

Tooth Growth Analysis

Analysis Summary

This document briefly explores the Tooth Growth dataset in R and draws simple statistical inferences from this dataset. The findings may be summarized as follows: Evidence shows that at the two lower dosages (0.5 mg/day and 1 mg/day) of vitamin C, the orange juice delivery method has a statistically significant more pronounced effect on tooth length (units unspecified) than ascorbic acid delivery method. Additionally, it is found to be statistically significant in all cases that an increased dosage level for both types of delivery results in increased tooth growth. All code for this document is collected in the appendix at the end to allow clarity of discussion in the document.

Overview and Exploration of Tooth Growth Dataset

The Tooth Growth dataset is a set of 60 observations of guinea pig odontoblast length (units unspecified). Each of 60 guinea pigs was given one of three dose levels (0.5, 1, or 2 mg/day) via one of two delivery methods: orange juice (OJ) or ascorbic acid (VC). In Fig. 1 below, a series of boxplots is shown to compare each delivery method for a given dose level. Notice the inclusion of a simple linear regression of the data shows that the delivery method in the first two cases shows a negative correlation with length. This will lead to our first hypothesis that in simple terms reads: “For orange juice delivery of vitamin C results in a greater tooth growth at a given dosage level.”

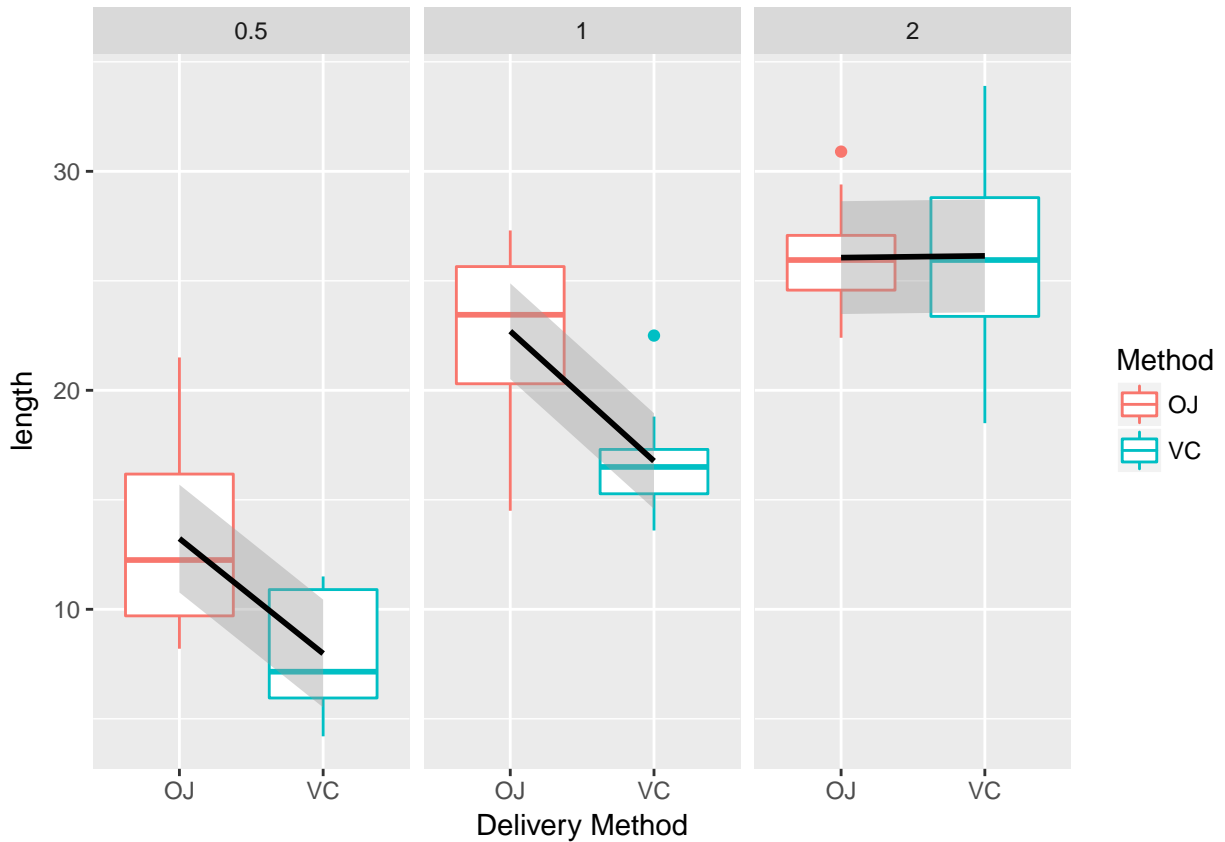


Figure 1: Boxplots of length given the delivery method for each dosage level.

