

Statistical Inference: Course Project Part 2

Anne Strader

28. July 2020

Introduction

The ToothGrowth dataset (available in the R datasets package) is analyzed as follows:

1. A basic exploratory data analysis is performed, visually identifying potential impacts of supplement type and dosage level on tooth growth.
2. Two-sample t-tests are performed to investigate whether supplement types' impacts on tooth growth significantly differ at each dosage level.

Basic Exploratory Data Analysis

First, the data set is loaded and a summary is provided:

```
# load dataset
data(ToothGrowth)
# display a summary of the dataset
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 ...
```

```
# display the first five rows of the dataset
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
```

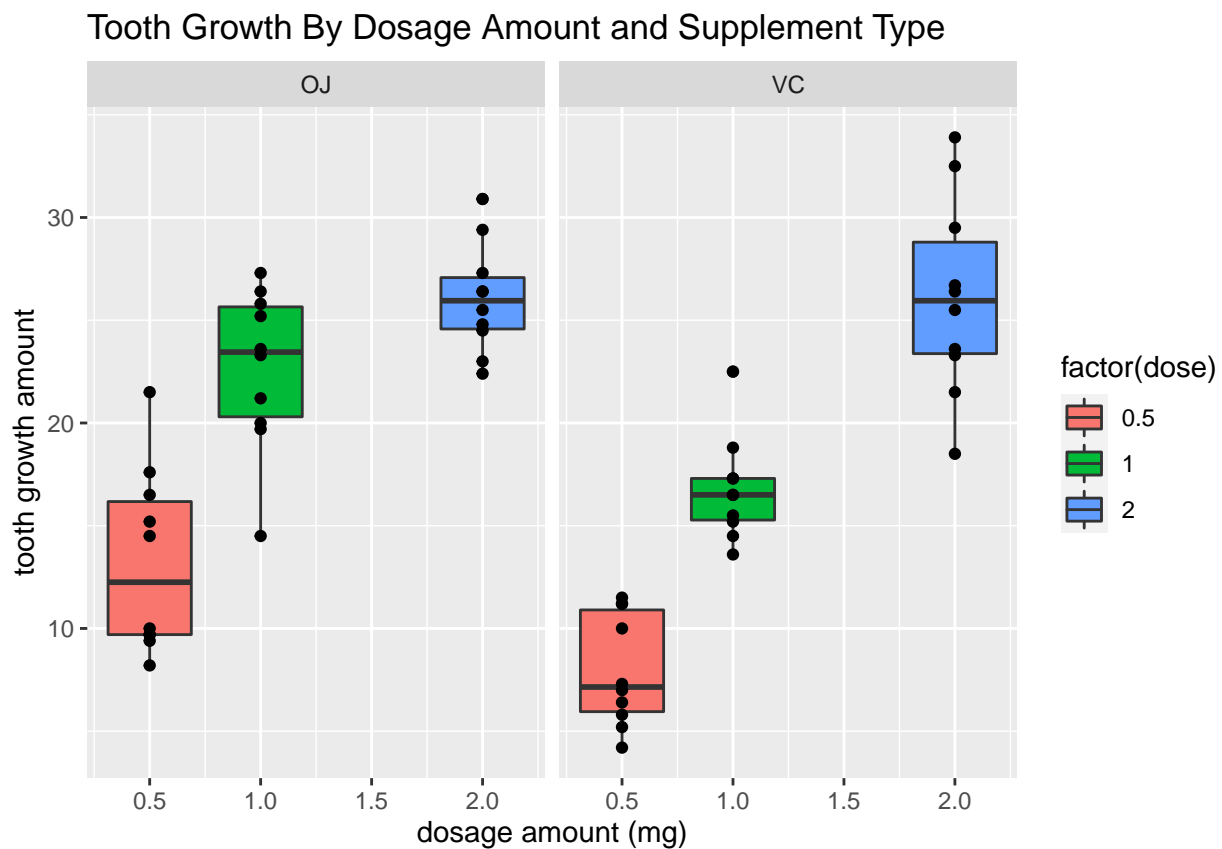
“len” and “dose” are numerical variables indicating amount of tooth growth and dosage level (mg), respectively. “supp” is a categorical variable indicating supplement type. There are two levels: “OJ” = orange juice and “VC” = ascorbic acid.

