

Statistical Inference Course Project Part 2

JSchneyer

1/3/2021

Introduction

The purpose of this project is to investigate and analyze the ToothGrow dataset in the R datasets package. In this part, we will:

- load the data
- provide a summary of the data
- use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose
- use the results of our test to make a conclusion about our hypothesis

Part 1: Load dataset and summarize data

First, load the library “datasets” and assign the data to the data frame TG.

```
library(datasets)
TG <- ToothGrowth
```

If we look at the documentation for the TG dataset, we see that the data describes the effect of Vitamin C on tooth growth in Guinea pigs. The response variable is the length of odontoblasts (cells responsible for tooth growth). The subjects ($n = 60$ in this study) were given one of three doses of vitamin c (0.5, 1, or 2 mg/day) by one of two delivery methods, orange juice (OJ) or ascorbic acid (VC). There are three fields in this dataset:

- “len” > tooth length
- supp > supplement type (VC or OJ)
- dose > dose in mg/day

First, let’s check the dimensions of the dataset to ensure we have all the data, and check the column data types.

```
str(TG)
```

```
## '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 ...
```

We see that there are 60 observations of three variables. “Len” is numeric, which is the measure of tooth length, “supp” is a factor with two levels (OJ and VC), and dose is numeric, which is the amount of vitamin C given to each subject.

To better understand the data, we can see take a look at the summary metrics for the data set.

```
summary(TG)
```

```
##           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
```

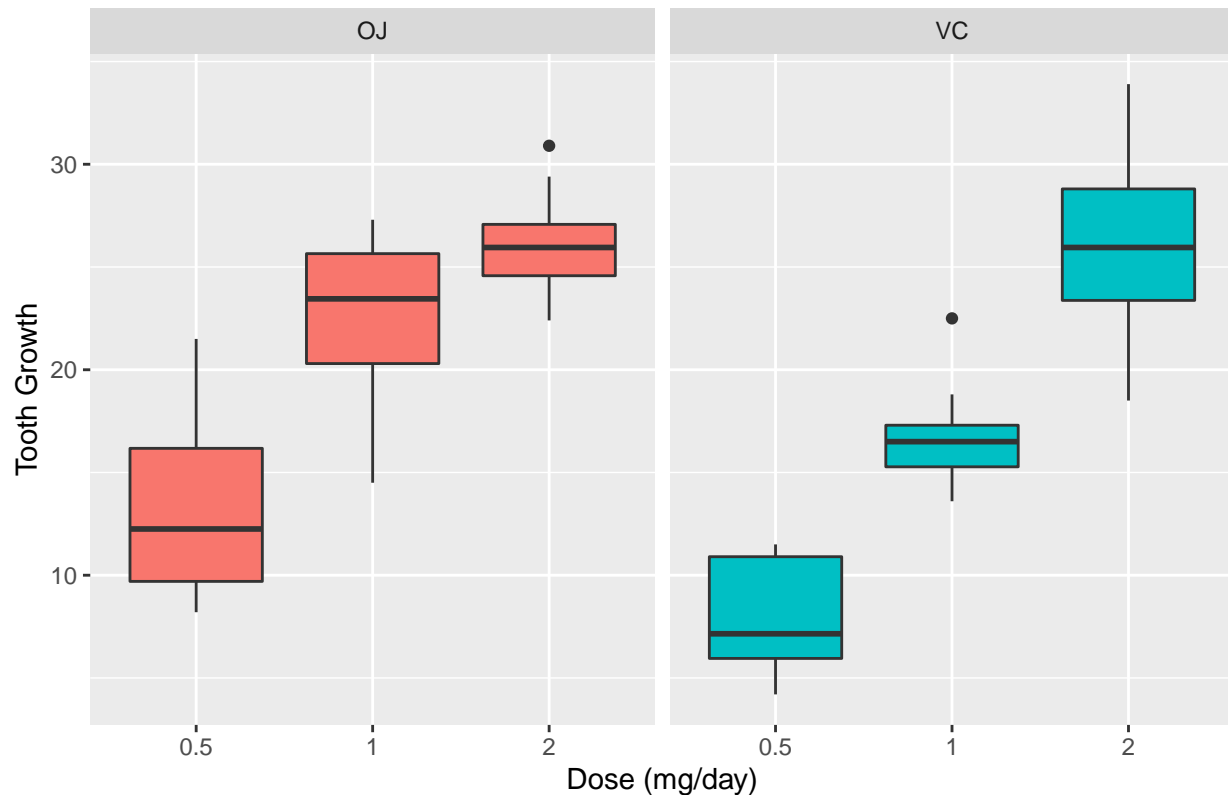
We see that, for our entire dataset, tooth length ranges from 4.20 to 33.90 with a mean and median length of 18.81 and 19.25, respectively. Thirty of the subjects received OJ and 30 subjects received VC. The doses ranged from 0.5 to 2.0 mg/day (we knew this from the documentation).

