Guinea Pig Tooth Growth Analysis

Miguel Couto

4/24/2020

Overview

This report concerns the second part of the final project of the "Statistical Inference" course. Here we investigated data on tooth growth in guinea pigs, carried out some exploratory data analysis and performed a few hypothesis tests, the conclusions of which are presented at the end of the report.

Data

We investigated the database 'ToothGrowth' from the 'datasets' package on R. This data set comprises the measurement of tooth growth length on 60 guinea pigs that were given vitamin C on one of two methods and one of three different doses. More specifically, the data set contains the following 3 variables:

- len: the length of tooth growth measured
- supp: the method of delivery of vitamin C. It can either be orange juice (OJ) or ascorbic acid (VC).
- dose: the dose level of vitamin C: either 0.5, 1 or 2 mg/day.

More information about this data can be found here.

All code for this analysis was written in R and can be found in the appendix at the end of the report.

Summary of data

We started our study of this data set by performing some exploratory data analysis.

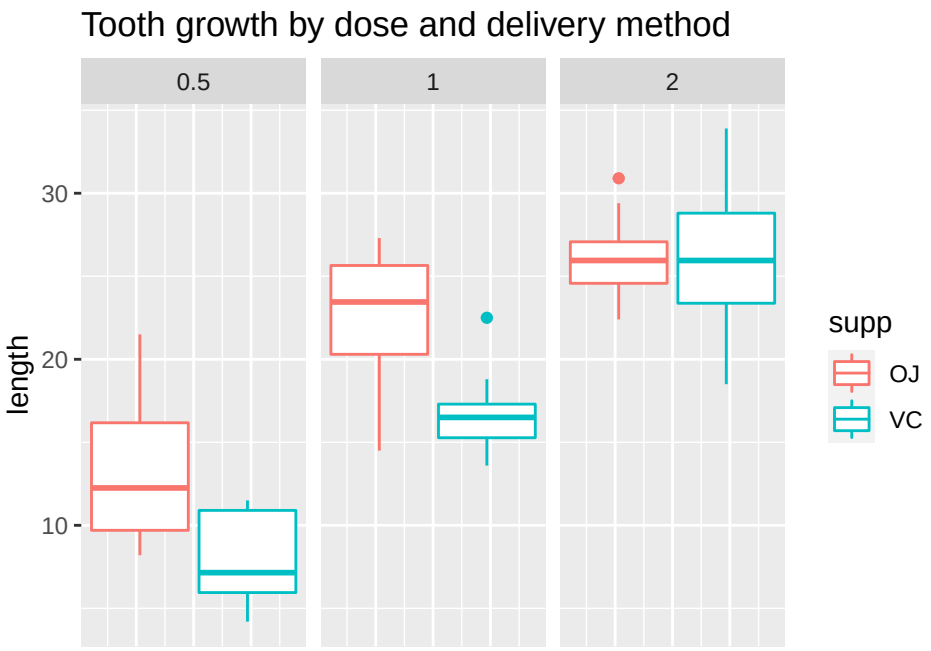
The first thing we wanted to know was the number of subjects that were given a specific dose on a specific method. As the following table shows, the 60 subjects were evenly split into six groups of 10 each receiving a specific dose on a specific delivery method.

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

Next we computed the mean and standard deviation for each of these six groups, which we present in the following table:

```
## # A tibble: 6 x 4
## # Groups:   supp [2]
##   supp   dose  mean    sd
##   <fct> <dbl> <dbl> <dbl>
## 1 OJ      0.5 13.2   4.46
## 2 OJ      1   22.7   3.91
## 3 OJ      2   26.1   2.66
## 4 VC      0.5  7.98  2.75
## 5 VC      1   16.8   2.52
## 6 VC      2   26.1   4.80
```

The average tooth growth seems to be larger for delivery method OJ when compared to VC. In order to visualize this better, we also plotted each of these groups in a box-and-whisker plot.



At first glance, on average the dose of 2mg/day seems to be linked to higher tooth growth and the dose of 0.5mg/day to be associated to lower tooth growth for both delivery methods. Moreover, the guinea pigs that were given either 0.5 or 1 mg/day of vitamin C in orange juice (OJ) seem to show a larger tooth growth than the ones given vitamin C by ascorbic acid (VC).

