows the rodent inspections after June 19, 2020 which is the earliest submission in Open Restaurant Application. Hence, for the dataset Rodent Inspection, the potential variables are BBL, Result, Inspection date. For the dataset Open Restaurant Application, the potential variables are BBL, NTA and time of submission.

The most challenging thing is that we should prevent the occurrence of omitted variable bias. Therefore, we should dimmish the effect of garbage, the tendency of rodent inspection in specific area. In the project, we should explore more dataset about these two indicators to show the relations between number of restaurants and the number of rats in the area.

The way for us to solve this challenge is to try to consider every possible dependent variable carefully and connect our findings to our main question.

![图表

描述已自动生成]()

Fig. 1, The number of re-opened restaurant in each region in NYC

地图

描述已自动生成

Fig. 2, The rodent inspection in NYC where color indicate different result

Diagram

Description automatically generated