Dengue Statistics

We looked at precipitation and temperature measures in our dataset as these are known to affect mosquito populations. We looked at the different climate readings between Iquitos and San Juan since there is not only a difference in the case patterns but of their climate patterns as well.

The average air temperature between the two cities, we see a stronger amplitude delta for San Juan, Puerto Rico than in Iquitos, Peru, meaning we have a wider variation in average air temp in San Juan than what is observed in Iquitos. Looking at the distribution of this feature, we see San Juan is bimodally distributed and Iquitos seems more normally distributed. This is good to know for when building models, should this feature be incorporated.

To get an idea of how well this is representative of the population, I used bootstrapping to draw replicates from these samples to observe what a mean value of this feature could be for each city. I split each city into its own separate subset since the pattern and volume of case numbers appears to be separate as does the climate data behavior as seen in the EDA portion of this project.

Drawing replicates and utilizing the Central Limit Theorem validated the difference in mean values for the average air temperature between the cities, suggesting that the observed data is representative of the difference of the population from these two cities. I’m glad to see this result, because I can be more confident there is a distinct difference these two cities’ mean average air temperature when modeling, to know that this difference isn’t observed by chance.

Also using this approach, I looked at the precipitation values in these locations as well. Standing water is a hotbed of mosquito activity, which is the vector of spreading the Dengue viruses. I was able to find some information on what happened during those peak outbreak seasons in San Juan, but less so for the peak infection times in Iquitos. I wanted to look at how these precipitation measurements were different between the two cities as I suspected that San Juan being in the Caribbeans had precipitation records closer to that of Iquitos than our sample data showed. By utilizing a permutation test of means, the results showed that the two mean replicate precipitation values were much closer together than shown in our data, suggesting that both locations had close to equal amounts of precipitation on average.

As seen in the EDA portion, there doesn’t seem to be a feature variable that is strongly correlated with our target variable of total cases per week, so knowing how these two features that can affect mosquito populations strongly are related and represented is important to know.