

**BOOTCAMP DATA ANALYSIS**

**PATRICIO AUGUSTO PATIÑO ESPAÑA**

**CHALLENGE MODULE4**

**Pymaceuticals Analysis**

**WRITTEN REPORT**

**July 14th, 2024**

Objective:

Analyze the data we get from the Pymaceuticals.jpynb from the csv’s study\_results and mouse\_metadata.

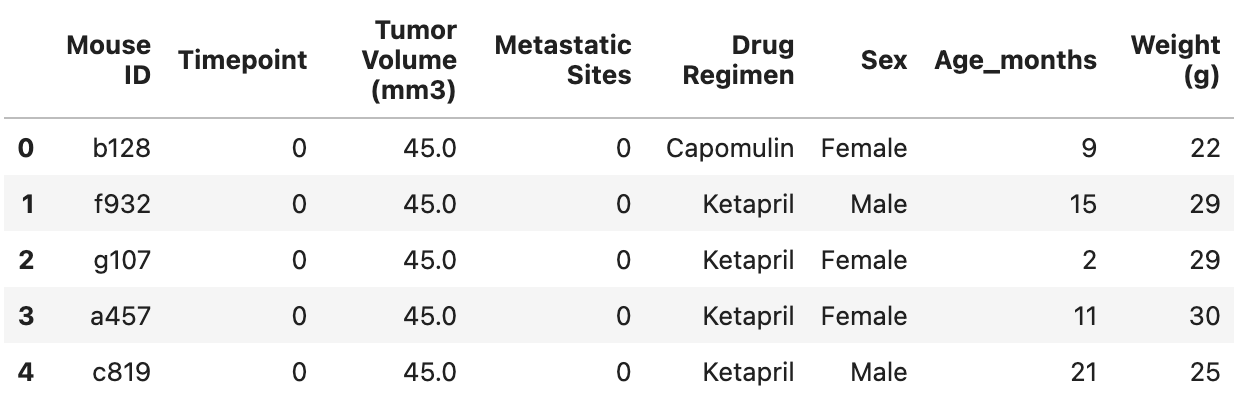
This study is based on a company that specializes in anti-cancer medication, for squamous cell carcinoma (SCC), from skin cancer. This study is based on 249 mice who are identified with SCC tumors, and it analyze over the course of 45 days, analyzing 10 drugs, been Campulin the Drug of interest, and with the data we obtain the following.

**Data**

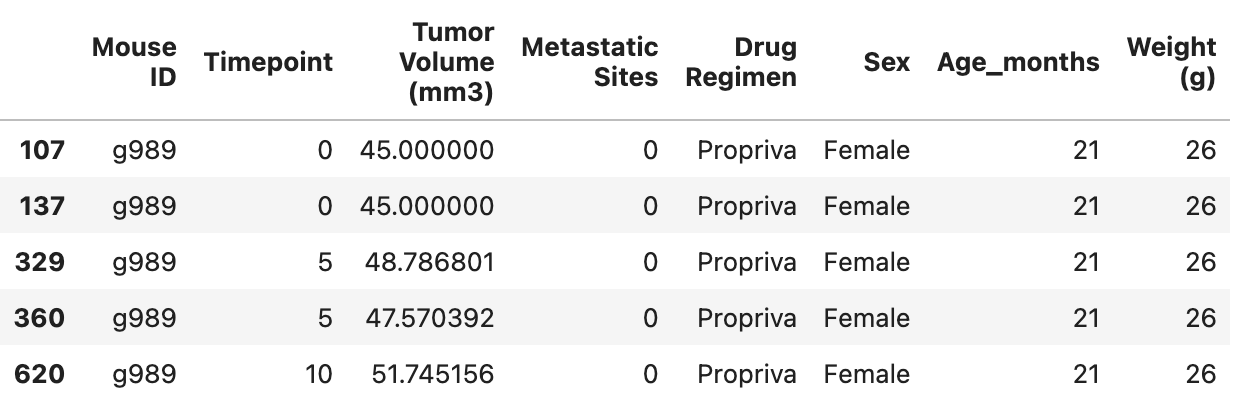
**General Data**

The first data we get:

