Analysis of Guinea Pig Tooth Growth for Different Doses of Vitamin C and Delivery Methods

Course 6 (Statistical Inference) / Assessment 1 Part 2

Paul Ringsted, 20th January 2019

Synopsis

In this report we analyze a study of tooth growth in guinea pigs, each given three different dose levels of Vitamin C (0.5, 1, and 2 mg/day), administered by one of two different delivery methods (orange juice - OJ, or ascorbic acid - VC). The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Analysis of the data concludes that higher doses of Vitamin C result in higher average tooth growth; and that use of orange juice as delivery method results in higher growth at lower doses (0.5-1.0 mg/day). However, there was no evidence to suggest that delivery method makes any difference at highest dose of 2.0 mg/day. This may be the result of a limit to growth rate already achieved via the higher dosage of Vitamin C, reducing any further efficacy of different delivery methods, which is a possible area for further study.

Loading Study Data - Initial Observations

This analysis is on the ToothGrowth dataset provided in the R datasets package. We first load the data and gather some basic statistics to understand the data.

Dose	Supp	Mean	Min	Max	SD	Var
0.5	OJ	13.23	8.2	21.5	4.45971	19.88900
1.0	OJ	22.70	14.5	27.3	3.91095	15.29556
2.0	OJ	26.06	22.4	30.9	2.65506	7.04933
0.5	VC	7.98	4.2	11.5	2.74663	7.54400
1.0	VC	16.77	13.6	22.5	2.51531	6.32678
2.0	VC	26.14	18.5	33.9	4.79773	23.01822

Two plots of the data are shown overleaf:

- Figure 1 - Boxplot by dose, with coloration and regression line by delivery method
- Figure 2 - Violin plot, grouped by dose and delivery method

Initial observations on the data are that it is reasonably distributed (no significant outliers or data cleanup warranted), and both plots show a trend of higher growths for higher doses of Vitamin C. Orange Juice delivery method appears to increase growth at lower doses of 0.5-1.0 mg/day, but the impact of delivery method is less obvious at the high dose of 2.0 mg/day.

Statistical Model and Assumptions

Given the low sample size we will use two-sided t-tests at the 95% confidence interval, and assume population variances are equal. No adjustment for multiplicity was performed as the number of tests performed in this analysis was low (2 for dose, 3 for delivery method).

