Part 2: Basic Inferential Data Analysis

Ivan Gimenez

January 7, 2017

Assignment Scope

Mean tooth length of Guinea Pigs is a raising cause of concern because its complex socioeconomic and political implications. The effect of the C Vitamin on Guinea Pigs odontoblasts is becoming a much debated subject among scientific community. Nevertheless, brave new studies are bringing light to this worrisome problem. In this second part of the assignment project, we're going to analyze the ToothGrowth data in the R datasets package.

Dataset load

In order to load the ToothGrowth dataset, the 'datasets' library along with ggplot2 must be loaded, and then after, load the ToothGrowth dataset from the 'datasets' library, with the following commands

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

As stated by the 'datasets' library ToothGrowth help documentation, the dataset documents the effect of vitamin C on tooth growth in guinea pigs, in particular the response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice (coded as OJ) or ascorbic acid, a form of vitamin C (coded as VC).

Initial Summary of the Data

To provide an initial overview of the dataset, let's take a peek into the dataset with str()

```
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 ...
```

The dataset consists in a data frame of 60 observations on 3 variables.

Variable	Type	Description
len	numeric	*Tooth length*
supp	factor	*Supplement type: VC='Vitamin C' or OJ='Oranje Juice'*
dose	numeric	*Dose in mg/day* (converted to factor)

Let's display a statistical summary

```
# Display a summary of the data
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
```

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

Let's check how observations are distributed depending on dose and supplement delivery method,

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

so therefore, it looks an even sampling: two sets of 30 guinea pigs per supplement delivery method, and groups of 10 randomly chosen guinea pigs within the set receive diferent dosages (0.5, 1 and 2 mg) of C-Vitamin.

Exploratory Analysis

To provide an initial overview of the properties of the dataset, we will plot how the dose and supplement delivery method of the C vitamin affects the tooth length.

First, we will explore how the tooth length depends on the C Vitamin dose, per each supplement delivery method.

```
# Plot tooth length vs dose, by supplement delivery method
ggplot(aes(x=dose, y=len), data=ToothGrowth) +
                    geom_boxplot(aes(fill=dose)) +
                    xlab("Dose (mg/day)") +
                    ylab("Length") +
                    facet_grid(~ supp) +
                    ggtitle("Tooth Length vs Dose by Delivery Method") +
                    theme(legend.position='none')
```

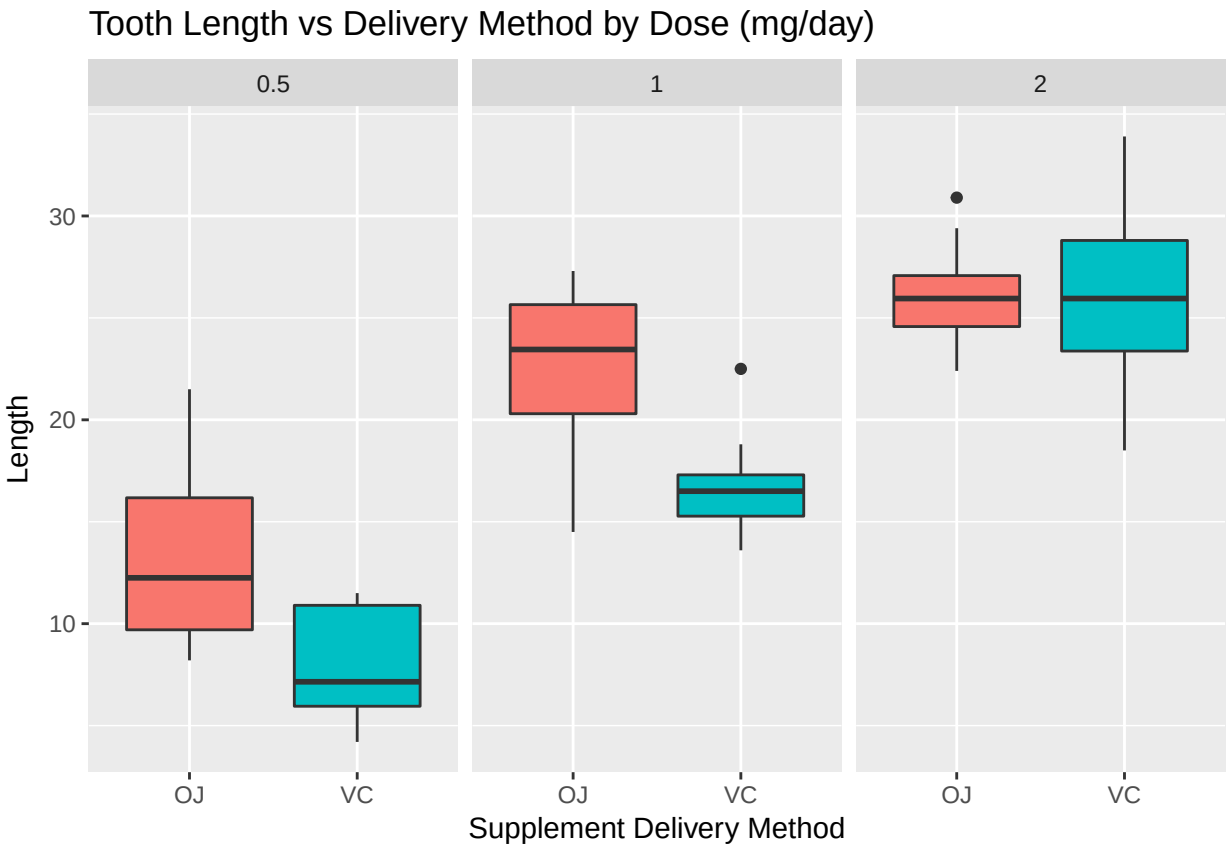
Apparently, it looks like there is a positive correlation between the mean tooth lenght and the dose of C vitamin, regardless the supplement delivery method.

```
OJ<-subset(ToothGrowth, supp=="OJ")
VC<-subset(ToothGrowth, supp=="VC")
```

