To: Neighborhood Health

Quality and Evaluation Intern

Public Health Solutions

From: Brian Holliday

Applied Mathematics

New York City College of Technology

Subject: New York State COVID – 19 infection rates

Figure:

Date: 5/7/20

Summary:

In this mini project we are looked at the rate of infection per 100,000 among 6 different areas across the state of New York for COVID - 19. The data consists of the county infection rate in these areas. These areas are the New York City metro area, the Buffalo Metro Area, the Syracuse metro area and the Rochester metro area. We broke up the New York City metro area into three different areas and we did not include areas that are in the New York City metro area that are not a part of New York State. The areas included are Long Island, which includes Nassau and Suffolk counties, the five counties of New York City, as well as five counties in Hudson Valley, including Westchester and Rockland counties. We know that New York City has been one of the hardest hit areas in the world by COVID – 19. How has COVID – 19 spread throughout the rest of the state? We want to see whether there is a difference in the mean infection rate per 100,000 between these different areas. We ran a One – Way ANOVA test on the infection rate to answer this question. We used Excel to put this data into a useable format and SAS to run the test.

The results of the ANOVA test show that the spread of COVID-19 is most prevalent in the New York City area, extending throughout the Hudson Valley and Long Island. The mean spread in the Buffalo, Syracuse and Rochester areas are statistically the same. The spread through these areas in Upstate New York is much lower than the New York City area.

Figure 1: Metro Area Means

The MEANS Procedure

metro=buff

Analysis Variable : rate				
N	Mean	Std Dev	Minimum	Maximum
3	255.6533333	182.8112607	70.2800000	435.7900000

metro=cuse

Analysis Variable : rate				
N	Mean	Std Dev	Minimum	Maximum
3	204.7766667	135.0545821	59.3700000	326.2900000

metro=hud

Analysis Variable : rate				
N	Mean	Std Dev	Minimum	Maximum
6	2036.15	1245.79	774.3600000	3747.06

metro=isld

Analysis Variable : rate				
N	Mean	Std Dev	Minimum	Maximum
2	2574.73	247.4095917	2399.78	2749.67

metro=nyc

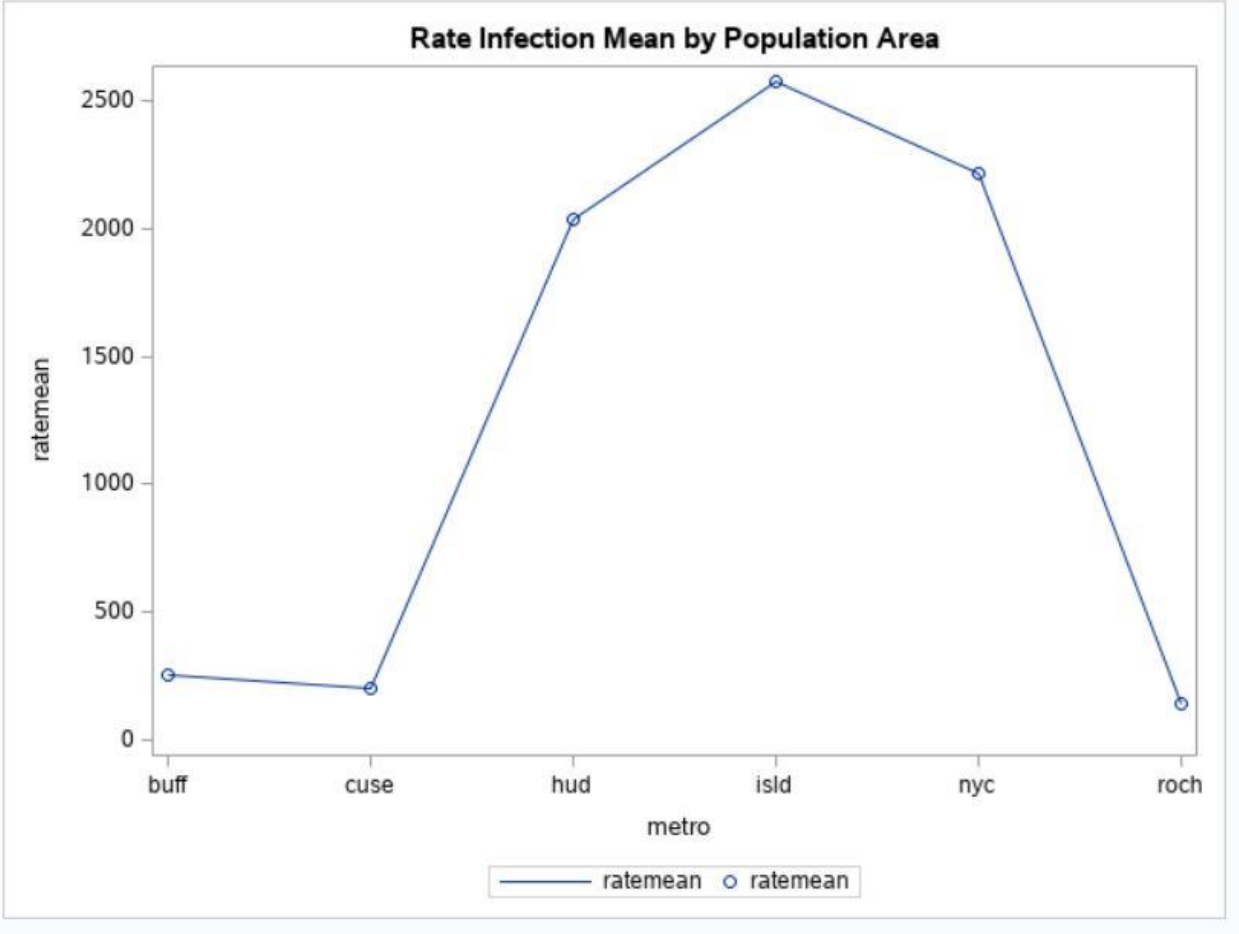
Analysis Variable : rate				
N	Mean	Std Dev	Minimum	Maximum
5	2215.92	560.6725716	1427.95	2781.80

