Pymaceuticals

January 19, 2018

1 Pymacueticals

1.1 Analysis

Observed Trend 1

- Capomulin outformed the Infubinol, Ketapril, and the placebo in tumor volume reduction, survival rate, and lower metastatic site occurences. It was the only treatment out of the 4 reported that saw average tumor size consistently decrease over the 45 day treatment and averagen tumor size decrease overall. The average number of metastatic sites was nearly 1.75 sites lower than the Placebo group and the Ketapril group. Capomulin also saw significantly higher survival rates over the course of the 45 days. Capomulin were the only group with a final survival rate of over 50% and a survival rate ~40% higher than any other group.

Observed Trend 2

- Ketapril seems to have a positive affect on tumor volume with the tumors seeing slightly greater average growth over the 45 day time span than the placebo group. The Ketapril group saw an average growth of about 5.7% higher than the Placebo group. This may or may not be statistically significant. The Ketapril group saw a slower average increase in metastatic sites than the Placebo group during the course of the 45 days, but overall the number of metatastic sites were about the same. Ketapril's surival rates are comparable to the placebo at the end of the 45 day treatment.

Observed Trend 3

- The average Infubinal group tumor growth was similar to that of the Placebo group, but the average increase metastatic sites was lower over the course of treatment and overall. This group still experienced average growth in metastatic sites higher than Capomulin. The Infubinal group, however, experienced the lowest surival rates of the two other treatments and the Placebo.

```
In [71]: # Dependencies
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns
         import os
```

```
In [72]: # name and read files

         data_file1 = os.path.join('Resources', 'clinicaltrial_data.csv')
         data_file2 = os.path.join('Resources', 'mouse_drug_data.csv')

         trial_data = pd.read_csv(data_file1)
         mouse_data = pd.read_csv(data_file2)

         merged_df = trial_data.merge(mouse_data, on = 'Mouse ID')
         #trial_data.head()
         #mouse_data.head()
         #merged_df.head()
```

```
In [73]: # Mouse ID g989 is listed with two different drugs but all other records are exact du
         # all of those records were excluded, this makes no difference as the Mouse excluded
         # as one of the four drugs in this analysis

         merged_df = merged_df[merged_df['Mouse ID'] != 'g989']

         #keep only the drugs we want to examine
         merged_df = merged_df[(merged_df["Drug"] == 'Capomulin') |
                               (merged_df["Drug"] == 'Infubinol') |
                               (merged_df["Drug"] == 'Ketapril') |
                               (merged_df["Drug"] == 'Placebo')
                              ]
```

```
In [74]: merged_df.head()
```

```
Out[74]:   Mouse ID  Timepoint  Tumor Volume (mm3)  Metastatic Sites       Drug
         0     b128          0           45.000000                 0  Capomulin
         1     b128          5           45.651331                 0  Capomulin
         2     b128         10           43.270852                 0  Capomulin
         3     b128         15           43.784893                 0  Capomulin
         4     b128         20           42.731552                 0  Capomulin
```

1.2 Tumor Volume Changes During Treatment

```
In [75]: #group by treatment for change in average tumor size calculations
         avg_volume = pd.DataFrame(merged_df.groupby(['Drug', 'Timepoint']).mean()['Tumor Volu
         volume_sem = pd.DataFrame(merged_df.groupby(['Drug', 'Timepoint']).sem()['Tumor Volume
```

```
In [76]: #reshaping of avg tumor volume data
         avg_volume = avg_volume.unstack(level = 0)
         avg_volume.columns = avg_volume.columns.get_level_values(1)
         avg_volume
```

