

Statistical Inference - Course Project Part 2

Arturo Equihua

April 24, 2015

Overview

This part of the course project presents a basic analysis of a dataset around experiments done on guinea pigs, to test the effectiveness of varying doses of two dietary supplements in the length of a tooth of the subjects. The intention is to determine whether there is a significant difference in the results that can be obtained by using one of the supplements vs. the other.

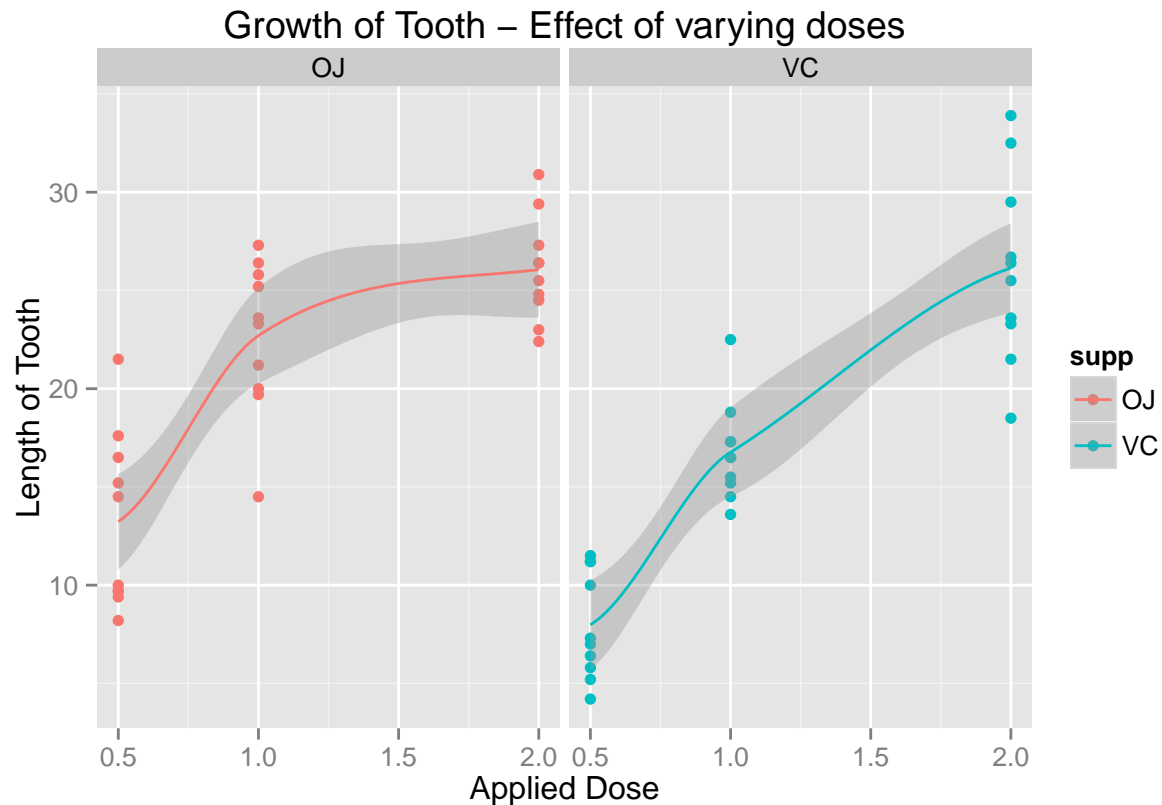
Basic characteristics of the data

The ToothGrowth data is composed of three columns: The dietary supplement code (supp), the applied dose in miligrams (dose) and the measured tooth length (len). There were 60 data points in total. They were taken by forming two groups of 30 guinea pigs (groups OJ, VC). Each of the two groups was subdivided in groups of 10, each of which received three different dose levels of the specific supplement. The tooth length was recorded for each specimen.

The following is a partial view of the data:

```
##      len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
```

As can be seen below, there seems to be a positive effect between the applied dose and the gain in tooth length:



However, from the data above it cannot be directly established if one of the supplements produce better results than the other. This is to be assessed in the next section.

Hypothesis Test

One way to compare the two supplements is to compare the average result in tooth length obtained by them, and determine if the “difference in averages” is zero or greater than zero. To do so, a t-test confidence interval is needed for each of the dose levels (the results of a lower dose might not have anything to do with the results obtained by a higher one, so it does not make sense to mix all the values). If the confidence intervals contain zero, then it is possible that the supplements have no significant difference in results between them; on the other hand, if the confidence interval does not contain zero, it could be concluded that one of the supplements is providing better results than the other.

The following is the combined result of the 95% t-test confidence intervals of the difference in averages for supplement OJ and VC, for the 0.5, 1.0 and 2.0 mg doses:

##	from	to	meanOJ	meanVC
## 1	1.719057	8.780943	13.23	7.98
## 2	2.802148	9.057852	22.70	16.77
## 3	-3.798070	3.638070	26.06	26.14

As can be seen, in the first two cases (lower doses), since the confidence interval does not contain zero, this suggests that the OJ supplement has a difference vs. the VC supplement. This does not hold for the third case (2.0 mg dose).

Conclusion

The experiment around tooth growth demonstrates that, for the two tested supplements in smaller doses, the OJ has an advantage vs. the VC, whereas for a higher dose, the two supplements behave similarly. This can be assessed with a 95% of confidence, that is, there is only a 5% of chance that these results are due to an exceptional circumstance and not to the real attributes of the specific substances used in the experiment.

This exercise has shown how t-estimates can be useful to determine the comparison between these kinds of experimental groups and determine their differences considering the randomness factor. This is more solid than simply taking samples and plotting charts to try to figure out visually how experiments compare.

Key assumptions made

For the analysis above, the following were the assumptions made given that no specific information was available apart from the dataset itself:

- The groups of guinea pigs are not paired, even though the samples were the same size it was better to consider those experiments as if they were fully independent and with different sample sizes.
- The guinea pig groups of the tests are independent, that is, the selection of each of the groups was fully random from an “infinite” population.
- The variances in each group are unequal, that is, there was no guarantee that, for a given dose-supplement group, its variance would have some relationship with the variance of another group.