Analytical Model for Estimating Tumor Growth Rate

# Abstract

Cancer is the most active research topic in the field of medical science. The estimation of tumor growth rate is helpful for the doctors to study and understand the behavior of the tumor in a particular patient before proceeding for the treatment. To develop analytical models for estimating tumor growth rates from the observation of tumor volumes taken at different dates. To fit mathematical models like Gompertz law and its improvements on the temporal data set. Though the data sets are small and there are biases due to the way the samples were ascertained, useful information can still be extracted after careful study.

# INTRODUCTION

Cancer is a wide range of diseases that has in common and unusual cells proliferation of the organism itself. It is an uncontrolled proliferation that provokes the formulation of a cellular mass called Tumor. Tumor growth rates are approximately exponential which means that the rate of growth depends on developmental stage. To describe the growth rate of a tumor, it is best to describe the growth rate in terms of doubling time, or the time it takes for the population of cells – and tumor volume – to increase by 100 percent. As there is a lack of clinical data at non-symptomatic stages, it can be assumed that two to three decades can elapse between the first carcinogenic stimulus and the emergence of the neoplasm. Typical doubling times for a tumor are expected to be between 60 days for very aggressive tumors and 100 days for non-aggressive tumors.

A mathematical model is a function of time which is derived from an Ordinary Differential Equation. A number of Ordinary Differential Equation models (ODE) have been proposed to represent tumor growth and are regularly used to make predictions about the efficacy of cancer growth. Unfortunately, the choice of a growth model is often driven by ease of mathematical analysis rather than whether it provides the best model for growth of a tumor. Some researchers have attempted to find the best ODE growth model by fitting various models to a small number of experimental data sets of tumor growth. Seven ordinary differential equation (ODE) models of tumor growth (exponential, logistic, linear, surface, Gompertz) have been proposed. Among all the models, we have used Gompertz model which was originally created in 1825 to explain human mortality curves. In 1938, The model became a generalization of the logistic model with a sigmoidal curve that is asymmetrical with the point of inflection. The curve was eventually applied to model growth in size of entire organisms and more recently, was shown to provide the best fits for breast and lung cancer growth.

# DATASET

We have searched a lot of datasets and finally looked up onto the two datasets. First one is the temporal dataset of Breast Cancer and the second one is the dataset of mice which are under observations in labs.

**\*\*\*** Refer [Appendix-1](#_gqk4d0twjujy) for all datasets that we have studied.

## Breast Cancer Dataset

This temporal dataset is for Breast Cancer. The population of tumors comprising this study is 109 primary breast cancers discovered in 108 women that were undergoing screening for breast cancer at the Breast Cancer Detection and Demonstration Project at the University of Louisville.

All of the 108 women had received mammography as a screening modality and it was speculated that some of these cases could have views of the tumor on one or more mammograms before biopsy diagnosis. Fortyfive of the cancers were diagnosed on the initial screening exam so no previous mammographic record of the tumor was available. However, of the 64 remaining cancers, 32 had two or more serial mammograms on which the tumor nucleus shadow could be seen in retrospect.

The dataset is taken from the results of the demonstration project that was carried out at the University of Louisville.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Accession number** | **Date of initial volume determination** | **major axis (mm)** |  | **minor axis (mm)** | **Date of second volume determination** | **major axis (mm)** |  | **minor axis (mm)** |
| 3210 | 25/04/75 | 12 | x | 11 | 06/01/76 | 30 | x | 14 |
| 5607 | 15/10/74 | 22 | x | 17 | 15/01/75 | 27 | x | 20 |
| 2147 | 30/01/74 | 10 | x | 8 | 28/01/75 | 22 | x | 15 |
| 5264 | 14/10/74 | 7 | x | 5 | 10/07/75 | 16 | x | 8 |

**Accession number:** It is a unique patient id.

**Date of initial volume determination :** It is the date on which first observation of

tumor had taken in a patient.

**major axis (mm) :** It is the length of tumor cell aggregated with the longest side.

**minor axis (mm) :** It is the length of tumor cell aggregated with the smallest side.

**Date of second volume determination :** It is the date on which second observation of

tumor had taken in a patient.

## Mice Dataset

We have found out the mice dataset from the book Mixed Models: Theory and Applications with R Second Edition by PROFESSOR EUGENE DEMIDENKO.

The dataset has data for 28 mice and for each mice we have 7-12 reading and here the data given is time instance and the tumVolume at that instance , for simplicity we have assumed that this time instance is equal time intervals.