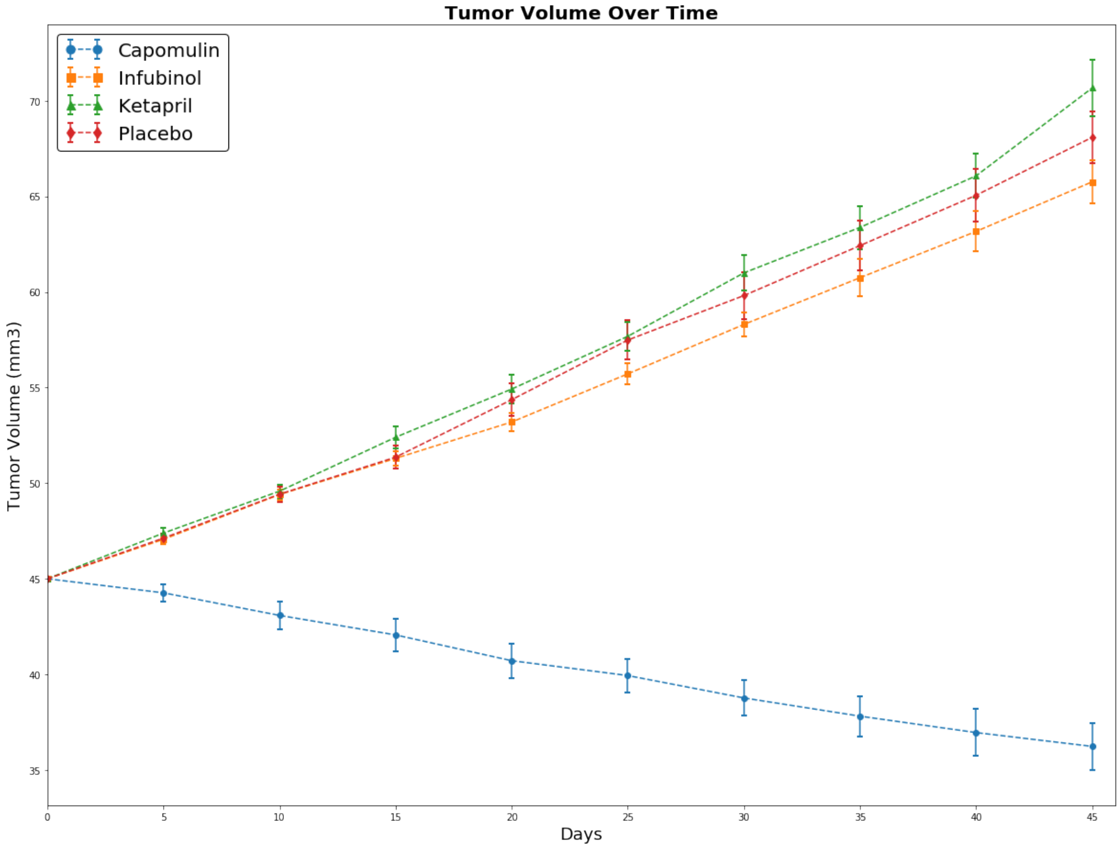
```
Out[76]: Drug       Capomulin  Infubinol   Ketapril    Placebo
         Timepoint
         0          45.000000  45.000000  45.000000  45.000000
```

```
5          44.266086  47.062001  47.389175  47.125589
10         43.084291  49.403909  49.582269  49.423329
15         42.064317  51.296397  52.399974  51.359742
20         40.716325  53.197691  54.920935  54.364417
25         39.939528  55.715252  57.678982  57.482574
30         38.769339  58.299397  60.994507  59.809063
35         37.816839  60.742461  63.371686  62.420615
40         36.958001  63.162824  66.068580  65.052675
45         36.236114  65.755562  70.662958  68.084082
```

```
In [77]: #reshaping of avg tumor volume data - std error
         volume_sem = volume_sem.unstack(level=0)
         volume_sem.columns = volume_sem.columns.get_level_values(1)
         volume_sem
```

```
Out[77]: Drug       Capomulin  Infubinol  Ketapril   Placebo
         Timepoint
         0           0.000000   0.000000  0.000000  0.000000
         5           0.448593   0.235102  0.264819  0.218091
         10          0.702684   0.282346  0.357421  0.402064
         15          0.838617   0.357705  0.580268  0.614461
         20          0.909731   0.476210  0.726484  0.839609
         25          0.881642   0.550315  0.755413  1.034872
         30          0.934460   0.631061  0.934121  1.218231
         35          1.052241   0.984155  1.127867  1.287481
         40          1.223608   1.055220  1.158449  1.370634
         45          1.223977   1.144427  1.453186  1.351726
```

```
In [78]: #plot options
         plt.figure(figsize = (20,15))
         plt.title('Tumor Volume Over Time', fontdict = {'fontsize': 20, 'fontweight': 'bold'}
         plt.xlabel('Days', fontdict = {'fontsize': 18})
         plt.ylabel('Tumor Volume (mm3)', fontdict = {'fontsize': 18})
         plt.xticks(np.arange(0, avg_volume.index.max()+3 , 5)) # location of separators for t
         plt.xlim(0, avg_volume.index.max() + 1)

         #markers used throughout
         markers = ['o', 's', '^', 'd']

         # independent variable
         xvals = avg_volume.index #timepoints

         #count for markers in plot loop
         count = 0

         #plot each tumor volume column with standard error
         for c in avg_volume.columns:
             plt.errorbar(xvals,
                          avg_volume[c],
                          volume_sem[c], #std error
                          linestyle = '--',
                          marker =  markers[count],
                          capthick = 2, #for caps on error bars
                          capsize = 3)  # for caps on error bars
             count += 1

         #legend options
         lg = plt.legend(numpoints = 2, # gives two symbols in legend
                         frameon = True,
                         markerscale = 1.5,
                         edgecolor = 'black',
                         fontsize = '20',
                         framealpha = 1)
         plt.show()
```



