Part 2: Basic Inferential Data Analysis

Sergio Paz

Overview

Analyze the ToothGrowth data in the R datasets package.

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
data("ToothGrowth")
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
```

```
dim(ToothGrowth)
```

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

```
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 ...
```

Provide a basic summary of the data

```
summary(ToothGrowth)
```

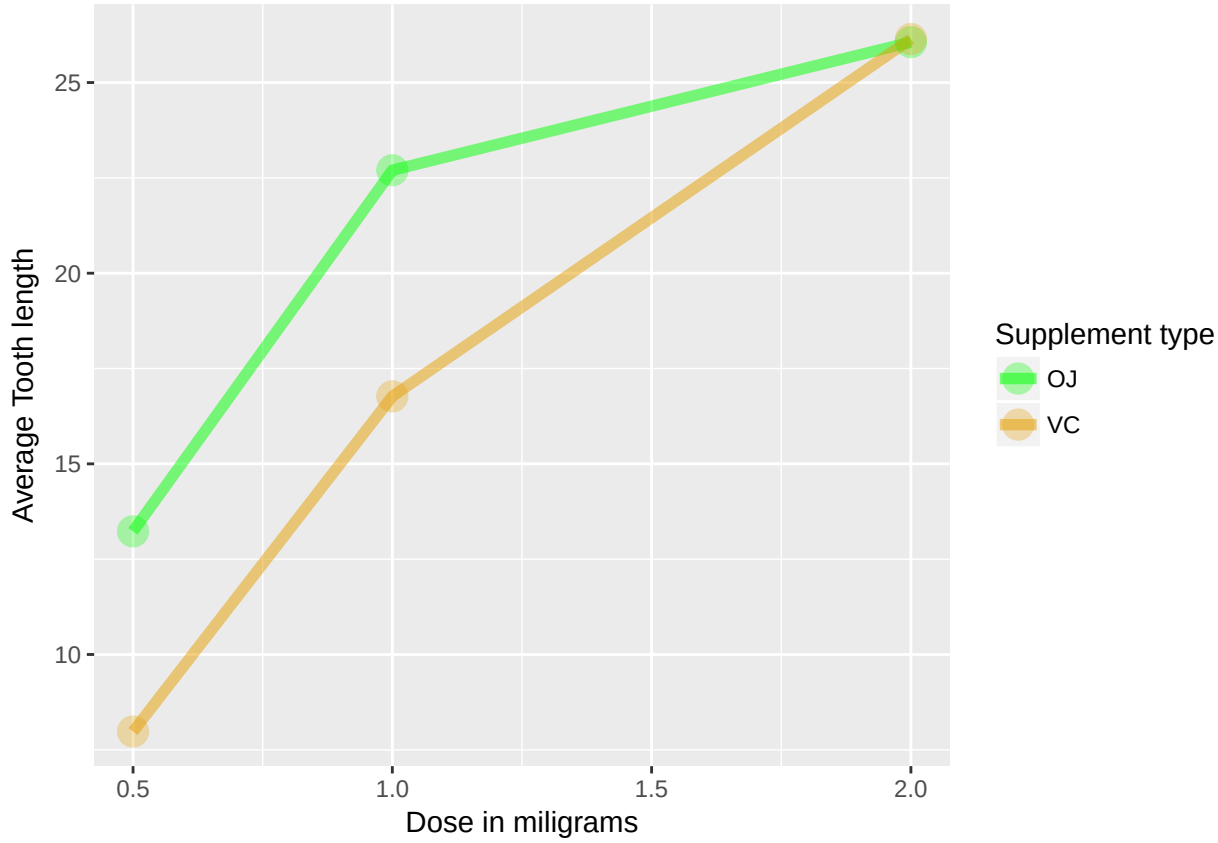
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

```
library(ggplot2)
library(plyr)
agg <- aggregate(len ~ dose + supp, ToothGrowth, mean)
ggplot(agg, aes(x=dose, y=len, colour = supp)) +
    geom_line(size=2, alpha=.5) + geom_point(size=5, alpha=.3) +
    xlab("Dose in miligrams") + ylab("Average Tooth length") +
    guides(colour=guide_legend(title="Supplement type")) +
    scale_color_manual(values = c("green", "#E69F01"))
```



There seems to be a correlation between the dose and tooth growth, where Orange Juice is more effective for lower doses, where the 2 milligram seems to be the maximum effect point.

```
ddply(ToothGrowth, dose ~ supp,function(x)
      c(mean=mean(x$len), sd=sd(x$len),
        conf.int=t.test(x$len)$conf.int))
```

```
##   dose supp  mean       sd conf.int1 conf.int2
## 1  0.5   OJ 13.23 4.459709 10.039717 16.420283
## 2  0.5   VC  7.98 2.746634  6.015176  9.944824
## 3  1.0   OJ 22.70 3.910953 19.902273 25.497727
## 4  1.0   VC 16.77 2.515309 14.970657 18.569343
## 5  2.0   OJ 26.06 2.655058 24.160686 27.959314
## 6  2.0   VC 26.14 4.797731 22.707910 29.572090
```

We observe that in 95% confidence interval the `Ascorbic Acid(VC)` intervals are pairwise disjoint so we can claim with high level of confidence that the length means are distinct, moreover there is a clear growth correlation between dose & length means.

By now we can also immediately identify with high level of confidence that For 0.5 and 1 milligrams Orange Juice have has greater impact on tooth growth (On the merit that for those 2 doses there confidence interval are pairwise disjoint).

For Orange Juice(OJ) supplement type, however, there is an overlap for dose 1 and 2 milligrams, and we are forced to look deeper.

```
t.test(len ~ dose, paired=FALSE, var.equal=TRUE,
       data=subset(ToothGrowth, dose %in% c(1.0,2.0) & supp == 'OJ'))
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -2.2478, df = 18, p-value = 0.03736
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5005017 -0.2194983
## sample estimates:
## mean in group 1 mean in group 2
##           22.70           26.06
```

The t value `-2.2477612` being less than **`qt(.025, 18)`** == `-2.100922` allows us to assert that the mean length for 2 milligrams as greater than the for the 1 milligram dose.

In the 2.0 milligram dose there is an overlap between `Orange Juice (OJ)` and `Ascorbic Acid (VC)` let's dig deeper

```
t.test(len ~ supp, paired=FALSE, var.equal=FALSE,
       data=subset(ToothGrowth, dose == 2.0))
```

```
##
##  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
```

The confidence Interval includes 0 and hence difference between the supplements types vis-a-vis mean lengths is insignificant.

State your conclusions and the assumptions needed for your conclusions

The analysis has shown with high confidence that the there is a correlation between the supplement type used and teeth growth in guinea pigs, when for small doses of 0.5 and 1 milligrams, Orange Juice, clearly has an advantage.