

Tooth Growth Analysis

Michael Szczepaniak

September 2015

Synopsis

This report explores the relationship between the length of teeth in 10 guinea pigs to three Vitamin C dosages and two delivery methods using the R ToothGrowth data set. The data provided evidence that an increased mean tooth length was observed with increased dosage with both delivery methods under a Type II error rate α of 0.05. The data also provided evidence that an increased mean tooth length was observed with the orange juice delivery method for the 0.5 and 1.0 mg doses under a Type II error rate α of 0.05.

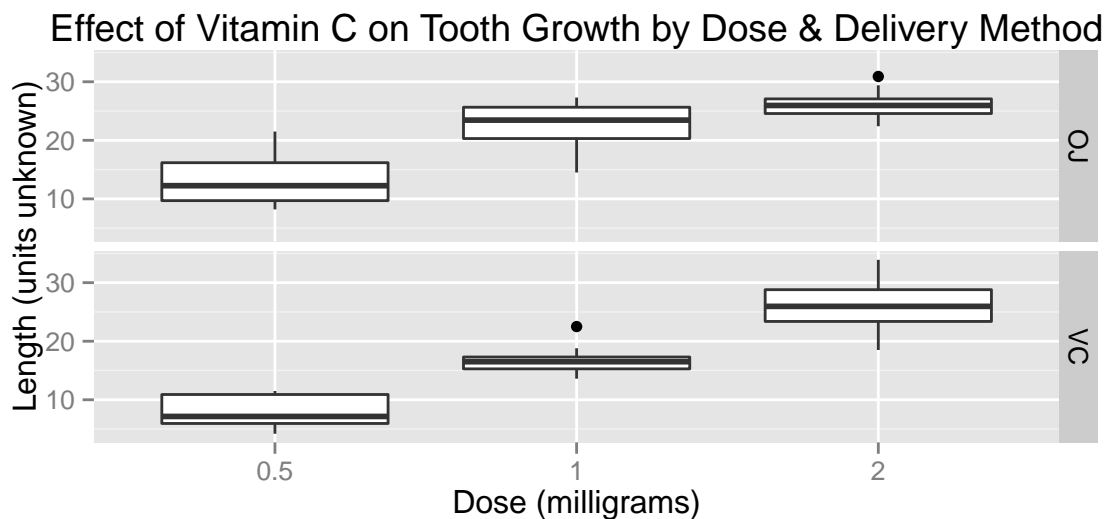
Loading Data & Exploratory Data Analysis

Note: All code used for the report are contained the **Code Used for the Analysis** section.

The *loadToothData* function was used to load the data. This function was first passed to the *getToothSummary* to create the following summary table. The **Group** variable concatenates the **Delivery** and **Dose** variables.

Group	Mean.Length	Std.Dev.Length
OJ0.5	13.23	4.46
OJ1.0	22.70	3.91
OJ2.0	26.06	2.66
VC0.5	7.98	2.75
VC1.0	16.77	2.52
VC2.0	26.14	4.80

The *loadToothData* function was then passed to the *exploreBox* function to create the boxplot below.



From this initial look at the data, three questions emerge to guide further exploration:

1. Does increased vitamin C dosage lead to an increase or decrease in tooth length?
2. Does vitamin C delivery method lead to an increase or decrease in tooth length?
3. Are the effects of dosage and delivery method on tooth length independent or coupled?

The plot above suggests a possible relationship between increasing dosage and tooth length for both delivery methods. Dosage and delivery method were then explored by hypothesis testing. The independence of the impact of dosage and delivery as stated in the third question above was not explored in this report.

Analysis of the Data

Because the data does not contained ids for each animal, a paired analysis is not possible without assuming the data was ordered in a particular way. For this reason, the analysis will focus on group comparisons. The data contains 6 groups each with 10 measurments and are defined in the summary table.

To answer the first two questions, the null hypothesis was defined as $H_0 : \mu_A = \mu_B$ (Mean A = Mean B in the testing table) and the alternative hypothesis as $H_a : \mu_B > \mu_A$ (Mean B > Mean A in the testing table) where A and B are the groups defined in the testing table (created by the function *createTestingTable*) shown below.