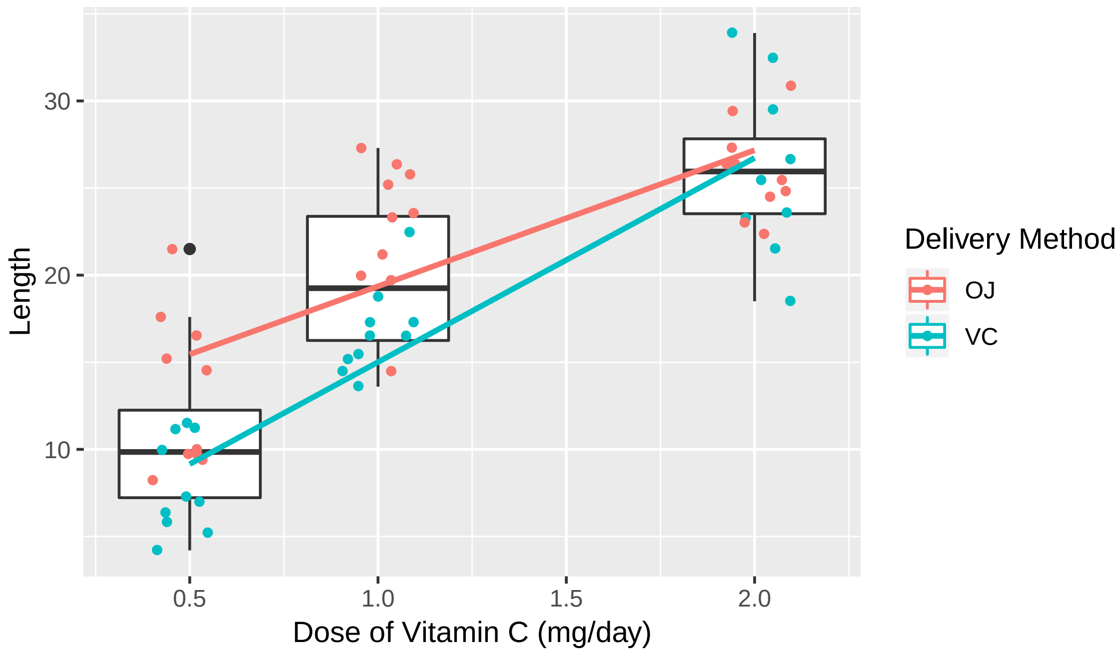

Figure 1: Tooth Growth by Dose

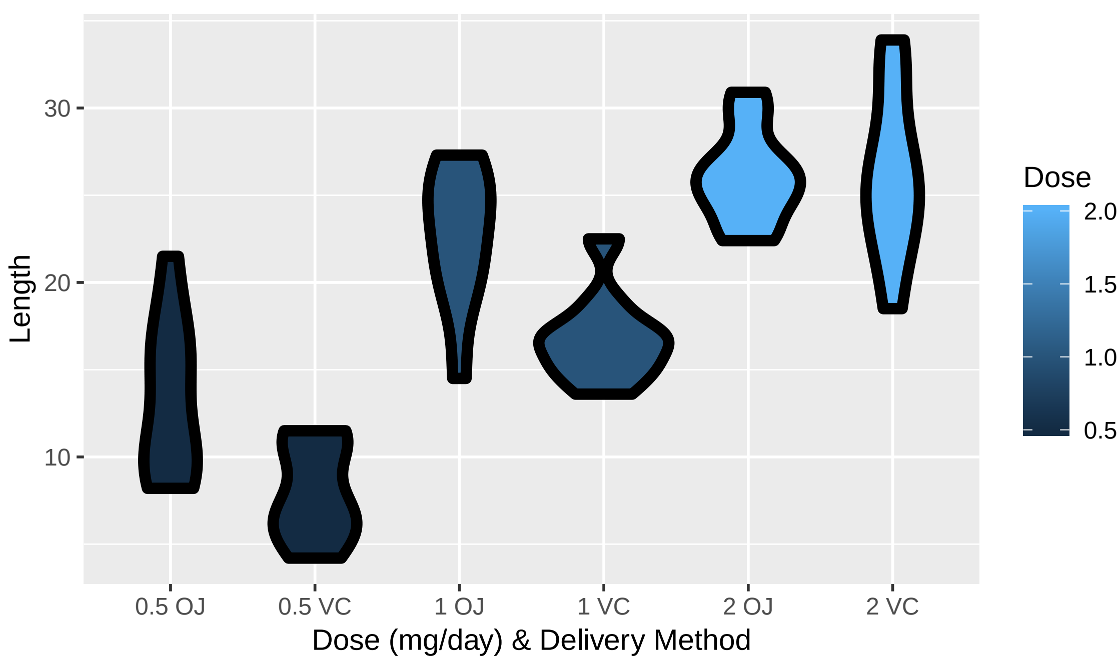

Figure 2: Tooth Growth, Grouped by Dose and Delivery Method

Analysis of Tooth Growth by Dose

We first examine the observation that tooth growth increases with higher doses of Vitamin C, regardless of delivery method.