Method	Mean (.5,1,2 mg/day)	Variance (.5,1,2 mg/day)
Oranje Juice(OJ)	13.23, 22.7, 26.06	19.889, 15.2955556, 7.0493333
Ascorbic Acid(VC)	7.98, 16.77, 26.14	7.544, 6.3267778, 23.0182222

Second, we will explore how the tooth length depends on the supplement delivery method, per each C Vitamin dosage.

```
# Plot tooth length vs supplement delivery method, by dose
ggplot(aes(x=supp, y=len), data=ToothGrowth) +
                    geom_boxplot(aes(fill=supp)) +
                    xlab("Supplement Delivery Method") +
                    ylab("Length") +
                    facet_grid(~ dose) +
                    ggtitle("Tooth Length vs Delivery Method by Dose (mg/day)") +
                    theme(legend.position='none')
```



From the figure, it looks like at low C Vitamin dosages (.5 and 1 mg/day), the Oranje juice supplement delivery method looks more effective than Ascorbic Acid in terms of attained tooth length.

```
dose05<-subset(ToothGrowth, dose==.5)
dose10<-subset(ToothGrowth, dose==1)
dose20<-subset(ToothGrowth, dose==2)
```

Dose	Mean (OJ,VC)	Variance (OJ,VC)
0.5 mg/day	13.23, 7.98	19.889, 7.544
1.0 mg/day	22.7, 16.77	15.2955556, 6.3267778
2.0 mg/day	26.06, 26.14	7.0493333, 23.0182222

Inference Analysis

First, let's analyse the alternative hypothesis that the tooth length of the Guinea Pigs depends on supplement delivery methods. The null hypothesis is that different supplement types have no effect on tooth length. We will be under the assumption of unequal variance as these two groups are not paired.

A t test for the difference will test this hypothesis:

```
t.test(VC$len,OJ$len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  VC$len and OJ$len
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

The p-value of this test was 0.06, which is close but greater to the significance level of 5%. Furthermore, the confidence interval of the test contains zero (0). This indicates that ***we can not reject the null hypothesis that the different supplement types have no effect on tooth length***.

Second, let's test the alternative hypothesis that the tooth length of the Guinea Pigs increases as C Vitamin dosage grows. The null hypothesis is that increasing C Vitamin dosages have no effect on tooth length.

A t test for the difference between all dosage pairs will test this hypothesis:

```
t.test(dose05$len, dose10$len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose05$len and dose10$len
## 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 of x mean of y
##    10.605    19.735
```

```
t.test(dose05$len, dose20$len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose05$len and dose20$len
## 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 of x mean of y
##    10.605    26.100
```

```
t.test(dose10$len, dose20$len, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  dose10$len and dose20$len
## 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 of x mean of y
##    19.735    26.100
```

We can confirm on all 3 tests that p-value was much smaller than significant level of 5%. Moreover, the confidence intervals were all below zero. This indicates that ***we can reject the null hypothesis that increasing C Vitamin dosages have no effect on tooth length***

Conclusions

- Supplement Delivery Method has no effect on the tooth length.
- Increasing levels of C Vitamin dosages leads to longer Guinea Pig tooth lengths, which is great.