Pymaceutical Study – Tumour Treatment in Mice

Introduction

As a component of the Data Analytics Bootcamp at the University of Western Australia, a comprehensive analytical review was undertaken concerning pharmaceutical studies on tumour treatments in mice. The dataset encompasses mouse metadata as well as the results of the study. This report articulates the methodologies adopted for data analysis, deliberates on the emergent findings, and synthesises conclusions and recommendations to potentially propel future research endeavours.

Methodology

The challenge provided the desired analysis methodology using Pandas and Matplotlib libraries in Python, to perform the following tasks:

1. Data Preparation
2. Summary Statistics
3. Pie and Pie Charts
4. Quartiles, Outliers and Boxplots
5. Line and Scatter Plots
6. Correlation and Regression

In addition to this, the following analyses were performed:

- **Average Tumour Volume by Drug:** Calculate the mean tumour volume for each drug regimen at the latest timepoint available to see which drugs are associated with the smallest tumour sizes.
- **Metastatic Spread:** Calculate the average number of metastatic sites for each drug regimen to evaluate how well the drug prevents cancer spread.

Results

The dataset underpinning the pharmaceutical study presents a rich array of attributes that facilitate a comprehensive examination of tumour treatments in mice. Each data point is associated with a unique mouse, identified by a 'Mouse ID'. The attributes encompass:

- **Mouse ID:** A distinct identifier allocated to each mouse participating in the study.
- **Timepoint:** The specific stages of the treatment, quantified in days, capturing the temporal aspect of the study.

- **Tumour Volume (mm^3):** The measured size of the tumour, serving as a primary indicator of treatment efficacy.
- **Metastatic Sites:** The number of detected metastases, which provides information on the cancer's spread.
- **Drug Regimen:** Denotes the specific drug administered, with regimens including Capomulin, Ketapril, Infubinol, Naftisol, and others.
- **Sex:** The gender of the mouse, allowing for gender-specific efficacy analysis.
- **Age_months:** Represents the age of the mouse in months, which may correlate with treatment response.
- **Weight (g):** The mouse's weight in grams, a factor that could potentially affect drug dosage and efficacy.

The dataset outlines a variety of drug regimens, enabling a comparison across a spectrum of pharmaceutical treatments regarding their impact on tumour growth and metastatic progression. Notably, all mice begin the study with an initial tumour volume set at 45.0 mm^3 at the zero timepoint, establishing a uniform baseline for subsequent observations.

Sex distribution within the dataset is balanced, encompassing both male and female subjects, thereby providing a gender-inclusive assessment of the drug regimens' effectiveness. Additionally, the dataset exhibits variation in both age and weight among the mice, which introduces the potential for stratified analysis to understand how these factors may influence drug efficacy.

The data is longitudinal, chronicling multiple observations for each mouse across various timepoints. This structure is particularly conducive to longitudinal analysis, shedding light on individual and aggregate trends concerning tumour volume dynamics and the development of metastatic sites over the course of the treatment period. Such an analysis promises to yield valuable insights into the temporal patterns of tumour behavior and treatment response.

Despite being a comprehensive dataset, it is important to consider the following, as they may impact results:

- **Missing Data:** Any gaps in the dataset, such as missing timepoints or incomplete records for some mice, could skew analysis and affect the integrity of the conclusions.
- **Outliers and Anomalies:** The presence of outliers in tumour volume or metastatic sites could indicate experimental error or unique responses to treatment, requiring careful scrutiny to avoid misleading results.
- **Bias in Initial Selection:** If the selection of mice was not random or lacked diversity in terms of genetic background, age, or pre-existing health conditions, it could introduce bias.
- **Standardisation of Conditions:** Ensuring that all mice are subject to identical conditions is crucial. Variations in environment, diet, or handling could affect the results.

