# Pymaceuticals Analysis and Takeaways

Background

Pymatecuticals is a pharmaceutical company that has been conducting experiments to figure out the efficacy of drug regimens in terms of treating squamous cell carcinoma, a commonly occurring form of skin cancer. In the experiment analyzed below, 249 mice with squamous cell carcinoma were given different drug regimens, including a control group with a placebo, to compare the occurrence of metastatic sites and growth of tumors across the treatments. All mice selected had tumors that were 45 mm3 and sample sizes were 25 for each treatment, except for Ceftamin, which had 24 mice. The drug of interest in the analysis below is Capomulin.

Data Cleaning Methodology

The data used for analysis was comprised of two separate datasets – one with mouse metadata, and the other with study results. These sets were combined based on the unique mouse ID’s and five duplicate entries were dropped. Duplicate entries were defined based on unique mouse identifier and time point.

Key Takeaways

**Capomulin and Ramicane should have further experimentation.** Both Capomulin and Ramicane have lower variance in terms of tumor volume and lower standard errors. The lower variance suggests that the results for the drug treatments are more consistent and might be more applicable for a general population. The lower standard error suggests that the sample of mice for Capomulin and Ramicane are more likely to be representative of the results of the larger population of people with squamous cell carcinoma.

Additionally, Capomulin and Ramicane are the only drug treatments that had average and median tumor volumes that were lower than the initial tumor volume.

A screenshot of a cell phone

Description automatically generated

If the mice weren’t given any treatment, the mortality rate would be expected to look similar to the placebo treatment:

. A screenshot of a cell phone

Description automatically generated

Ramicane and Capomulin both have lower mortality rates than the placebo, which is encouraging:

A screenshot of a cell phone

Description automatically generatedA picture containing drawing

Description automatically generated

**Ramicane and Capomulin might not be as effective for people who are heavier.**  When looking at the average tumor volume for a given weight, there are strong positive correlations for sets of mice in the Capomulin treatment and the Ramicane treatment. The correlation coefficient was 0.95 for the Capomulin mice and 0.94 for the Ramicane mice. The linear regression model generated for both treatments were descriptive with r-squared values of .90 and .89 for Capomulin and Ramicane, respectively.

**A close up of a map

Description automatically generated**

**A close up of a map

Description automatically generated**

Note that in this experiment, the weights of the mice stayed the same throughout, so it isn’t possible to draw an accurate conclusion about whether prescribing a change in diet and exercise in addition to Ramicane and Capomulin would be effective and increase success rates.

Limitations and Alternative Analysis

**Given Capomulin treatment, what’s the correlation of weight vs average tumor volume by sex?**

Since treatment impacts often vary for males vs females, it seemed like a logical next step to look at the correlation for weights and average tumor volume by sex. The correlation coefficient for females was 0.97 suggesting a strong positive correlation between weight and tumor volume. Foe male mice on Capomulin, the correlation coefficient was much lower at .80, but still suggests a positive correlation.

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

The fit of the linear regression models varies wildly for each sex. For females the r-squared value is 0.95, versus 0.65 for males. However, this variance in r-squared values can be attributed to the fact that male mice on Capomulin had less variation in their weights compared to their female counterparts.

Note that there isn’t a takeaway about the correlation between weight and average tumor volume by sex. The sample sizes are extremely small and so the confidence interval for these sets of numbers would be extremely low.

**Sample size.** While 249 mice are a large sample, the issue with this experiment is that it has ten different treatments. This sample is good for an exploratory look at what drugs to hone in on, but each treatment has around 25 mice , which doesn’t seem large enough to represent the population with high confidence considering that squamous cell carcinoma is the second most common skin cancer and one million cases are diagnosed annually in the United States alone [cite].