The dataset contains: ID(Unique), Time and TumVol.

|  |  |  |
| --- | --- | --- |
| id | Time | TumVol |
| 1 | 0 | 120.1 |
| 1 | 1 | 162.8 |
| 1 | 2 | 269.8 |
| 1 | 3 | 470.6 |
| 1 | 4 | 536.6 |
| 1 | 5 | 860.7 |
| 1 | 6 | 795.9 |
| 1 | 7 | 1013.3 |

# SOLUTION APPROACH

In Order to calculate the growth rate we need at least two time instances and the volume for the corresponding time instance , we have applied gompertz law to predict the volume as the time increases.

## Approach 1:

This approach is for the breast cancer dataset.In this dataset, Major and Minor axis observations are given for tumor size based on this axis, tried to guess tumor volume and it’s shape. Also from the Initial date and Second date, the time difference needs to be calculated.

Steps to Calculate Tumor Volume:

Let and be the minor and major radii.

Let and be the geometric mean.

Sphere :

Cylinder :

Spheroid-1 :

Spheroid-2 :

The cylinder and the first spheroid volumes differ by a constant, so the rate estimates will be the same. We do not expect drastic differences between the other three methods, so we will work with the first spheroid formula.

Hence, after volume estimation, these values can be applied to a Gompertz’s mathematical model.

**Conclusion**: Only two data points for a patient were not enough to study the growth rate behavior, so we looked out for another dataset.

## Approach 2:

In mice dataset, tumor volume and its time instance is given to us. But we have observed some of the Id’s having uncertainty in tumor volume with respect to time because of the behavior of cells grown in a laboratory setting where they always have an ample supply of nutrients is not the same as that of tumors in human body. We pre-processed the dataset and removed that id’s from our experiment.

## Gompertz’s Function

Where,

is the volume of tumor cells,

is the constant intrinsic growth of cells with ,

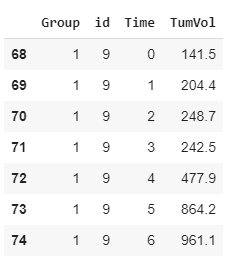
Steps followed -

1. Tumor Volume v/s Time curve was not good so we tried optimizing the value of constants(A,V0) present and also tried drawing relations between volume , rate and time.

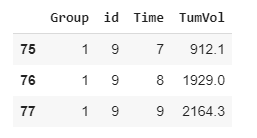
We chose curve fit function from scipy.optimize library of python to optimize the constant values and also tried drawing cardinality between k (as the carrying capacity of the tumor, that is, the maximum size that it can achieve with the available nutrients), but with no background of nutrient conditions and as observation time was relative and not absolute, and weight and age and gender not given no solid conclusions were drawn.

1. Divided the data in training and testing dataset(divided in 80-20 ratio)

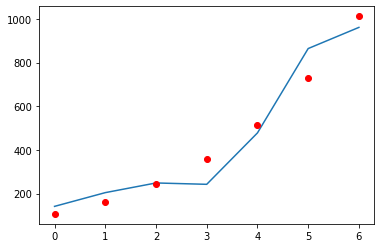
**Training Data:**

****

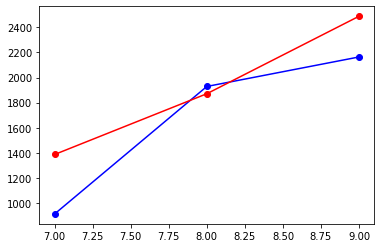
**Testing Data:**

****

**Graph for Training Data after curve\_fit**

****

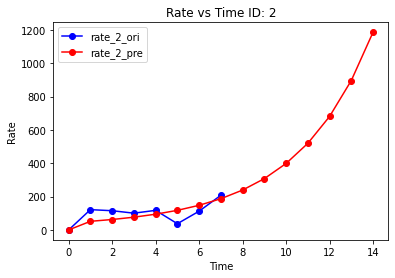
**Graph for Testing Data after Predicting Volume**

