ToohGrowthAnalysis

S Fraser

June 3, 2016

Overview

This paper will use the ToothGrowth dataset available in R to see what effect a delivery method of Orange Juice (OJ) or Ascorbic Acid (VC) used to deliver doses of Vitamin C in amounts of .5, 1.0 or 2.0 mg/day has on growth of teeth in guinea pigs.

A description of the data can be found at https://bugs.r-project.org/bugzilla3/show_bug.cgi?id=15953.

Assumptions

It is assumed that we are dealing with 60 guinea pigs. Each guinea pig received one of three dose levels of Vitamin C (0.5, 1.0, or 2.0 mg/day) in one of two delivery methods (supp) - orange juice (OJ) or an aqueous solution of ascorbic acid (VC). It is also assumed that each guinea pig received the same delivery method and dose each day.

Exploratory Data Analysis

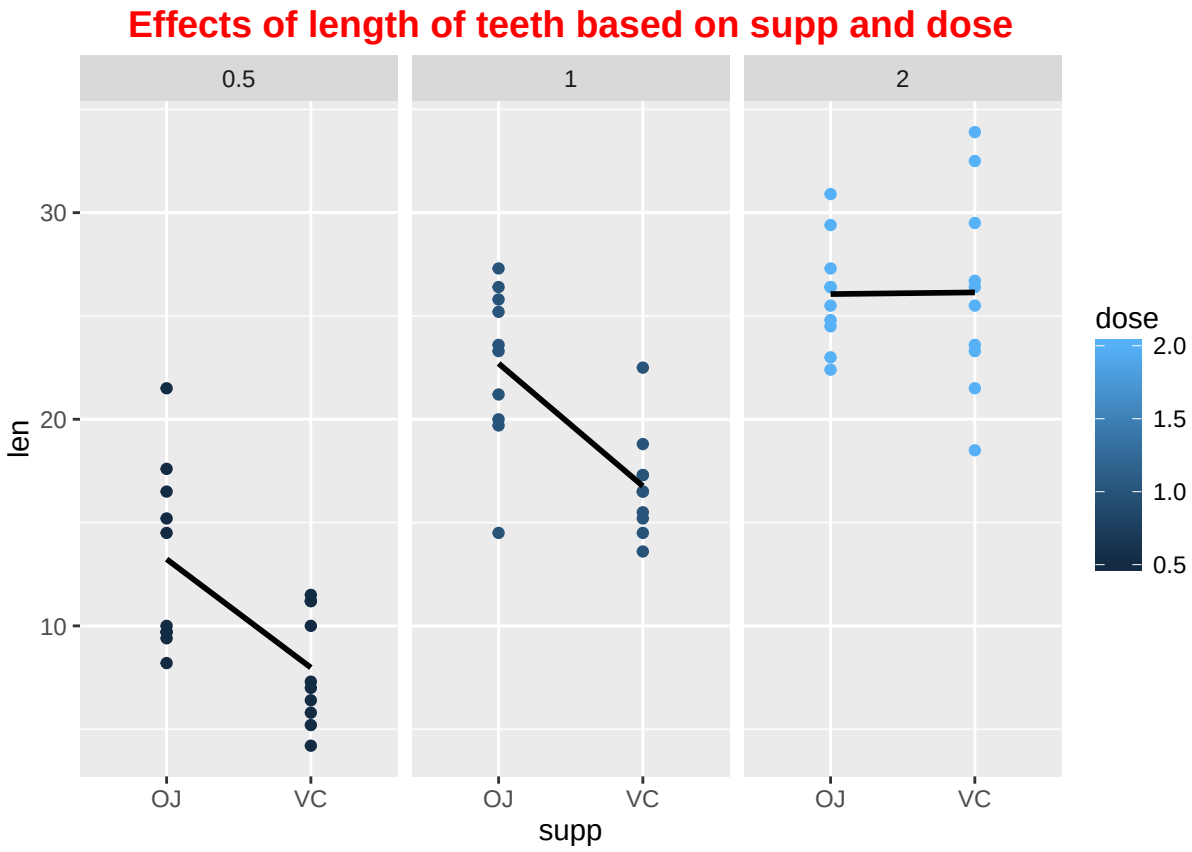
1. Load the ToothGrowth data and perform some basic exploratory data analysis

```
library(data.table)
library(ggplot2)
data(ToothGrowth)

## First copy the ToothGrowth data set to a local variable
TG <- ToothGrowth
```

