ToothGrowth analysis

Gill Mundin

15 August 2019

1 - Overview

This project investigates the ToothGrowth data and tests whether type and dose of two supplements influence tooth length.

4 - Data exploration

Looking at the head and the dimensions of the dataframe.

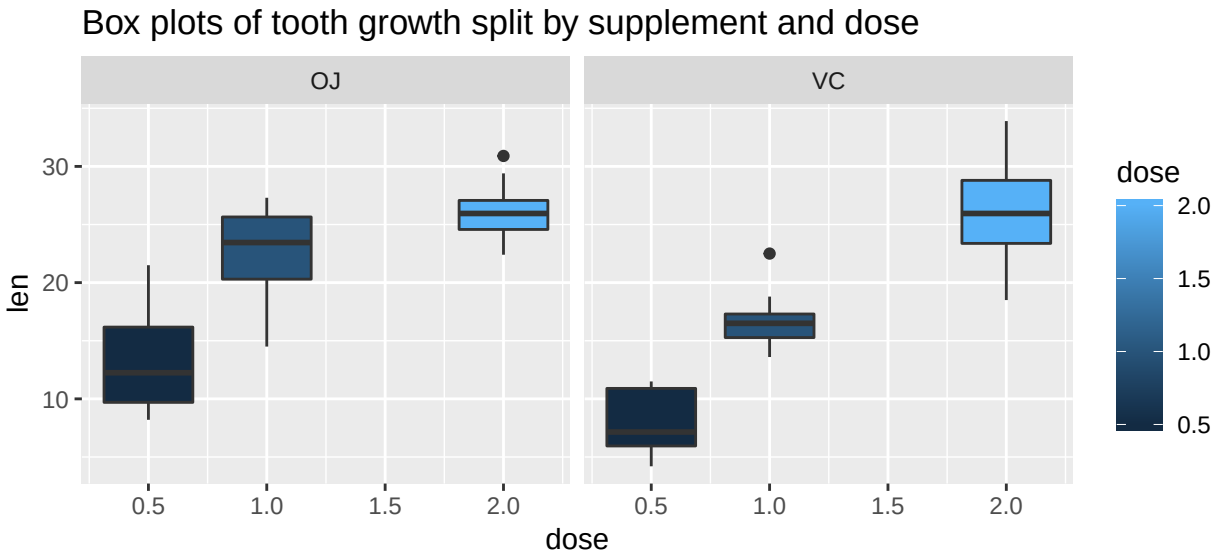
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

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

Summary of means and SDs for each supplement split by dose:

```
##   Group.1 Group.2    x.Mean      x.SD
## 1      OJ     0.5 13.230000  4.459709
## 2      VC     0.5  7.980000  2.746634
## 3      OJ     1.0 22.700000  3.910953
## 4      VC     1.0 16.770000  2.515309
## 5      OJ     2.0 26.060000  2.655058
## 6      VC     2.0 26.140000  4.797731
```

Exploratory plots



2 - Assumptions

There are a number of comparisons that could be made and each comparison increases the probability of a type 1 error. I have kept the comparisons to a minimum, but have not controlled the type 1 error to be 5% overall. The conclusions may therefore not be valid, and alternative tests (eg, ANOVA) may be more appropriate.

Based on the summary statistics and differences in the SDs between the supplements, we cannot assume equal variance. There is a small n number, so I have chosen to use Gosset's (Student's) t-test and confidence intervals rather than a normal distibution. The data are treated as unpaired.

I am making the assumption that the desirable outcome is a greater increase in tooth length, and that you would want to choose the supplement and dose that achieves this.

3 - Hypotheses

First possible hypothesis

H0 : Supplement dose does not affect tooth length Ha : Tooth length increases with supplement dose

Second possible hypothesis

H0 : There is no difference between VC and OJ in relation to tooth length Ha : There is a difference between VC and OJ in relation to tooth length, in either direction

Testing first hypothesis

The code and outcome of the ttests are included in the appendix.

Compare the following subsets, H0 is that there is no difference between doses:

- OJlo vs OJhi
- OJmid vs OJhi
- VClo vs VChi
- VCmid vs VChi

For each of the comparisons, the 95% confidence interval does not contain 0, and the p-values are small and <0.05, meaning that the H0 should be rejected, and that there is a statistical difference between the doses compared (recognising that we haven't controlled the type 1 error rate for multiple testing.)

Testing the second hypothesis

Given that there is a difference in tooth length between the mid and highest doses for each supplement, I have chosen to compare just the highest dose of the supplements to test the hypothesis that there is no difference between VC and OJ in relation to tooth length.

- OJhi vs VChi

At the highest dose, comparing the OJ and VC supplements results in a confidence interval that contains 0 and a p-value of 0.9, meaning that there is no evidence to reject H0, and we conclude that the two supplements have the same effect on tooth length at the highest dose.

Conclusions

Two supplements were tested at 3 doses. Under the assumption that greater tooth length is desirable, it is beneficial to give the highest dose of 2 mg of either supplement, however, it does not appear to matter which supplement is chosen. From the box plots and summary statistics, the data for OJ suggests that it is less variable than VC, however the sample size is small and this may change with more observations.

Appendix

Creating data subsets

```
tg <- ToothGrowth
VClo <- tg$len[1:10]
VCmid <- tg$len[11:20]
VChi <- tg$len[21:30]
OJlo <- tg$len[31:40]
OJmid <- tg$len[41:50]
OJhi <- tg$len[51:60]
```

Dose comparisons

OJ lo vs hi

```
t.test(OJlo, OJhi, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJlo and OJhi
## t = -7.817, df = 14.668, p-value = 1.324e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.335241  -9.324759
## sample estimates:
## mean of x mean of y
##     13.23     26.06
```

OJ mid vs hi

```
t.test(OJmid, OJhi, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJmid and OJhi
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean of x mean of y
##     22.70     26.06
```

VC lo vs hi

```
t.test(VClo, VChi, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VClo and VChi
## t = -10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.90151 -14.41849
## sample estimates:
## mean of x mean of y
##      7.98     26.14
```

VC mid vs hi

```
t.test(VCmid, VChi, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VCmid and VChi
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean of x mean of y
##     16.77     26.14
```

Supplement comparison

OJhi vs VChi

```
t.test(OJhi, VChi, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  OJhi and VChi
## 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
```