Study of the relationship between Tooth Growth and Vitamin C in Guinea Pigs

Data Summary

```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000  
```

The ToothGrowth dataset contains three variables; *len*, *supp* and *dose*.

- The *len* variable corresponds to the measured length, in microns, of the guinea pigs' odontoblast, ranging from 4.20 to 33.90 microns, with a mean of 18.81.
- The *supp* variable corresponds to the supplement type used. It can take two nominal values, *OJ* or *VC*, for *Orange Juice* and *Ascorbic Acid* respectively.
- The *dose* variable corresponds to the dose of Vitamin C, in milligram, of either supplement. Three set doses have been used, 0.5 mg, 1.0 mg and 2 mg.

It contains 60 observations, 10 for each combination of *dose* and *supp*.

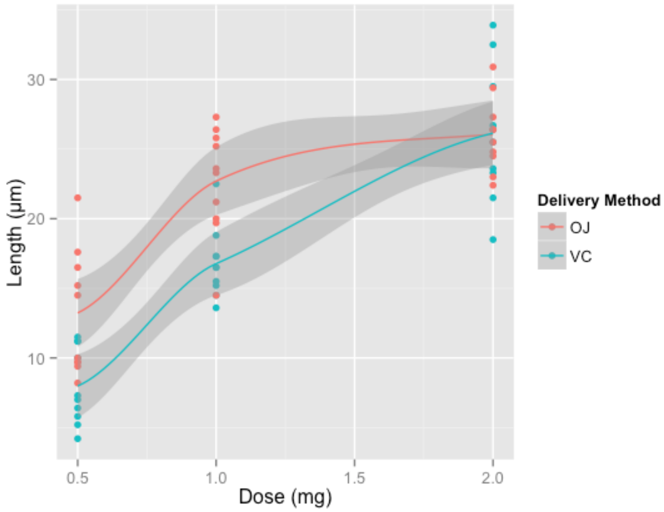
Exploratory Data Analysis

The following graphs and metrics aim at observing certain patterns between variables to propose relevant hypothesis testing.

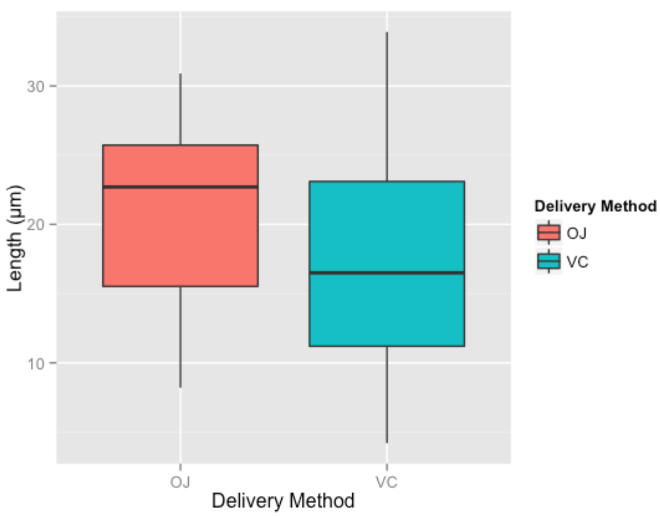
Metrics

In a first instance, the linear relationship between tooth growth in guinea pigs and the dose of ascorbic acid appears to be strong, with a correlation of 0.8. This linear relationship remains strong, and even gets stronger, within the delivery methods, with a correlation of 0.75 and 0.9 for orange juice and ascorbic acid respectively.

