Appendix Q
Chi-Square Methodology Notes

Population data used as a benchmarking reference were sourced from the Vintage 2023 Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic published July 1, 2024, by the US Census Bureau. Additional details regarding the data are available in Appendix M.

The chi-square test is a non-parametric (can be applied to any distribution) statistical test used to examine relationships between categorical variables. It helps determine whether observed frequencies differ significantly from expected frequencies by comparing the observed frequency distribution with the expected distribution under the null hypothesis, which assumes that any difference or association between the observed and expected frequencies are the result of random chance.

- *Null hypothesis (H_0):* There is no association between the variables (they are independent), or the observed distribution matches the expected one.
- *Alternative hypothesis (H_1):* There is an association, or the distribution differs.

The chi-square statistic is calculated:

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

Analysis of the CPA traffic stop data applied the following methodology to both searches and arrests pursuant to traffic stops involving Black and Hispanic drivers compared to White drivers.

Using the specific case of searches involving Black drivers,

- The **expected** distribution was the number of stops that resulted in a search if the ratio of White *search* subjects to Black *search* subjects were exactly the same as the ratio of White *stop* subjects to Black *stop* subjects.

$$\frac{\text{\#Searches of White Drivers}}{\text{\#Searches of Black Drivers}} = \frac{\text{\#Stops of White Drivers}}{\text{\#Stops of Black Drivers}}$$

- The **observed** distribution was the number of actual stops that resulted in searches for White and Black drivers, respectively.

$$\frac{\text{Observed \#Searches of White Drivers}}{\text{Observed \#Searches of Black Drivers}} \quad \textit{Compared to} \quad \frac{\text{Observed \#Stops of White Drivers}}{\text{Observed \#Stops of Black Drivers}}$$

For each agency meeting the testing threshold criteria[1], the chi-square test was calculated:

$$\chi^2 = \sum \frac{(\text{Observed \#Searches of Black Drivers} - \text{Expected \#Searches of Black Drivers})^2}{\text{Expected \#Searches of Black Drivers}}$$

The resulting difference between the observed number of searches of Black drivers compared to the expected number of searches of Black drivers if the ratio of searches to stops for Black drivers was equal to that of White drivers is referred to as the residual. The square of the residual divided by the expected number of searches yields the chi-square statistic.

If the probability of the observed result (distribution) occurring by chance was less than one percent (i.e., $p < 0.01$), chance was not considered a plausible explanation, and the test results were deemed statistically significant.

Statistically significant results indicated that random chance was not a plausible explanation for the difference in an agency's search rates for White and Black drivers. Significance also indicated that White and Black drivers had different likelihoods of being searched by officers of a given law enforcement agency as a result of some difference between the two racial groups. In contrast, a lack of statistical significance in test results indicated that the difference in search results *could* be the result of chance or some difference between the two groups.

It is important to note that statistical significance is not indicative of, nor does it prove, a specific causation. Statistical significance indicated that *some* difference between White and Black drivers affected their likelihood of being searched/arrested, but did not prove that racial identity (or, more precisely, racial bias) was responsible for the difference in search/arrest rates. Other variables that aligned with White or Black racial identity may well have contributed to the differences in search rates.

[1] In this context, chi-squared testing required that expected values (i.e., White drivers searched, White drivers not searched, Black drivers searched, and Black drivers not searched) be greater than or equal to five. Some agencies did not meet the minimum criteria for testing and therefore could not be tested for statistical significance.