Data Mining and Statistical Learning: Exercise 3

Milo Opdahl

April 7, 2019

Question 2: What Causes What?

1.

Although there may be a correlation between "Crime" and "Police", there is not a guarantee that there is a direct causation between the two variables on their own. In other words, there can be other factors that can influence "Police" levels within a city that will subsequently affect the level of "Crime."

2.

Researchers wanted to isolate the effect of "Police" on "Crime", so they took data from Washington, D.C. to observe what other factors can create higher levels of "Police" that are unrelated to "Crime." They found that when the Terrorist Alert System was on high-alert, the number of cops present in D.C. goes up as well; and since the level of "Police" goes up, the level of "Crime" goes down in D.C. Therefore, a direct causal relationship was found between "Police" and "Crime" by introducing a "High-Alert" variable. In the table below, we see the effect that "High-Alert" has on "Crime." In column (1), the recorded model shows the interaction of "High-Alert" by itself on "Crime." The interpretation of column (1) is: Holding all else constant, the presence of a High-Alert Day in Washington, D.C. results in a 7.316 point decrease in the Total Daily Crime level; this regression is statistically significant at the 5 percent level, and "High-Alert" is able to predict 0.14 of the changes in "Crime." In column (2), the researchers decided to control for METRO ridership, the public transportation of Washington, D.C. Column (2) interpretation, from top to bottom: Holding all else constant, the presence of "High-Alert" and an one percentage point increase in "Ridership", there will be a 6.046 point decrease in the level of "Crime", which is statistically significant at the 5 percent level; for the latter control variable, a one percent increase in "Ridership" will lead to a 17.341 increase in the Total Daily Crime level, which is statistically significant at the 1 percent level, holding all else constant; finally, the combined effect of "High-Alert" and "Ridership" is able to predict 0.17 of the changes in "Crime." The results from column (2) show that controlling for both "High-Alert" and "Ridership" will still result in a decreased level of "Crime."

EFFECT OF POLICE ON CRIME

TABLE 2

TOTAL DAILY CRIME DECREASES ON HIGH-ALERT DAYS

	(1)	(2)
High Alert	−7.316*	−6.046*
	(2.877)	(2.537)
Log(midday ridership)		17.341**
		(5.309)
R^2	.14	.17

Figure 1: The dependent variable is the daily total number of crimes in D.C. This table present the estimated coefficients and their standard errors in parenthesis. The first column refers to a model where the only variable used in the High Alert dummy whereas the model in column (2) controls form the METRO ridership. * refers to a significant coeficient at the 5% level, ** at the 1% level.

3.

The researchers controlled for METRO ridership as an indicator of how much pedestrian and tourist traffic there is on High-Alert Days. This was done to observe if the level of "Crime" would still increase when controlling for both "Ridership" and "High-Alert." In other words, the researchers wanted to know if crime still took place in D.C. on High-Alert Days.

4.

The model being estimated in the first data column of Table 4 is the effect on "Crime" by "High-Alert" and some dummy variables that describe the certain police district areas of Washington, D.C. The "District 1" dummy variable represents the area of D.C. around The National Mall, while "Other Districts" controls for all other districts in D.C. Again, "Ridership" is used again to represent pedestrian and tourist traffic, while "Constant" simply accounts for all the other variables in the data when the previous variables are fixed at 0. The purpose of the estimation is to observe how much "High-Alert" affects different "Crime" levels in two separate areas of D.C. Based on the results, it can be concluded that "High-Alert" affects one district of D.C. more so than the rest of the city, and that the effect of "High-Alert" on "Crime" is not as strong as we previously observed in Table 2. Specifically: Holding all else constant, the presence of a High-Alert Day in Washington, D.C. results in a 2.621 point decrease in the Total Daily Crime level in "District 1"; in addition, the presence of a High-Alert Day in Washington, D.C. leads to a 0.571 point decrease in the Total Daily Crime level throughout the rest of the city districts.

TABLE 4

REDUCTION IN CRIME ON HIGH-ALERT DAYS: CONCENTRATION ON THE NATIONAL MALL

	Coefficient (Robust)	Coefficient (HAC)	Coefficient (Clustered by Alert Status and Week)
High Alert × District 1	−2.621**	−2.621*	−2.621*
	(.044)	(1.19)	(1.225)
High Alert × Other Districts	−.571	−.571	−.571
	(.455)	(.366)	(.364)
Log(midday ridership)	2.477*	2.477**	2.477**
	(.364)	(.522)	(.527)
Constant	−11.058**	−11.058	−11.058^{+}
	(4.211)	(5.87)	(5.923)

Figure 2: The dependent variable is the daily total number of crimes in D.C. District 1 refers to a dummy variable associated with crime incidents in the first police district area. This table present the estimated coefficients and their standard errors in parenthesis.* refers to a significant coeficient at the 5% level, ** at the 1% level.

Question 1