Statistical Inference Course Project - ToothGrowth Dataset Analysis (Part 2)

Andrey Fedyna

Synopsis

We're going to analyze the ToothGrowth data in the R datasets package. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice **(OJ)** or ascorbic acid (a form of vitamin C and coded as **VC**).

The research was conducted in four directions.
1. Load the ToothGrowth data and perform some basic exploratory data analyses.
2. Provide a basic summary of the data.
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).
4. State your conclusions and the assumptions needed for your conclusions.

ToothGrowth data analysis

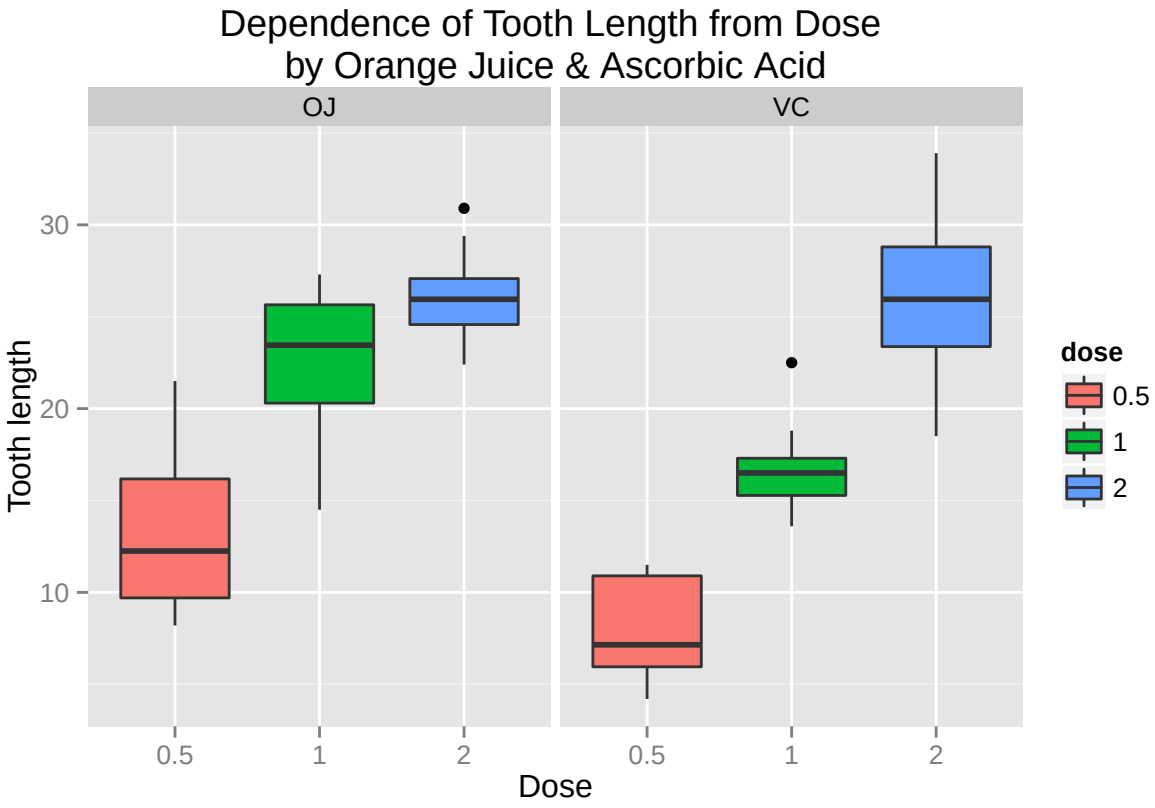
```
library(ggplot2)
library(datasets)

data("ToothGrowth")
```

Looking at the data, it has 60 observations on 3 variables.
len - Tooth length (numeric)
supp - Supplement type (VC or OJ) (factor)
dose - Dose in milligrams (numeric)

All calculations can be found in the appendix at the end of the document.

Construct a graph$^{(0)}$ corresponding to our dataset:



Answer for Q1: Load the ToothGrowth data and perform some basic exploratory data analyses.

Hypothesis: above from the graph shows that with same dose level of Vitamin C (0.5, 1, 2), in average (tooth lenght), OJ (orange juice) delivery method has bigger effect than VC (ascorbic acid) for the tooth growth.

Answer for Q2: Provide a basic summary of the data.

For both Supplement type and Doses, the number of observations are the same [(1)], but its not clear if the two methods were tested on the same set of 30 guinea pigs or not?

The samples are independent[(2)] and we should use **paired=FALSE** in our t.test.

Variances two sets of data responsible for the different supply methods are different[(3)]. Accordingly **var.equal=FALSE**.

Answer for Q3: 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)

For a dose of 0.5, our 95% confidence interval **(1.72, 8.78)**[(4)] does not contain zero. Also the evaluation of the mean **(13.23(OJ), 7.98(VC))**[(4)], we can conclude that at doses of 0.5, OJ has a high statistical significance for the growth of tooth than VC.

For a dose of 1.0, our 95% confidence interval **(2.80, 9.06)**[(4)] also does not contain zero. And from the evaluation of the mean **(22.70(OJ), 16.77(VC))**[(4)], we can conclude that at doses of 1.0, OJ has a high statistical significance for the growth of tooth than VC.

For a dose of 2.0, our 95% confidence interval **(-3.80, 3.64)**[(4)] contains zero and same evaluation of means. We can conclude that at 2.0 dose level, both OJ and VC has no difference in tooth growth.

Answer for Q4: State your conclusions and the assumptions needed for your conclusions.

Both Orange Juice and Ascorbic Acid seems giving growth for guinea pig, but with different dose levels. To be accurate, OJ has a significant growth at a dose of 1.0 while for VC shows good results in a dose level - 2.0.

Thus, we can conclude that the orange juice shows a higher growth rate at a lower dosage, showing their superiority over ascorbic acid.

Appendix: (i)

(0)

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ggplot(ToothGrowth, aes(x = dose, y = len, fill = dose)) +
     geom_boxplot() + facet_grid(. ~ supp) +
     labs(title = "Dependence of Tooth Length from Dose by
          Orange Juice & Ascorbic Acid", x = "Dose", y = "Tooth length")
```

(1)

```
summary(ToothGrowth)
```

```
##       len        supp     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
```

```
OJdata <- ToothGrowth[ToothGrowth$supp=="OJ",]
VCdata <- ToothGrowth[ToothGrowth$supp=="VC",]
dim(OJdata); dim(VCdata)
```

```
## [1] 30  3
```

```
## [1] 30  3
```

(2)

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

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

(3)

```r
var(OJdata$len); var(VCdata$len)
```

```
## [1] 43.63344
```

```
## [1] 68.32723
```

(4)

```r
Dose_0.5 <- ToothGrowth[ToothGrowth$dose==0.5, ]
tt05 <- t.test(len~supp, data=Dose_0.5, paired=FALSE, var.equal=FALSE)
tt05$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```r
tt05$estimate
```

```
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
Dose_1 <- ToothGrowth[ToothGrowth$dose==1, ]
tt1 <- t.test(len~supp, data=Dose_1, paired=FALSE, var.equal=FALSE)
tt1$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```r
tt1$estimate
```

```
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
Dose_2 <- ToothGrowth[ToothGrowth$dose==2, ]
tt2 <- t.test(len~supp, data=Dose_2, paired=FALSE, var.equal=FALSE)
tt2$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```r
tt2$estimate
```

```
## mean in group OJ mean in group VC
##            26.06            26.14
```