****

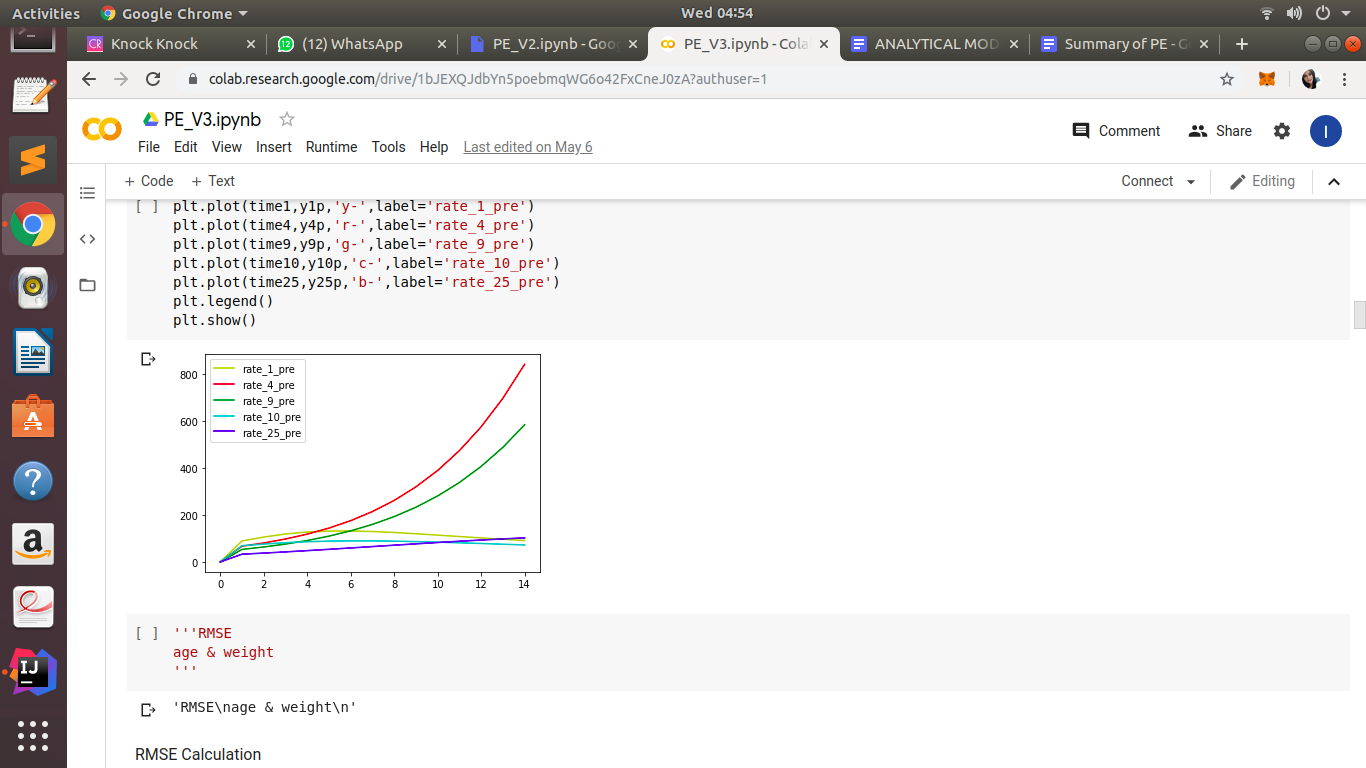
**Conclusion** - most of the ids follow gompertz law.

1. Calculated rate and plotted it against time
2. When considered wrt to previous volume -
3. When considered wrt to initial volume-