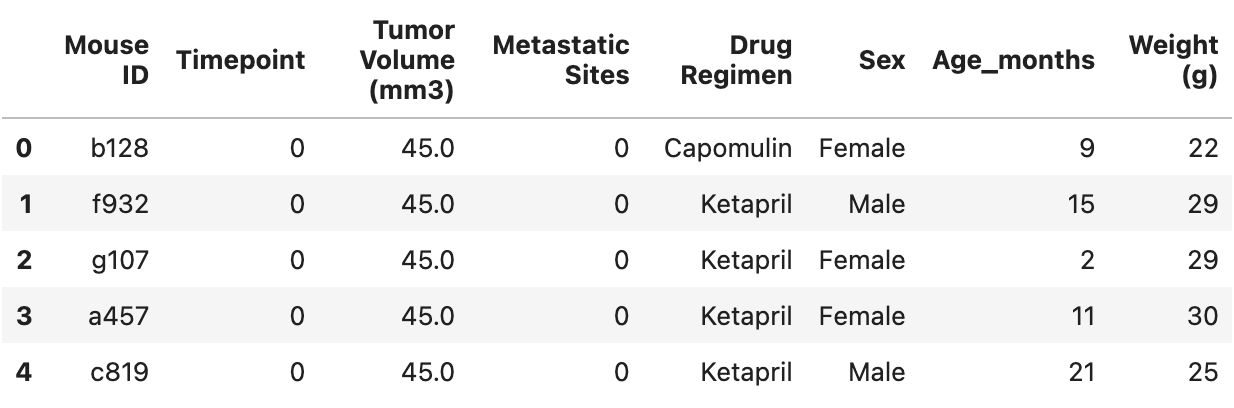
* Number of mice: 249
* Number of mice no duplicate: 248
* Mouse duplicate: g989

In this first general data, we get the data frames for the mice, we get the two csv in one by the Mouse ID.

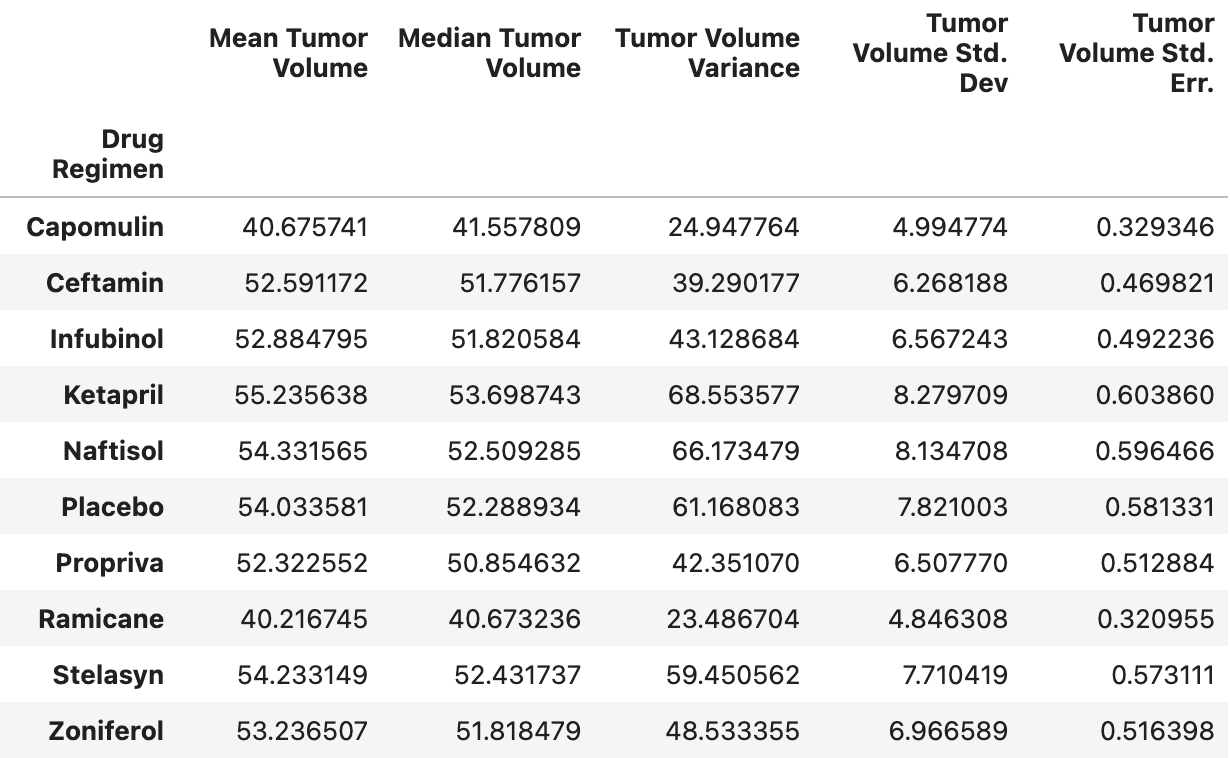
Then we get the Mouse ID that it’s duplicated.



