HOMEWORK05 – MATPLOTLIB

CANCER ANALYSIS

GOAL

The exercise aims to study the influence of a certain amount of Drugs Regimen on mice with cancer. The scientists tracked the disease's evolution during a maximum of 45 days, where different mice were under a different Drug Regimen.

DATA CLEANING & TREATMENT

In what concerns the data used, it is divided into two different datasets. One with information about the mice themselves, like Age (month), Sex, Weight, and to which Drug Regimen they were subjected. The other dataset contains all the tracking records of the scientists, composed by Timepoint, Tumor Volume, and # of metastatic sites.

Thus, the first step was to merge both datasets based on the Mouse ID. Below you can find a small sample of each table.

TABLE 2: DATASET #1 - MICE METADA

	Mouse ID	Drug Regimen	Sex	Age_months	Weight (g)
0	k403	Ramicane	Male	21	16
1	s185	Capomulin	Female	3	17
2	x401	Capomulin	Female	16	15
3	m601	Capomulin	Male	22	17
4	g791	Ramicane	Male	11	16

TABLE 1: DATASET #2 STUDY RESULTS

	Mouse ID	Timepoint	Tumor Volume (mm3)	Metastatic Sites
0	b128	0	45.0	0
1	f932	0	45.0	0
2	g107	0	45.0	0
3	a457	0	45.0	0
4	c819	0	45.0	0

Before moving forward to the analysis, the data required some data clearing due to a specific mouse who had duplicated study data – where I am not able to separate which of the two data points are corrected for this analysis.

Thus, the decision was to remove all mouse information on the review related to their drug regimen or in any aggregated data where Drug Regimen is the main feature to be analyzed.

Data Analysis

Summary Analysis

The study was performed over 249 mice, presenting different results and outcomes based on the Drug Regimen they had been submitted to. Worthwhile to mention that all drug treatments were equally tested in 25 different mice – a small exception to Stelasyn, which was only applied to 24 mice.

Before moving into a more detailed view of the data is always better to understand the get overall information about the main objective of the analysis – assuming the efficiency of each drug regimen on cancer (in the dataset defined as tumor volume). The below table illustrates the aggregate information about it.

TABLE 3: DRUG REGIMEN SUMMARY TABLE

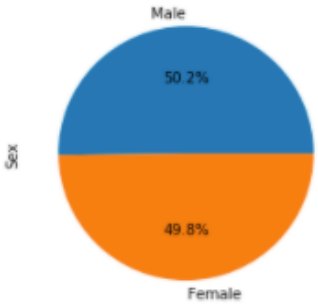
Drug Regimen	Mean Tumor Volume	Median Tumor Volume	Tumor Volume Variance	Tumor Volume Std. Dev.	Tumor Volume Std. Err.
Capomulin	40.68	41.56	24.95	4.99	0.33
Ceftamin	52.59	51.78	39.29	6.27	0.47
Infubinol	52.88	51.82	43.13	6.57	0.49
Ketapril	55.24	53.70	68.55	8.28	0.60
Naftisol	54.33	52.51	66.17	8.13	0.60
Placebo	54.03	52.29	61.17	7.82	0.58
Propriva	52.32	50.45	43.85	6.62	0.54
Ramicane	40.22	40.67	23.49	4.85	0.32
Stelasyn	54.23	52.43	59.45	7.71	0.57
Zoniferol	53.24	51.82	48.53	6.97	0.52

Based on this table, I would like to highlight both Drug Regimen: "Ramicane and Capomulin." Across all treatments, those are the two that present better results on the overall Average Tumor Volume and the Standard Error Associated with it compared with the population. It sets those to a part of the competition and can be considered as the best treatment.

Population Demographics

Despite the fact, I have not gone further on analyze how the demographics could affect the Drug Regimen results, and it was essential to understand how balanced was the mice sample for this study. In the case of having a fair sample, we could assume the Sex should not get a significant influence on the results itself.