**Conclusion**- got better reading and good curve fitted graphs when plotted against initial volume



4. Plotted rate vs time for few ids on a single graph

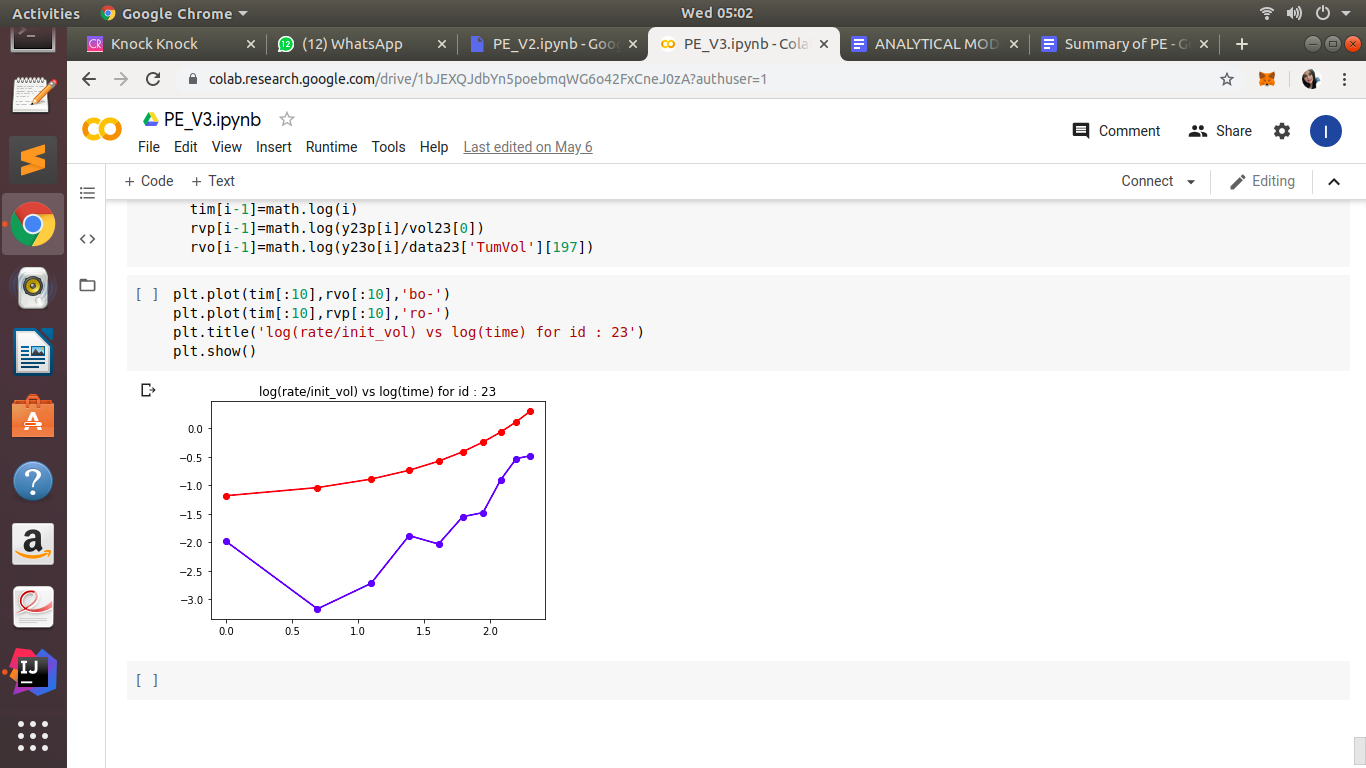


5. Plotted graph for volume v/s time and rate v/s time for all the mice

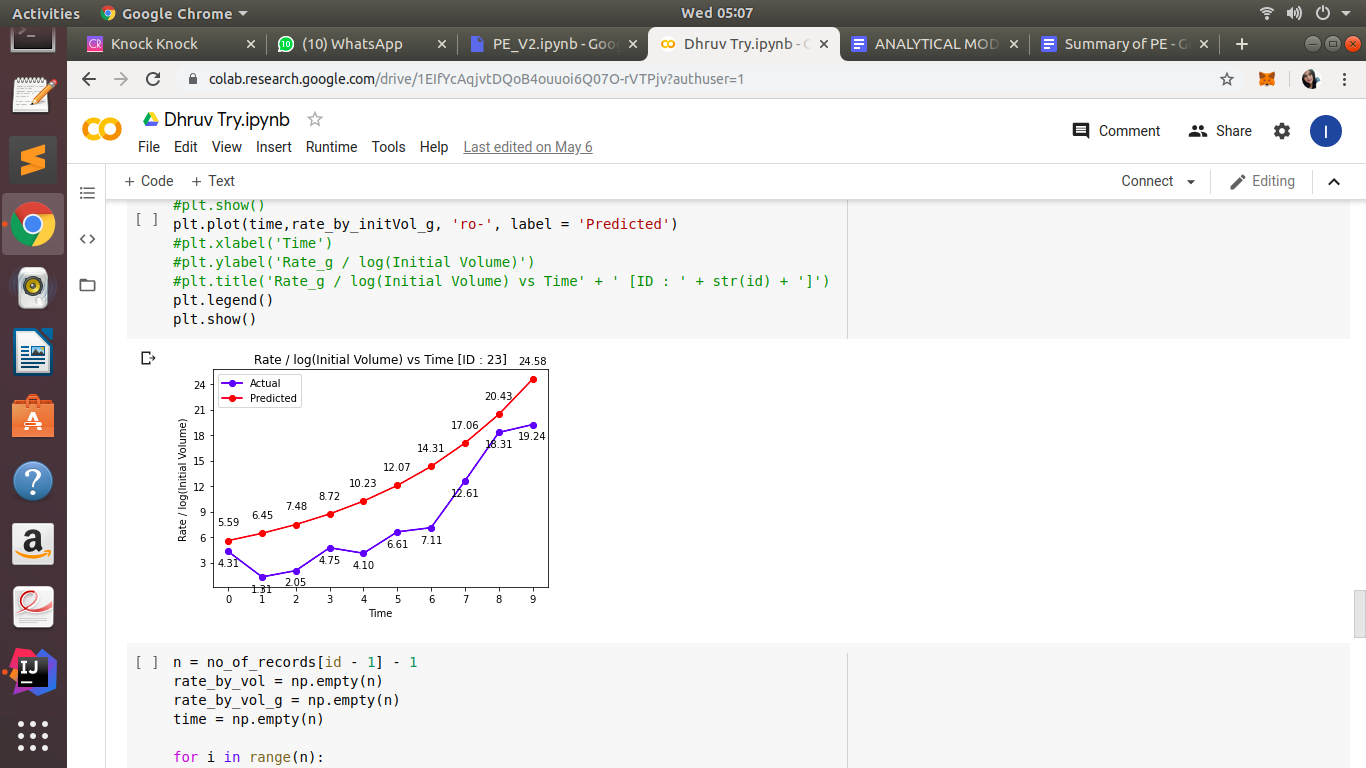
**Conclusion** - almost all followed the gompertz law equation for growth rate prediction.

