Statistical Inference: Basic inferential data analysis

Jacques Botes

September 2014

```
## Loading required package: ggplot2
## Loading required package: plyr
```

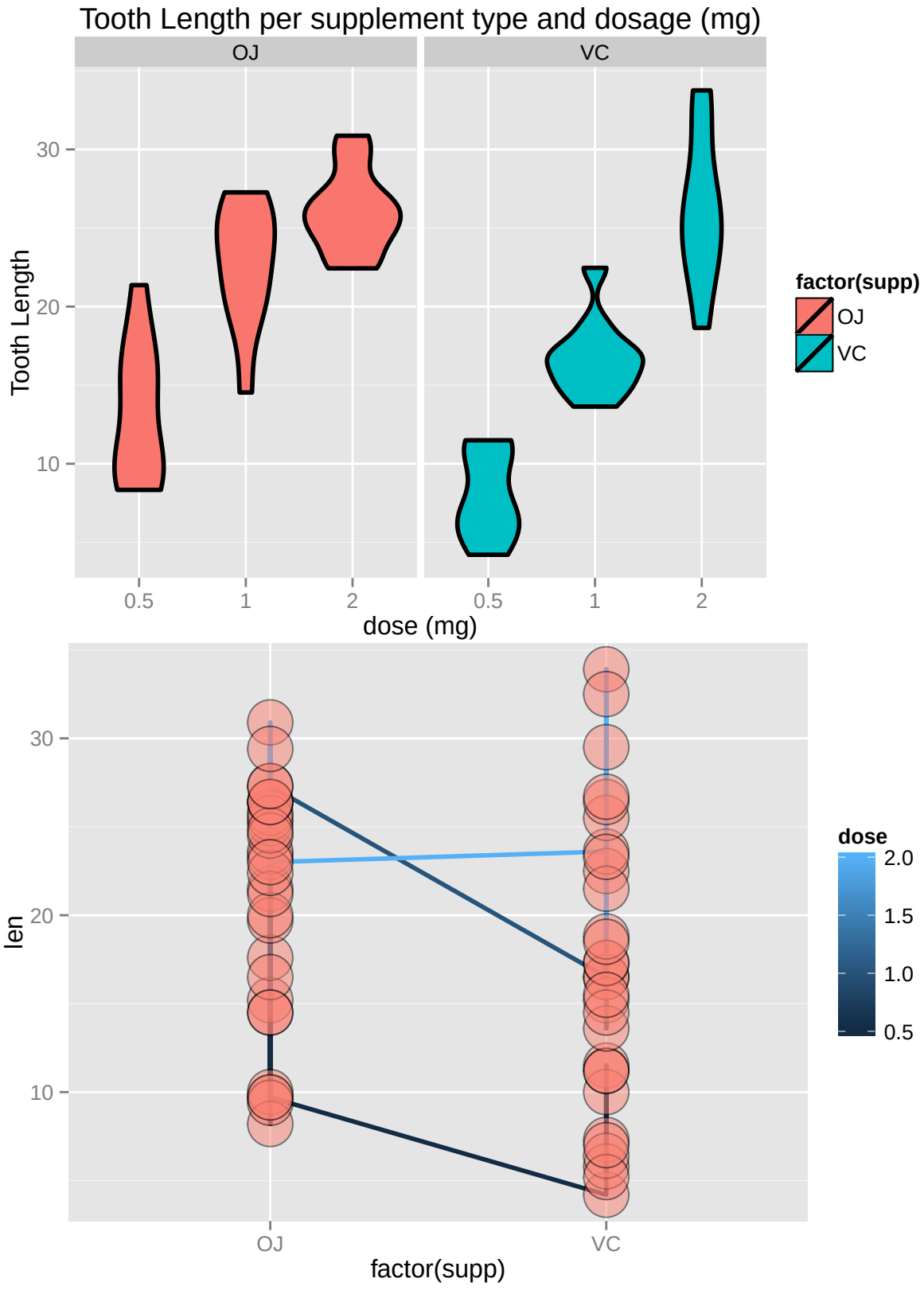
Problem Statement

Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering) State your conclusions and the assumptions needed for your conclusions.