FIGURE 1: SAMPLE DEMOGRAPHICS - SEX

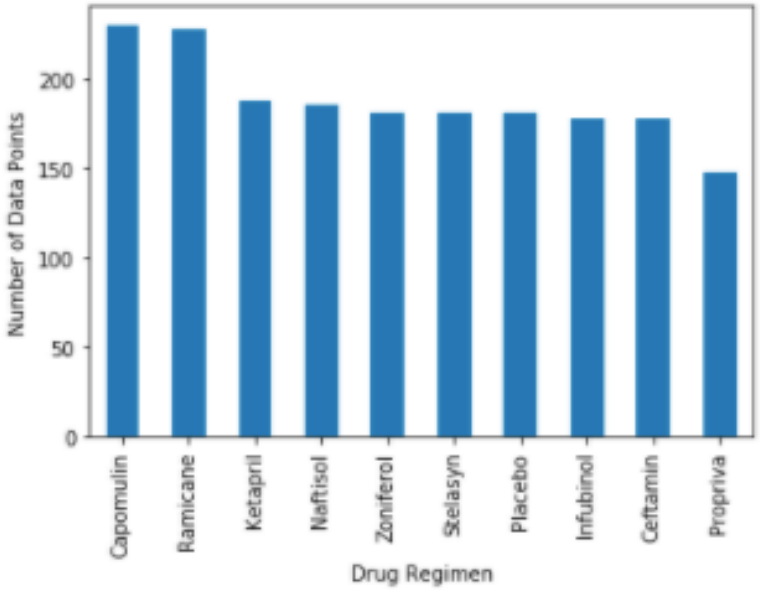


Long-Lasting Drug Regimen

Since all treatments have the same number of mice, we have decided to understand which treatment accounts have a significant amount of data points. This metric is vital because it connects with mice who reacted better to Drug Regimen, therefore that last longer during it.

Figure 1 shows illustrate once again that Capomulin and Ramicane set apart of the competitors, where the mice subjected to their drugs lasted longer than the others.

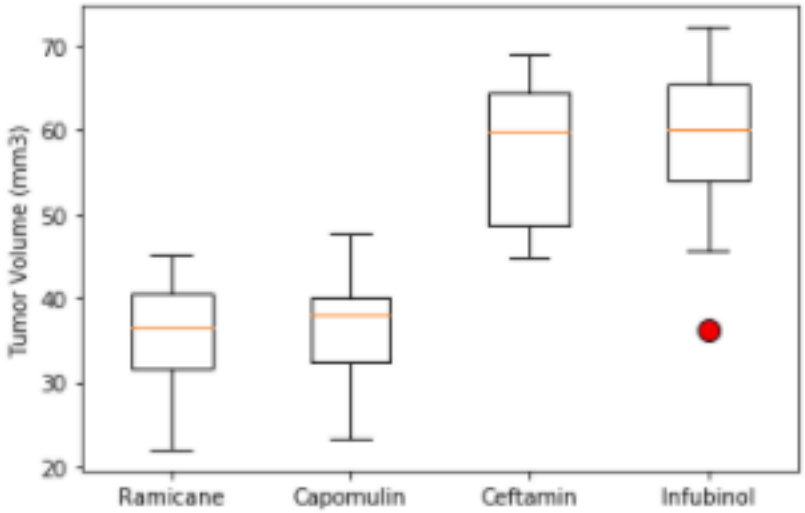
FIGURE 2: NUMBER OF DATAPOINTS BY DRUG REGIMEN



End of Study Results Analysis

This section will be analyzed the last measure under the four drugs regimen with better results – Ramicane, Capomulin, Ceftamin, and Infubinol. The selection was made using as a factor the Total Std Error of each treatment.