6. Tried drawing cardinality between volume and time, took log of rate/vol and plotted against log of time to scale up the observation.

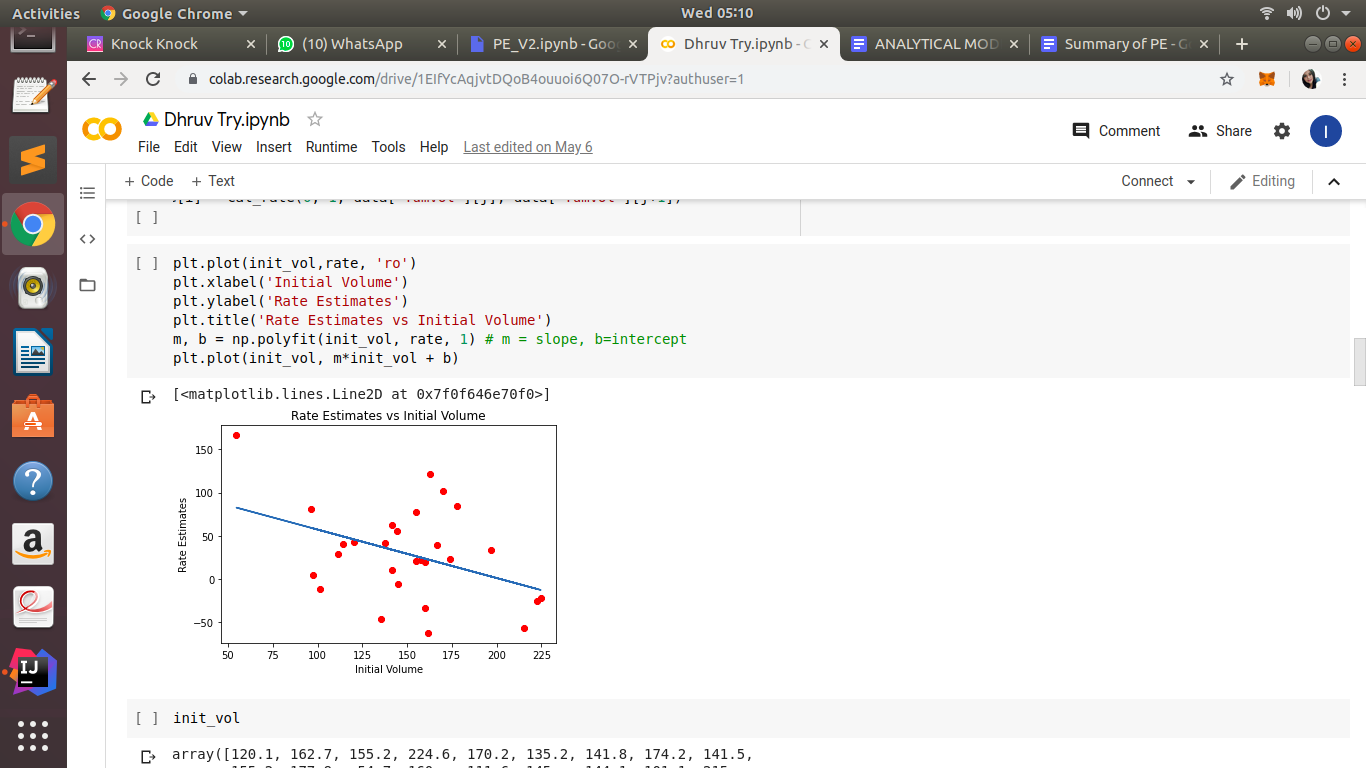
Here red line is predicted and blue is observed log(rate/init\_vol) vs log(time)