- **Dosing Differences:** If the dosage of drugs was not strictly controlled or varied among mice, it would be difficult to compare the effects of the drug regimens accurately.
- **Longitudinal Consistency:** Maintaining consistency in measurement techniques and intervals over time is challenging but necessary for reliable longitudinal analysis.
- **Interaction Effects:** Potential interactions between the mouse's characteristics (such as weight and age) and the drug's efficacy are difficult to isolate and interpret.

The results obtained are as follows:

Table 1. The Overall Summary of Statistics for the Study Results.

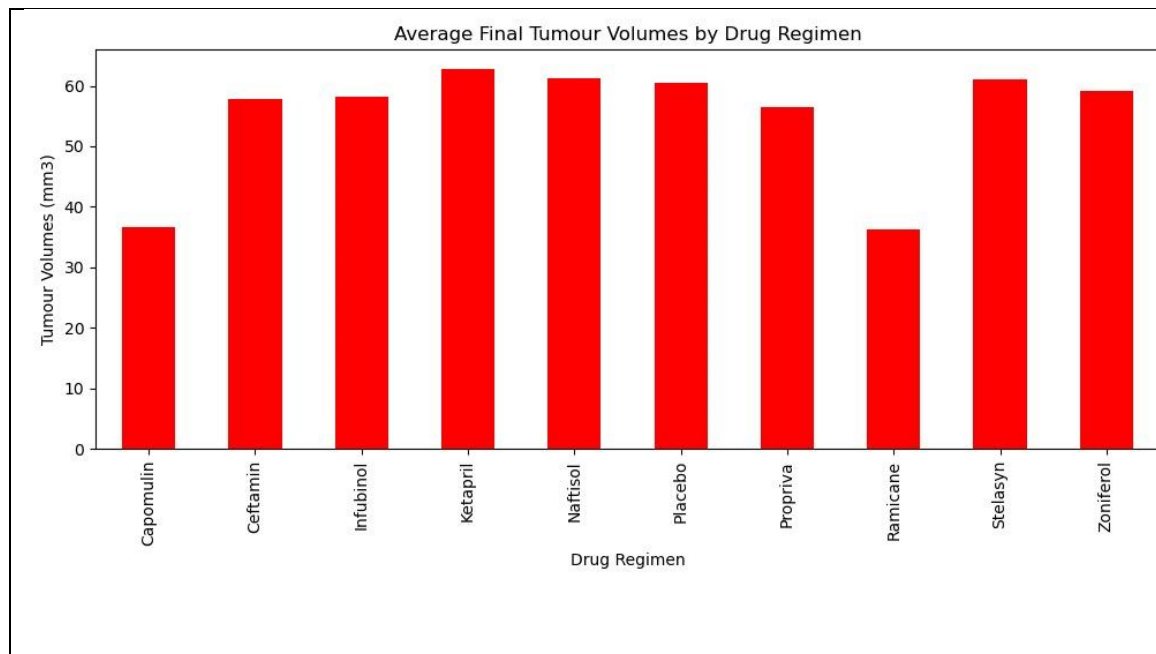| Tumour Volume (mm3)					
Drug Regimen	Mean	Median	Variance	Std_Dev	SEM
Capomulin	40.675741	41.557809	24.947764	4.994774	0.329346
Ceftamin	52.591172	51.776157	39.290177	6.268188	0.469821
Infubinol	52.884795	51.820584	43.128684	6.567243	0.492236
Ketapril	55.235638	53.698743	68.553577	8.279709	0.603860
Naftisol	54.331565	52.509285	66.173479	8.134708	0.596466
Placebo	54.033581	52.288934	61.168083	7.821003	0.581331
Propriva	52.320930	50.446266	43.852013	6.622085	0.544332
Ramicane	40.216745	40.673236	23.486704	4.846308	0.320955
Stelasyn	54.233149	52.431737	59.450562	7.710419	0.573111
Zoniferol	53.236507	51.818479	48.533355	6.966589	0.516398