And a clean Data Frame without a duplicated mouse.



**Summary Statistics**

****

In this we are going to analyze by Drug Regimen, we have 10, this are:

* Campulin
* Ceftamin
* Infubinol
* Ketapril
* Naftisol
* Placebo
* Propriva
* Remicane
* Stelasyn
* Zoniferol

Mean

In the mean we have that Campulin and Ramicane has the lower mean and Ketapril and Naftisol have the highest mean.

Median

In the median we have also Campulin and Ramicane with the lower median and Ketapril and Naftisol with the highest median.

Variance

In the variance we also have Campulin and Ramicane with the lower variance and Ketapril and Naftisol with the highest variance.

Standard Deviation

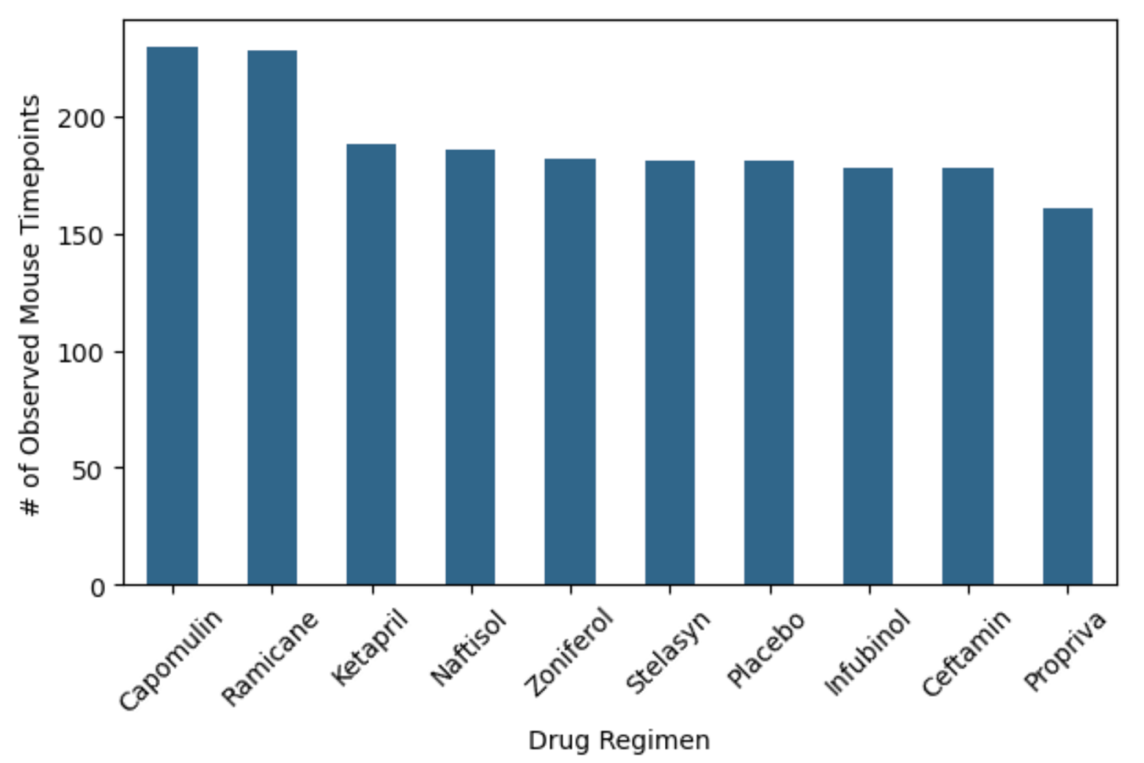
In the standard deviation we also have Campulin and Ramicane with the lower standard deviation and Ketapril and Naftisol with the highest standard deviation.

