Statistical Inference Project Part 2

Contents

0.1 Synopsis 1
0.2 Data Description 1
0.3 Data Processing 1
0.4 Anaylsis 2
0.5 Assumptions and conclusions 8

0.1 Synopsis

This is the project for the statistical inference class. This is the second part of the project which is an analysis of the ToothGrowth data.

In the second portion of the class, I am going to analyze the ToothGrowth data in the R datasets package.

0.2 Data Description

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

[,1] len numeric Tooth length
[,2] supp factor Supplement type (VC or OJ).
[,3] dose numeric Dose in milligrams.
(source: R Doumentation)

0.3 Data Processing

0.3.1 Loading packages and libraries

```
library(ggplot2)
library(datasets)
library(reshape2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

0.3.2 Loading dataset

```
toothGrowth = tbl_df(ToothGrowth)
```

0.4 Anaylsis

0.4.1 Summarizing Data

```
glimpse(toothGrowth)
```

```
## Variables:
## $ len  (dbl) 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16....
## $ supp (fctr) VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, ...
## $ dose (dbl) 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1....
```

```
summary(toothGrowth)
```

```
##       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
```

```
head(toothGrowth,3); tail(toothGrowth,3)
```

```
## Source: local data frame [3 x 3]
##
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

```
## Source: local data frame [3 x 3]
##
##     len supp dose
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

0.4.2 Exploratory Data Analysis

0.4.2.1 Tooth length vs supplement type Fisrt, I will explore the relaion between the supplement type and the lenth of tooth

```
ggplot(data=toothGrowth, aes(x=len)) +
  geom_histogram(binwidth=2) +
  facet_grid( . ~ supp) +
  xlab("Tooth Length") +
  ylab("Count")
```

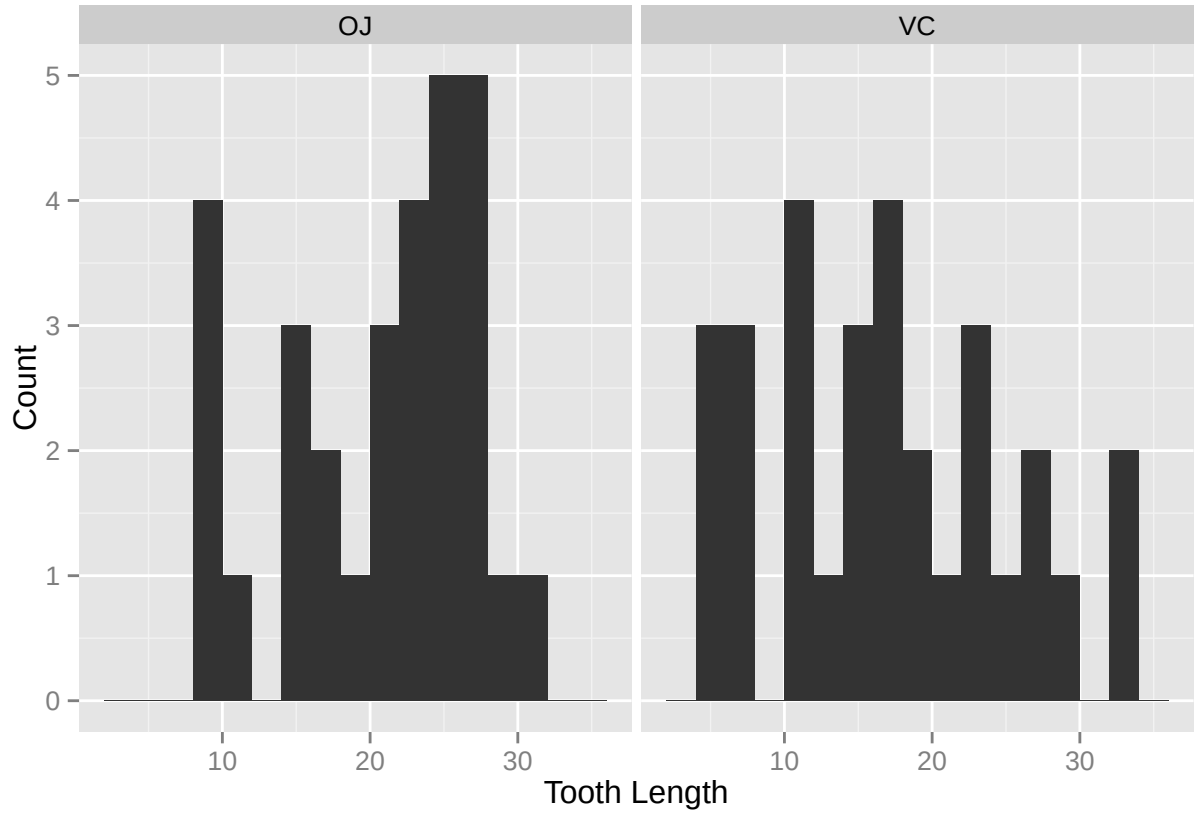

Figure 1:

It seems that both supplements have similar effects on the length of tooth, however, the orange juice, OJ, has more profound effect on most of the teeth

0.4.2.2 Tooth length vs dosage I will then explore the effect of the dosage quantity on the length of the tooth while examinning this effect for both of the supplements.

