Drug Regimens Report
For
Skin-Cancer

Written by

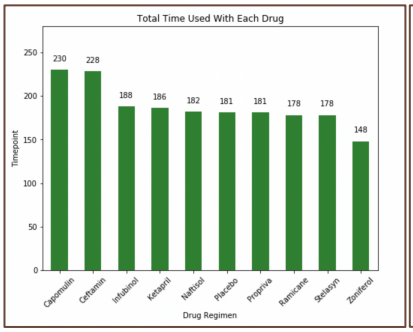
HUAN JHE LIN

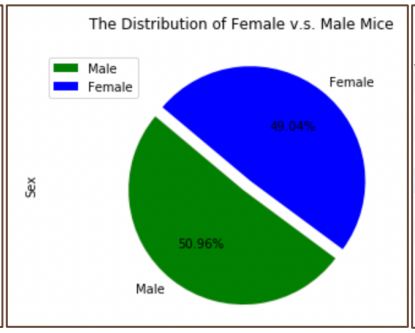
Clean Data

	Mouse ID	Drug Regimen	Sex	Age_months	Weight (g)	Timepoint	Tumour Volume (mm3)	Metastatic Sites
908	g989	Propriva	Female	21	26	0	45.000000	0
909	g989	Propriva	Female	21	26	0	45.000000	0
910	g989	Propriva	Female	21	26	5	48.786801	0
911	g989	Propriva	Female	21	26	5	47.570392	0
912	g989	Propriva	Female	21	26	10	51.745156	0
913	g989	Propriva	Female	21	26	10	49.880528	0
914	g989	Propriva	Female	21	26	15	51.325852	1
915	g989	Propriva	Female	21	26	15	53.442020	0
916	g989	Propriva	Female	21	26	20	55.326122	1
917	g989	Propriva	Female	21	26	20	54.657650	1
918	g989	Propriva	Female	21	26	25	56.045564	1
919	g989	Propriva	Female	21	26	30	59.082294	1
920	g989	Propriva	Female	21	26	35	62.570880	2

In the dataset, I found the mouse ID g989 has duplicated test results at the same time point. As I do not know which data is correct, I delete it to keep the data consistent.

Analysis





Drug Regimen	MEAN	MEDIAN	VARIANCE	STD	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

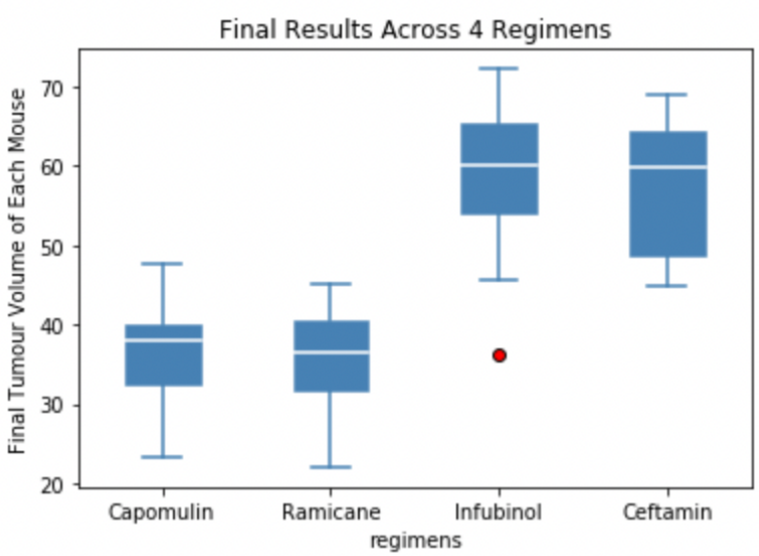
PIC LEFT: The amount of using time in each drug is quite evenly. As The amount of using time for Zoniferol is 30 less than others, I will not go deep on that.

PIC MID: The distribution of female and male mice is almost the same. Therefore, the factor of sex will not be an issue if the test drug is going to be tested in clinical experiments afterward.

PIC LEFT: That is a table to show the statistical summary for each drug regimen. Based on the calculation, I will focus on drugs with SEM lower than 0.5.

