Statistical Inference - Tooth Growth

Jeroen Zonneveld

May, 2015

Overview

In this project we will investigate the `ToothGrowth` data in the R data sets package. The data shows the result of measuring the effect of different dosage amounts of Vitamin C on the length of the teeth of ten guinea pigs. There are two supplement types of Vitamin C tested, Orange Juice (`OJ`) and Ascorbic Acid (`VC`), and they are given in three different milligram dosage amount, 0.5, 1.0, and 2.0. The report goes through the process of cleansing the dataset, conducting exploratory analysis, and a statistical inference around different categorizations to the length of the teeth.

Data Exploration

Loading the data and exploring basic build.

```
data(ToothGrowth)
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 ...
```

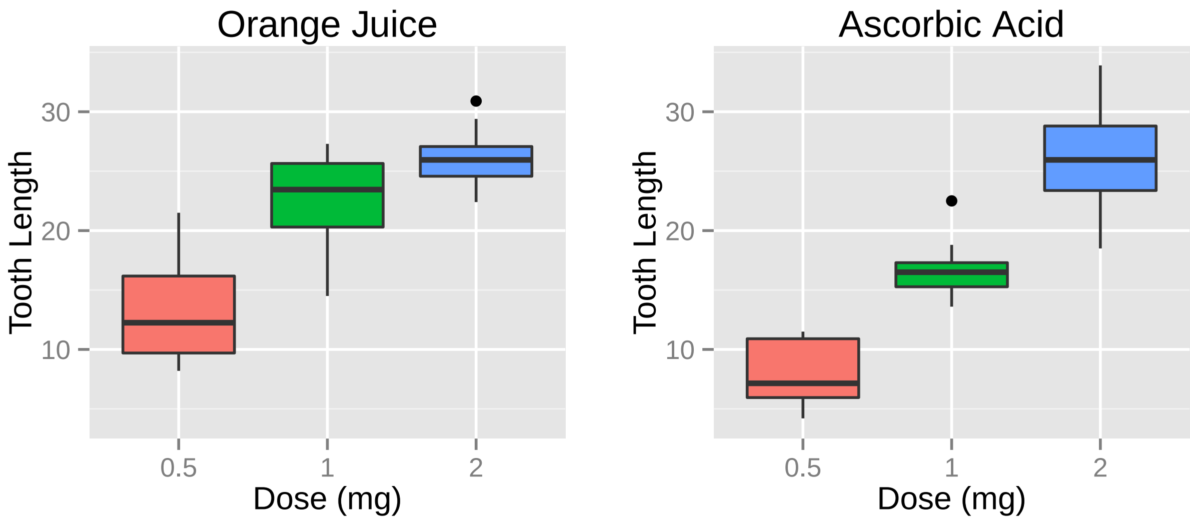
So the data set consists of 60 measurements and 3 variables. We know that the dosage amount is only in 3 set amounts so we can change the dosage to a factor variable. Also, we will change the names of the columns to something meaningful.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
colnames(ToothGrowth) <- c('Length','Supplement','Dose')
summary(ToothGrowth)
```

```
##      Length      Supplement  Dose
##  Min.   : 4.20   OJ:30      0.5:20
##  1st Qu.:13.07   VC:30      1  :20
##  Median :19.25              2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Data Summary

We've seen that there are a total of 60 records; 30 for each supplement; 20 for each dose; this boils down to 10 records for each of the 6 supplement/dose combinations. Let's break this down visually to explore the supplement methods and dosages independent effects on tooth growth.



It appears that the tooth lengths are substantionally higher for Orange Juice in the lower dosage cases, and that tooth length increases along with the dose of Vitamin C for both Ascorbic Acid and Orange Juice supplement methods.

Let's dive in to see whether our hypotheses here are valid.

Comparison of Tooth Growth by Supplement and Dose

To test our ideas we will use the `t.test` function to perform a Student's T-test on the data with a 95% confidence level and equal variances.

Our null hypothesis in all cases is that the mean of the two groups is equal, in other words that the observed difference in tooth length can not be attributed to the supplement/dosage. The alternative hypothesis is that the observed difference in tooth lengths is statistically significant, in other words the supplement/dosage has an effect on the tooth length.