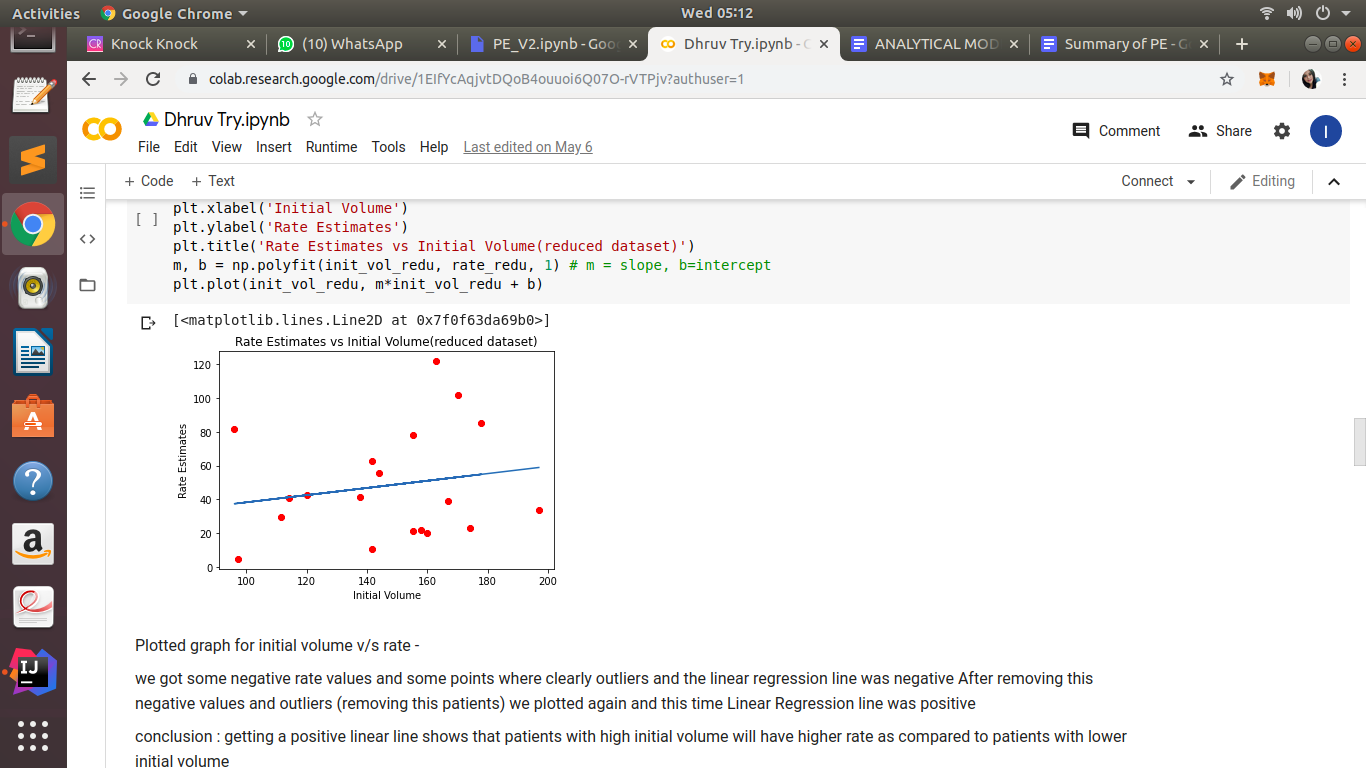
7. Graph for rate/log(vol) vs time -



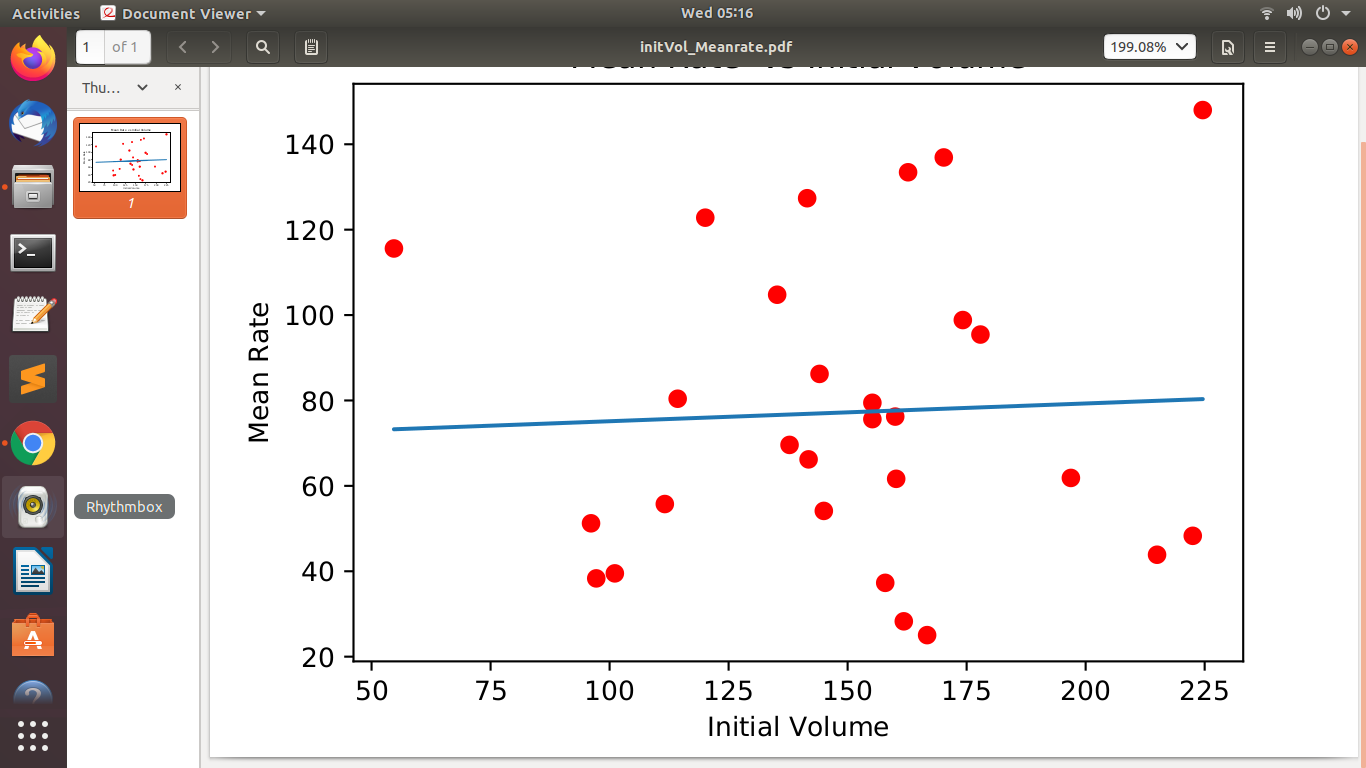
8. Tried plotting initial rate of all the patients vs initial vol



But it is sloping down but we assumed that rate increases as volume increase.Outliers are sloping it down as observed in regression line,after we remove outliers , we got graph like this-



If we consider all rate estimated reading, do we need to drop any patient? The slope of the regression line is not significantly different from 0. so no need to discard any reading for a patient.