```
ggplot(data=toothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_bar(stat="identity") +
  xlab("Dosage (mg)") +
  ylab("Tooth Length") +
  guides(fill=guide_legend(title="Supplement Type"))
```

As the dosage increases so does the length of the tooth, however, again the orange juice, the OJ supplement has a stronger effect on the tooth length.

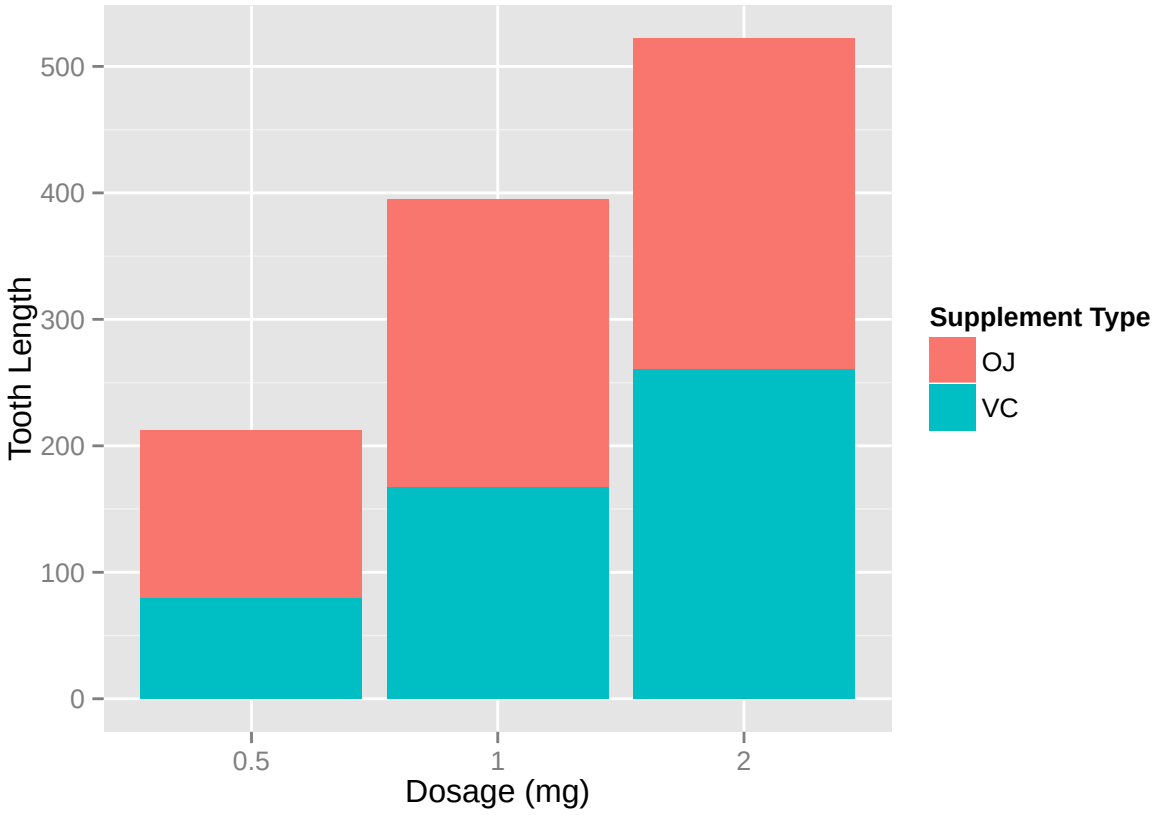
500
400
300
200
100
0
Tooth Length
0.5
1
2
Dosage (mg)
Supplement Type
OJ
VC

Figure 2:

0.4.3 Comparison of tooth growth by supp and dose

0.4.3.1 Groups based on supplement

```
oj= subset(toothGrowth, supp=="OJ")$len
vc= subset(toothGrowth, supp=="VC")$len
```

0.4.3.1.1 Extracting Groups

```
var(oj);var(vc)
```

0.4.3.1.2 Are groups of different variance ?