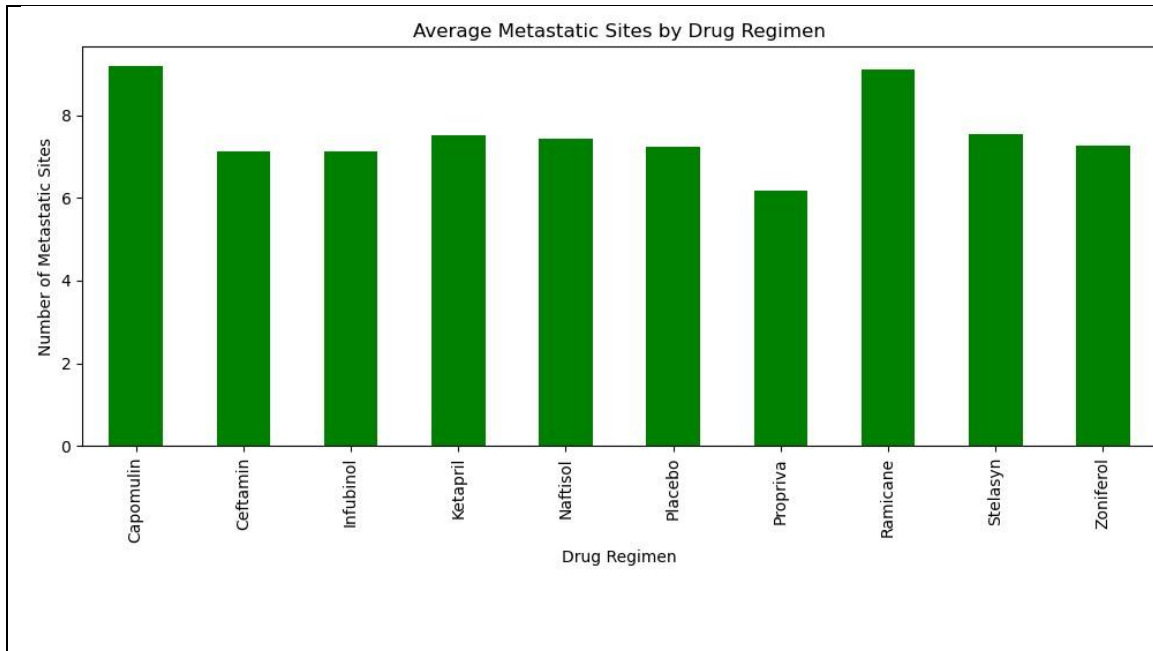**Figure 1.** Average Final Tumour Volumes by Drug Regimen.

Figure 2. Average Metastatic Site Count by Drug Regimen.

Discussion

The following observations are drawn from the analysis results presented above:

- Capomulin and Ramicane show the most promising results, with relatively low average tumour volumes (38.0 mm^3 and 39.0 mm^3 respectively) and moderate control over metastatic spread (1 metastatic site each).
- Infubinol has a higher average tumour volume but interestingly shows no new metastatic sites at the 20-day mark, indicating effective control of cancer spread.
- Ketapril and Naftisol exhibit less efficacy in tumour volume control and a higher number of metastatic sites, suggesting they are less effective compared to Capomulin and Ramicane.

Recommendations

Based on the observations above, it may be prudent to consider the following:

- Enhance monitoring and data entry protocols to reduce the occurrence of missing data, ensuring a comprehensive dataset for more accurate analysis.
- Establish a consistent protocol for identifying and handling outliers, assessing their impact on study outcomes to inform decisions regarding their inclusion.
- Strengthen randomisation procedures and consider stratification by sex, age, and weight to control for potential confounding factors that could affect the drug efficacy outcomes.
- Maintain strict control of experimental conditions to ensure that variations in tumour growth or metastasis are due to the drug treatments and not external variables.

- Apply advanced statistical methods designed for longitudinal data, like mixed-effects models, to appropriately analyse the progression of tumour growth and metastasis over time.