Based on Analysis section, I choose to analyse further on Capomulin, Ramicane, Infubinol, and Ceftamin.

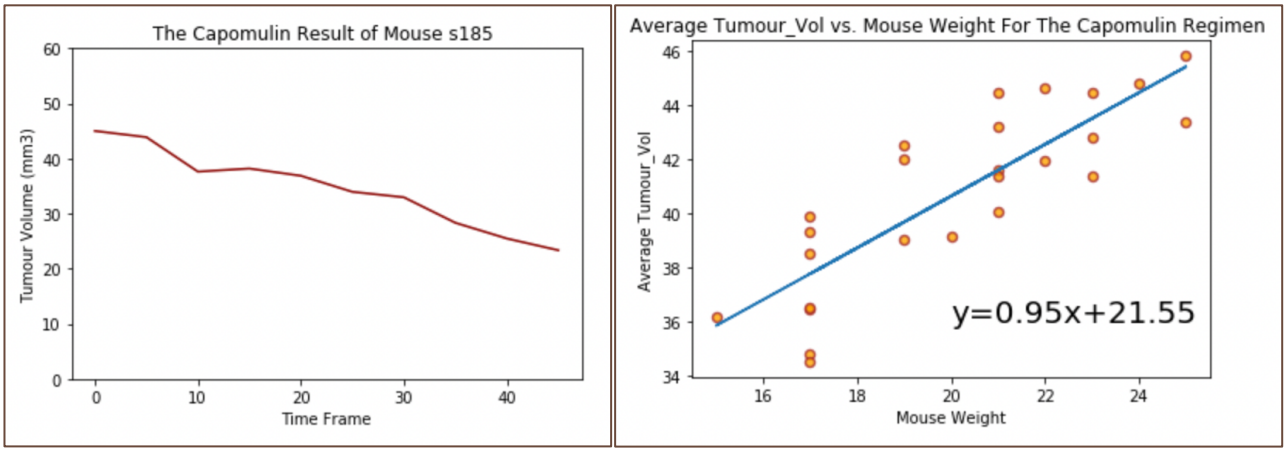
Advanced Analysis



The column of median in the PIC LEFT (data for whole timepoints), Capomulin is 40.67, Ramicane is 40.21, Infubinol is 52.88, and Ceftamin is 52.59. Compared to the above picture of the final result of four drugs usage, we can find out that the median of Infubinol and Ceftamin are increase from around 52 to 60, while the median of Capomulin and Ranicane are decrease from around 40 to 37.

Furthermore, compared to mice which were treated with Capomulin, mice which were treated with Ramicane have better results on tumour size. However, the distribution of the latter's final result is a negatively skewed distribution. It means we need to get more testing data of Ramicane, then the distribution of the new dataset might be close to natura distribution. The LEFT PIC is a proof of my conclusion about the testing group of Ramicane as group Capomulin has 230 timepoints and group Ramicane only has 178.

In this experiment of ten anti-cancer drugs, each drug had been used with similar test time and similar number of mice. At this stage, in my opinion, Capomulin is relatively the best anti-cancer pharmaceutical for squamous cell carcinoma (SCC).



PIC LEFT: The picture is showing the Capomulin testing result of Mouse S185. We can see that the tumour size in S185 shrinks from around 48 to 25. It is almost half the size compared to the first day.

PIC RIGHT: That is a scatter plot with a regression line. It shows that mouse weight is a positive factor for the size of tumour in testers' bodies. Based on the regression line, it might help us in deciding how many doses we should use on our patients.

Summary

1. Statistically, the dataset is too small as each group (based on medicine) only has around 200 samples. If each group of population is only that much, then the standard deviation and standard error might be too large. In another word, the analysis based on this dataset might have test error.
2. Due to finding out that the dataset contains duplicate data and the number of test times scientists using in each drug group is not the same, I recommend to my client that they have to control variables carefully in the future experiments because control accuracy can increase the data credibility.
3. Based on this dataset, Capomulin is relatively the best anti-cancer pharmaceutical for squamous cell carcinoma (SCC). However, if our client is willing to do extra tests on the Ramicane group. Ramicane could be the best drug as its performance on curing cancer is better than Capomulin's.