Tooth Growth Analysis

Tomas Gogorza

Tuesday, September 15, 2015

Summary

This document shows a quick analysis and statistical comparison of 2 different supplements (orange juice or ascorbic acid) and 3 different doses (0.5, 1 and 2 mg) used to analyze their influence in tooth growth on guinea pigs.

Data Analysis

Load data

```
library(ggplot2)
library(dplyr)

data("ToothGrowth")
data <- as.data.frame(ToothGrowth)
summary(data)
```

```
##       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  
```

From a quick overview of the data, we can see that there are 2 different supplements (OJ & VC) and the dosage range from 0.5 to 2.
Figure 1 in the Appendix shows the samples distribution to give a graphical sense of the difference between supplements and dosage.

Comparison

In order to determine which kind of supplement type is more efficient (if any), we'll perform a dose by dose comparison between them. We'll say that the difference of means between both supplements is 0, meaning there's no difference between the supplements, and take that hypothesis as H0.

Small Dose (0.5 mg)

We perform a T test on the small dosage samples for both supplements:

```r
sdoseoj <- select(filter(data, supp == "OJ" & dose==0.5),len)
sdosevc <- select(filter(data, supp == "VC" & dose==0.5),len)
smalltest <- t.test(sdoseoj,sdosevc,paired = FALSE,var.equal = FALSE)
smalltest
```

```
##
##  Welch Two Sample t-test
##
## data:  sdoseoj and sdosevc
## 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 of x mean of y
##     13.23      7.98
```

We can see the confidence interval, ranging from **1.7190573** to **8.7809427**, doesn't contain 0, and so we can reject H0 with 95% confidence, therefore one of the supplements is more effective than the other when using a 0.5 mg dose. Moreover, with a p value of **0.0063586%**, a lot smaller than 5%, we can assert that the null hypothesis if false.

Medium Dose (1 mg)

We perform a T test on the medium dosage samples for both supplements:

```r
mdoseoj <- select(filter(data, supp == "OJ" & dose==1),len)
mdosevc <- select(filter(data, supp == "VC" & dose==1),len)
medtest <- t.test(mdoseoj,mdosevc,paired = FALSE,var.equal = FALSE)
medtest
```

```
##
##  Welch Two Sample t-test
##
## data:  mdoseoj and mdosevc
## 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 of x mean of y
##     22.70     16.77
```

Again, we can see the confidence interval, ranging from **2.8021482** to **9.0578518**, doesn't contain 0, and so we can reject H0 with 95% confidence, therefore one of the supplements is also more effective than the other when using a 1 mg dose. Moreover, with a p value of **0.0010384%**, a lot smaller than 5%, we can assert that the null hypothesis if false.

Large Dose (2 mg)

We perform a T test on the large dosage samples for both supplements:

```
ldoseoj <- select(filter(data, supp == "OJ" & dose==2),len)
ldosevc <- select(filter(data, supp == "VC" & dose==2),len)
largetest <- t.test(ldoseoj,ldosevc,paired = FALSE,var.equal = FALSE)
largetest
```

```
##
##  Welch Two Sample t-test
##
## data:  ldoseoj and ldosevc
## 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 of x mean of y
##     26.06     26.14
```

In this case, with a CI in the range of **-3.7980705** to **3.6380705**, we can see that the mean difference of 0 falls within our interval, meaning we cannot rule out H0 for the case of tests with a 2 mg dose. A p value **0.9638516%** shows support for the null hypothesis as well.

OJ vs VC

Finally, an overall test between samples taking VC and samples using OJ for tooth growth:

```
oj <- rbind(sdoseoj,mdoseoj,ldoseoj)
vc <- rbind(sdosevc,mdosevc,ldosevc)
ojvsvc <- t.test(oj,vc,paired = FALSE,var.equal = FALSE)
ojvsvc$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
ojvsvc$p.value
```

```
## [1] 0.06063451
```

Again, with a CI in the range of **-0.1710156** to **7.5710156**, we can see that the mean difference of 0 falls within our interval, meaning we cannot rule out H0 in the overall comparison. A p value **0.0606345%** also shows support for the null hypothesis in this case.