1.3 Metastatic Data Changes During Treatment

```
In [79]: #group by treatment for metastatic site counts and std error
         avg_num_spread = pd.DataFrame(merged_df.groupby(['Drug', 'Timepoint']).mean()['Metast
avg_spread_err = pd.DataFrame(merged_df.groupby(['Drug', 'Timepoint']).sem()['Metasta
```

In [80]:
```
#reshaping of metastatic site data
avg_num_spread = avg_num_spread.unstack(level = 0)
avg_num_spread.columns = avg_num_spread.columns.get_level_values(level = 1)
avg_num_spread
```

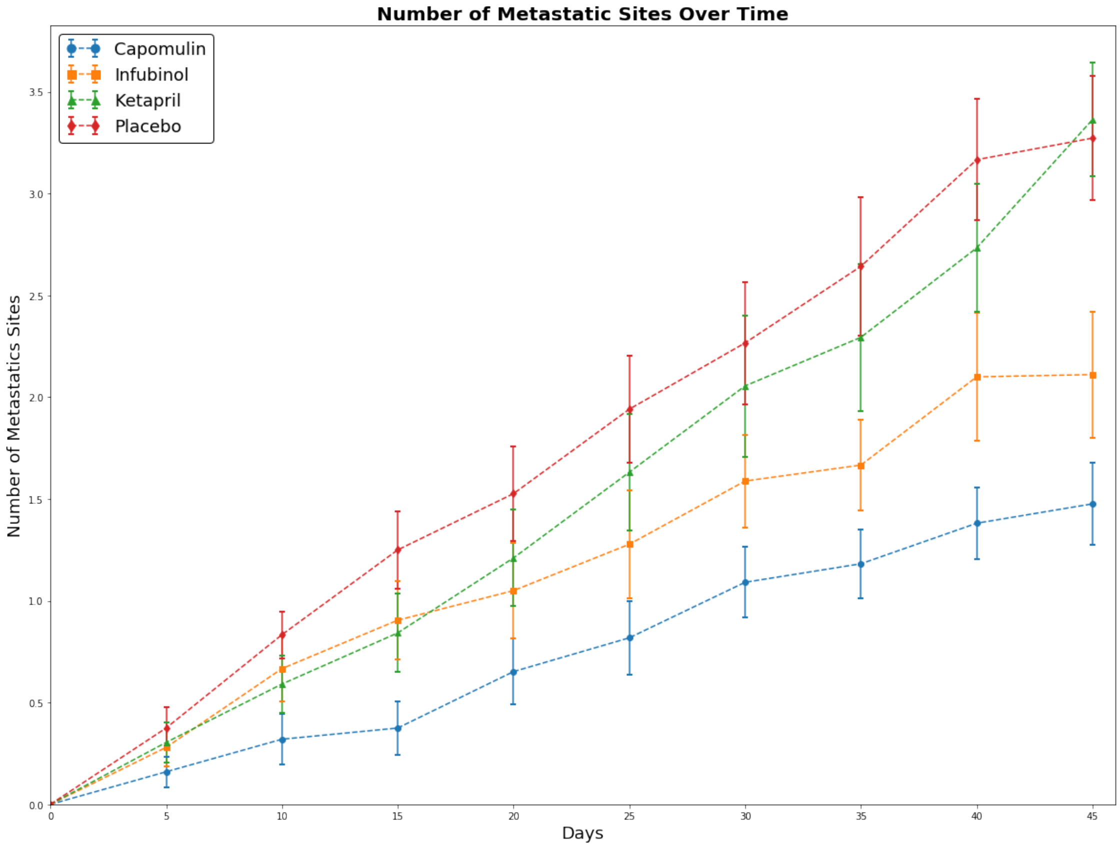
Out[80]:
```
Drug       Capomulin  Infubinol  Ketapril   Placebo
Timepoint
0           0.000000   0.000000  0.000000  0.000000
5           0.160000   0.280000  0.304348  0.375000
10          0.320000   0.666667  0.590909  0.833333
15          0.375000   0.904762  0.842105  1.250000
20          0.652174   1.050000  1.210526  1.526316
25          0.818182   1.277778  1.631579  1.941176
30          1.090909   1.588235  2.055556  2.266667
35          1.181818   1.666667  2.294118  2.642857
40          1.380952   2.100000  2.733333  3.166667
45          1.476190   2.111111  3.363636  3.272727
```

