

Inferential Data Analysis of Tooth Growth

Omar Nooreddin

11/21/2018

Synopsis

This paper will endeavour to perform basic inferential data analysis on the Tooth Growth data set. More precisely it will assess the effects of vitamin C on tooth growth in guinea pigs, and will try and conclude if there's a strong relationship or not.

Data Set Summary

The data set this study will be using is the ToothGrowth dataset provided in R. The following is the description of the data given with dataset (in R):

"The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC)."

We can ascertain the above by looking at the dimension of the dataset:

```
data("ToothGrowth")  
dim(ToothGrowth)
```

```
## [1] 60 3
```

Delving deeper into each variable of the data:

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ  
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

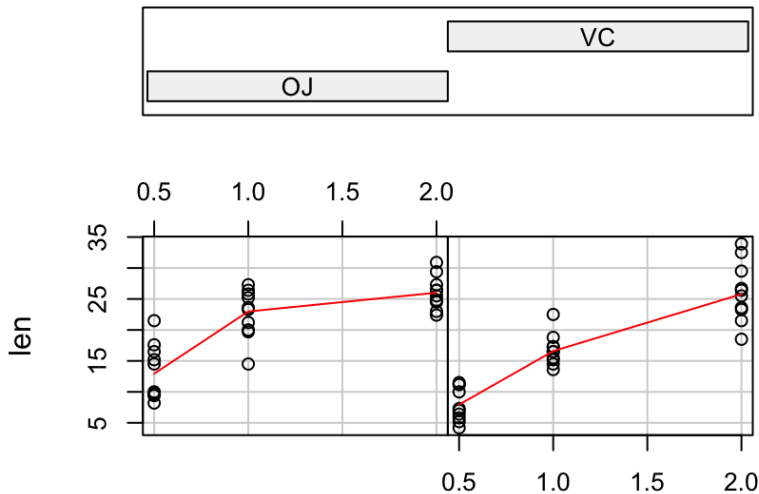
We can see from the above, that there are two supplements(delivery methods): VC (Vitamin C) and OJ (Orange Juice). As for the dosages, it is indeed as per description either: 0.5, 1.0 or 2.0 mg/day of Vitamin C.

Exploring the Dataset

To get a general idea of the effects of the different dosages on teeth length (split by delivery method/supplement) we are going to consider the following plot:

```
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth, xlab = "Length vs Dose, by supplement")
```

Given : supp

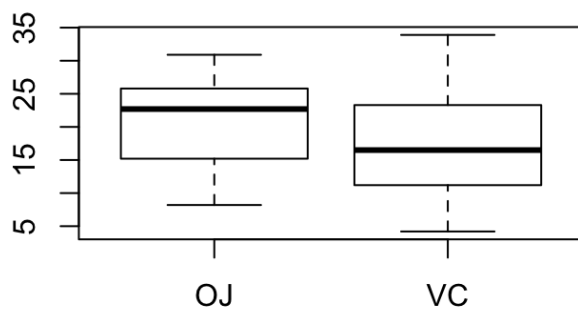


Length vs Dose, by supplement

It can be noted that the higher the dosage the more effect it has on tooth growth. Furthermore we can see the different effects for each delivery method on tooth growth, which can be further verified by a boxplot:

```
boxplot(ToothGrowth$len ~ ToothGrowth$supp, main="Boxplot of Length split by Supplement")
```

Boxplot of Length split by Supplement



It is evident that length of teeth that had OJ as a delivery method is higher.

As such we will be investigating the following two hypotheses:

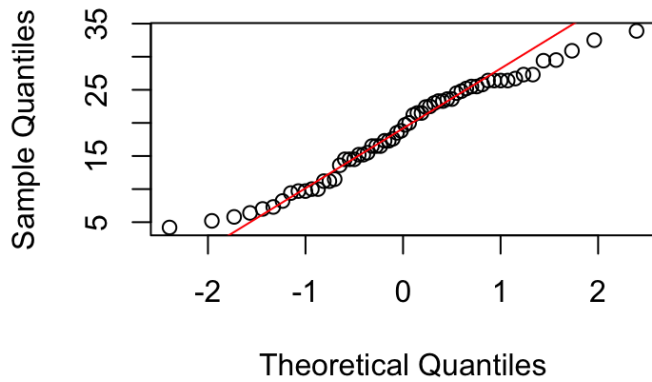
- Mean toothgrowth by OJ is more than VC, while holding the dose constant
- The higher the dose, the more it promotes tooth growth, while holding the delivery method constant

Hypothesis

To ascertain our assumptions above and considering the sample is small (when dissected) and we don't have the population's standard deviation, we will perform confidence test using the **T-Test**. Though for this we will check that this data is normally distributed, and we will ascertain this by a Q-Q plot:

```
qqnorm(ToothGrowth$len)
qqline(ToothGrowth$len, col="red")
```

Normal Q-Q Plot



Following from above we will be testing whether method of delivery/supp (OJ vs VC) has any effect while holding the dose constant at 0.5, 1.0 and 2.0.