FIGURE 3: BOXPLOT OF THE 4 BEST DRUG REGIMEN



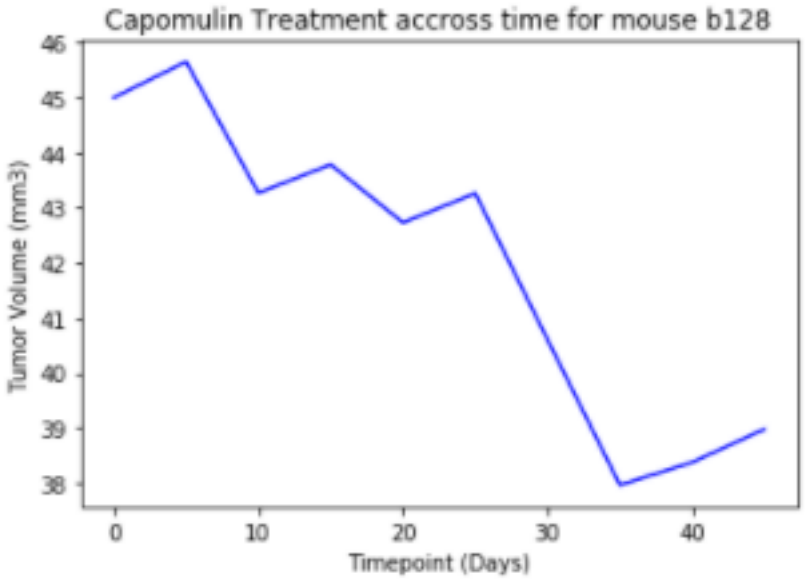
The boxplot graphs in figure 3 illustrate the fact that at the end of the study, the Tumor Volume mean is lower in Ramicane treatment, closely followed by Capomulin. Moreover, we can understand there is no much variance in the results since the interquartile distance is smaller on those two treatments when compared with Ceftamin and Infubinol. These points just reinforce all the major conclusions drawn previously.

Another interesting point to highlight is the manifest presence of a possible outlier under Infubinol treatment (marked in red in the graph). Due to the lack of further information, I cannot take it out or conclude that it is an outlier and should not be considered, although it is a point that calls attention.

Tracking Evolution

Just for a matter of example, it is possible to see on the below image the evolution of the Tumor Volume for a specific mount under the Capomulin Drug Regimen.

FIGURE 4: TREATEMENT EVOLUTION FOR B128 MOUSE



There is a negative trend of the Tumor Volume with the time, with this decreasing based on the application of the Drug Regimen. It provides a sound signal of the efficiency of it. Although it is not possible to conclude the fact that this behavior will happen this way since it is just one case; nevertheless, it sums to the findings before where Capomulin is one of the best performers.

A new characteristic to explore in the future analysis would be the fact to understand if there is any period where the Drug is more effective. Just looking for this example, we can define that between the 25 and the 35 days was where the treatment got a significant impact. It is crucial since we can articulate several therapies based on their time efficiency.

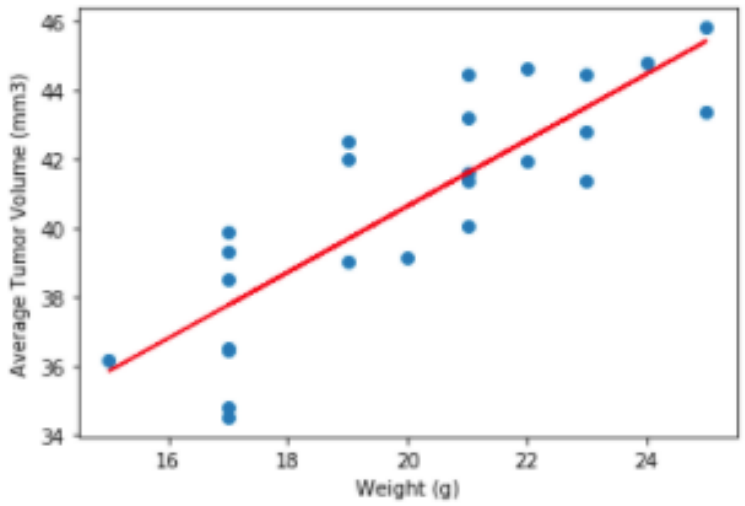
Weight Dependency

Another point for further analysis comes from the fact of studying the Dependency of the tumor volume on the weight. Below we have drawn it for one of the treatments – Capomulin.

With a correlation of 0.84, it is possible to state that those two variables are somehow connected. In my personal opinion, it is expected since mice with more weight probably are more prominent, and therefore have triggered more growth hormones that also help the Tumor growing.

FIGURE 5: WEIGHT VS TUMOR VOLUME FOR CAPOMULIN REGIMEN

```
The correlation between mouse weight and the average tumor volume is 0.84
```



CONCLUSION

The overall conclusion of this study is the fact that Ramicane and Capomulin treatments perform better than competitors.

However, this study also highlights the fact the much more analysis and reviews should be performed around the data, like understanding Treatment trends and efficiency periods, further investigation on the weight of the mice, there is any age influence on the results, just to mention a couple of them.