Coursera Statistical Inference Project - Part 2

Brian Altman

December 26, 2015

Overview

The project analyzes the ToothGrowth data in the R datasets package. The project

```
1) Load the ToothGrowth data
2) Perform some basic exploratory data analyses and a basic summary of the data.
3) Uses confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4) Provides conclusions and the assumptions needed
```

This file and related code can be found here: https://github.com/beaisis/Statistical-Inference ***

1) Load the ToothGrowth data Tooth growth data is part of the base R package install.

```
library(ggplot2)
data(ToothGrowth)
```

2) Perform some basic exploratory data analyses related to supp and dose There are 60 observations of 3 variables:

```
str(ToothGrowth)
```

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

```
head(ToothGrowth)
```

```
##    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
```

```
summary(ToothGrowth)
```

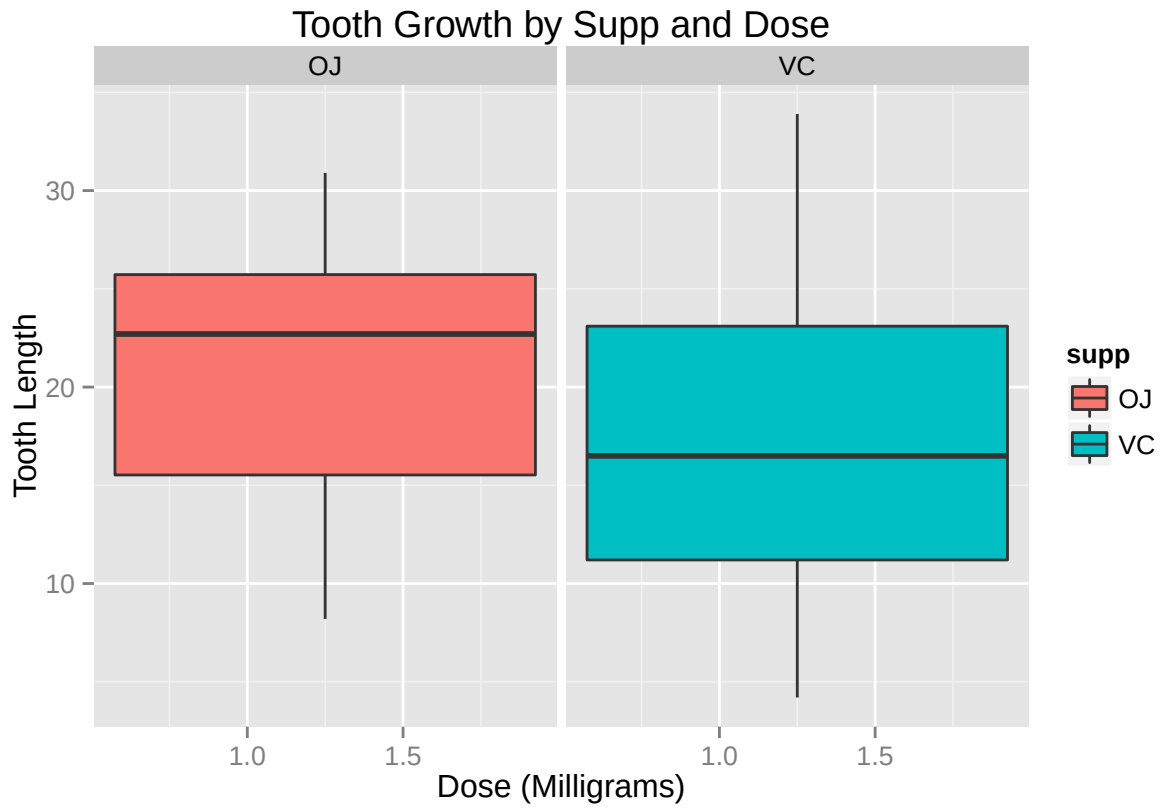
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Means of tooth grown by dose and supp(delivery method)

```
tapply(ToothGrowth$len, list(ToothGrowth$dose,ToothGrowth$supp),mean)
```

```
##        OJ    VC
## 0.5 13.23  7.98
## 1   22.70 16.77
## 2   26.06 26.14
```

Plot showing tooth growth by dose and supp:



The summary tables and boxplot suggests tooth length is positively correlated with the dosage for both supplements.

