**ToothGrowth - Guinea pigs**

In this dataset I will check if the guinea pig has better tooth growth with orange juice or with the ascorbic acid, does it follow the same curve or is one better than the other? And I also find a way if it is possible to predict the tooth growth without the vitamin c intake.

The data set is the length of the teeth in each guinea pig at three Vitamin C dosage levels which are 0.5, 1, and 2 mg. The intake for the guinea pigs are with two methods, one is orange juice (OJ) or ascorbic acid (VC).

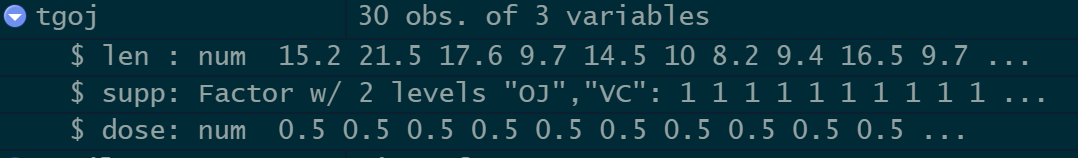
This data set contains 60 observations of three different variables,

* len = tooth length
* supp = supplement type (VC or OJ)
* dose = dose in milligrams

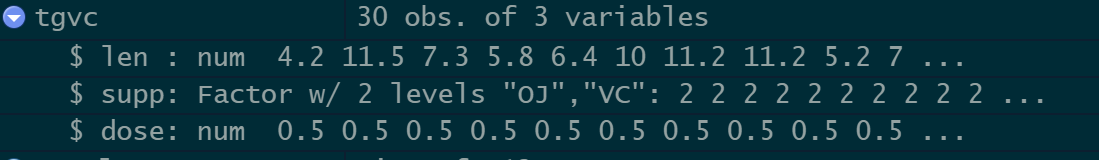
**Data overview**

With the dataset, I have started to split the data to read OJ and VC values separately as this will show us later on in this report why we have done this. With the split data, we can see that there are 30 guinea pigs receiving OJ supplement and 30 guiniea pigs with VC supplement.

*Image 1*

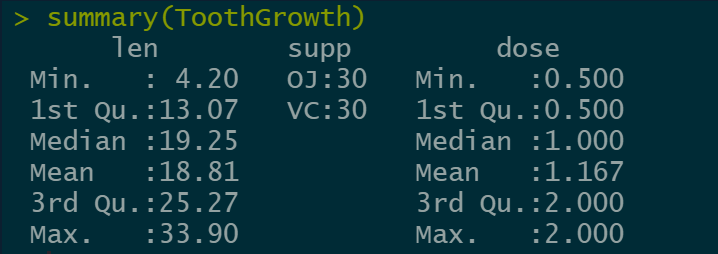


*Image 2*



Continuing, I have requested a summary of the data to find out the minimum, average and maximum values for dose intake and the length of the tooth growth.

