Statistical Inference: Basic inferential data analysis

Jacques Botes

September 2014

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

Problem Statement

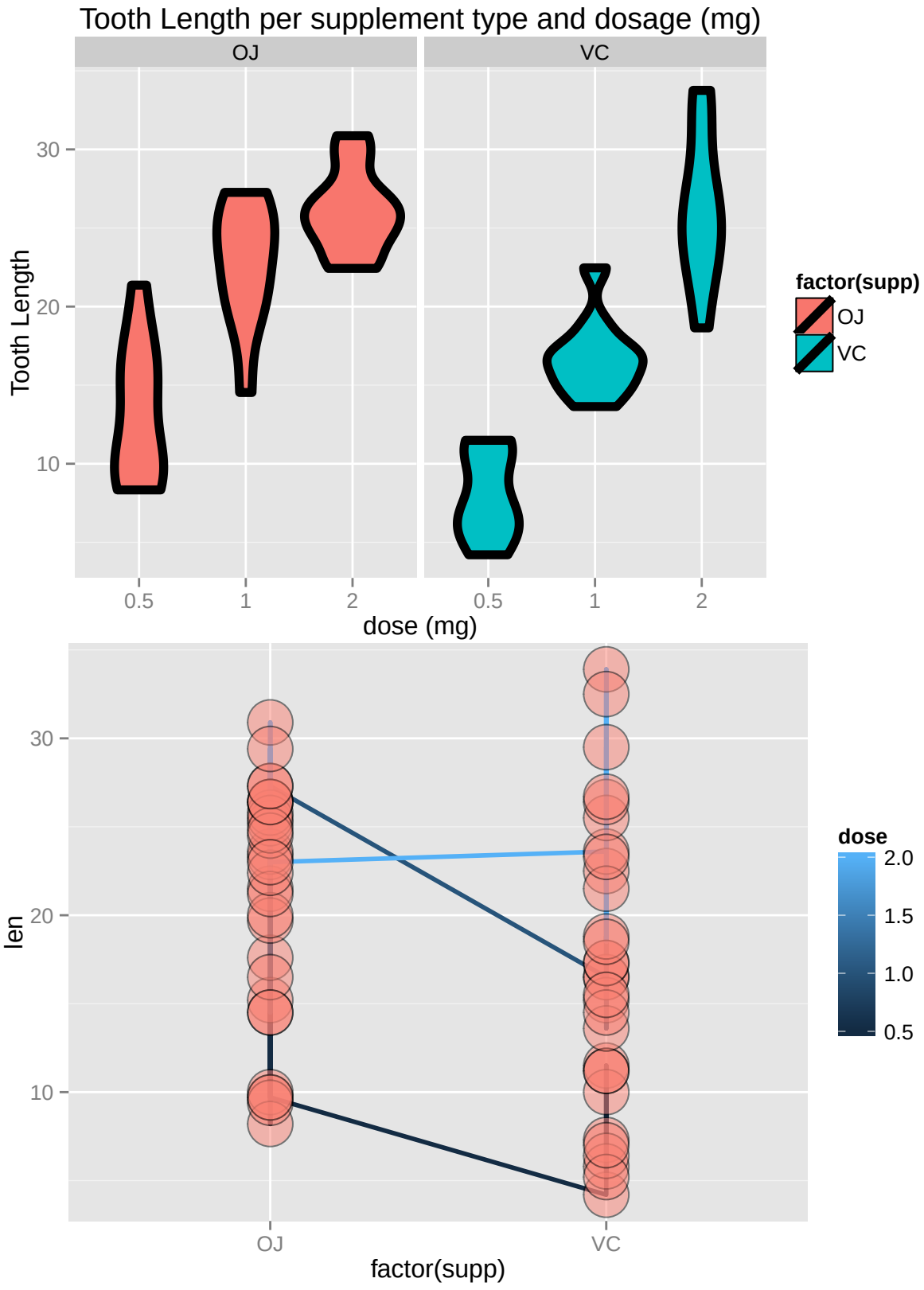
Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering) State your conclusions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

```
##    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       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

Assumptions

- The variances are different for the populations

Analysis Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

```
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
```

We cannot reject the null hypothesis as there are a significant difference in tooth length betwen OJ and VC supplement types.

Lets look at it for specific dosages

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

rbind(
c('0.5mg conf interval',t.test(len ~ supp, paired = F, var.equal= F,data = tooth.D05)$conf),
c('1.0mg conf interval',t.test(len ~ supp, paired = F, var.equal= F,data = tooth.D1)$conf),
c('2.0mg conf interval',t.test(len ~ supp, paired = F, var.equal= F,data = tooth.D2)$conf)
)
```

```
##      [,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"
```

In Conclusion:

If we ignore the supplement types and just look at the difference in dosages we notice we can reject the null hypothesis for 0.5 and 1.0mg levels and not for the 2.0mg level. for the 2.0mg there is a significant difference in tooth length.

Full markdown file: