22 de abril del 2021, Ciudad de México.

Over the course of 45 days, we observed the performance of several drugs on 248 mice. We compared the change in size of their tumor with different drug regimens, including a placebo and our drug of interest.

In order to register our findings, we created to dataframes. In the first one, we mouse’s ID, the drug we gave it, their sex, their age in months, and their weight. In the second one, we wrote the ID, the timepoint, the tumor’s volume, and if there was metastasis. We joined both of this dataframes into one, using the mouse’s ID. We also cleaned the dataframe, dropping the duplicates. Once we had a clean dataframe, we did the analysis. This are our findings:

Without taking into consideration the presence of outliers and extreme values our statistical summary dataframe shows us that the drug Ramicane had a lower average. Which could possibly mean that the mice that had Ramicane as treatment had a better response. However, this is a first glance to the data.

Our analysis of the Standard Deviation concluded that both Capomulin and Ramicane had the most similar response to each drug. This means that since the variation is not high, the tumors responded similarly to the drug. Both drugs also had the lowest average, which could possibly mean that they had the best and most uniform response.

We also find that both these drugs had the most amount of measurements taken and given they had the lowest variation from the average, our conclusion of the best and most uniform response stands.

Looking at the boxplot we concluded that Ramicane is far more effective at reducing the size of the tumors. In contrast, Infubinol is the least effective and we should look into that outlier case, in order to better understand why it is so far from the rest of the distribution.

The analysis of Capolumin was done first with a line plot to further demonstrate the diminution of the tumor through time. We can observe the dramatic decrease in size of the tumor in a specific mouse as the days went by.

In our scatter plot we observed a clear positive correlation between the weight of the mouse and the volume of the tumor. When we did the linear regression, we confirmed the positive correlation between these two variables, which means that as one grows, the other will too. So, the higher the weight of the mouse, the higher the volume of the tumor.

Finally, we have the Pearson Correlation Coefficient, which is 0.84. This allows us to make two conclusions; the first one is that the correlation between the two variables is positive and it is a high correlation. But as we already know, correlations doesn’t mean causation.