Statistical Inference Course Project - Part 2

Aaron Cole

8 February 2016

Contents

1 Synopsis

The following text is quoted from the assignment information page on Coursera.

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

Evaluation criteria as follows.

- Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?
- Did the student perform some relevant confidence intervals and/or tests?
- Were the results of the tests and/or intervals interpreted in the context of the problem correctly?
- Did the student describe the assumptions needed for their conclusions?

2 Exploratory analysis and basic summary

2.1 Load libraries

```
suppressMessages(library(ggplot2))
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
suppressMessages(library(dplyr))
```

```
## Warning: package 'dplyr' was built under R version 3.2.3
```

2.2 Load tooth growth data and configure groupings

```
data(ToothGrowth)
tg <- tbl_df(ToothGrowth)
tg <- group_by(tg, supp, dose)
```

2.3 Summary and explanation of the data

```
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
```

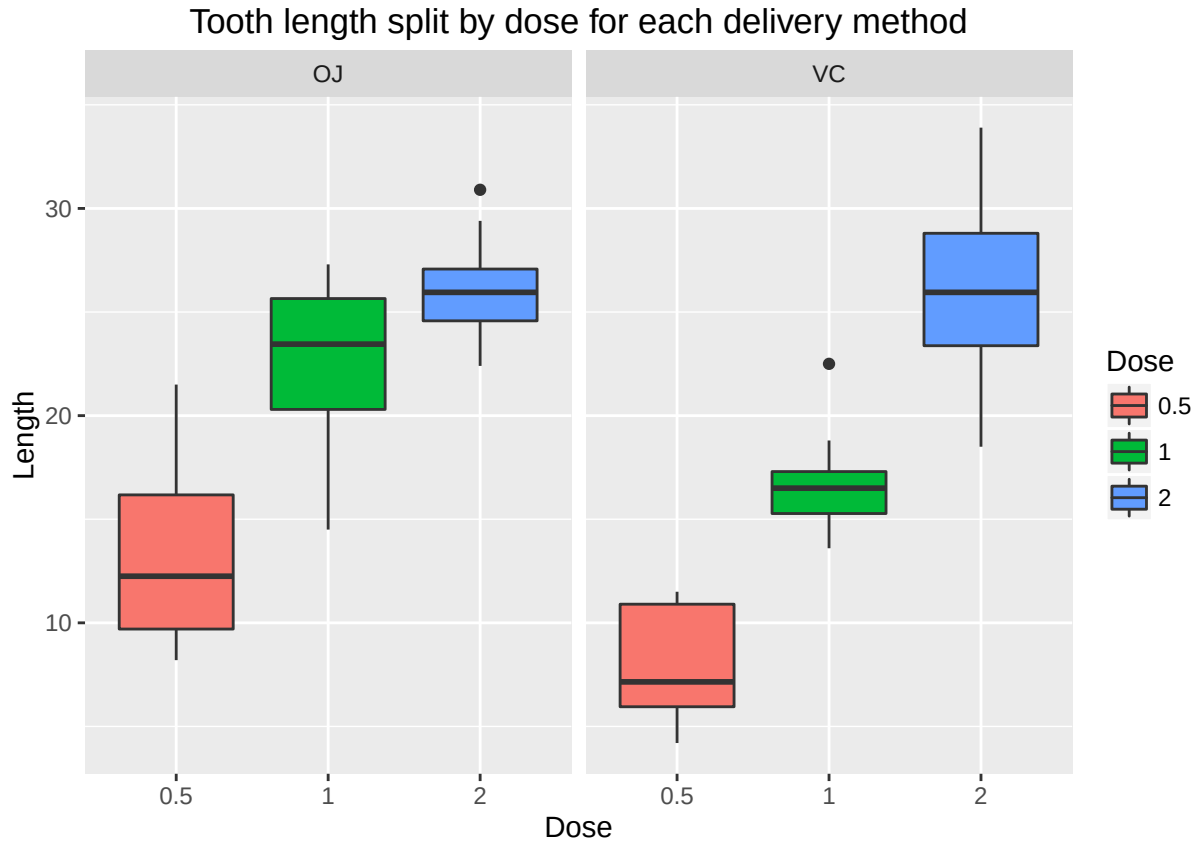
The data is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid). The data frame has 60 observations on 3 variables.