Standard Error

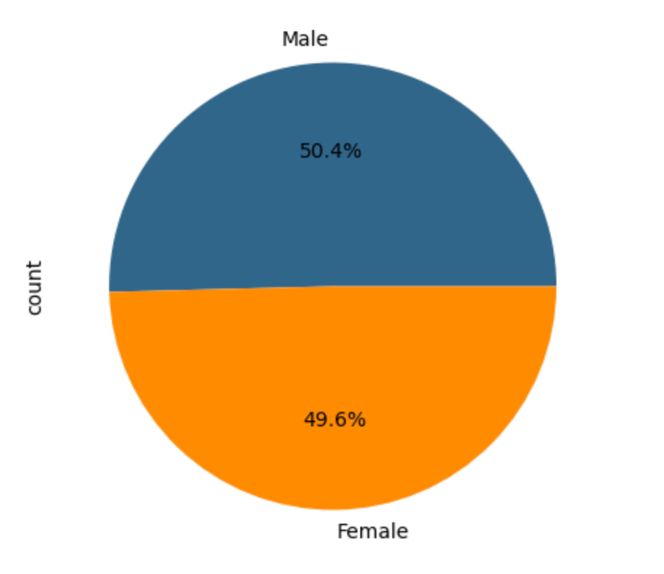
In the standard error we also have Campulin and Ramicane with the lower standard error and Ketapril and Naftisol with the highest standard error.

With this data that we obtain, the most trustable results are for Campulin and Ramicane, has the lower variance and standard deviation from the mean, so we don’t have a lot of spread data, and has the lower standard error, so the data will not variate a lot, the mean and the median, are the lower value for the tumor volume, so in this case, with this drug the tumor volume didn’t increase a lot, contrary to Ketapril and Naftisol that obtain the highest values, so we can say that are not trustable data, and the Tumor volume register a bigger size.