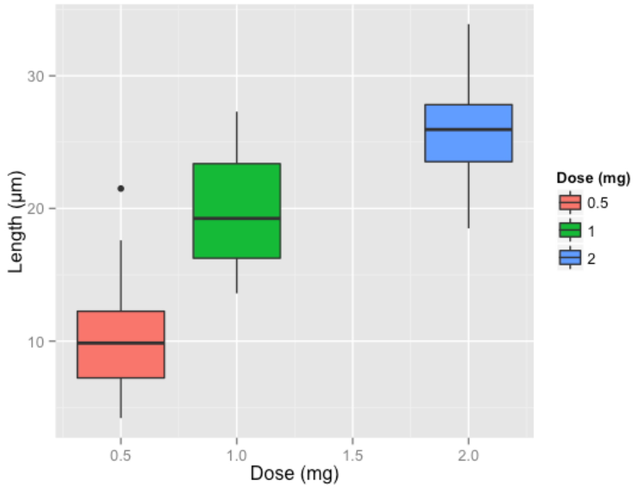
Exploratory Graphs



From this graph, the overall positive aspect of the relationship between the dose and the length of the odontoblast can be fully appreciated. By it's overall look, the ascorbic acid (*VC*) data displays a stronger linearity than the orange juice (*OJ*).



This graph is meant to compare both delivery method and showcase any significant difference between both.



Finally, this graph intends to portray the relationship between the growth of odontoblast in guinea pigs, with no distinction between delivery method.

Hypothesis Testing

In the light of the previous exploratory graphs, two questions will be examined in this experiment, both pertaining to the link between the growth of odontoblasts in guinea pigs and ascorbic acid.

1. Does the ascorbic acid delivery method impact the growth of odontoblasts?

2. Does the growth of odontoblasts in respect of ascorbic acid is dose dependent?

The Alpha value selected for this experiment is 0.05, for a 95% confidence level.

Supply Method

Hypothesis Formulation and Presentation

Ascorbic acid delivery generates more odontoblast growth than orange juice delivery.

Ascorbic acid supply = VC Orange Juice delivery = OJ

- $H_0: \mu_{OJ} = \mu_{VC}$
- $H_a: \mu_{OJ} < \mu_{VC}$

The data provided does not allow to reject the null hypothesis, as the confidence interval (-0.171, 7.571) includes 0, suggesting that no change from the mean is possible. Furthermore, the p-value (0.0606) suggests that there is weak evidence that H_a is true, rendering the test result not statistically significant. Finally, the test statistic (1.92) is not part of the rejection region (-2, 2)

Dose of Ascorbic Acid

Hypothesis Formulation and Presentation

The length of odontoblasts in guinea pigs is positively linked to the dose of ascorbic acid provided.

- $H_0: \mu_{0.5mg} = \mu_{1.0mg} = \mu_{2.0mg}$
- $H_a: \mu_{0.5mg} < \mu_{1.0mg} < \mu_{2.0mg}$

The data provided does allow to reject the null hypothesis in all three cases (Doses 1.0 vs 0.5, 2.0 vs 1.0 and 2.0 vs 0.5), as their respective confidence interval (6.28, 3.73 and 6.28) do not include 0, suggesting that a difference from the mean is likely. Furthermore, the p-values (1.3×10^{-7}, 1.9×10^{-5} and 4.4×10^{-14}) suggest that there is overwhelming evidence that H_a is true, rendering the test result to be highly significant. Finally, all three test statistics (6.48,4.9 and 11.8) are part of the rejection regions of all three tests (~[-2.025;2.025] < test statistic).

Assumptions and Conclusions

Assumptions

In order be valid, the following assumptions with regards to the data, the population and the experiment must be fulfilled. The population must be independent and unpaired, it must be normally distributed. Concerning the samples, its variances must be unequal as the test for unequal variances has been used (Welch Two Sample t-test) and finally, the sampling from the population must be random.

According to the original experiment, one guinea pig was sacrificed for each observation provided, meeting assumption regarding independence. Despite the small size of the data set, some normality can be observed surrounding the length of odontoblasts (see Annex B). It is also assumed that the guinea pigs were randomly selected from their population and the variances have been automatically calculated by R using *t.test* function.

Conclusions

Assuming that the preceding assumptions are true, it may be infered that there is a highly significant difference between tooth length and the dose levels across delivery methods. The data used in this analysis provides overwhelming evidence that higher dosage of Vitamin C led to longer odontoblasts no matter the delivery method.