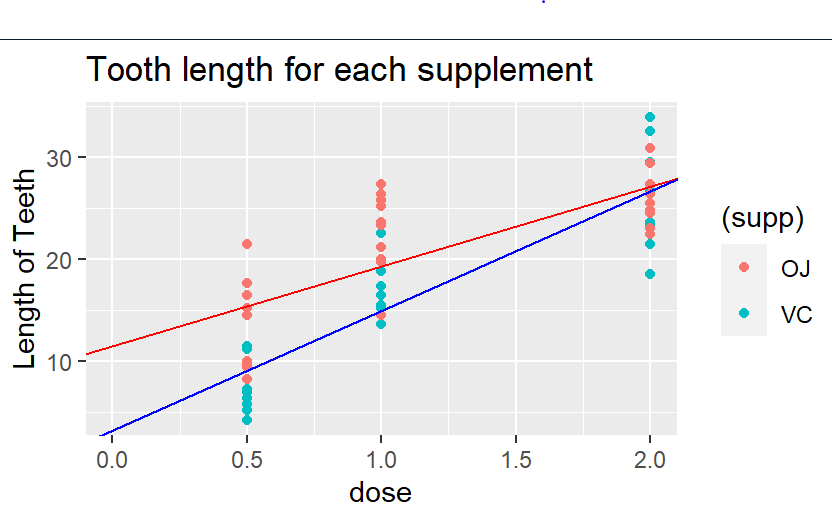
*Image 3*



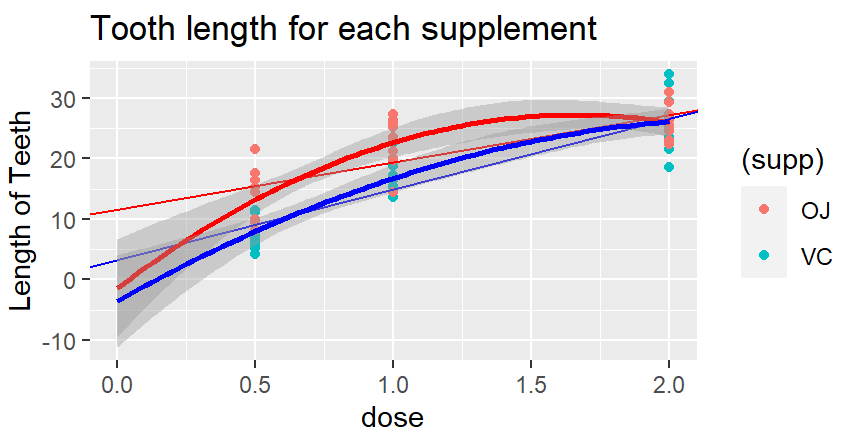
**Visualizing the data with linear and quadratic fit**

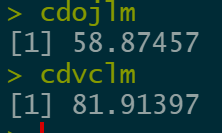
As we can see on the image below, it appears that the length of the tooth increases as the dosage increases. The lower dosages (0,5 and 1.0) with OJ supplement seem to lead to more tooth growth than the lower dosages of VC delivery.

*Image 4*



*Image 5*



However, in my opinion, choosing the linear fit to visualize and analyze the data isn’t ideal to predict what the teeth length would be without vitamin C, as the difference is due to the fact that there are errors in the interpolation. In this case, I have calculated the coefficient of determination *Image 6* ****and whilst the VC plot showed high coefficient of determination (cd) the OJ plot cd was smaller which means that the fitting is not as accurate. In order to mitigate the fit errors, I have tried the quadratic method and with this method it did converge but minus value 0 shows that these are not adequate curves either altho they did converge with less error. Hence either a better approximate or more data is required in order to have a better fit to predict the tooth growth without the supplement.

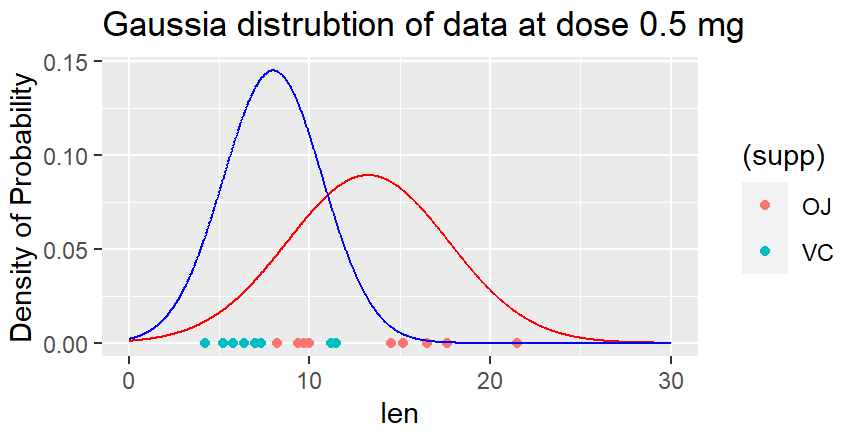
**Gaussinan distribution plots for each dose level**

For the following images, I have taken the data set and split it for each dose where I have taken the subset and splitted it into OJ and VC and calculated two sigmas to see how it spreads. With the two sigmas, I have calculated the mean and also the confidence intervals meaning what is the minimum and maximum values for which 95% the whole values fits in the distribution. If taken as a random point e.g., if x1 is 5 and x2 is 20, then you can see that 95% of all the values are between 5 and 21.

