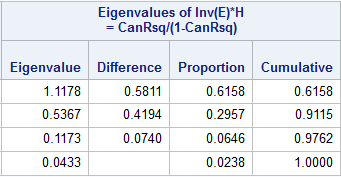
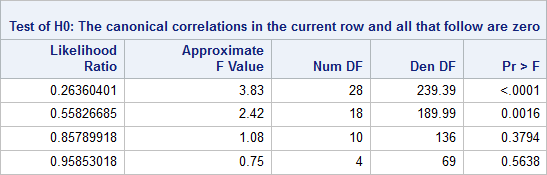
## Discriminant Analysis

Using our dataset, we can perform a discriminant analysis to identify which of the seven variables helps contribute the most to the definition of the 5 NYC boroughs. Using the formula , we can see that there are going to be four non-zero eigenvalues for **E-1H**, indicating there would be four uncorrelated discriminant functions describing the separation of the NYC boroughs based on felony crime statistics. In Figure 1, we can see there are indeed four non-zero eigenvalues describing the matrix **E-1H**, and that only the first two eigenvalues are needed to account for over 90% of the separation.



Figure

Utilizing Wilks’ test, we can perform an iterative test to examine each discriminant function and their significance to the separation of the five boroughs. In Figure 2, we can see the values in the “Likelihood Ratio” column and their significance based on approximate F values in the “Pr > F” column.

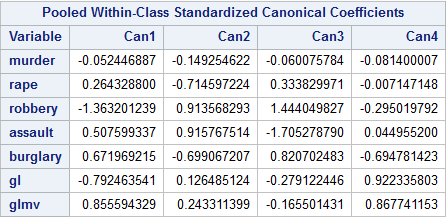


Figure

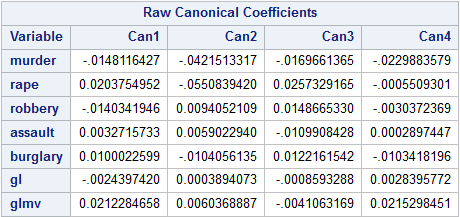
Although the p-value of is below our threshold of 5%, the p-value is based off an approximate F value. Using the table of the lower critical values for Wilks , we can confirm that of is significant since . We used since is not made available in the table and since is a more conservative comparison, it is sufficient for our test.

In Figure 3, we can find the standardized coefficients for the four discriminant functions, which we will use for comparison and interpretation. Figure 4 is included to show the raw coefficients of the discriminant functions. Since only the first two eigenvalues are significant, we will only focus on analyzing the first two discriminant functions, Can1 and Can2. Using the magnitudes of the coefficients, we can see that “robbery” (-1.36) makes the largest contribution for Can1, but there are also some non-insignificant contributions from the variables “grand larceny of motor vehicles” (0.86), “grand larceny” (-0.79), “burglary” (0.67), and “assault” (0.51). However, if we take the sign into consideration, we can see the weight of robberies and grand larcenies (-2.15) is similar to the weight of grand larcenies of motor vehicles, burglaries, and assaults (2.04).

In the second discriminant function, we see the variables “assault” (0.92) and “robbery” (0.91) providing the largest contributions to the separation of the boroughs. “Rape” (-0.71) and “burglary” (-0.70) also provide significant amount of contribution to the separation of the boroughs.