Variable	Class	Comments
len	numeric	Tooth length.
supp	factor	Supplement type (VC or OJ).
dose	numeric	Dose in milligrams.

2.4 Exploratory analysis and basic summary

```
ggplot(ToothGrowth, aes(x=factor(dose),
                        y=len,
```

```
                    fill=factor(dose))) +
    geom_boxplot(notch=F) +
    facet_grid(.~supp) +
    ggtitle("Tooth length split by dose for each delivery method") +
    xlab("Dose") +
    ylab("Length") +
    scale_fill_discrete(name="Dose")
```




```
# Text summary of the means by supp and dose.
summarise_each(tg, funs(mean))
```

```
## Source: local data frame [6 x 3]
## Groups: supp [?]
##
##     supp  dose   len
##   (fctr) (dbl) (dbl)
## 1     OJ   0.5 13.23
## 2     OJ   1.0 22.70
## 3     OJ   2.0 26.06
## 4     VC   0.5  7.98
## 5     VC   1.0 16.77
## 6     VC   2.0 26.14
```

```
# Text summary of the total growth (len) by supp.
summarise_each(select(group_by(tg, supp), supp, len), funs(sum))
```

```
## Source: local data frame [2 x 2]
##
##     supp   len
##   (fctr) (dbl)
## 1     OJ 619.9
## 2     VC 508.9
```

- The plot and text summary appear to indicate that higher dose levels result in higher growth length.
- Supplement type "OJ" would appear to result in higher average growth at the lower dose levels of 0.5 and 1.0.
- Supplement type "OJ" also has higher growth across all observations for all dose levels, when compared with "VC".
- The average growth is similiar for both supplement types at a dose level of 2.0.

3 Using confidence intervals to compare tooth growth by supp and dose

3.1 Assumptions

- Our data is roughly normal.
- The variance between groups is unequal.

3.2 Testing the supplements for all dose levels

```
t.test(len ~ supp,
       data = filter(tg),
       paired = FALSE,
       var.equal = TRUE,
       alternative = "two.sided")
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

- P-value is above 5%.
- Confidence interval contains 0.

3.3 Testing the supplements at a dose level of 0.5

```
t.test(len ~ supp,
       data = filter(tg, dose == 0.5),
       paired = FALSE,
       var.equal = TRUE,
       alternative = "two.sided")
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

- P-value is below 5%.
- Confidence interval does not contain 0.
- Indicates with 95% confidence that supplement type OJ results in higher growth rates at a dose of 0.5.

3.4 Testing the supplements at a dose level of 1.0

```
t.test(len ~ supp,
       data = filter(tg, dose == 1.0),
       paired = FALSE,
       var.equal = TRUE,
       alternative = "two.sided")
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

- P-value is below 5%.
- Confidence interval does not contain 0.
- Indicates with 95% confidence that supplement type OJ results in higher growth rates at a dose of 1.0.

3.5 Testing the supplements at a dose level of 2.0

```
t.test(len ~ supp,
       data = filter(tg, dose == 2.0),
       paired = FALSE,
       var.equal = TRUE,
       alternative = "two.sided")
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.722999  3.562999
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

- P-value is above 5%.
- Confidence interval contains 0.

3.6 Conculsions

- Our t-tests support a difference between the supplements only for dose levels of 0.5 and 1.0, with a 95% confidence.
- At a dose level of 0.5 and 1.0 supplement type OJ appears to promote higher growth.
- Our t-tests do not support any difference between the supplements at a dose level of 2.0, or across the total sample for all dose levels.