For each category in the dataset:

1. Count: the number of occurrences of each category in the sample (you can find from Dataset)

2. Proportion (p): The relative frequency of each category calculated as:

3. Standard Error (SE): The measure of variability of the proportion, calculated as:

where *n* - sample size, *p* – proportion;

4. The formula for the 95% confidence interval (CI) of a proportion p is

where Z - for the desired confidence level (1.96 for 95%).

5. The intraclass correlation coefficient (rho) and average cluster size () are assumed for each category as follows:

 category\_specific\_params = {

"White": {"rho": 0.015, "m": 6},

"Black": {"rho": 0.025, "m": 5},

"Mexican": {"rho": 0.030, "m": 4},

"Other": {"rho": 0.020, "m": 5},

"Other Hispanic": {"rho": 0.018, "m": 6},

}

When clustering is assumed in the sampling design, the Design Effect (DEFF) is typically greater than

where - Intraclass correlation coefficient (ICC), representing the degree of similarity within clusters;

*m* - Average number of observations per cluster

Adjusting for Clustering :

6. To compute proportions and standard errors while accounting for clustering:

7. Update confidence intervals with the adjusted standard errors.