In [81]:
```
#reshaping of metastatic data - std error
avg_spread_err = avg_spread_err.unstack(level = 0)
avg_spread_err.columns = avg_spread_err.columns.get_level_values(level = 1)
avg_spread_err
```

Out[81]:
```
Drug       Capomulin  Infubinol  Ketapril   Placebo
Timepoint
0           0.000000   0.000000  0.000000  0.000000
5           0.074833   0.091652  0.098100  0.100947
10          0.125433   0.159364  0.142018  0.115261
15          0.132048   0.194015  0.191381  0.190221
20          0.161621   0.234801  0.236680  0.234064
25          0.181818   0.265753  0.288275  0.263888
30          0.172944   0.227823  0.347467  0.300264
35          0.169496   0.224733  0.361418  0.341412
40          0.175610   0.314466  0.315725  0.297294
45          0.202591   0.309320  0.278722  0.304240
```

In [82]:
```
#Plot options

plt.figure(figsize = (20,15))
plt.title('Number of Metastatic Sites Over Time', fontdict = {'fontsize': 20, 'fontwe
plt.xlabel('Days', fontdict = {'fontsize': 18})
plt.ylabel('Number of Metastatics Sites', fontdict = {'fontsize': 18})
plt.xticks(np.arange(0,avg_num_spread.index.max() + 3 ,5))
plt.xlim(0, avg_num_spread.index.max()+1)
plt.ylim(0, avg_num_spread.max().max() + avg_spread_err.max().max() + .1)
```

```
#creates plot for each treatment, using different markers for each

count = 0
xvals = avg_num_spread.index #timepoints

for c in avg_num_spread:
    plt.errorbar(xvals,
                 avg_num_spread[c],
                 avg_spread_err[c],
                 linestyle = '--',
                 marker =  markers[count], #markers defined above
                 capthick = 2,
                 capsize = 3)
    count += 1

#legend options
lg = plt.legend(numpoints = 2,
                frameon = True,
                markerscale = 1.5,
                edgecolor = 'black',
                fontsize = '18',
                framealpha = 1)

plt.show()
```