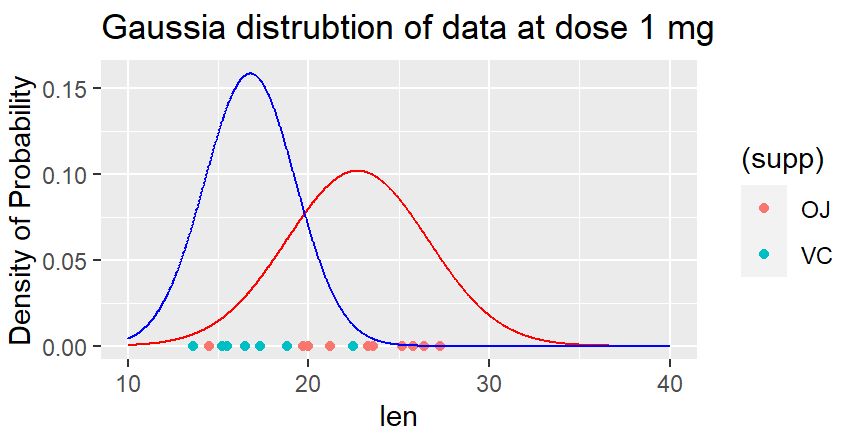
When calculating the difference, you can see which mean is bigger than the other one, e.g. at 0.5 the OJ will have higher mean meaning on average the guinea pigs teeth will grow more than if having VC as supplement.

As shown on the graph for dose 1, we can see that the same applies here where the OJ has higher mean and grows the teeth more than if they would’ve been given the VC instead.

*Image 7*

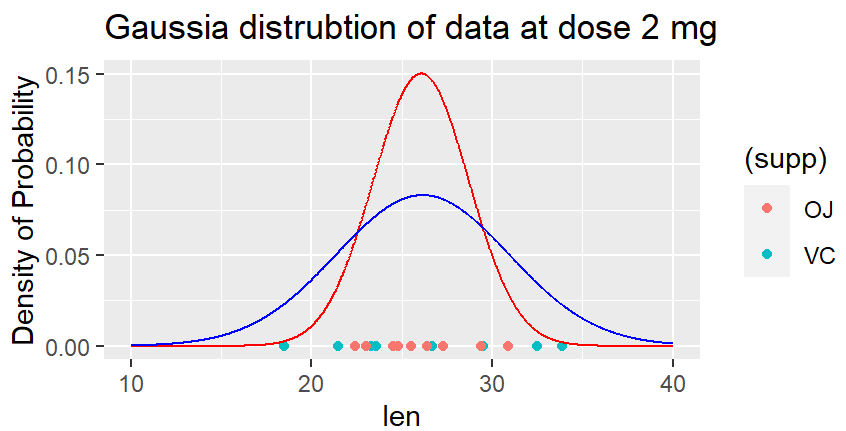


*Image 8*



However, when increasing the supplement to 2 mg instead, the curves has changed places and as you can see the mean of the VC is higher than OJ which tells us that the teeth grow more with this supplement in the higher dose than the lower.

*Image 9*



**Conclusion**

From this data and visualization we can come to the conclusion that dosage as well as supplements have a great impact on the teeth growth. Supplement has a positive association with teeth growth, meaning it grows when administering more supplements.

* The correlation determination is bigger on the VC data hence better fit is obtained on VC data
* If the independent sample assumption is correct, random guinea pig chosen then the difference at 0 should be minimal for the extrapolation data.
* The tooth growth is higher with OJ supplement in lower doses than the VC method.
* The effects of VC are equal to the effects of OJ at 2 mg. The break even is visible on image 4 and 5 where the lines correlate.
* As seen on the positive slopes in the data, tooth length increases regardless of supplement method.
* The error fit on the linear and quadratic is not enough to be able to predict the tooth growth from point 0.

**Assumption**

* Assuming the populations are independent, independent data struggle meaning to not have guinea pigs from the same family that has natural teeth growth that is bigger than the rest of the population
* Independent data samples that the guinea pigs are chosen at randomly from the data dataset without relation to the specimens
* If the independent sample assumption is correct (random guinea pig chosen) then the difference at 0 should be minimal for the extrapolation data.