metro=roch

Analysis Variable : rate				
N	Mean	Std Dev	Minimum	Maximum
6	145.0816667	76.1584878	76.4900000	253.6200000

The means procedure shows that the means between Rochester, Buffalo, and Syracuse look similar and the means throughout New York City areas look similar.

Figure 2: Rate Infection Mean Plot



As we can see from the graph, the rate of infection per 100,000 in the areas closest to New York City have a much higher infection rate than Buffalo, Syracuse and Rochester. We will still need to run the ANOVA procedure to see if there is a statistical difference.

Figure 3: GLM COVID-19

The GLM Procedure

Dependent Variable: rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	25439496.63	5087899.33	10.50	<.0001
Error	19	9210883.92	484783.36		
Corrected Total	24	34650380.55			

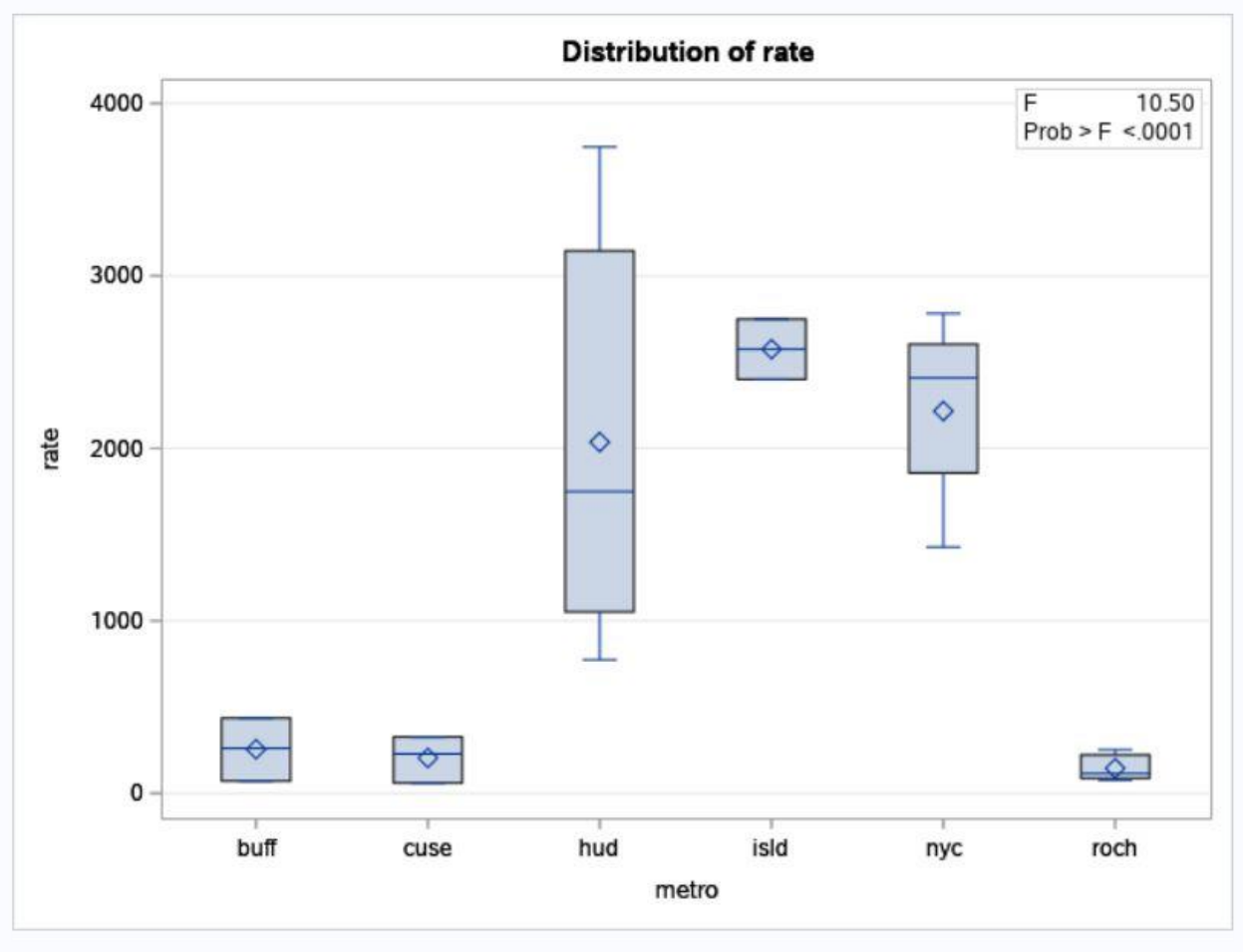
R-Square	Coeff Var	Root MSE	rate Mean
0.734177	56.70322	696.2639	1227.909