Hypothesis tests and results

We now perform a few hypothesis tests to ascertain the veracity of the previous statements.

We assume throughout that the tooth growth under each dose and each delivery method follows a normal distribution, all with unknown and possibly different means and variances.

Tooth growth versus delivery method, with fixed dose

For each dose d, we consider the following hypotheses:

H0: 'the average tooth growth for dose d is the same under delivery methods OJ and VC.'

H1: 'the average tooth growth for dose d is larger under the delivery method OJ than under the delivery method VC.'

Assuming H0 is true, the standardized mean difference follows a t-distribution. We must then perform a 2-sample one-sided t-test for 2 populations with unknown different variances (also known as a Welch's t-test).

```
## $`0.5`
## [1] 0.003179303
##
## $`1`
## [1] 0.0005191879
##
## $`2`
## [1] 0.5180742
```

Conclusion: we have strong evidence to conclude that the average tooth growth is larger on guinea pigs that were given either 0.5 or 1 mg/day of vitamin C by orange juice than by ascorbic acid.

It is not surprising that we were not able to conclude the same for the dose of 2mg/day. Note that, as shown in a table above, the sample averages are approximately equal (to 26.1) for both delivery methods.

Tooth growth versus dose, with fixed delivery method

We now compare the average tooth growth against the three different doses for a specific delivery method. We start by comparing the doses of 1 and 2mg/day. Let s be a delivery method and consider the hypotheses:

H0: 'the average tooth growth for method s is the same for doses of 1 and 2mg/day.'

H1: 'the average tooth growth for method s is smaller for the dose of 1mg/day than for 2mg/day.'

Once again, we perform a 2-sample one-sided t-test for 2 populations with different variances, and obtain the following p-values:

```
## $OJ
## [1] 0.01959757
##
## $VC
## [1] 4.577802e-05
```

Similarly, consider the hypotheses:

H0: 'the average tooth growth for method s is the same for doses of 0.5 and 1mg/day.'

H1: 'the average tooth growth for method s is smaller for the dose of 0.5mg/day than for 1mg/day.'

We perform a Welch's t-test and obtain the following p-values:

```
## $OJ
## [1] 4.39246e-05
##
## $VC
## [1] 3.405509e-07
```

Conclusion: we have strong evidence that, out of the three doses studied here, the average tooth growth is larger when guinea pigs are given a dose of 2mg/day and smaller when given a dose of 0.5mg/day. This holds for both delivery methods studied here.

Appendix

Code for loading the data and needed packages:

```
library(datasets)
library(ggplot2)
library(dplyr)

data <- ToothGrowth
```

Code for distribution of subjects per dose and delivery method:

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

Code for table on means and standard deviations per dose and delivery method:

```
data %>% group_by(supp, dose) %>% summarise(mean = mean(len), sd = sd(len))
```

Code for box-plot graph:

```
ggplot(data, aes(y=len)) + geom_boxplot(aes(color=supp)) + facet_grid(. ~ dose) +
    labs(y='length', title='Tooth growth by dose and delivery method') +
    theme(axis.title.x=element_blank(),
          axis.text.x=element_blank(),
          axis.ticks.x=element_blank())
```

Code for hypothesis tests of tooth growth against delivery method, with fixed dose:

```
lapply( split(data, data$dose) , function(df)
    t.test(len ~ supp, data=df, paired=FALSE, var.equal=FALSE,
           alternative='greater')$p.value
)
```

Code for hypothesis tests of tooth growth against delivery method, with fixed delivery method:

```
data2 <- data[data$dose %in% c(1,2), ]

lapply( split(data2, data2$supp) , function(df)
    t.test(len ~ dose, data=df, paired=FALSE, var.equal=FALSE,
           alternative='less')$p.value
)

data3 <- data[data$dose %in% c(.5,1),]

lapply( split(data3, data3$supp) , function(df)
    t.test(len ~ dose, data=df, paired=FALSE, var.equal=FALSE,
           alternative='less')$p.value
)
```