Statistical Inference Course Project: Part 2

Juan Agustin Melendez

July 21, 2015

Description

The second part of the course project will focus on performing an inferential analysis on the ToothGrowth data provide in the R datasets package. Confidence intervals and hypothesis testing will be performed to determine the effects of vitamin C on tooth growth in guinea pigs.

Loading and Summarizing Data

```
library(datasets)
TG <- ToothGrowth
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
```

```
str(TG)
```

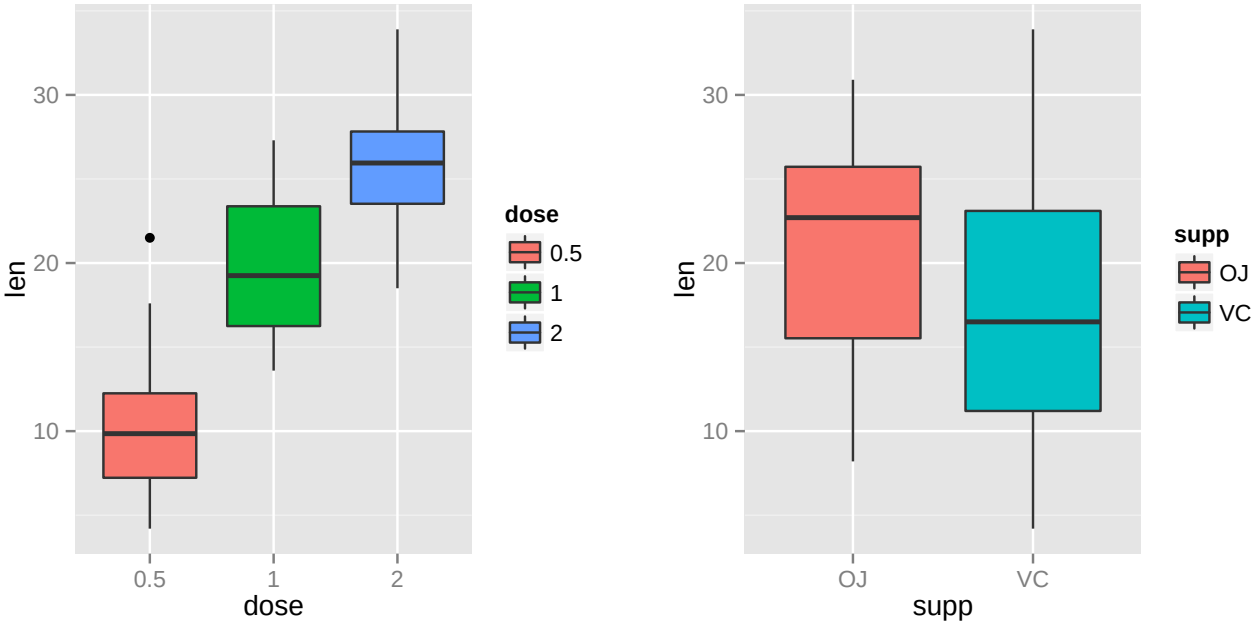
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
# Verifying groups by supplement and dose levels.
TG$dose <- as.factor(TG$dose)
table(TG$dose, TG$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

Visualization of Data

```
library(ggplot2);library(gridExtra)
# Plotting data by dose levels
d <- ggplot(data=TG, aes(x=dose, y=len)) + geom_boxplot(aes(fill=dose))
# Plotting data by supp levels
m <- ggplot(data=TG, aes(x=supp, y=len)) + geom_boxplot(aes(fill=supp))
grid.arrange(d, m, ncol=2)
```



Hypothesis testing using T tests.

First, a comparison is going to be made to determine if the delivery method has any effects on tooth growth. Then tooth growth is going to be evaluated by dosage, comparing doses 0.5-1, 0.5-2, and, 1-2. It is assumed that the variance between the tested groups are not equal.

1. Testing for supplement type vs length

```
t.test(len ~ supp, data=TG)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

As can be seen, the confidence interval [-0.171, 7.571] includes 0 and p values is 0.06. This indicates that there is not enough evidence to reject the null hypothesis that the supplement type has no effects on tooth growth.

2. Testing for dose vs length

```
# subsetting ToothGrowth data by dosage groups
TG_0.5.1 <- subset(TG, dose %in% c(0.5, 1))
TG_0.5.2 <- subset(TG, dose %in% c(0.5, 2))
TG_1.2 <- subset(TG, dose %in% c(1, 2))
# Testing dose group 0.5-1
t.test(len ~ dose, data=TG_0.5.1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
# Testing dose group 0.5-2
t.test(len ~ dose, data=TG_0.5.2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
# Testing dose group 1-2
t.test(len ~ dose, data=TG_1.2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

P values for every group tested are below 0.05 and the confidence interval does not include zero. This indicates that there is positive correlation between dose and length. Further more the value of t is very small (t = -11.799) when testing between the groups with dose 0.5 and 2 than in the other two groups suggesting that the evidence is pretty strong.

Summary

When comparing delivery method with tooth length, it was found that there is not sufficient evidence to reject the null hypothesis that supplement type has no effects on tooth growth. The confidence interval included zero and p was above 0.05.

When testing between dosage levels, all the tests performed between each group had p values below 0.05 and confidence intervals did not included zero. This suggests that on average, tooth length in guinea pigs increased with an increase in dosage for the guinea pigs in the experiment.

There is little documentation provided for the data in terms of how the experiment was performed therefore, any analysis performed on this data set is not deterministic. There is no data provided for a control group (what is the tooth growth rate on guinea pigs with no supplement?). For the purpose of this analysis, it is assumed that the only difference in guinea pig diet was that of orange juice vs vitamin C supplement at different dose levels. Also, different variance for each group were assumed when performing the t test.