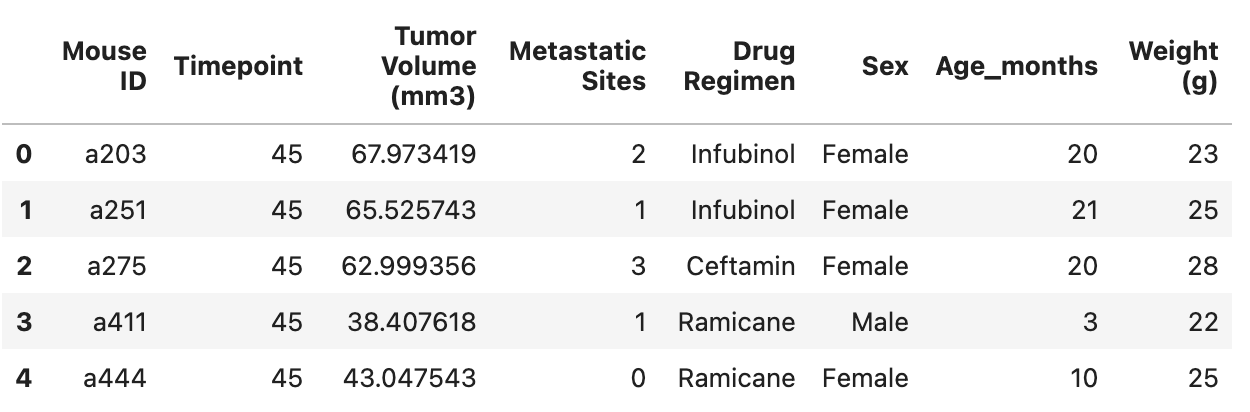
**Bar and Pie Charts**



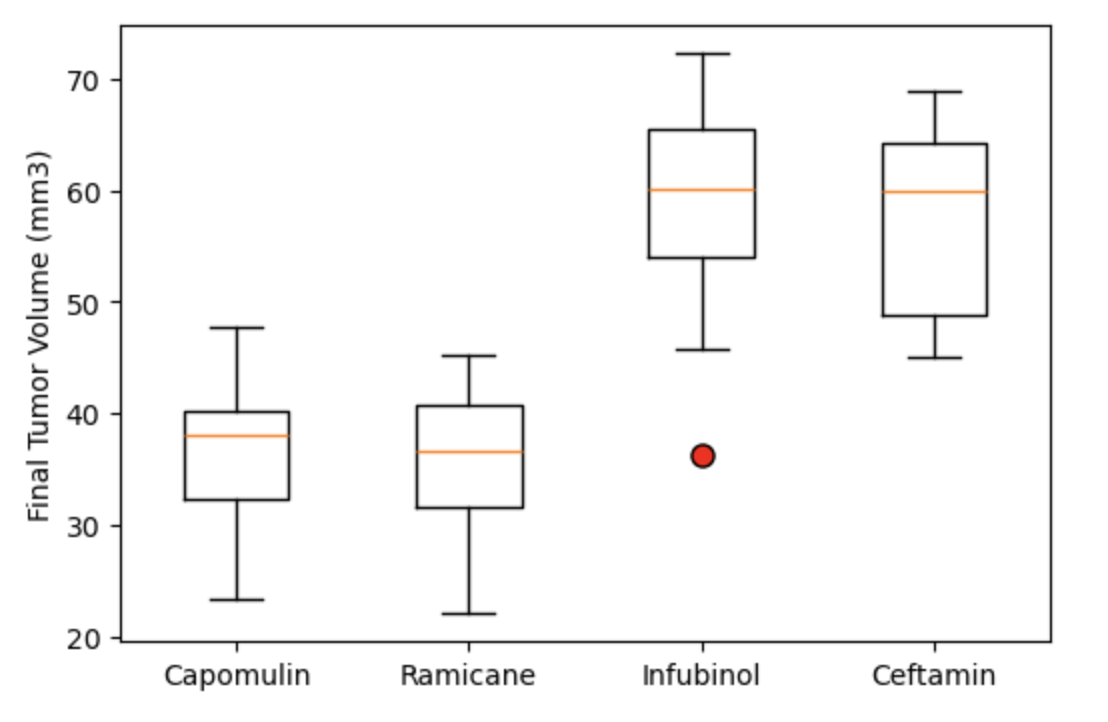
In this bar plot, we can observe each Drug Regimen and the Number Observed Mouse Timepoints, and we can observe that Campulin and Ramicane have more number of mouse observed timepoints, contrary to Propriva that has the lowest number of observed mouse timepoints.



In this Pie charts we obtain that 50.4% of the mice are male and the 49.6% are female.