Figure

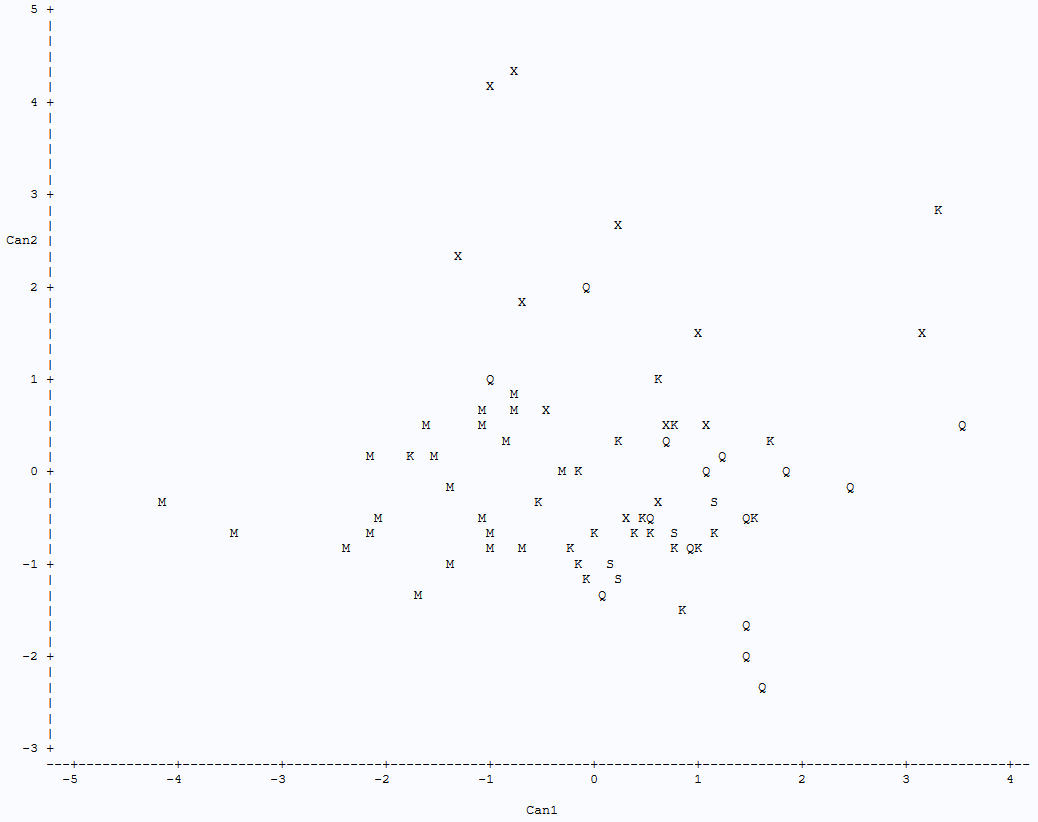


Figure

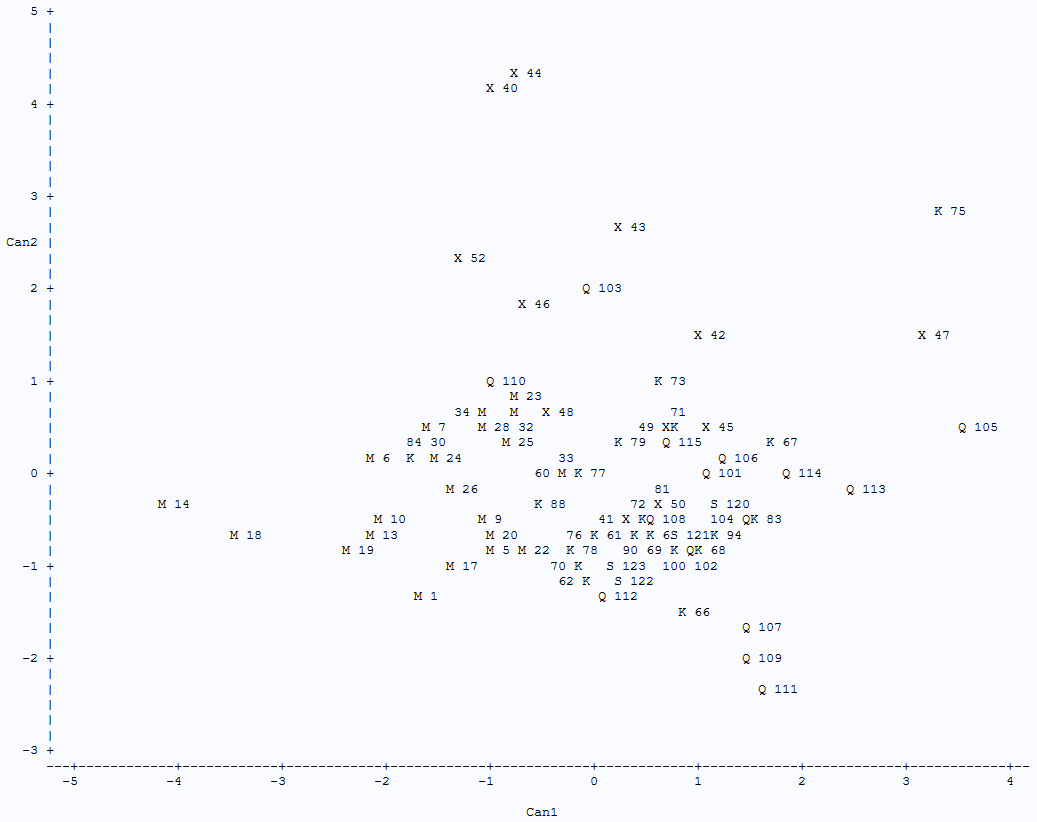
Since only two discriminant functions are significant to the separation of the five boroughs, we can visualize the flattening of our 7-variable and 5-group data onto a 2-dimensional plane of our group separation. In Figure 5, we plot the transformed observations of crime in each police precinct onto the axes of the first two discriminant functions.

There is an easily distinguishable section that represents the Manhattan police precincts. However, the other 4 boroughs do not have such clear separation. Although Bronx seems to have a clear section described by the transformed observations, there are several Bronx precincts that appear much further away from that area. Between the Brooklyn and Queens precincts, there is a large overlap between these two boroughs. Finally, Staten Island is fully encased by either the Brooklyn or Queens area.

While this analysis does not provide us with any information on the overall safety of the five boroughs, we can see how the major felony crime statistics in each neighborhood help to describe the borough as a whole. Anecdotally, Bronx had reputation of being dangerous and our analysis does show that the crime statistics seem to group their neighborhoods together. However, we do see some Bronx precincts (48, 49, 50) have crime statistics more closely related to a Brooklyn neighborhood. There are also 2 non-Bronx precincts (75, 103) that have crime statistics like some of Bronx’s neighborhood. These precincts are located in Brownsville and Jamaica, both having a poor reputation of being dangerous.



Figure



Figure

## SAS Code for Discriminant Analysis

data felony;

infile "/folders/myfolders/MV.DAT";

input borough $ precinct $ murder rape robbery assault burglary gl glmv;

run;

proc candisc out=nyc;

class borough;

run;

proc print data=nyc;

run;

proc plot data=nyc;

plot can2\*can1=borough;

plot can2\*can1=borough $ precinct;

run;