3) Confidence intervals and null hypothesis tests to compare tooth growth by supp and dose
Null hypothesis: No correlation between method of delivery and tooth length.

The following t-test was conducted:

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"], ToothGrowth$len[ToothGrowth$supp=="VC"], paired = FALSE
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

A P-value of .06 and confidence interval of [-0.171, 7.571] that includes zero, suggests the null hypthesis can not be rejected. There does not appear to be a correlation between method of delivery and tooth length.

Null Hypothesis: No correlation between dosage and Tooth Length. The following models by dose and supp were run:

```
t1<-t.test(len[dose==.5 & supp == 'VC'],len[dose==1& supp == 'VC'], paired = FALSE, var.equal = TRUE)
t_tests <- rbind(t_tests, list('.5~1 mg - VC',  t1$p.value,t1$conf.int[1], t1$conf.int[2]))
t1<-t.test(len[dose==1 & supp == 'VC'], len[dose==2 & supp == 'VC' ], paired = FALSE, var.equal = TRUE)
t_tests <- rbind(t_tests, list(' 1~2 mg - VC',  t1$p.value, t1$conf.int[1], t1$conf.int[2]))
t1<-t.test(len[dose==.5 & supp == 'VC'], len[dose==2 & supp == 'VC'], paired = FALSE, var.equal = TRUE)
t_tests <- rbind(t_tests, list('.5~2 mg - VC',  t1$p.value,t1$conf.int[1], t1$conf.int[2]))

t1<-t.test(len[dose==.5 & supp == 'OJ'],len[dose==1 & supp == 'OJ'], paired = FALSE, var.equal = TRUE)
t_tests <- rbind(t_tests, list('.5~1 mg - OJ',  t1$p.value,t1$conf.int[1], t1$conf.int[2]))
t1<-t.test(len[dose==1 & supp == 'OJ'], len[dose==2 & supp == 'OJ' ], paired = FALSE, var.equal = TRUE)
t_tests <- rbind(t_tests, list(' 1~2 mg - OJ',  t1$p.value,t1$conf.int[1], t1$conf.int[2]))
t1<-t.test(len[dose==.5 & supp == 'OJ'], len[dose==2 & supp == 'OJ'], paired = FALSE, var.equal = TRUE)
t_tests <- rbind(t_tests, list('.5~2 mg - OJ',  t1$p.value,t1$conf.int[1], t1$conf.int[2]))
colnames(t_tests) = c('Model','p.value','Ci.lower','ci.upper')
t_tests <- t_tests[2:6,]
```

Summary of the 6 models that represented the 3 Dose and 2 delivery method combinations:

```
##  Model          p.value      Ci.lower  ci.upper
##  ".5~1 mg - VC" 6.492265e-07 -11.26435 -6.315654
##  " 1~2 mg - VC" 3.397578e-05 -12.96896 -5.77104
##  ".5~2 mg - VC" 4.957286e-09 -21.83284 -14.48716
##  ".5~1 mg - OJ" 8.357559e-05 -13.41081 -5.529186
##  " 1~2 mg - OJ" 0.0373628    -6.500502 -0.2194983
```

For all dosage combinations, the p-values are less than .05 and the confidence intervals do not contain zero.The null hypothesis is rejected: Dosages of .5, 1 and 2 do contribute to tooth length increases.

4)Conclusions Assuming random assignment of guinea pigs to different dose level categories and supplement types:

```
Supplement type has no effect on tooth growth.

Increasing the dose level leads to increased tooth growth.
```