Pymaceuticals is a new pharmaceutical company that specializes in anti-cancer medications. Recently, it began screening for potential treatments for squamous cell carcinoma (SCC), a commonly occurring form of skin cancer.

* This analysis is based on the complete data from their most recent animal study.
* In this study, 249 mice who were identified with SCC tumors received treatment with a range of drug regimens.
* Over the course of 45 days, tumor development was observed and measured.
* The purpose of this study was to compare the performance of Pymaceuticals’ drug of interest, Capomulin, against the other treatment regimens.
* The task of this python project is to generate all of the tables and figures needed for the technical report of the clinical study. Final analysis, deductions and conclusions can be found at the end of this notebook.

**Project Outline & Steps**

1. Prepare the data.
2. Generate summary statistics.
3. Create bar charts and pie charts.
4. Calculate quartiles, find outliers, and create a box plot.
5. Create a line plot and a scatter plot.
6. Calculate correlation and regression.

**Prepare the Data**

This sub-section:

1. Runs the provided package dependency and data imports, and then merges the "mouse\_metadata" and "study\_results" DataFrames into a single DataFrame.
2. Displays the number of unique mice IDs in the data, and then checks for any mouse ID with duplicate time points.
3. Displays the data associated with that mouse ID, and then creates a new DataFrame where this data is removed.
4. Displays the updated number of unique mice IDs.

**Summary Statistics**

This sub-section:

1. Creates a DataFrame of summary statistics including:
   * A row for each drug regimen and
   * A column for each of the following statistics: mean, median, variance, standard deviation, and SEM of the tumor volume.

**Bar Charts and Pie Charts**

This sub-section:

1. Generates two bar charts which are identical and show the total number of time points for all mice tested for each drug regimen throughout the study.
   * 1st bar chart: the Pandas DataFrame.plot() method.
   * 2nd bar chart: Matplotlib's pyplot methods.
2. Generates two pie charts which are identical and show the distribution of female versus male mice in the study.
   * 1st pie chart: Pandas DataFrame.plot() method.
   * 2nd pie chart: Matplotlib's pyplot methods.

**Quartiles, Outliers, Box Plots**

This Sub-Section:

1. Calculates the final tumor volume of each mouse across four of the most promising treatment regimens: Capomulin, Ramicane, Infubinol, and Ceftamin.
2. Then, calculates the quartiles and IQR, and determines if there are any potential outliers across all four treatment regimens by using the following substeps:
   * Creating a grouped DataFrame that shows the last (greatest) time point for each mouse.
   * Merging this grouped DataFrame with the original cleaned DataFrame.
   * Creating a list that holds the treatment names as well as a second, empty list to hold the tumor volume data.
   * Looping through each drug in the treatment list, locating the rows in the merged DataFrame that correspond to each treatment.
   * Appending the resulting final tumor volumes for each drug to the empty list.
   * Determining outliers by using the upper and lower bounds, and then printing the results.
3. Using Matplotlib, generates a box plot that shows the distribution of the final tumor volume for all the mice in each treatment group. Highlights any potential outliers in the plot by changing their color and style.

**Line Plot and Scatter Plot**

This Sub-Section:

1. Selects a mouse that was treated with Capomulin, and generates a line plot of tumor volume versus time point for that mouse.
2. Generates a scatter plot of tumor volume versus mouse weight for the Capomulin treatment regimen.

**Correlation and Regression**

This Sub-Section:

1. Calculates the correlation coefficient and linear regression model between mouse weight and average tumor volume for the Capomulin treatment.
2. Plots the linear regression model on top of the previous scatter plot.

**Analysis**

This project saw the analysis of test results involving 249 mice divided in 10 drug treatments, including a placebo/control test regimen based on specific parameters help determine trends, and patterns as well as to draw conclusions that can be used to make informed decisions about anti-cancer medications.

Conclusions drawn from the thorough analyses applied to the complete data from Pymaceuticals' most recent animal study suggests:

1. There is a **positive correlation** between the weight of the mice and the tumor size with a **correlation value of 0.84**.
2. Of the 4 drug regimens analyized, **Capomulin** and **Ramicane** came up top as the most effective at reducing tumor sizes.
3. Of the 4 drug regimens analyzed, the least effective were the bottom two drugs **Infubinol** and **Ceftamin**.