Next, the qualitative impacts of supplement type and dosage level are visualized using a boxplot:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
# initialize plot
g <- ggplot(data=ToothGrowth, aes(x = dose, y = len))
# generate boxplot and add data points to boxplot
g <- g + geom_boxplot(aes(fill = factor(dose))) + geom_point()
# generate boxplots for each supplement type ("supp")
g <- g + facet_grid(.~supp)
# add plot title and x, y axis labels
g <- g + labs(title = "Tooth Growth By Dosage Amount and Supplement Type",
              x = "dosage amount (mg)", y = "tooth growth amount")
# show plot
g
```



The plot shows three different dosage amounts: 0.5 mg, 1.0 mg and 2.0 mg. At a dosage level of 2.0 mg, supplement type does not appear to have an impact on tooth growth. However, at dosage levels of 0.5 and 1.0 mg, tooth growth amount appears to increase when the supplement type is orange juice.

Impact of Supplement Type on Tooth Growth

For each dosage amount, the following null (H_0) and alternative (H_a) hypotheses are defined:

- H_0 : Using orange juice as the supplement does not result in significantly increased tooth growth compared to using ascorbic acid as the supplement.
- H_a : Using orange juice as the supplement results in significantly increased tooth growth compared to using ascorbic acid as the supplement.

One-sided two-sample t-tests are used to test this hypothesis at each dosage level. It is assumed that the data are not paired and that the variances for each group are not necessarily equal. Additionally, sample means are assumed to be normally distributed, and each sample is assumed to be representative of its population.

A confidence level of 0.95 is used. Therefore, if the p-value is less than 0.05, the null hypothesis is rejected in favor of the alternate.

For a dosage level of 0.5 mg:

```
# define tooth growth data where the dosage level equals 0.5 mg
ToothGrowth_subset <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5))
# run two-sample t-test
t.test(len~supp, data = ToothGrowth_subset, alternative="greater")

##
## Welch Two Sample t-test
##
## data: len by supp
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604      Inf
## sample estimates:
## mean in group OJ mean in group VC
##           13.23           7.98
```

Since $p < 0.05$, the null hypothesis is rejected in favor of the alternate: that using orange juice as a supplement results in significantly increased tooth growth amount when the dosage level is 0.5 mg.

For a dosage level of 1.0 mg:

```
# define tooth growth data where the dosage level equals 1.0 mg
ToothGrowth_subset <- subset(ToothGrowth, ToothGrowth$dose %in% c(1))
# run two-sample t-test
t.test(len~supp, data = ToothGrowth_subset, alternative="greater")

##
## Welch Two Sample t-test
##
## data: len by supp
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean in group OJ mean in group VC
##           22.70           16.77
```

Since $p < 0.05$, the null hypothesis is rejected in favor of the alternate: that using orange juice as a supplement results in significantly increased tooth growth amount when the dosage level is 1.0 mg.

For a dosage level of 2.0 mg:

```
# define tooth growth data where the dosage level equals 2.0 mg
ToothGrowth_subset <- subset(ToothGrowth, ToothGrowth$dose %in% c(2))
# run two-sample t-test
t.test(len~supp, data = ToothGrowth_subset, alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -3.1335 Inf
## sample estimates:
## mean in group OJ mean in group VC
## 26.06 26.14
```

Since $p > 0.05$, the null hypothesis is not rejected, and it cannot be stated that using orange juice as a supplement causes significantly increased tooth growth at a dosage level of 2.0 mg.

Conclusions

It can be concluded that, for dosage levels less than 2.0 mg, supplement type significantly affects tooth growth amount. In this case, using orange juice as the supplement type results in significantly greater tooth growth than using ascorbic acid.