To confirm this statistically, we performed two-sided 95% t-tests, between doses of 0.5 and 1, and between doses of 1 and 2 (i.e. two tests, each with 2 groups of 20 observations). Our null hypothesis in both cases is that, average tooth growth is unaffected by dose of Vitamin C. The results are given in Table 1:

Table 1: Results of T-Tests Between Doses

T-Test	T-Statistic	DoF	Conf Int Low	Conf Int High	P-Value %
Dose 0.5 vs 1.0 mg/day	-6.477	38	-11.98	-6.28	0.00001
Dose 1.0 vs 2.0 mg/day	-4.900	38	-8.99	-3.74	0.00181

Conclusions:

- These tests have confidence intervals that do not bound zero, and very-low P-values, indicating we reject the null hypothesis of no effect on tooth growth in all cases, and conclude that **higher doses of Vitamin C increase tooth growth.**

Analysis of Tooth Growth by Delivery Method

To analyze whether delivery method affects tooth growth, we will perform two-sided 95% t-tests between the different delivery methods, for each level of dosage (i.e. three tests, each with 2 groups of 10 observations). Again our null hypothesis in each case is that, delivery method has no impact on the average tooth growth. The results are given in Table 2:

Table 2: Results of T-Tests Between Delivery Methods

T-Test	T-Statistic	DoF	Conf Int Low	Conf Int High	P-Value %
Delivery OJ vs VC @ 0.5 mg/day	3.170	18	1.77	8.73	0.53037
Delivery OJ vs VC @ 1.0 mg/day	4.033	18	2.84	9.02	0.07807
Delivery OJ vs VC @ 2.0 mg/day	-0.046	18	-3.72	3.56	96.37098

Conclusions:

- For dose level at 0.5 and 1.0 mg/day, these tests have confidence intervals that do not bound zero, and very-low P-values, leading us to reject the null hypothesis of no effect on tooth growth in these cases, and conclude that **Orange Juice delivery method results in an increase in average growth at doses of 0.5-1.0 mg/day.**
- For dose level at 2.0 mg/day, the confidence interval bounds 0 with a very high P-value, so there is no reason to reject the null hypothesis, and we conclude that **there is no evidence to suggest that delivery method affects growth at dose of 2.0mg/day.** This may be the result of a limit to growth rate already achieved via the higher dosage of Vitamin C, reducing any further efficacy of different delivery methods, which is a possible area for further study.