**Graph for mean rate vs initial volume** 

**Observations**

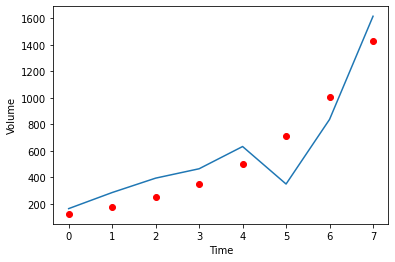
Graph of Tumor volume vs Time and Rate vs Time has plotted which follows the Gompertz curve.

**\*\*\* NOTE:** Blue Lines/Dots for Actual Data points

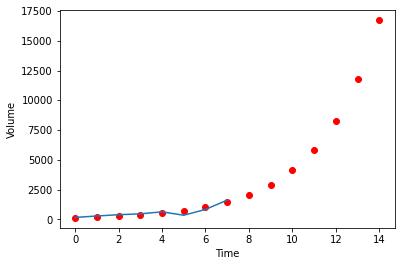
Red Lines/Dots for predicted Data points

Example taken for ID: 2

**Fitted Gompertz law to the data points:**



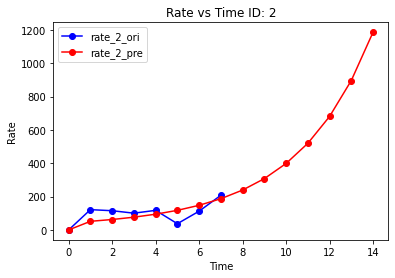
**Predicted Tumor Volume**

****

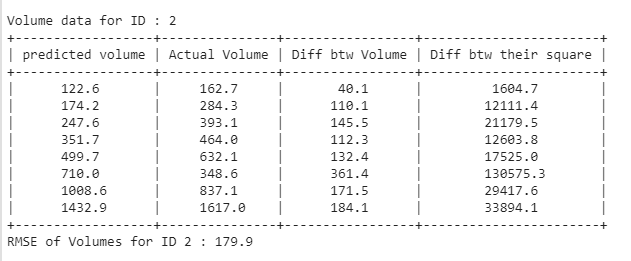
**Conclusion:** From this we can say that, on 10th day the tumor volume might be around

5000.

