README

Module 5: Data Visualization

Company request by: Pymaceuticals Inc.

Data Analyst: Lisa Miller

The repository URL:

Subject: Study of pharmaceuticals effect on common skin cancer, SCC (squamous cell carcinoma).

Study Duration: 45 days

Request by company: Generate figures and tables to display technical reporting of the clinical study with a top-level summary.

Data Source: Data from the company included two DataFrame .csv files.

* Analyzation began by merging both files into one DataFrame.
* Data was rid of duplications by way of mouse ID and time points.

Summary of statistics structure of the DataFrame allowed for:

* A row of four specific drug treatments were selected by the company and said to be the most promising regimens.

[Capomulin, Ramicane, Infubinol, and Ceftamin]

* Columns display the mean, median, variance, standard deviation, and SEM of the tumor volume for each of those four regimens.

Data from the study was sorted into many charts and graphs. A short description of those visualizations is given here.

* Mouse ID and Timepoint for each drug regimen throughout the study are shown in a bar chart.
* The distribution of unique male and female mice was shown in a pie chart.
* Last, a box plot shows the distribution of the final tumor volume by treatment group of all mice. Outliers of the final tumor volume results were identified with the distinction of circles within the same regimen column box.
  + The box plot is where the quartiles are calculated.
  + Note: Only drug regimens Capomulin and Ramicane proved to have outliers.

Separately, one specific mouse given the treatment Capomulin regimen, chosen randomly (Mouse ID “y685”) showed in a line chart its tumor volume by timepoint through the entire study.

Together, the weight and tumor volume of the entire Capomulin treatment regimen were captured by scatter plot.

The final plot calculates the correlation and regression Capomulin regimen between mouse weight and average observed tumor volume.

A copy of the final plot is marked by a solid red line to show the linear regression model.

Disclosure: Only resources provided by Boot Camp Spot and/or for use of Data Analytics Boot Camp were used in this analysis.