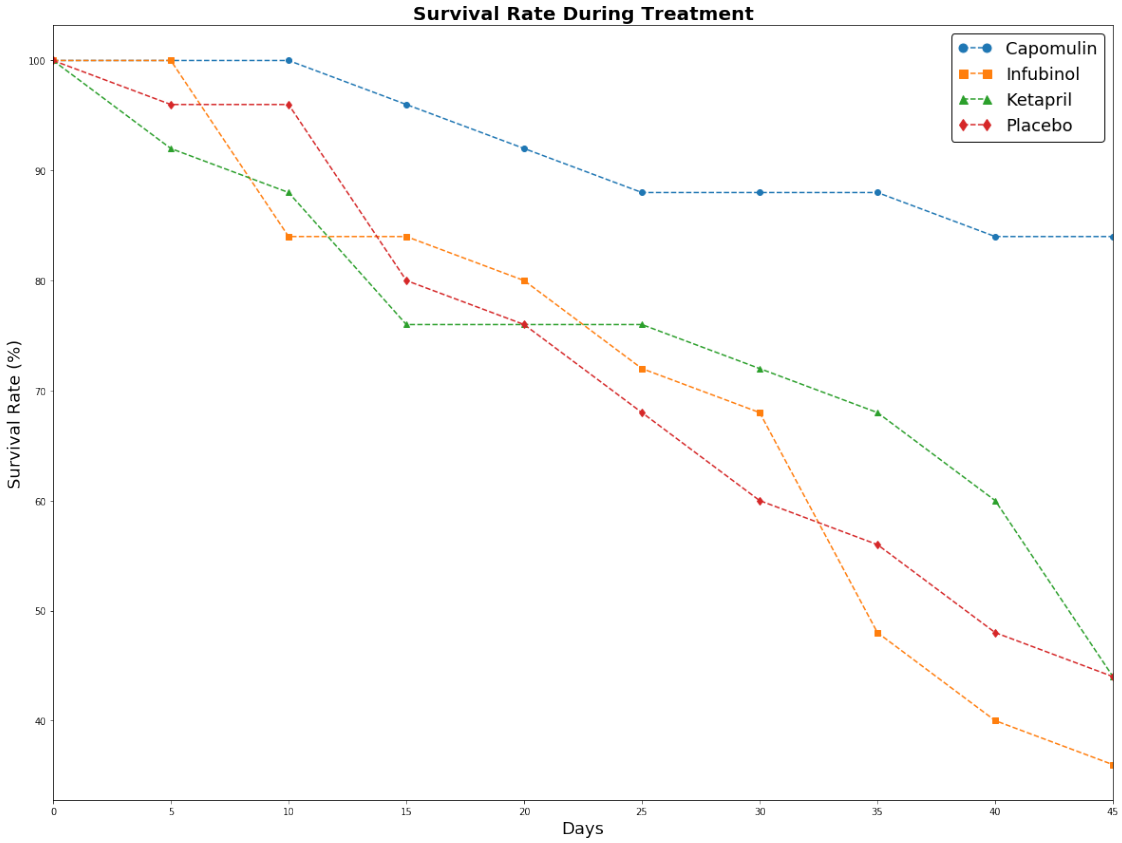


1.4 Survival Rate Changes During Treatment

```
In [83]: #gets dataframe of data counts at each time point per drug
         #this corresponds to how many mice are still in study and therefore, alive
         mice_count = merged_df.groupby(['Drug', 'Timepoint']).count()['Mouse ID']
         mice_count = mice_count.unstack(level = 0)
         mice_count
```

```
Out[83]: Drug       Capomulin  Infubinol  Ketapril  Placebo
         Timepoint
         0                 25         25        25       25
         5                 25         25        23       24
         10                25         21        22       24
         15                24         21        19       20
         20                23         20        19       19
         25                22         18        19       17
         30                22         17        18       15
         35                22         12        17       14
         40                21         10        15       12
         45                21          9        11       11
```

```
In [84]: #plot options
         plt.figure(figsize = (20,15))
         plt.title('Survival Rate During Treatment', fontdict = {'fontsize': 20, 'fontweight':
         plt.xlabel('Days', fontdict = {'fontsize': 18})
         plt.ylabel('Survival Rate (%)', fontdict = {'fontsize': 18})
         plt.xlim(0, mice_count.index.max())


         xvals = mice_count.index #timepoints

         count = 0

         for c in mice_count:
             yvals = mice_count[c]/mice_count.loc[0,c] * 100 #calculates survival rate for eac
             plt.plot(xvals,
                      yvals,
                      linestyle = '--',
                      marker =  markers[count], #markers list from above
                      )
             count += 1

         #legend options
         lg = plt.legend(numpoints = 2,
                         frameon = True,
                         markerscale = 1.5,
                         edgecolor = 'black',
                         fontsize = '18',
                         framealpha = 1)

         plt.show()
```



