Statistical Inference Project - Part 2: Inferential Data Analysis

S C Jayaprakash

Friday, January 23, 2015

Overview

This report analyses the ToothGrowth data in the R data sets package. The data has 60 observations, 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).

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
# load the dataset
library(datasets)
data(ToothGrowth)

# number of rows
nrow(ToothGrowth)
```

```
## [1] 60
```

```
# convert variable dose from numeric to factor since it has only 3 levels
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

# look at the 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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

2. Basic summary of the data.

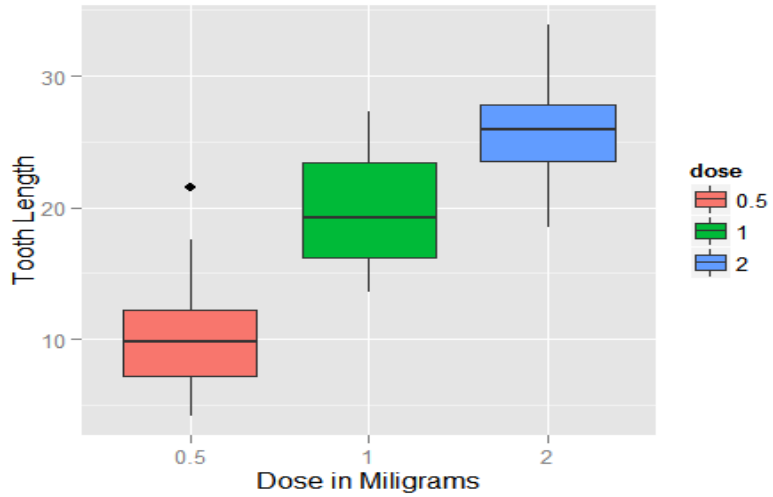
```
# summary statistics for all variables
summary(ToothGrowth)
```

```
##       len       supp     dose
##  Min.   : 4.2   OJ:30   0.5:20
##  1st Qu.:13.1   VC:30   1  :20
##  Median :19.2           2  :20
##  Mean   :18.8
##  3rd Qu.:25.3
##  Max.   :33.9
```

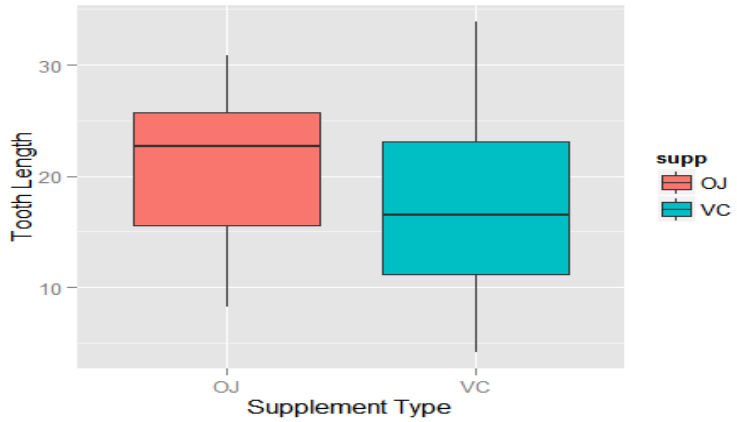
```r
# breakdonw of cases by dose levels and delivery methods
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

```r
# Exploratory Data analysis
library(ggplot2)

# Tooth growth Vs Vitamin C dose

ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) +
xlab("Dose in Miligrams") +ylab("Tooth Length")
```




```r
# Tooth growth Vs Supplement type

ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) +
xlab("Supplement Type") +ylab("Tooth Length")
```



3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

[i] Check for group differences due to different supplement type

Perform t-test for the difference of the means of two independent groups.

Assume unequal variances between the two groups. (var.equal = FALSE, conf.level = 0.95)

```
# Simplyfy variable references;
len<-ToothGrowth$len
supp<-ToothGrowth$supp
dose<-ToothGrowth$dose

t.test(len ~ supp, 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
```

p-value is 0.06 and the confidence interval contains zero. So we do not have enough evidence to reject the null hypothesis that the different supplement types have no effect on tooth length.

[ii] Check for group differences due to different Vitamin C dosage level

Perform t-test for the difference of the means of 2 dosage combinations.

Start with the test whether the mean tooth length of the of the group with vitamin C dose of 2mg, is equal to the group with vitamin C dosage of 1mg.

Assume unequal variances between the two groups. (var.equal = FALSE, conf.level = 0.95)

```
t.test(len[dose==2] , len[dose==1])

##
##  Welch Two Sample t-test
##
## data:  len[dose == 2] and len[dose == 1]
## t = 4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.734 8.996
## sample estimates:
```

```
## mean of x mean of y
##     26.10     19.73
```

p-value of this test is 0 and the confidence interval does not contain zero. The mean tooth length increases on raising the dose level from 1mg to 2mg. This indicates that we can reject the null hypothesis and establish that increasing the dose level leads to an increase in tooth length.

Next test whether the mean tooth length of the of the group with vitamin C dose of 1mg, is equal to the group with vitamin C dosage of 0.5mg.

```
t.test(len[dose==1] , len[dose==0.5])
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dose == 1] and len[dose == 0.5]
## t = 6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276 11.984
## sample estimates:
## mean of x mean of y
##     19.73     10.61
```

p-value of this test is 0 and the confidence interval does not contain zero. Again, the mean tooth length increases on raising the dose level from 0.5mg to 1mg. This indicates that we can reject the null hypothesis and establish that increasing the dose level leads to an increase in tooth length.

4. Conclusions and assumptions needed for the conclusions.

Conclusions

```
Supplement type has no effect on tooth growth.
Increasing the dose level leads to increase in tooth growth.
```

Assumptions

```
The experiment design suggest that each Guinea pig was randomly assigned to a
combination of dosage and supplement type, so the test that we performed used
the independent samples methodology.
The sample of 60 Guinea pigs is assumed to be representative of all Guinea pi
gs, so that conclusions can be generalized to the population.
For the t-tests, the variances are assumed to be different (unequal) for the
two groups being compared.
```