Test	t1	t2	t3	t4	t5	t6	t7
Addresses	Question 1.	Question 1.	Question 1.	Question 1.	Question 2.	Question 2.	Question 2.
Group A	OJ0.5	OJ1.0	VC0.5	VC1.0	VC0.5	VC1.0	VC2.0
Group B	OJ1.0	OJ2.0	VC1.0	VC2.0	OJ0.5	OJ1.0	OJ2.0
Mean A	13.23	22.7	7.98	16.77	7.98	16.77	26.14
Mean B	22.7	26.06	16.77	26.14	13.23	22.7	26.06

Hypothesis Testing

There are 7 testing scenarios that arise from this data set which can be described as follows:

The above is a table

Assumptions & Conclusions

Assumptions

1. Variation between the populations tested were not equal.
2. The guineau pigs were IID normal in their selection.
3. The effects of dosage and delivery method on tooth length are independent.

Conclusions

1. We reject H_0 in favor of H_a for all tests except **t7**.
2. Based on 1., the evidence of the data supports the hypothesis that an increased mean tooth length will be observed with increased dosage with both delivery methods under a Type II error rate α of 0.05.
3. Also based on 1., the evidence of the data supports the hypothesis that an increased mean tooth length will be observed using the orange juice deliver method for the 0.5 and 1.0 mg dose levels under a Type II error rate α of 0.05.

Code Used for the Analysis

```
## Load the R ToothGrowth dataset and returns an updated data frame
loadToothData <- function() {
  suppressMessages(suppressWarnings(require(dplyr)))
  library(datasets)
  library(dplyr)
  data("ToothGrowth")
  # rename variables/columns to make them clearer
  tooth.data <- rename(ToothGrowth, Length=len, Delivery=supp, Dose=dose)
  # add group id
  tooth.data <- mutate(tooth.data,
                       Group=factor(paste0(Delivery, sprintf("%.1f", Dose))))
  tooth.data <- mutate(tooth.data, Dose = factor(Dose))
  return(tooth.data)
}

## Summarizes the tooth data by group
getToothSummary <- function(tooth.data = loadToothData()) {
  suppressMessages(suppressWarnings(library(dplyr)))
  suppressMessages(suppressWarnings(library(knitr)))
  library(dplyr)
  tooth.summ <- group_by(tooth.data, Group)
  tooth.table <- summarize(tooth.summ, Mean.Length = mean(Length),
                           Std.Dev.Length = sd(Length))

  library(knitr)
  kable(tooth.table, digits=2)
}

## Creates the first explorer box plot of tooth length by dose and delivery
exploreBox <- function(tooth.data = loadToothData()) {
  library(ggplot2)
  # exploratory panel plot
  p <- ggplot(tooth.data, aes(Dose, Length))
  p <- p + geom_boxplot()
  p <- p + facet_grid(Delivery ~ .)
  p <- p + xlab("Dose (milligrams)") + ylab("Length (units unknown)")
  p <- p + ggtitle("Effect of Vitamin C on Tooth Growth by Dose & Delivery Method")
  print(p)
}

##
createTestingTable <- function(tooth.data = loadToothData()) {
  suppressMessages(suppressWarnings(library(dplyr)))
  suppressMessages(suppressWarnings(library(knitr)))
  library(dplyr)
  tooth.summ <- group_by(tooth.data, Group)
  tooth.table <- summarize(tooth.summ, Mean.Length = mean(Length),
                           Std.Dev.Length = sd(Length))
  df <- data.frame(Test=c("Addresses", "Group A", "Group B", "Mean A", "Mean B"),
                   t1=c("Question 1.", "OJ0.5", "OJ1.0",
                        as.character(tooth.table[1,2]),
                        as.character(tooth.table[2,2])),
                   t2=c("Question 1.", "OJ1.0", "OJ2.0",
                        as.character(tooth.table[2,2]),
                        as.character(tooth.table[3,2])),
```

```

t3=c("Question 1.", "VC0.5", "VC1.0",
     as.character(tooth.table[4,2]),
     as.character(tooth.table[5,2])),
t4=c("Question 1.", "VC1.0", "VC2.0",
     as.character(tooth.table[5,2]),
     as.character(tooth.table[6,2])),
t5=c("Question 2.", "VC0.5", "OJ0.5",
     as.character(tooth.table[4,2]),
     as.character(tooth.table[1,2])),
t6=c("Question 2.", "VC1.0", "OJ1.0",
     as.character(tooth.table[5,2]),
     as.character(tooth.table[2,2])),
t7=c("Question 2.", "VC2.0", "OJ2.0",
     as.character(tooth.table[6,2]),
     as.character(tooth.table[3,2]))

library(knitr)
kable(df, digits=2)
}

```