A division of Georgia Tech Data Science and Analytics Boot Camp

Georgia Tech Data Science Bootcamp
800 Spring St NW, Atlanta, GA 30308

Internal Memo

To: Georgia Tech Data Science Bootcamp Instructors Team
Georgia Tech Data Science Bootcamp Teaching Assistants Team

From: Ivan Escalona-Faria

Subject: Homework 1 (Report and Jupyter Notebook attachment)

Date: March 31, 2019

Distinguished instructors and teaching assistants:

As requested, I am attaching the Jupyter notebook that contains the data analysis for the cancer drug trials. As requested I analyzed 4 out of 10 drugs from the study, namely Capomulin, Infubinol, Ketapril, and the placebo.

I am also providing you with a summary of the highlights that can be derived from this data, along with some inputs for further study that could be implemented in a future study.

The drug labeled “Capomulin” resulted the best drug of the four drugs based on three key indicators: Tumor volume (mm^3), Metastatic Sites, and Survival Rates of mice subject to the clinical trials.

Tumor Volume (mm^3):

The volume of the tumor clearly shows a decrease through the timepoints by -19%, which is even more significant, considering that the Placebo, which serves as control, increased its size by 51%. This means that Capomulin not only stopped the growth of the tumor, but it also decreased it. Infubinol slowed the tumor growth by 5% compared to the placebo, Ketrapil on the other hand enhanced the growth rate of the tumor.

The error bars were small in all cases, and all timepoints. This is the result of having great accuracy at the time of measuring the tumor volume, having a relatively small standard deviation, and having enough points to show a small enough error compared to the values of the mean.

Metastatic Sites:

The metastatic sites data shows a similar overall trend for Capomulin, in which it has the least number of metastatic sites among the studied drugs. In this case the Placebo appears to develop greater amount of metastatic sites than with the other drugs.

This data might be misleading because of the big error bars on all the points, due to the small values of the mean of each point and drug. I would recommend a deeper study, probably with more mice, in order to decrease the standard error of the mean and get a better idea of the consistency of these results.

The good news is that in spite of this, the data seems consistent with the tumor volume time series, particularly for Capomulin.

Survival Rates:

Regarding survival rates, this is also good news, because it shows that Capomulin is the drug with the highest survival rate across the board throughout the timepoints with greater than 80% after 45 days.

As you know, the survival rate is not as directly correlated to the other factors (tumor volume and metastatic sites) than how they are among themselves. This means that two seemly independent parameters are working in the right direction in the betterment of the health of the test subjects.

Further steps:

I would recommend that to make the study more robust:

1) Either increase the dosage of Capomulin and study the same factors with a range of dosis.
2) Test with different types of cancerous cells to test the effectiveness of Capomulin against a variety of cancer types.
3) A closer study of the chemical structure of both the cancerous cell and the drug in an attempt to both identify improvements that could be made on the drug, or other drugs that could better the outcome obtained with Capomulin in the present study.

Please let me know of any additional requirements or follow up questions you may have.

Yours truly,

Ivan Escalona-Faria
Data Scientist - Pymaceuticals