```
## [1] 43.63344
```

```
## [1] 68.32723
```

The variance is different for all groups created based on supplement

```
test= t.test(len ~ supp, paired = FALSE, val.equal = FALSE, data = toothGrowth)
```

0.4.3.1.3 Comparing the supplement groups using T Test The p-value is 0.0606345 and the 95% confidence interval is -0.1710156, 7.5710156 which suggests that there is no substantial difference between the tooth growth of those subjected to the two types of supplements, namely, OJ and VC, regardless of the dosage.

0.4.3.2 Groups based on supplement and dosage

```
oj05= subset(toothGrowth, dose ==0.5 & supp=="OJ")$len
oj1= subset(toothGrowth, dose ==1 & supp=="OJ")$len
oj2= subset(toothGrowth, dose ==2 & supp=="OJ")$len


vc05= subset(toothGrowth, dose ==0.5 & supp=="VC")$len
vc1= subset(toothGrowth, dose ==1 & supp=="VC")$len
vc2= subset(toothGrowth, dose ==2 & supp=="VC")$len
```

0.4.3.2.1 Extracting Groups

```
var(oj05);var(oj1);var(oj2)
```

0.4.3.2.2 Are groups of different variance ?

```
## [1] 19.889
```

```
## [1] 15.29556
```

```
## [1] 7.049333
```

```
var(vc05);var(vc1);var(vc2)
```

```
## [1] 7.544
```

```
## [1] 6.326778
```

```
## [1] 23.01822
```

The variance is different for all groups created based on supplement and dosage

```
testOJ05_1= t.test(x= oj1, y=oj05,  val.equal = FALSE)
testOJ05_1$p.value; testOJ05_1$conf.int
```

0.4.3.2.3 Comparing the supplement groups using T Test

```
## [1] 8.784919e-05
```

```
## [1]  5.524366 13.415634
## attr(,"conf.level")
## [1] 0.95
```

```
testOJ1_2= t.test(x= oj2, y=oj1,  val.equal = FALSE)
testOJ1_2$p.value; testOJ1_2$conf.int
```

```
## [1] 0.03919514
```

```
## [1] 0.1885575 6.5314425
## attr(,"conf.level")
## [1] 0.95
```

```
testOJ05_2= t.test(x= oj2, y=oj05,  val.equal = FALSE)
testOJ05_2$p.value; testOJ05_2$conf.int
```

```
## [1] 1.323784e-06
```

```
## [1]  9.324759 16.335241
## attr(,"conf.level")
## [1] 0.95
```

The p-values and the 95% confidence intervals suggest that as we increase the dosage the teeth growth increases for both supplements, however, the increase in the tooth growth is more aparent in the increase of the dosage from 0.5 to 1 or from 0.5 to 2. The increase in the dosage from 1 to 2 results in more increase in the tooth growth with the VC supplement more than that of the OJ supplement.

0.4.3.3 Groups based on dosage

```
dose05_1= subset(toothGrowth, dose ==0.5|dose ==1)
dose05_1$supp = NULL
dose1_2= subset(toothGrowth, dose ==1|dose ==2 )
dose1_2$supp = NULL
dose05_2= subset(toothGrowth, dose ==0.5|dose ==2)
dose05_2$supp = NULL
```

0.4.3.3.1 Extracting Groups

```
var(dose05_1);var(dose1_2);var(dose05_2)
```

0.4.3.3.2 Are groups of different variance ?

```
##            len       dose
## len  40.736000 1.17051282
## dose  1.170513 0.06410256
```

```
##            len      dose
## len  26.825583 1.6320513
## dose  1.632051 0.2564103
```

```
##            len      dose
## len  78.366660 5.9596154
## dose  5.959615 0.5769231
```

The variance is different for all groups created based on dosage

```
testD05_1= t.test(len ~ dose, paired = FALSE, val.equal = FALSE, data = dose05_1)
testD05_1$p.value; testD05_1$conf.int
```

0.4.3.3.3 Comparing the supplement groups using T Test

```
## [1] 1.268301e-07
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
testD1_2= t.test(len ~ dose, paired = FALSE, val.equal = FALSE, data = dose1_2)
testD1_2$p.value; testD1_2$conf.int
```

```
## [1] 1.90643e-05
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

```
testD05_2= t.test(len ~ dose, paired = FALSE, val.equal = FALSE, data = dose05_2)
testD05_2$p.value; testD05_2$conf.int
```

```
## [1] 4.397525e-14
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

The p-values and the 95% confidence intervals suggest that there is no substantial difference between the tooth growth of the different dosages regardless of the supplement type.

0.5 Assumptions and conclusions

```
- The variance differes among the different groups, as calculated.
- The distribution of the observations in each group are normally distributed.
- These groups are not paired.
- It is obvious that the OJ supplement is more effective with the tooth growth than the VC supplement es
-It is also obvious that an increase in dosage shows an increase in the tooth growth with both supplemen
```