***“Matplotlib Homework Analysis”***

1. ***From regimen\_summary Table:***

It is obvious that “Ketapril” drug with the highest amount of all statistical measures (mean, median, variance, standard deviation, and SEM) among all regimens is the most effective drug regimen in estimating the Tumor Volume of the tested mice even though the numbers of mice in all tests are approximately equal to 25.

Also, the “Remicane” drug with the lowest amount of all statistical measures is the least effective drug regimen among others with the same number of tested mice.

1. ***From Tumor Volume (mm3) across four Regimens boxplot:***

It can be stated that there is just one outlier in the “Infubinol” drug regimen, which shows data out of the acceptable range of measurement. Except for this, all other obtained data from this drug and other drug regimens are between the range of lower\_bound and upper\_bound of all data. So, the gathered data are reliable, and the research results are logical and trustworthy.

1. ***From different plots:***

In the Tumor Volume vs. Timepoint line plot, we can see a downward trend between Timepoint and Tumor Volume variables for a sample mouse selected randomly. It proves that over time with increasing the Timepoint levels, the amounts of Tumor Volume are decreasing. So, it concludes that passing the time for this sample mouse helps reduce the Tumor Volume. But in the Average Tumor Volume (mm3) vs Mouse Weight (g) scatter plot, there is an upward trend between Mouse Weight and Average Tumor Volume which means that with increasing the weight of the mouse, the volume of the tumor increases, and it proves another disadvantage of the gaining weight.

This issue can also be seen in the amount of “Pearson-r correlation coefficient“ with the amount of 0.84 that is more than 0.7 and therefore shows a strong relationship between these two variables (Mouse Weight as “X,” Average Tumor Volume as “Y”). Furthermore, this positive relationship can be proven in the Average Tumor Volume (mm3) versus Mouse Weight (g) regression line equation and the plot. Because the calculated slope is +0.95, that is positive and also near 1, which indicates a positive and strong relation between X, Y variables.