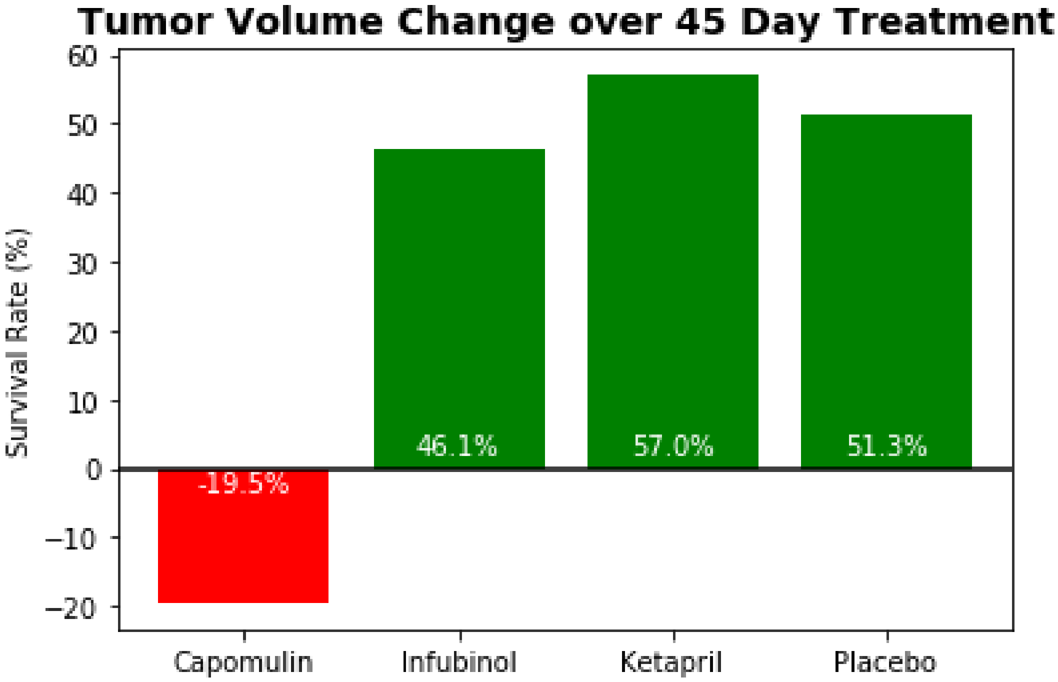
1.5 Tumor Volume % Change

```
In [85]: #calculates tumor change for each drug
         tumor_change = (avg_volume.loc[45, :] - avg_volume.loc[0, :])/avg_volume.loc[0, :] *
         tumor_change
```

```
Out[85]: Drug
         Capomulin   -19.475303
         Infubinol    46.123472
         Ketapril     57.028795
         Placebo      51.297960
         dtype: float64
```

```
In [86]: #plot options
         plt.title('Tumor Volume Change over 45 Day Treatment', fontdict = {'fontsize': 14, 'f
         plt.ylabel('Survival Rate (%)')
         plt.axhline(y=0, color = 'black') #adds a horizontal line at zero
         xlabels = tumor_change.index
         #puts drug names as labels
         plt.xticks(np.arange(len(xlabels)), xlabels)
```

```
#graphs % change of tumor as bar graph
plt.bar(np.arange(4),
        tumor_change,
        # colors each bar appropriately based on + or - values by using a conditional
        color = ['green' if tumor_change[r] > 0 else 'red' for r in np.arange(len(xlab
       )

#count for x coordinate location of value label
count = 0

#loops through tumor change data and places the value in the correct position based o
for r in tumor_change:
    if r < 0:
        y_coor = -3.5
    else:
        y_coor = 2
    plt.text(count, y_coor, str(round(r, 1)) + '%', ha = 'center', color = 'white')
    count += 1

plt.show()
```


Tumor Volume Change over 45 Day Treatment
Survival Rate (%)
Capomulin
Infubinol
Ketapril
Placebo