We will reject the null hypothesis in favor of the alternative hypothesis when we find a p-value of less than 0.05.

Effects of Supplement

First, we split the data into subsets for clarity in the analyses below.

```
Dose05 <- ToothGrowth[ToothGrowth$Dose == 0.5,]
Dose10 <- ToothGrowth[ToothGrowth$Dose == 1.0,]
Dose20 <- ToothGrowth[ToothGrowth$Dose == 2.0,]
OJ <- ToothGrowth[ToothGrowth$Supplement == "OJ",]
VC <- ToothGrowth[ToothGrowth$Supplement == "VC",]
```

Now we'll test the effect of the supplement type on tooth length while holding dosage constant.

```
# Perform T-test on VC vs OJ at 0.5 mg dose
Test.Dose05 <- t.test(Length ~ Supplement, data=Dose05, var.equal=FALSE)
# Perform T-test on VC vs OJ at 1.0 mg dose
Test.Dose10 <- t.test(Length ~ Supplement, data=Dose10, var.equal=FALSE)
```

```r
# Perform T-test on VC vs OJ at 2.0 mg dose
Test.Dose20 <- t.test(Length ~ Supplement, data=Dose20, var.equal=FALSE)
```

Let's take a quick look at a summary of the results:

Dose.Level	p.value	Conf.Int.min	Conf.Int.max
0.5 mg	0.0063586	1.7191	8.7809
1.0 mg	0.0010384	2.8021	9.0579
2.0 mg	0.9638516	-3.7981	3.6381

Effects of Dosage

Now we'll test the effect of dosages on tooth length:

```r
# Perform t test on 0.5 mg vs 1.0 mg, within each supplement
TestVC0510<-t.test(Length ~ Dose, data=VC[VC$Dose == 0.5 | VC$Dose == 1.0,], var.equal=T)
TestOJ0510<-t.test(Length ~ Dose, data=OJ[OJ$Dose == 0.5 | OJ$Dose == 1.0,], var.equal=T)
# Perform t test on 1.0 mg vs 2.0 mg, within each supplement
TestVC1020<-t.test(Length ~ Dose, data=VC[VC$Dose == 1.0 | VC$Dose == 2.0,], var.equal=T)
TestOJ1020<-t.test(Length ~ Dose, data=OJ[OJ$Dose == 1.0 | OJ$Dose == 2.0,], var.equal=T)
# Perform t test on 0.5 mg vs 2.0 mg, within each supplement
TestVC0520<-t.test(Length ~ Dose, data=VC[VC$Dose == 0.5 | VC$Dose == 2.0,], var.equal=T)
TestOJ0520<-t.test(Length ~ Dose, data=OJ[OJ$Dose == 0.5 | OJ$Dose == 2.0,], var.equal=T)
```

The p-values and lower confidence intervals for each of these three cases follow here:

Doses.Compared	Supplement	p.value	Conf.Int.min	Conf.Int.max
0.5 mg vs 1.0 mg	VC	0.0000006	-11.2643	-6.3157
0.5 mg vs 1.0 mg	OJ	0.0000836	-13.4108	-5.5292
1.0 mg vs 2.0 mg	VC	0.0000340	-12.9690	-5.7710
1.0 mg vs 2.0 mg	OJ	0.0373628	-6.5005	-0.2195
0.5 mg vs 2.0 mg	VC	0.0000000	-21.8328	-14.4872
0.5 mg vs 2.0 mg	OJ	0.0000003	-16.2782	-9.3818

Conclusion

Based on the above analysis, there is a definite dependence of the dosage level on tooth growth. The p-values for all tests are less than the threshold and the confidence intervals do not include zero, indicating that the increase in tooth growth when the supplement dose is increased is significant.

Furthermore, we see there is a correlation between the supplement type (Orange Juice vs Ascorbic Acis) and tooth growth that disappears at a higher dose. For 0.5 mg and 1.0 mg doses, we see that the p-values are below our threshold and the confidence interval does not include zero, indicating that the increase in tooth growth that's seen with OJ vs VC is statistically significant. However, at 2.0 mg this discrepancy vanishes and our null hypothesis at this level is accepted.

