**Data Sampling**

Camilo Hoyos

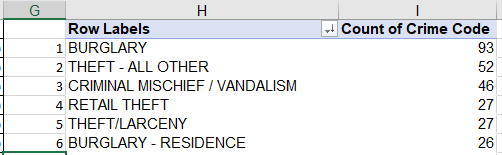
DAT-375 Data Analysis Techniques

Jacqueline U. Clauss, M.S.

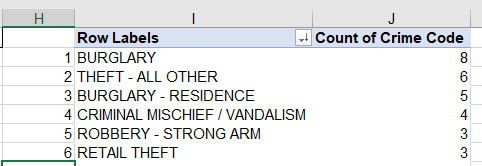
February 7, 2021

For the purpose of this analysis, we will call the entire data set the “population” and every tenth record will be defined as the “sample”. The top three most frequent crimes from the sample set are Burglary, Theft-other, and Burglary - Residence. The top three most frequent crimes from the population are Burglary, Theft-other, and Criminal Mischief / Vandalism.

*Population top 6*



*Sample top 6*

**

The results from the sample are pretty close to the population in such a way that the top two entries match between population and sample, yet the third entry has a variation. It should be noted that in the population entries, the ‘Burglary - Residence’ was ranked 6th. In the sample, the ‘Criminal Mischief / Vandalism’ was ranked 4th. This goes to show that sampling will not always be one-hundred percent accurate to the population, but it is close enough to draw conclusions.

The purpose of taking a sample from an analysis is to be able to draw inferences by more efficient means. It is cost efficient and more practical than working with an entire population. ( Gangwal, 2019) This is especially true if the data set is of immense proportions such as over a trillion records. Sampling is also reasonable as it is not always possible to query or get feedback on an entire population. So, individuals gather a strong sample that represents a population and then to draw conclusions of the population based on the sample.

References

Gangwal, R. (2019, September 11). What is sampling: Types of sampling techniques. Retrieved February 05, 2021, from https://www.analyticsvidhya.com/blog/2019/09/data-scientists-guide-8-types-of-sampling-techniques/