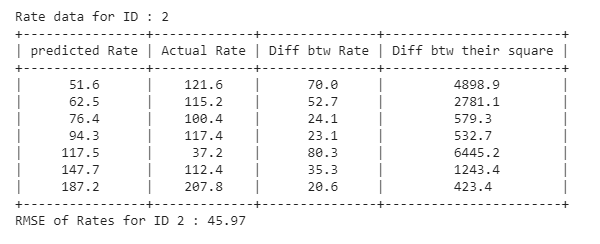
**Calculated Rate vs Time**



**Calculated RMSE value for Volume**

****

**Calculated RMSE value for Rate**



# CONCLUSION

The Gompertz equation for estimating tumor growth rate best fits the dataset and the Volume vs Time curve follows the data points. The tumor growth rate is different for different Patient IDs based upon patient’s supply of nutrients and other factors like age, gender and weight. The prediction of tumor volume can be done by fitting the curve to the data points. We can also estimate the average groth rate of tumor in a patient.

# REFERENCES

# Appendix-1

## Datasets:

# **Heuser et al (Breast cancer tumors)**

Heuser et al discovered 109 breast cancer tumors in 108 women in a screening population of 10,120 women receiving over 30,000 mammograms over three years. For each tumor they reported the size of the **major axis** “*b”* and the **minor axis** “*a”* measured in *mm*, e.g., 22 × 17.

* 45 cancers diagnosed on initial screening.
* 32 cancers with measured growth rates from serial mammograms.
* **15** cancers with evidence of their presence at initial exam but growth rates could not be determined by the method employed.
* **17** cancers without evidence of their presence at the initial exam, growth rates could not be determined by the method employed.

**E. M. Laasonen and H. Troupp (acoustic neuroma)**

Of 79 acoustic neurinomas seen between June 1980 and June 1984 by Laasonen and Troupp, These numbers include patients in whom the primary CT scan was done in another hospital. no operation was performed on 21 of these patients, 6 men and 15 women, or it was delayed for at least six months, so a second CT scan was available for these patients. The reasons for not operating were as follows: 7 patients had bilateral tumors so there was a delay on the operation for the other one, 9 patients wanted more time to decide in favor or against an operation, 4 were too old or too ill with some other disease, and 1 had a 0.38 *cm*3 tumor not diagnosed at first in another hospital. leaving **23** neurinomas for our study. Volume measurements were done with a program built into the scanner. They report the initial and final **volume in *cm*3**.

**Nakamura et al (Meningioma)**

At the Nordstate Hospital between 1978 and 2000, a total of 1954 patients seen had meningiomas and 1700 were operated on. Between 1990 and 2001, a total of 80 asymptomatic patients were diagnosed by computed tomography or MRI. Among them were 7 patients with associated neurofibromatosis Type 2, 4 patients had multiple meningiomas, 22 patients underwent surgery immediately after diagnosis, and 6 had surgery later due to significant tumor growth. Nakamura et al. examined the natural history of the remaining **41** “incidental” meningiomas, which occurred at a wide variety of different locations in the brain. Again the initial and final volumes were reported in ***cm*3**.

**Nakajima et al (liver cancer - HCC)**

Nakajima et al studied **34** hepatocellular carcinomas (HCCs) in patients who initially refused therapy, giving data, as did, on the **major** and **minor** axis. The tumors varied in their clinical stage: 18 were stage I, 14 stage II and 3 stage III, and histology: 19 were well-differentiated, 9 moderately differentiated, and 6 poorly differentiated.

**Yoshifumi Saito (liver cancer - HCC)**

Saito et al studied the tumor volume doubling times of **21** HCCs. Patients were only selected if their tumors were less than 3 *cm* in diameter at the start of observation, and two abdominal ultrasounds were available. This occurred because 3 patients refused treatment, 6 had clinical complications that prevented surgery, and in 12 cases the initial diagnosis was uncertain. They talk about the **major** and **minor** axes when they discuss the volume.

**Mice Dataset**

The mice dataset has been taken from the book Mixed Models: Theory and Applications with R Second Edition by PROFESSOR EUGENE DEMIDENKO.

The dataset has data for 28 mice and for each mice we have 7-12 readings. Here the data given has several time instances and the tumor volume at that instance. For simplicity we have assumed that this time instance is in equal time intervals.

The dataset contains:

* ID(Unique) : For each id, we have around 7-12 readings.
* TumVol : Volume of tumor at a given time instance
* Time : Time in Days on which tumor volume has measured.