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metro	5	25439496.63	5087899.33	10.50	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metro	5	25439496.63	5087899.33	10.50	<.0001

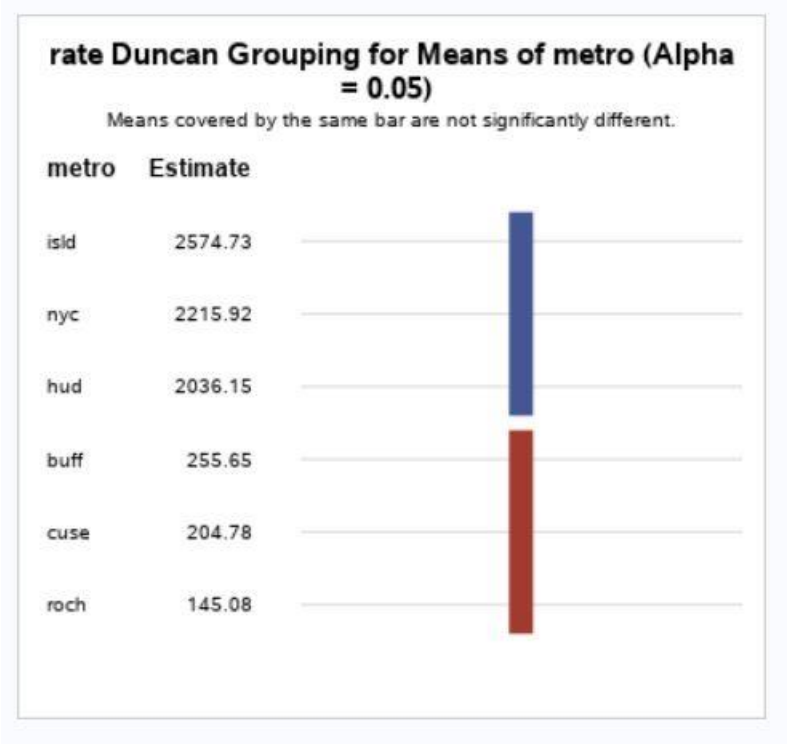
We can see from the GLM procedure that runs our ANOVA test the p-value is well under 5 percent for our model, this means that we reject the null hypothesis. In this case, the null hypothesis is the means between the metro areas the same. Since we reject that idea, we can say that we have strong evidence that mean infection rate between the different areas throughout the state is different. Let's check the boxplots to confirm this idea.

Figure 4: Metro Boxplots



From the boxplots we can see that there might be a difference between the means of Buffalo, Syracuse and Rochester when contrasted with the areas closer to New York City. The areas that have the counties with the largest variance is the Hudson Valley region. From the evidence of the p-value of the model and the boxplot, we can stay that there is strong evidence that the rate of infection is different throughout the state, but which rates are the same between the metro areas?

Figure 5: Duncan Test COVID -19 Rates



This is a Duncan test which tests for means that are statistically the same between the means of the ANOVA test, which. We can see that the mean of the infection rate between the counties of Long Island, New York City, and the Hudson Valley are statistically the same. This is shown by the blue bar. The red bar, which includes the counties of the Buffalo, Syracuse, and the Rochester metro areas are statistically the same. This means that the rate of infection in Long Island, Hudson Valley and New York City are considered the same, while the means between Buffalo, Syracuse and Rochester are considered the same. This shows that there is a clear divide in the rate of infection in Upstate New York and the New York City area.

DATA:

The data for this project was manipulated in excel.

The county populations:

https://worldpopulationreview.com/us-counties/ny/

The COVID – 19 infection data:

https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/