About the data

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). A data frame with 60 observations on 3 variables. (taken from `?ToothGrowth` in R)

```
##       len       supp         dose       
##  Min.   : 4.2   OJ:30   Min.   :0.50  
##  1st Qu.:13.1   VC:30   1st Qu.:0.50  
##  Median :19.2           Median :1.00  
##  Mean   :18.8           Mean   :1.17  
##  3rd Qu.:25.3           3rd Qu.:2.00  
##  Max.   :33.9           Max.   :2.00
```


Tooth Length per supplement type and dosage (mg)
OJ
VC
Tooth Length
30
20
10
0.5
1
2
dose (mg)
factor(supp)
OJ
VC
len
dose
2.0
1.5
1.0
0.5
factor(supp)

Summary of dataset:

Interpreting the first graph we can see that the Orange Juice (OJ) supplement at lower doses results in a longer tooth length. With the 2mg dose the difference in tooth lengh is closely matched and seems to be working equally well.

Testing

Analysis

Test 1 Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. First test is to test the data only looking at the supplement and not dosage

```
t.test(len ~ supp, paired = F,var.equal=F, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

The interval contains the zero value, we cannot reject the null hypothesis that there isn't a significant difference in tooth length betwen OJ and VC supplement types.

Test 2 Lets compare specific dosage levels and see if we notice a significant tooth growth difference. We will test 0.5mg, 1mg and 2mg independently. We create 3 different subsets of the data and do t level tests for each one.

```
tooth.D05 <- subset(ToothGrowth, dose == 0.5)
tooth.D1 <- subset(ToothGrowth, dose == 1.0)
tooth.D2 <- subset(ToothGrowth, dose == 2.0)
```

```
##      [,1]                  [,2]                [,3]
## [1,] "0.5mg conf interval" "1.71905727146767"  "8.78094272853233"
## [2,] "1.0mg conf interval" "2.80214824916537"  "9.05785175083463"
## [3,] "2.0mg conf interval" "-3.79807046333516" "3.63807046333515"
```

```
##                       mean in group OJ mean in group VC
## [1,] "0.5mg mean est" "13.23"          "7.98"
## [2,] "1.0mg mean est" "22.7"           "16.77"
## [3,] "2.0mg mean est" "26.06"          "26.14"
```

At 2mg the p-value is 0.9639, we cannot reject the null hypothesis that there isnt a difference in tooth length. The interval also containst 0. For the other two tests at 0.5 and 1mg we can reject the null hypothesis. The mean estimate also shows significant differences betwen OJ and VC at the lower dosage levels.

Test 3 Lets test the differences in dosage levels and see if we notice a significant tooth growth difference. We will subset the dataset and test 0.5mg with 1mg and 2mg then 1mg with 2mg.

```
tooth.D05_01 <- subset(ToothGrowth, dose %in% c(0.5,1.0))
tooth.D05_02 <- subset(ToothGrowth, dose %in% c(0.5,2.0))
tooth.D01_02 <- subset(ToothGrowth, dose %in% c(1.0,2.0))
```

```
##      [,1]          [,2]                        [,3]
## [1,] "Description" "Confidence Interval Lower" "Confidence Interval Upper"
## [2,] "0.5mg-1mg"   "-11.9837812579016"         "-6.27621874209841"
## [3,] "0.5mg-2mg"   "-18.1561665388306"         "-12.8338334611694"
## [4,] "1mg-2mg"     "-8.99648051689202"         "-3.73351948310799"
```

The p-values for all 3 tests (1.2683 $\times$ 10-7, 4.3975 $\times$ 10-14, 1.9064 $\times$ 10-5) was very small and therefore we can reject the null hypothesis that there isnt a significant difference. The higher dosage level in each test resulted in a longer tooth length.

Conclusion:

We have conducted three different tests and found the following

- When ignoring dose levels: No significant difference in tooth length noted
- Taking dosage and supplement into account: Orange Juice (OJ) resulted in increased tooth length at lower dosage levels with a insignificant difference in length between OJ and VC at the 2mg dosage.
- When ignoring supplement type: Higher dose levels had a significant impact on tooth length. There is also a significant impact between the different dosages

Assumptions

- No other outside factor had a significant impact on the study.
- The variances are different for the seperate populations.

The full report including all R code: https://github.com/JacquesBot/StatisticalInference/blob/master/Q2.md