The same conclusion cannot be drawn concerning the two delivery methods studied. Although the first two graphs hint for a difference in tooth length between the delivery methods, the present data does not provide sufficient evidence to support this, when analyzed as a whole. Larger samples of data, following the same research protocols and methodology, could probably narrow the confidence interval and potentially allow to reject the null hypothesis.

Annex A - The Code

```
#fetching the data
data("ToothGrowth")
#loading ggplot2 to make some neat graphs
require(ggplot2)

summary(ToothGrowth)

#Some more summaries of the data
head(ToothGrowth)

##    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

str(ToothGrowth)

## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

#Exploration of potential relationships between variables
OverallLenDose <- cor(ToothGrowth$len,ToothGrowth$dose)

OJ <- subset(ToothGrowth,ToothGrowth$supp =="OJ")
VC <- subset(ToothGrowth,ToothGrowth$supp =="VC")

OJLenDose <- cor(OJ$len,OJ$dose)
VCLenDose <- cor(VC$len,VC$dose)

#Overall portrait of the data
qplot(ToothGrowth$dose,ToothGrowth$len, colour=ToothGrowth$supp, ylab =
"Length (µm)", xlab="Dose (mg)") + geom_smooth() +
scale_colour_discrete("Delivery Method")

#Supply methods compared
qplot(ToothGrowth$supp,ToothGrowth$len,geom="boxplot",
fill=ToothGrowth$supp,ylab = "Length (µm)", xlab="Delivery Method")+
scale_fill_discrete("Delivery Method")

#Dose provided compared
qplot(ToothGrowth$dose,ToothGrowth$len,geom="boxplot",
fill=factor(ToothGrowth$dose), ylab = "Length (µm)", xlab="Dose (mg)") +
scale_fill_discrete("Dose (mg)")

#Setting the alpha in order to establish the confidence level
alpha <- 0.05
```

```
#Subsetting the data by delivery method
OJ <- subset(ToothGrowth,ToothGrowth$supp =="OJ")
VC <- subset(ToothGrowth,ToothGrowth$supp =="VC")

#Testing for any statistical difference between them
Hypo1 <- t.test(OJ$len,VC$len,paired=FALSE,conf.level=1-alpha)

#Establishing the rejection regions for the t-tests
RejRegionHypo1 <- qt(1-alpha/2,df=Hypo1$parameter)*c(-1,1)

#Subseting the different dosage of Vitamin C given to guinea pigs
Dose0.5 <- subset(ToothGrowth,ToothGrowth$dose == 0.5)
Dose1 <- subset(ToothGrowth,ToothGrowth$dose == 1)
Dose2 <- subset(ToothGrowth,ToothGrowth$dose == 2)

#Testing for any statistical difference between them
Hypo2.1 <- t.test(Dose1$len,Dose0.5$len,paired=FALSE,conf.level=1-alpha)
Hypo2.2 <- t.test(Dose2$len,Dose1$len,paired=FALSE,conf.level=1-alpha)
Hypo2.3 <- t.test(Dose2$len,Dose0.5$len,paired=FALSE,conf.level=1-alpha)

#Establishing the rejection regions for the t-tests
RejRegionHypo2.1 <- qt(1-alpha/2,df=Hypo2.1$parameter)*c(-1,1)
RejRegionHypo2.2 <- qt(1-alpha/2,df=Hypo2.2$parameter)*c(-1,1)
RejRegionHypo2.3 <- qt(1-alpha/2,df=Hypo2.3$parameter)*c(-1,1)
```

Annex B - Distribution Graph

```
## Graph 1 - Distribution of Odontoblast Length
ggplot(ToothGrowth, aes(x=ToothGrowth$len))+geom_histogram(aes(y =
..density.., fill=..count..),binwidth=3.5) + stat_function(fun = dnorm,
colour="blue", arg = list(mean =
mean(ToothGrowth$len),sd=sd(ToothGrowth$len)), geom="line", size=2) +
geom_vline(aes(xintercept = mean(ToothGrowth$len),colour="red"),size=1)+
ylab("Density") + xlab("Odontoblast Length (µm)") +
labs(title="Odontoblast Length Distribution")
```