Basic Inferential Data Analysis

Marçal Boix

Nov 21, 2015

In the second part of the project, we analyze the ToothGrowth data in the R datasets package. The data is set of 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).

Initialization

In order to complete the project we need to initialize R with a few statements.

```
library(ggplot2)
library(pander)
```

Loading data

```
# ToothGrowth - The Effect of Vitamin C on Tooth Growth in Guinea Pigs.
library(datasets)
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
```

Basic Summary

```
# Structure of data frame.
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 ...
```

```
# Summary of columns. Few statistics for each column.
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
```

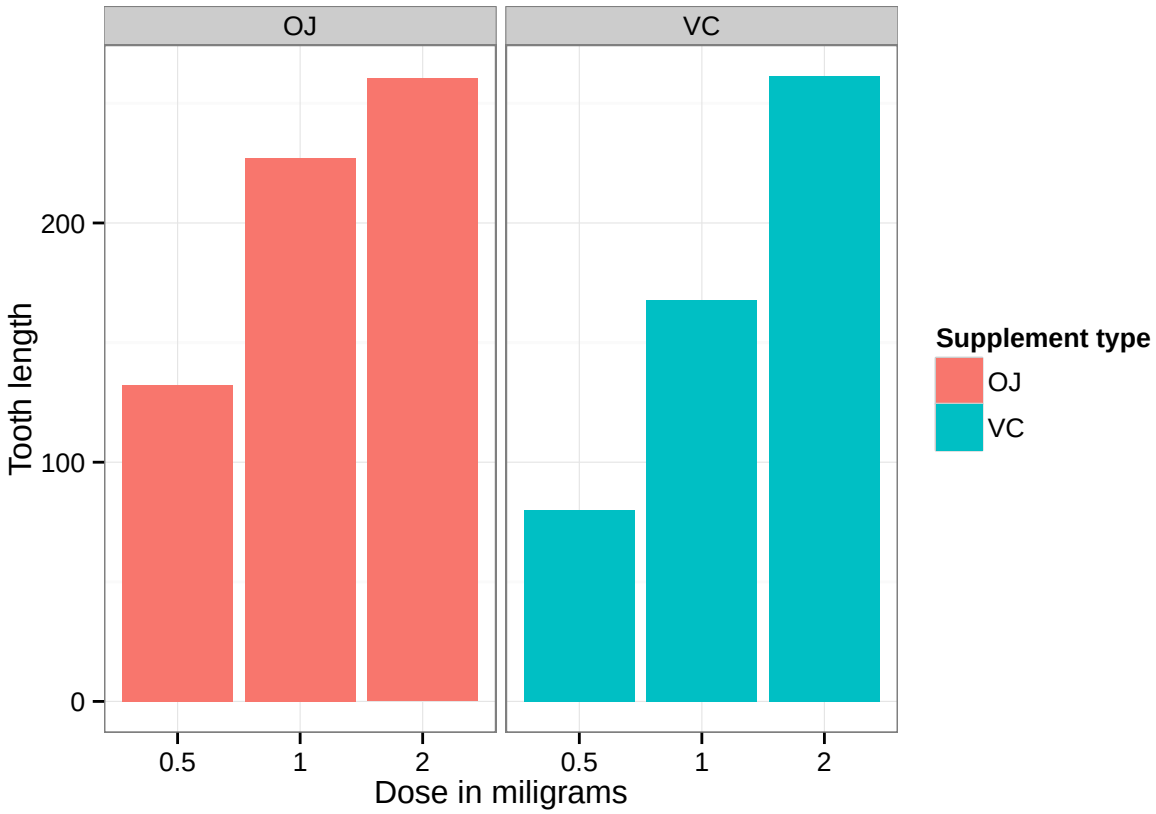
```
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
# Number of observations by supplement type and dose.
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

Correlation

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity") +
    facet_grid(. ~ supp) +
    xlab("Dose in miligrams") +
    ylab("Tooth length") +
    guides(fill=guide_legend(title="Supplement type")) +
    theme_bw()
```



As we can see in the graph above, there is a positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods.

For each supplement type and dose, we will calculate mean, sd and variance.

```
dose = as.numeric(levels(as.factor(ToothGrowth$dose)))
supp = levels(ToothGrowth$supp)
basicSt = list()
for (d in dose)
{
    for (s in supp)
    {
        subTG = ToothGrowth$len[ToothGrowth$dose==d & ToothGrowth$supp==s]
        basicSt <<- rbind(basicSt, list(supp = s, dose = d, mean=mean(subTG),
                                        sd=sd(subTG), var =var(subTG)))
    }
}
as.data.frame(basicSt)
```

```
##   supp dose  mean       sd      var
## 1   OJ  0.5 13.23 4.459709   19.889
## 2   VC  0.5  7.98 2.746634    7.544
## 3   OJ    1  22.7 3.910953 15.29556
## 4   VC    1 16.77 2.515309 6.326778
## 5   OJ    2 26.06 2.655058 7.049333
## 6   VC    2 26.14 4.797731 23.01822
```

Multiple hypothesis testing

To check if there is a real difference between the groups by dose level and delivery method, we will do a two-sided unpaired t-tests.
This test allow us to obtain the confidence intervals and p-values.
The null hypothesis, in all cases, is that there is no difference in the means between the two groups.

```
tests = list()
for (d in dose) {
  ojDose <- ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "OJ"]
  vcDose <- ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "VC"]
  # We assume unequal variance.
  t <- t.test(ojDose, vcDose, var.equal=FALSE, paired=FALSE)
  id <- paste("OJ", d, "-", "VC", d)
  tests <- rbind(tests, list(id=id, p.value=round(t$p.value,5),
                             ci.lo=round(t$conf.int[1],6), ci.hi=round(t$conf.int[2],6)))
}
tests
```

```
##      id                p.value ci.lo    ci.hi
## [1,] "OJ 0.5 - VC 0.5" 0.00636 1.719057 8.780943
## [2,] "OJ 1 - VC 1"     0.00104 2.802148 9.057852
## [3,] "OJ 2 - VC 2"     0.96385 -3.79807 3.63807
```

Conclusions

1. For 0.5 and 1 mg dose levels, there is a significant difference between the means of the OJ and VC groups (p-values < 0.05 and the 95% confidence intervals doesn't include zero).

2. For the 2 mg dose level, we fail to reject the null hypothesis (p-value > 0.5 and the 95% confidence interval includes zero). There is no significative influence of the delivery method on tooth growth in guinea pigs, for 2mg dose.