In Fig. 2 below, a series of boxplots is shown to compare each dosage level for a given delivery method. Notice the inclusion of a simple linear regression of the data shows that the in the dosage level vs length shows a correlation. This will lead to our second hypothesis that in simple terms is: “For a given delivery method, a higher dosage of vitamin C results in greater tooth growth.”

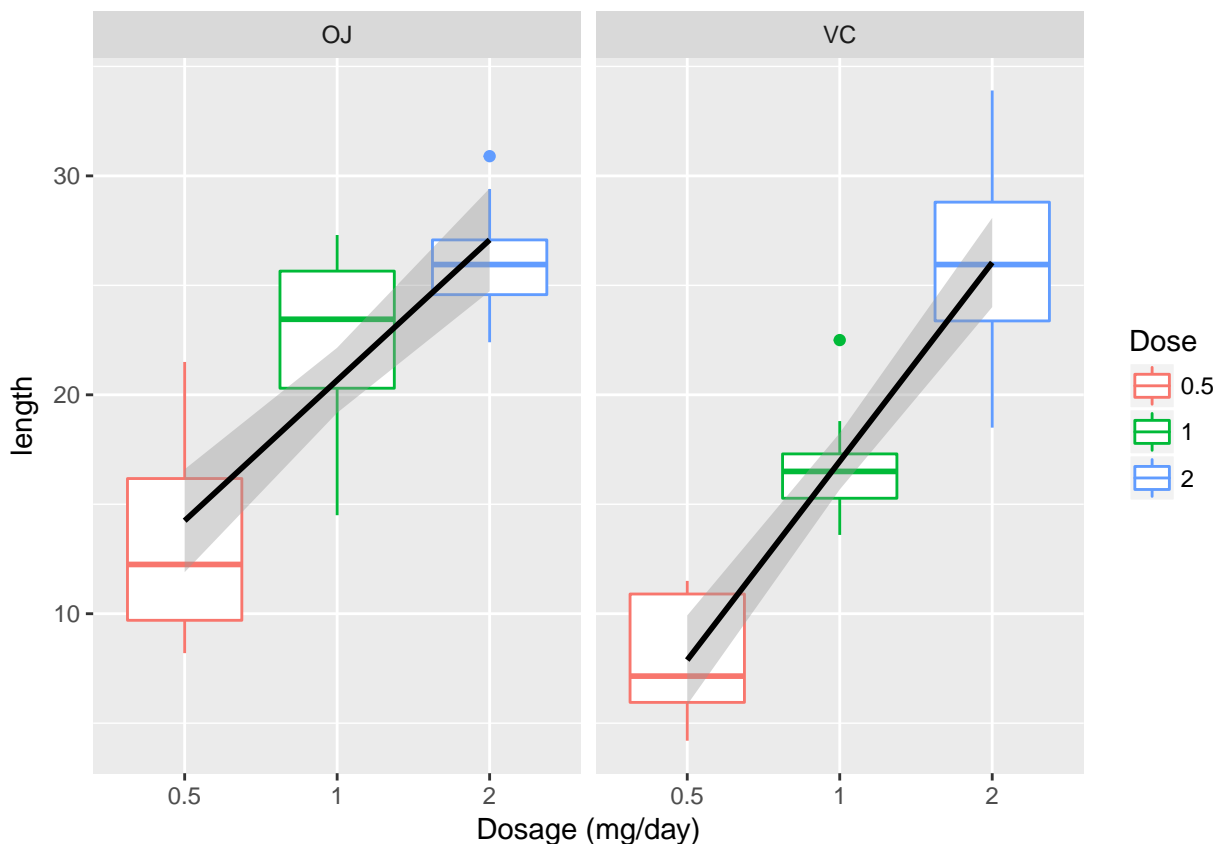


Figure 2: Boxplots of length vs dosage level for each delivery method.

Hypothesis Testing Results

First I subset the data by dosage and delivery method. Then I test my hypotheses. For each case, a Type I error rate of $\alpha = 0.5$ is used.

Sample Delivery Method Hypothesis

The first set of tests correspond to $H_a : \mu_{OJ} > \mu_{VC}$ for a given dosage level. I perform a one-sided t-tests assuming unequal variances for this versus the null hypothesis $H_0 : \mu_{OJ} = \mu_{VC}$ for each of the dosage levels. A summary of the results of these tests is in the table below.