Plot the data to show the effects of tooth length (len) against dose and delivery method (supp)

```
## plot the raw data
g <- ggplot(TG,aes(x=supp,y=len,colour=dose))
g <- g+ggtitle("Effects of length of teeth based on supp and dose") +
     theme(plot.title = element_text(size = 14, colour = 'red',face="bold"))
g <- g+geom_point()
g <- g+ stat_summary(aes(group = 1),fun.y = mean, geom="line",size=1,col="black")
g <- g+facet_grid(.~ dose)
g
```



From this plot we can see that increasing dose definitely increases tooth growth and that when the dose is lower, the delivery method of OJ increases tooth growth more than VC. We will do more work to prove this using confidence intervals and hypothesis testing.

2. Provide a basic summary of the data.

```
## Let's see how much data we have
dim(TG)
```

```
## [1] 60  3
```

```
## A summary look at the data
summary(TG)
```

```
##       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
```

```
## Take a look at the column names
names(TG)
```

```
## [1] "len"  "supp" "dose"
```

```
## Check out the values for the supplement and dose columns
table(TG$supp)
```

```
##
## OJ VC
## 30 30
```

```
table(TG$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

```
## Show some of the values of the tooth length (len) column
head(TG$len)
```

```
## [1]  4.2 11.5  7.3  5.8  6.4 10.0
```

```
## What are the data types of each column
str(TG)
```

```
## '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 ...
```

Hypothesis Tests / Confidence Intervals

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

Assume a null hypothesis that the delivery method (supp) of Orange Juice (OJ) creates greater growth in teeth in guinea pigs than ascorbic acid (VC) as seems to be shown in the graph above.

Tooth Growth as a result of Supplement Only

```
## t interval testing of supp vs. len
t.test(len ~ supp, paired = FALSE, var.equal = FALSE,data = TG)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Reject the null hypothesis: According to the confidence interval of '-0.1710156, 7.5710156', Which contains 0, and the p-value of 0.06063, which is greather then 0.05, we can reject the null hypothesis.

Tooth Growth by Supplement and Dose

```
## t interval testing of supp vs. len for each dose value
## Create subsets of the data using the dose values
D1 <- subset(TG,dose == 0.5)

t.test(len ~ supp,paired = FALSE,var.equal = FALSE,data=D1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

Accept the null hypothesis: We can say that there is a difference in the delivery method when delivering .5 mg/day of Vitamin C based on a probability lower than 0.05 (0.006359) and the confidence interval of '1.719057, 8.780943' not containing zero.

```
D2 <- subset(TG,dose == 1.0)

t.test(len ~ supp,paired = FALSE,var.equal = FALSE,data=D2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

Accept the null hypothesis: We can say that there is a difference in the delivery method when delivering 1.0 mg/day of Vitamin C based on a probability lower than 0.05 (0.001038) and the confidence interval of '2.802148, 9.057852' not containing zero.

```
D3 <- subset(TG,dose == 2.0)

t.test(len ~ supp,paired = FALSE,var.equal = FALSE,data=D3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Reject the null hypothesis: We cannot say that there is a difference in the delivery method when delivering 2.0 mg/day of Vitamin C. According to the confidence interval of '-3.79807, 3.63807', Which contains 0, and the p-value of 0.9639, which is greather then 0.05, we can reject the null hypothesis.

Conclusions

4. State your conclusions and the assumptions needed for your conclusions.

Assumptions can be found above under the heading 'Assumptions'.

From the graph and the hypothesis testing, it can be concluded that in lower doses of 0.5 or 1.0 mg/day, the delivery method of OJ contributes to greater growth in the length of guinea pig teeth than VC, but at dose of 2.0 mg/per day, the delivery method does not seemt to matter.

1. Higher doses of Vitamin C contribute to greater tooth growth.

2. OJ also contributes to greater tooth growth when used as a delivery method for lower doses of Vitamin C.