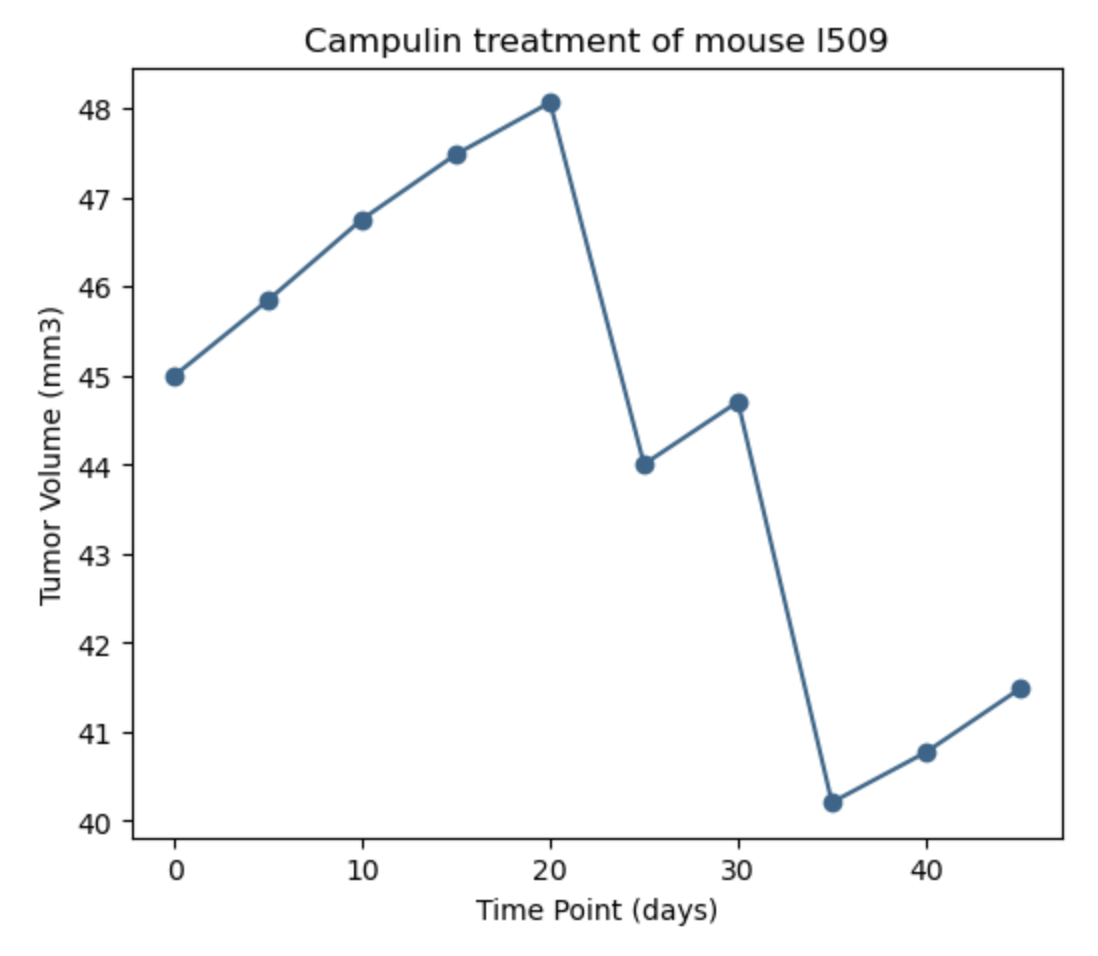


In this Data Frame we visualize all the final Tumor Volume (mm3) after the Timepoint of 45 days, with all of the values of Drug Regimen, Sex, Age\_months and weight, and has different values of the final tumor, with the different Drug Regimen.