Appendix

The following code was used to make the plots and tables in this report.

Multiplot function

Used to plot multiple plots in a single instance for ggplot2.

```
# Multiple plot function
#
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)
# - cols:   Number of columns in layout
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.
#
# If the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE),
# then plot 1 will go in the upper left, 2 will go in the upper right, and
# 3 will go all the way across the bottom.
#
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

 if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}
```

Boxplots

```
# Load and prepare Orange Juice data for plot
OJ <- ToothGrowth[ToothGrowth$Supplement == "OJ",]
plot.OJ <- ggplot(OJ, aes(x=factor(Dose),y=Length,fill=factor(Dose))) +
    geom_boxplot() +
    scale_x_discrete("Dose (mg)") +
    scale_y_continuous("Tooth Length", limits=c(4, 34)) +
    ggtitle("Orange Juice") +
    guides(fill=FALSE)

# Load and prepare Ascoribic Acid data for plot
VC <- ToothGrowth[ToothGrowth$Supplement == "VC",]
plot.VC <- ggplot(VC, aes(x=factor(Dose),y=Length,fill=factor(Dose))) +
    geom_boxplot() +
    scale_x_discrete("Dose (mg)") +
    scale_y_continuous("Tooth Length", limits=c(4, 34)) +
    ggtitle("Ascorbic Acid") +
    guides(fill=FALSE)

# Plot the OJ and VC plots created above.
multiplot(plot.OJ, plot.VC, cols=2)
```

Supplements Results Table

```
Supp <- data.frame(Dose.Level = 0, p.value = 0, Conf.Int.min = 0, Conf.Int.max = 0)
Supp[1,]<-c('0.5 mg',Test.Dose05$p.value,Test.Dose05$conf.int[1],Test.Dose05$conf.int[2])
Supp[2,]<-c('1.0 mg',Test.Dose10$p.value,Test.Dose10$conf.int[1],Test.Dose10$conf.int[2])
Supp[3,]<-c('2.0 mg',Test.Dose20$p.value,Test.Dose20$conf.int[1],Test.Dose20$conf.int[2])
Supp[,2]<-round(as.numeric(Supp[,2]), 8)
Supp[,3]<-round(as.numeric(Supp[,3]), 4)
Supp[,4]<-round(as.numeric(Supp[,4]), 4)
kable(Supp)
```

Dose Results Table

```
Dose<-data.frame(Doses.Compared = 0, Supplement = 0, p.value = 0, Conf.Int.min = 0,
                   Conf.Int.max = 0)
Dose[1,]<-c('0.5 mg vs 1.0 mg', 'VC', TestVC0510$p.value, TestVC0510$conf.int[1],
            TestVC0510$conf.int[2])
Dose[2,]<-c('0.5 mg vs 1.0 mg', 'OJ', TestOJ0510$p.value, TestOJ0510$conf.int[1],
            TestOJ0510$conf.int[2])
Dose[3,]<-c('1.0 mg vs 2.0 mg', 'VC', TestVC1020$p.value, TestVC1020$conf.int[1],
            TestVC1020$conf.int[2])
Dose[4,]<-c('1.0 mg vs 2.0 mg', 'OJ', TestOJ1020$p.value, TestOJ1020$conf.int[1],
            TestOJ1020$conf.int[2])
Dose[5,]<-c('0.5 mg vs 2.0 mg', 'VC', TestVC0520$p.value, TestVC0520$conf.int[1],
            TestVC0520$conf.int[2])
Dose[6,]<-c('0.5 mg vs 2.0 mg', 'OJ', TestOJ0520$p.value, TestOJ0520$conf.int[1],
```

```
            TestOJ0520$conf.int[2])
Dose[,3]<-round(as.numeric(Dose[,3]), 8)
Dose[,4]<-round(as.numeric(Dose[,4]), 4)
Dose[,5]<-round(as.numeric(Dose[,5]), 4)
kable(Dose)
```