Part 2: Provide Basic Summary of the Data

Next, we will use the ggplot2 package to create box and whisker plots of the data by supplement type.

```
library(ggplot2)
ggplot(TG, aes(x = factor(dose), y = len)) +
  facet_grid(~supp) +
  geom_boxplot(aes(fill = supp), show.legend = FALSE) +
  ggtitle("Comparison of Tooth Growth by Supplement and Dose") +
  xlab("Dose (mg/day)") +
  ylab("Tooth Growth")
```

Comparison of Tooth Growth by Supplement and Dose



We can see in the chart above that at lower doses (0.5 and 1 mg/day), tooth growth is greater in subjects given OJ. The spread of tooth growth lengths in subjects given OJ is greater at these doses. However, at higher doses, the supplement type does not have an effect on the median tooth growth, but the range of values in subjects given VC is greater.

Part 3:

To check for differences in tooth growth between the supplement types, we will use the Welch Two Sample t-test, which is perfect for comparing the means of two groups. The test can be used even if the variance of each group is not equal.

First, we will test to see if supplement type has an effect on tooth growth, regardless of dosage. Our null hypothesis, H_0 , is that the mean tooth growth of subjects given OJ is the same as that of subjects given VC ($\mu_{OJ} = \mu_{VC}$), and that supplement type has no effect on tooth growth. Our alternative hypothesis is that the means are not equal.

```
t.test(len ~ supp, data = TG)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1710156  7.5710156
```

```
## sample estimates:
## mean in group OJ mean in group VC
##      20.66333      16.96333
```

Our p-value (0.06) is greater than 0.05 and our confidence intervals contain the value zero, therefore we cannot reject the null hypothesis that supplement type has no effect on tooth growth.

Next, we will test to see if dosage has zero effect on tooth growth. We will subset the data into dose-pairs (0.5 and 1, 0.5 and 2, 1 and 2 mg/day), and use the Welch Two Sample t-test again. H_0 for each subset and test will be the same, that the mean tooth growth for one dose is the same as the tooth growth for the second dose. Our alternative hypothesis will be that dose impacts tooth growth.

```
TG_0.5_1 <- subset(TG, dose %in% c(0.5,1))
TG_0.5_2 <- subset(TG, dose %in% c(0.5,2))
TG_1_2 <- subset(TG, dose %in% c(1,2))

t.test(len ~ dose, TG_0.5_1)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983781 -6.276219
## sample estimates:
## mean in group 0.5 mean in group 1
##      10.605      19.735
```

```
t.test(len ~ dose, TG_0.5_2)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean in group 0.5 mean in group 2
##      10.605      26.100
```

```
t.test(len ~ dose, TG_1_2)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##           19.735           26.100
```

The p-values for each of the three t-tests above are less than 0.05 and the confidence intervals do not contain the value zero, therefore we can reject our null hypothesis that dosage has no effect on tooth growth.

Part 4: Conclusions and Assumptions

Based on the results of the Welch two sided t-tests run in the previous section, we can make the following conclusions:

1. We cannot conclude that the supplement type has an effect on tooth growth, as our p-value was greater than 0.05 and we failed to reject the null hypothesis.
2. We can conclude that the dosage amount does have an effect on tooth growth, as the p-values for all three dosage groupings (0.5 and 1, 0.5 and 2, and 1 and 2 mg/day) were less than 0.05 and the confidence intervals did not contain zero.

We made a number of assumptions in this analysis. First, we assumed that the study was performed with a randomness, aka there was no bias in which subjects were given which supplements and doses. We also are assuming that the subjects are representative of the population of guinea pigs. Finally, we are assuming that the variances between each group, for all four t-tests, are not equal.