Therefore our null hypotheses $H_0 : \mu_{oj} = \mu_{vc}$, while our alternative hypothesis will be $H_a : \mu_{oj} > \mu_{vc}$. We are going to carry out a t-test and set our significance level α to 0.05 (note all code is in the Appendix):

dose constant at	p-value OJ v VC
0.5	0.0031793
1.0	0.0005192
2.0	0.5180742

The above table informs us that when holding the dose constant, for example at 0.5 mg/day, when performing a t-test between μ_{oj} and μ_{vc} , it will give us a p-value as reported in the table.

For the second hypothesis, we will be testing whether a higher dose will have higher effect on mean toothgrowth, whilst holding the delivery method (supp) constant. Therefore, our hypotheses are as follows, null hypothesis: $H_0 : \mu_{0.5} = \mu_{1.0}$, our alternative hypothesis: $H_a : \mu_{0.5} < \mu_{1.0}$. We will repeat the test between dose 1.0 and 2.0 as well. As per previous test we are going to set our significance level α to 0.05:

supp method	p-value 0.5 v 1.0	p-value 1.0 v 2.0
OJ	4.39245952758075e-05	0.0195975710231221
VC	3.40550885143253e-07	4.57780152831932e-05

Conclusion

Following our T Tests as performed above, almost all of them have show a p-value lower than the our significance level α of 0.05. In other words, most tests have rejected the null hypothesis.

We can conclude the following, that OJ as a delivery method had a positive effect on tooth growth when compared to VC. The only exception is when the dose was constant at 2.0 mg/day, where the p-value was quite high, i.e confirming our null hypothesis.

Also for the dose: the higher the dose, the higher the tooth growth confirming our alternative hypotheses H_a .

In summary, OJ as a delivery method is better than VC (except when the dose is at 2mg/day). And a higher dose will have a higher effect on tooth growth.

Appendix

Code for OJ vs VC T-Test:

```
#Split data by supp
vc<-ToothGrowth[ToothGrowth$supp=="VC",]
oj<-ToothGrowth[ToothGrowth$supp=="OJ",]

suppTest<-data.frame("dose constant at"=numeric(), "p-value OJ v VC"=numeric())

suppTest<-as.data.frame(rbind(
c("dose constant at"=0.5,"p-value OJ v VC"=t.test(oj$len[oj$dose==0.5],vc$len[vc$dose==0.5]
, alternative = "greater")$p.value),
c("dose constant at"=1.0,"p-value OJ v VC"=t.test(oj$len[oj$dose==1.0],vc$len[vc$dose==1.0]
, alternative = "greater")$p.value),
c("dose constant at"=2.0,"p-value OJ v VC"=t.test(oj$len[oj$dose==2.0],vc$len[vc$dose==2.0]
, alternative = "greater")$p.value)))

library(knitr)
kable(suppTest,align = "l")
```

dose constant at	p-value OJ v VC
0.5	0.0031793
1.0	0.0005192
2.0	0.5180742

Code for 0.5 vs 1.0 vs 2.0 T-Test:

```
doseTest<-data.frame("supp method"=character(), "p-value 0.5 v 1.0"=numeric(),"p-value 1.0
v 2.0"=numeric())

doseTest<-as.data.frame(rbind(
c("supp method"="OJ","p-value 0.5 v 1.0"=t.test(oj$len[oj$dose==0.5],oj$len[oj$dose==1.0],
alternative = "less")$p.value,"p-value 1.0 v 2.0"=t.test(oj$len[oj$dose==1.0],oj$len[oj$d
ose==2.0], alternative = "less")$p.value),
c("supp method"="VC","p-value 0.5 v 1.0"=t.test(vc$len[vc$dose==0.5],vc$len[vc$dose==1.0],
alternative = "less")$p.value,"p-value 1.0 v 2.0"=t.test(vc$len[vc$dose==1.0],vc$len[vc$d
ose==2.0], alternative = "less")$p.value)))

kable(doseTest)
```

supp method	p-value 0.5 v 1.0	p-value 1.0 v 2.0
OJ	4.39245952758075e-05	0.0195975710231221
VC	3.40550885143253e-07	4.57780152831932e-05