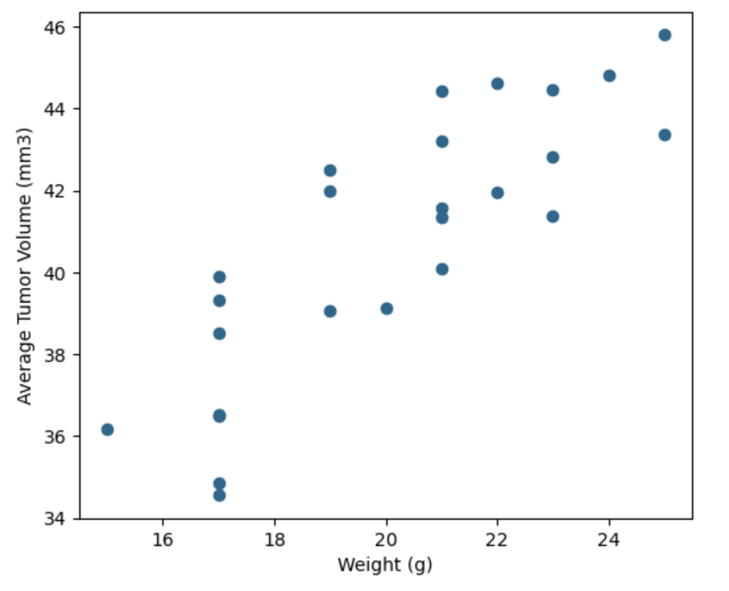


In this boxplot we are going to analyze only 4 Drug Regimen that are Campulin, Ramicane Infubinol and Ceftamin, and in this chart we can observe that Campulin and Ramicane doesn’t have a high increase in the tumor, it’s almost at the same of the median, and Infubinol and Ceftamin have a higher increase in the tumor volume than the other 2, because those two are higher than the median.

**Line and Scatter Plots**

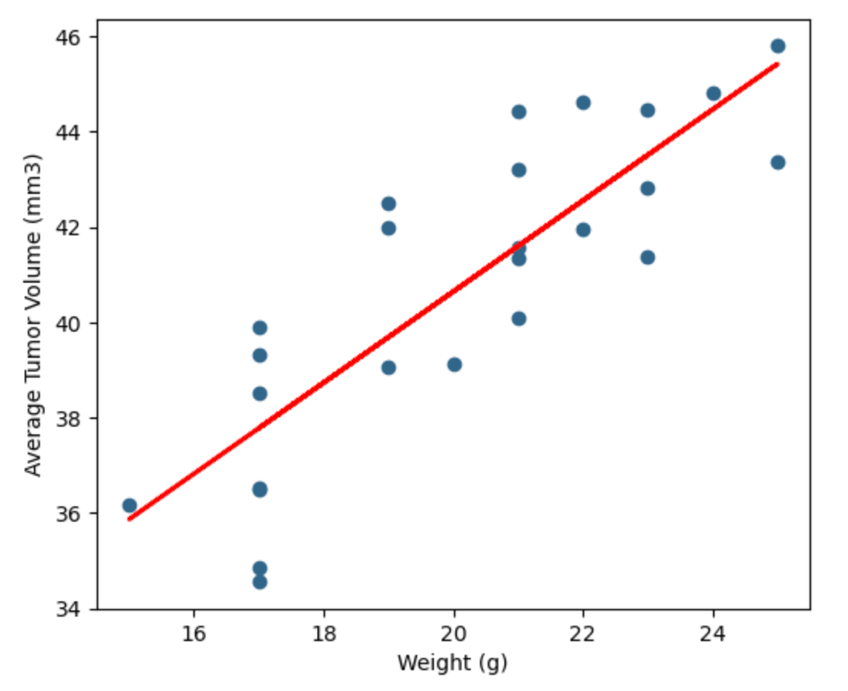


In this chart we are going to analyze only Campulin treatment on a specific mouse I509, in this line plot, we see that that the tumor volume starts at 45 and in the first 20 days increases at 48, but in the next days we saw a significant decrease, at the day 35 we saw the lower decrase at almost 40, but after the day 40 has an increase but not to high, it increase less than 2mm3.



In this chart we are going to analyze only the behavior of Campulin, the weight vs average tumor volume, so in this scatter plots we can see, that with more weight, more is the tumor volume.

**Correlation and Regression**

****

In this chart we can interpretate with the linear regression line that almost half of the data are positive and half are negative, so we are only analyzing Campulin, so the correlation is that more weight, more average tumor volume.

**Conclusion**

With all the Data we obtain we can say that Campulin and Ramicane have more effectivity with the Tumor volume, and Ketapril and Naftisol have the lowest effectivity in the treatment, and only analyzing Campulin with only one mouse we can observe that the treatment works and the mouse, and has a significant decrease in the first 20 days, and after the day 40 has a decrease but not the same way at the beginning, and another important data for this, is that mice that have more weight, their tumors have more volume.

**References**

Data generated by [MockarooLinks to an external site.](https://mockaroo.com/), LLC (2022). Realistic Data Generator.