Analysis of Simulation Distribution

Brady Chiu

May 2, 2016

Contents

1 Intro

The purpose of this project is to analyze the ToothGrowth data in the R datasets package.

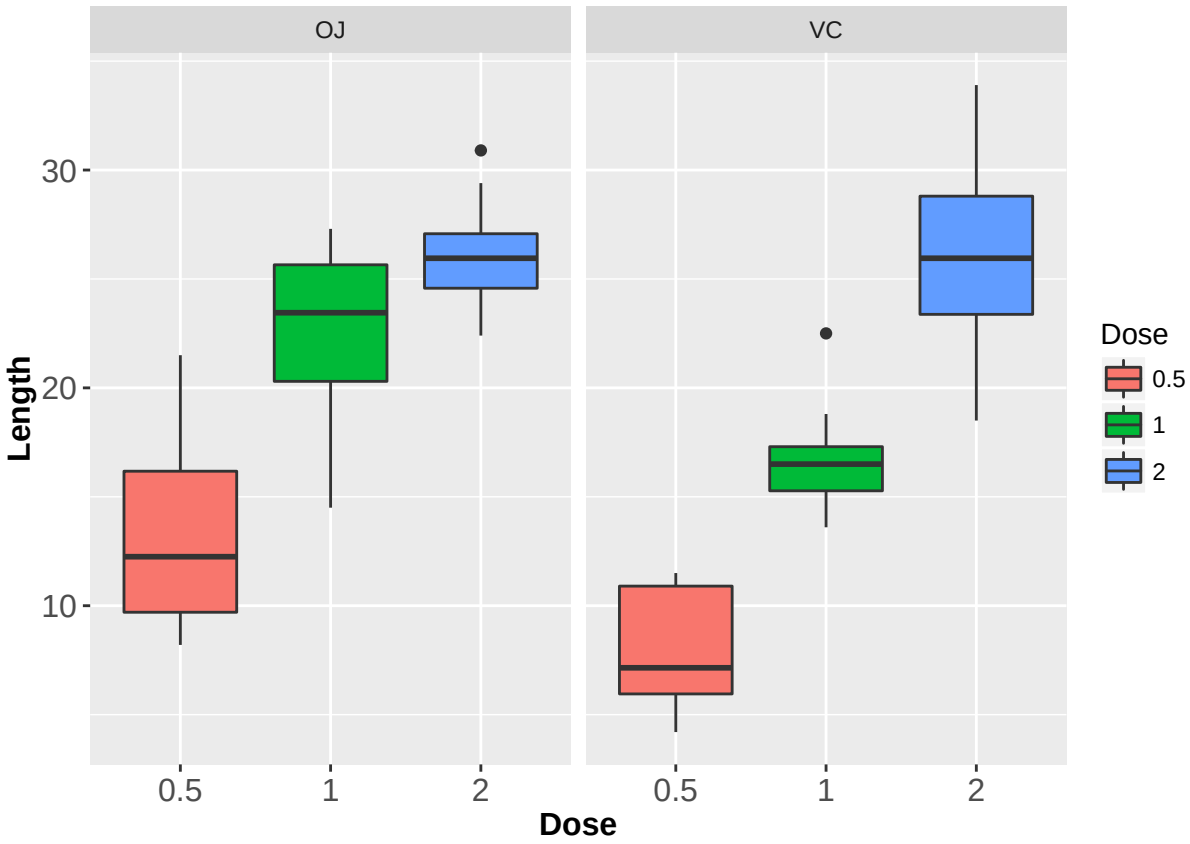
2 Data

- We take a high level overview of our dataset

```
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
```

- We create a boxplot to see the impact of varying dosages of orange juice vs. vitamin C on tooth growth



- We see tooth growth increases along with increased dosages of either orange juice or vitamin C

3 Analysis

3.1 Supplement

- We conduct a t-test to determine if a statistically significancant difference exists between our supplments

```
##
##  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
```

- Our p-value between 0.05 and 0.10 shows marginal statistical siginficance, with a 95% confidence interval that the differences is between -0.1670064 and 7.5670064

3.2 Dosage: 0.5 vs 1.0

- We conduct a t-test to determine if a statistically significancant difference exists between dosages of 0.5 vs 1.0

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

- Our p-value < 0.05 shows statistical siginficance, with a 95% confidence interval that the differences is between -11.983748 and -6.276252

3.3 Dosage: 0.5 vs 2.0

- We conduct a t-test to determine if a statistically significancant difference exists between dosages of 0.5 vs 2.0

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15352 -12.83648
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

- Our p-value < 0.05 shows statistical siginficance, with a 95% confidence interval that the differences is between -18.15352 and -12.83648

3.4 Dosage: 1.0 vs 2.0

- We conduct a t-test to determine if a statistically significancant difference exists between dosages of 1.0 vs 2.0

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

- Our p-value < 0.05 shows statistical siginficance, with a 95% confidence interval that the differences is between -8.994387 and -3.735613

4 Assumptions

Some assumptions we hold in order to form our conclusion(s) are:

- Population is iid normal gaussian
- Random samples were used for all treatments

5 Conclusion

- Suppment type has some effect on tooth growth
- Increasing dosage levels has a positive impact on tooth growth

6 Code

```
library(data.table, warn.conflicts=F)
library(dplyr, warn.conflicts=F)
library(ggplot2, warn.conflicts=F)
library(knitr, warn.conflicts=F)
library(lubridate, warn.conflicts=F)
library(tidyr, warn.conflicts=F)

# setwd("/Users/bradychiu/Dropbox (Uber Technologies)/R/Coursera/06_Statistical_Inference/stats_inferen
```

```
data(ToothGrowth)
```

```
ggplot(
  ToothGrowth
  ,aes(x=factor(dose),y=len,fill=factor(dose))
  )+
  geom_boxplot(aes(group=dose))+
  facet_grid(.~supp)+
  scale_x_discrete(name="Dose")+
  scale_y_continuous(name="Length")+
  scale_fill_discrete(name="Dose")+
  theme(
    axis.title=element_text(face="bold",size=12)
    ,axis.text=element_text(size=12)
    )
```

```r
summary(ToothGrowth)
```

```r
t.test(len~supp,data=ToothGrowth,var.equal=T,alternative="two.sided")
```

```r
t.test(len~dose,data=ToothGrowth %>% filter(dose %in% c(0.5,1.0)),var.equal=T,alternative="two.sided")
```

```r
t.test(len~dose,data=ToothGrowth %>% filter(dose %in% c(0.5,2.0)),var.equal=T,alternative="two.sided")
```

```r
t.test(len~dose,data=ToothGrowth %>% filter(dose %in% c(1.0,2.0)),var.equal=T,alternative="two.sided")
```