Dosage (mg/day)	μ_{OJ}	μ_{VC}	t-test p value	Hypothesis result
0.5	13.23	7.98	0.0032	H_a
1	22.7	16.77	0.001	H_a
2	26.06	26.14	0.52	H_0

So we see that we accept the alternate hypothesis $\mu_{OJ} > \mu_{VC}$ for dosages of 0.5mg/day and 1 mg/day, but fail to reject the null hypothesis $\mu_{OJ} = \mu_{VC}$ 2mg/day. Due to the relatively small dataset size and lack of

data for additional dosages, we are unable to make any additional statements about the 2mg/day case. For example, it could be a saturation effect (scientific/causal) or it could be small sample size (statistical) effects that cause us to fail to reject the null hypothesis in this case.

Dosage Level Hypothesis

The second set of tests correspond to $H_a : \mu_{HigherDose} > \mu_{LowerDose}$ for each pair of dosage levels for a given delivery method. Note that I do not compare different dosage and different delivery methods, but only different dosage levels within a given delivery method. The former case is an easy extension. I perform a one-sided t-tests assuming unequal variances for the alternative versus the null hypothesis $H_0 : \mu_{HigherDose} = \mu_{LowerDose}$ for each pair of dosage levels. A summary of the results of these tests is in the table below (a blank element in a column means that that dose level was not used for the given t-test). Note that p-values are rounded to 4 decimal places, so a p-value of zero is not identically zero but very small.

Delivery Method	$\mu_{dose=2}$	$\mu_{dose=1}$	$\mu_{dose=0.5}$	t-test p value (rounded)	Hypothesis result
OJ	26.06	22.7		0.0196	H_a
OJ		22.7	13.23	0	H_a
OJ	26.06		13.23	0	H_a
VC	26.14	16.77		0	H_a
VC		16.77	7.98	0	H_a
VC	26.14		7.98	0	H_a

So we see that we accept the alternate hypothesis $\mu_{HigherDose} > \mu_{LowerDose}$ for a given delivery method in all cases. The highest p-value corresponds to the case with delivery method OJ, higher dose = 2 mg/day, and lower dose = 1 mg/day as might have been expected from the exploratory data analysis shown in Fig. 2 above.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)
library(datasets)
library(ggplot2)
data("ToothGrowth")
g<-ggplot(ToothGrowth, aes(supp, len)) + geom_boxplot(aes(color=supp))
g<-g+ facet_wrap(~ dose)+geom_smooth(method = "lm", se=TRUE, color="black", aes(group=1)
,level=0.95)
g+labs(x="Delivery Method",y="length",color="Method")

g<-ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot(aes(color=as.factor(dose)))
g<-g + facet_wrap(~ supp)+geom_smooth(method = "lm", se=TRUE, color="black", aes(group=1)
,level=0.95)
g+labs(x="Dosage (mg/day)",y="length",color="Dose")
d1s1<-ToothGrowth[ToothGrowth$supp=="OJ"&ToothGrowth$dose==0.5,]$len
d1s2<-ToothGrowth[ToothGrowth$supp=="VC"&ToothGrowth$dose==0.5,]$len
d2s1<-ToothGrowth[ToothGrowth$supp=="OJ"&ToothGrowth$dose==1,]$len
d2s2<-ToothGrowth[ToothGrowth$supp=="VC"&ToothGrowth$dose==1,]$len
d3s1<-ToothGrowth[ToothGrowth$supp=="OJ"&ToothGrowth$dose==2,]$len
d3s2<-ToothGrowth[ToothGrowth$supp=="VC"&ToothGrowth$dose==2,]$len
D1<-t.test(d1s1,d1s2,paired=FALSE,var.equal = FALSE,alternative="greater")
D2<-t.test(d2s1,d2s2,paired=FALSE,var.equal = FALSE,alternative="greater")
D3<-t.test(d3s1,d3s2,paired=FALSE,var.equal = FALSE,alternative="greater")
S1<-t.test(d3s1,d2s1,paired=FALSE,var.equal = FALSE,alternative="greater")
S2<-t.test(d2s1,d1s1,paired=FALSE,var.equal = FALSE,alternative="greater")
S3<-t.test(d3s1,d1s1,paired=FALSE,var.equal = FALSE,alternative="greater")
S4<-t.test(d3s2,d2s2,paired=FALSE,var.equal = FALSE,alternative="greater")
S5<-t.test(d2s2,d1s2,paired=FALSE,var.equal = FALSE,alternative="greater")
S6<-t.test(d3s2,d1s2,paired=FALSE,var.equal = FALSE,alternative="greater")
```