Conclusions and Assumptions

After performing tests case by case, it can be stated that when using 0.5 mg and 1 mg doses, the orange juice (OJ) is much more effective than the ascorbic acid (VC), as we can also see in Figure1. In the case of large doses, we cannot determine which, if any, supplement works better. As a final note, some assumptions had to be made in order to perform the analysis. We assumed that that the variances between groups were unequal so when running t tests the parameter var.equal = FALSE was used. Also, for hypothesis testing, we took H0 to consider that if the mean of the difference between groups is 0, we cannot ensure one supplement to be more effective than the other.

Appendix

Figure 1

```
ggplot(data, aes(x=factor(dose),y=len)) + geom_boxplot() + facet_grid(.~supp) +
    labs(title="Dose vs Length by Supplement Type", y="Length (mm)", x="Dose (mg)")
```

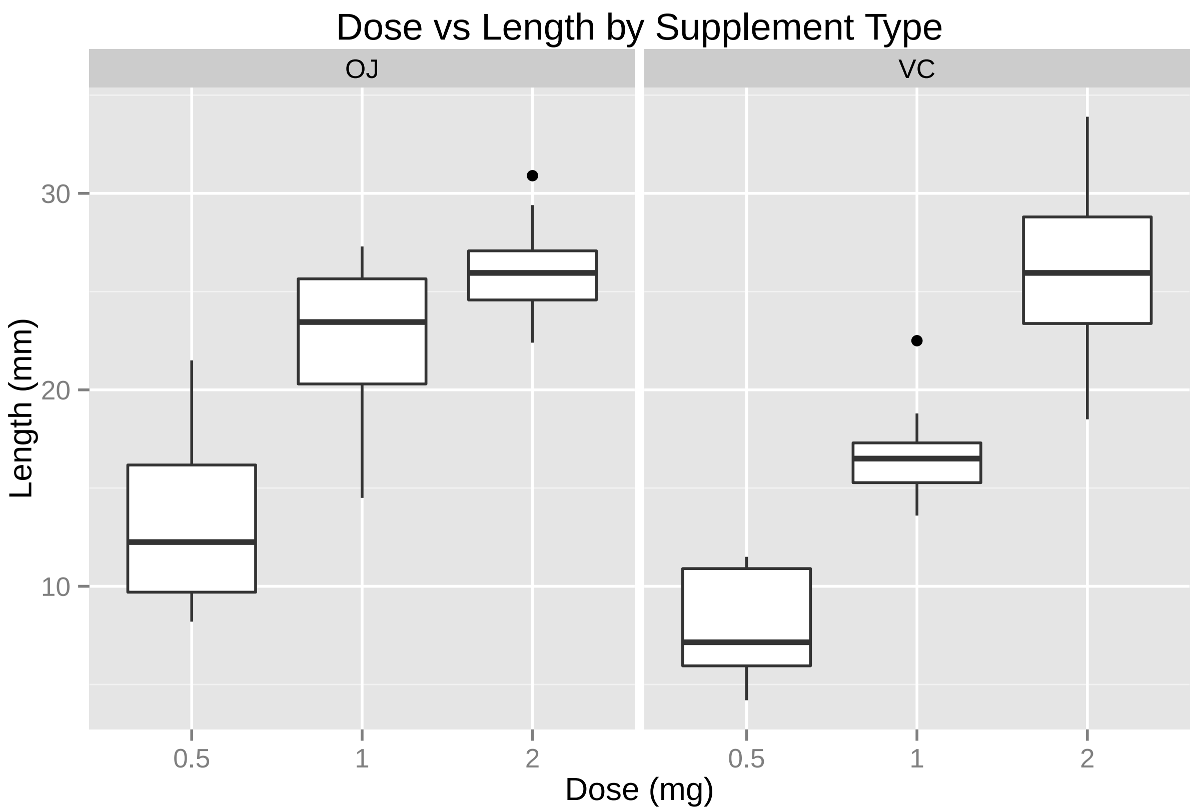

Figure 1. Distribution of tooth length by dosage, grouped by supplement type