STAT 210: Quiz #4 Name(s): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
Fall 2017
Points: 20 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Goal: Identify the “best” 5 and “worst” 5 precincts in the NYC Stop and Frisk data.

“Best” precincts are precincts whose race distribution in the NYC Stop and Frist data most closely matches the race distribution of residents who live within this precinct. “Worst” precincts are those were the race distribution in the NYC Stop and Frisk data is most extreme against the race distribution of residents who live within this precinct.

Answer the following questions.

1. Use the Chi-Square statistic shown below to measure the discrepancy between the race distribution from the NYC Stop and Frisk data to the race distribution of residents who live within this precinct.

$$\sum\_{all races}^{}\frac{\left(Observed-Expected\right)^{2}}{Expected}$$

Obtain a total value for each precinct. Sort the list by total. Use this sorted list to identify the “best” 5 and “worst” 5 precincts. (10 pts)

|  |  |  |
| --- | --- | --- |
| “Best” Precincts |  | “Bad” Precincts |
| Rank | Precinct # |  | Rank | Precinct # |
| 1. (Best Overall)
 |  |  | 1. (Worst overall)
 |  |
|  |  |  | 2. |  |
|  |  |  | 3. |  |
|  |  |  | 4. |  |
|  |  |  | 5. |  |

1. The race designations in the NYC Stop and Frisk data do not (exactly) match the race designations in the Census data. Explain what race designations were ignored or combined in order to compute the above measure of discrepancy between these two distributions. (4 pts)
2. Consider your list of 5 “best” precincts and 5 “worst” precincts. Use Google Maps to identify the location of these precincts.
	1. Do the 5 “best” tend to be in the same geographic area? Briefly discuss. (3 pts)
	2. Do the 5 “worst” tend to be in the same geographic area? Briefly discuss. (3 pts)

Link: <https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz/data>