Appendix - R Code

```
library(knitr)
opts_chunk$set(fig.width=6, fig.height=3.5, fig.pos = "H", echo=FALSE, eval=TRUE)
library(ggplot2)
library(kableExtra)

#-------------------------------------------------------------------------------------
# Load ToothGrowth data and copy to 'tg' dataframe for ease of reference
data(ToothGrowth)
tg<-ToothGrowth

# Add a factor combos of dose and supp for violin plot
tg$group <- as.factor(paste(tg$dose,tg$supp))

# Build and render a table of basic stats for the dataset
stats<-aggregate(tg$len,by=list(tg$dose,tg$supp),FUN=mean)
colnames(stats)<-c("dose","supp","mean")
stats<-cbind(stats,min=aggregate(tg$len,by=list(tg$dose,tg$supp),FUN=min)$x)
stats<-cbind(stats,max=aggregate(tg$len,by=list(tg$dose,tg$supp),FUN=max)$x)
stats<-cbind(stats,sd=aggregate(tg$len,by=list(tg$dose,tg$supp),FUN=sd)$x)
stats<-cbind(stats,var=stats$sd^2)
stats %>% kable(col.names=c("Dose","Supp","Mean","Min","Max","SD","Var"),
                  align=c("c","c",rep("r",5)),digits=c(2,2,2,2,2,5,5)) %>%
        kable_styling(latex_options = "hold_position")

#-------------------------------------------------------------------------------------
# Box plot by dose with linear regression line
g <- ggplot(data=tg,aes(x=dose,y=len,color=supp))
g <- g + geom_boxplot(aes(group=dose))
g <- g + geom_jitter(shape=16, position=position_jitter(0.1))
g <- g + geom_smooth(method="lm",se=FALSE)
g <- g + labs(x="Dose of Vitamin C (mg/day)",y="Length",color="Delivery Method")
print(g)

#-------------------------------------------------------------------------------------
# Violin plot by dose and delivery method
g <- ggplot(tg,aes(x=group,y=len,fill=dose))
g <- g+geom_violin(col="black",size=2)
g <- g + labs(x="Dose (mg/day) & Delivery Method",y="Length",fill="Dose")
print(g)

#-------------------------------------------------------------------------------------
# T-tests for different dosages
test_results<-data.frame(NULL)
dose1<-subset(tg,dose %in% c(0.5,1))            # Subset to test doses 0.5 and 1.0
dose2<-subset(tg,dose %in% c(1,2))              # Subset to test doses 1.0 and 2.0

#Perform t-tests of length by dose and gather results
test1<-t.test(len~dose,paired=FALSE,var.equal=TRUE,data=dose1)
test2<-t.test(len~dose,paired=FALSE,var.equal=TRUE,data=dose2)
c_lo<-c(test1$conf.int[1],test2$conf.int[1])
c_hi<-c(test1$conf.int[2],test2$conf.int[2])
```

```r
pval<-c(test1$p.value,test2$p.value)
tval<-c(test1$statistic,test2$statistic)
dval<-c(test1$parameter,test2$parameter)
testtype<-c("Dose 0.5 vs 1.0 mg/day",
            "Dose 1.0 vs 2.0 mg/day")

#Build a dataframe with the test statistics (p-values in % for display) and render it
pval<-pval*100
results<-data.frame(testtype,tval,dval,c_lo,c_hi,pval,stringsAsFactors = FALSE)
results %>% kable(col.names=c("T-Test","T-Statistic","DoF",
                              "Conf Int Low","Conf Int High","P-Value %"),
                  align=c("l",rep("r",5)),digits=c(2,3,0,2,2,5),booktabs=T,
                  caption="Results of T-Tests Between Doses") %>%
                kable_styling(latex_options = "hold_position")

#--------------------------------------------------------------------------------------
# T-tests for different delivery methods
test_results<-data.frame(NULL)
dose1<-subset(tg,dose == 0.5)           # Subset to OJ & VC, at dose 0.5
dose2<-subset(tg,dose == 1)             # Subset to OJ & VC, at dose 1.0
dose3<-subset(tg,dose == 2)             # Subset to OJ & VC, at dose 2.0

#Perform t-tests of length by delivery method and gather results
test1<-t.test(len~supp,paired=FALSE,var.equal=TRUE,data=dose1)
test2<-t.test(len~supp,paired=FALSE,var.equal=TRUE,data=dose2)
test3<-t.test(len~supp,paired=FALSE,var.equal=TRUE,data=dose3)
c_lo<-c(test1$conf.int[1],test2$conf.int[1],test3$conf.int[1])
c_hi<-c(test1$conf.int[2],test2$conf.int[2],test3$conf.int[2])
pval<-c(test1$p.value,test2$p.value,test3$p.value)
tval<-c(test1$statistic,test2$statistic,test3$statistic)
dval<-c(test1$parameter,test2$parameter,test3$parameter)
testtype<-c("Delivery OJ vs VC @ 0.5 mg/day",
            "Delivery OJ vs VC @ 1.0 mg/day",
            "Delivery OJ vs VC @ 2.0 mg/day")

#Build a dataframe with the test statistics (p-values in % for display) and render it
pval<-pval*100
results<-data.frame(testtype,tval,dval,c_lo,c_hi,pval,stringsAsFactors = FALSE)
results %>% kable(col.names=c("T-Test","T-Statistic","DoF","Conf Int Low",
                              "Conf Int High","P-Value %"),
                  align=c("l",rep("r",5)),digits=c(2,3,0,2,2,5),booktabs=T,
                  caption="Results of T-Tests Between Delivery Methods") %>%
                kable_styling(latex_options = "hold_position")
```