Coursera Satistical Inference - Course Project PART 2

Author: Dino Chioda Date: May 23, 2015

Overview

In this project, which is the second part of the Course Project for the Statistical Inference course, we're going to analyze the ToothGrowth data in the R datasets package. We will provide a basic summary of the data ans use confidence intervals to test hypotheses for tooth growth by supplement and dose.

Question 1: Load the ToothGrowth data and perform some basic exploratory data analyses

```
# Load the data
library(datasets)
data(ToothGrowth)

# Look at the data
nrow(ToothGrowth)
```

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

```
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 ...
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
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 ...
```

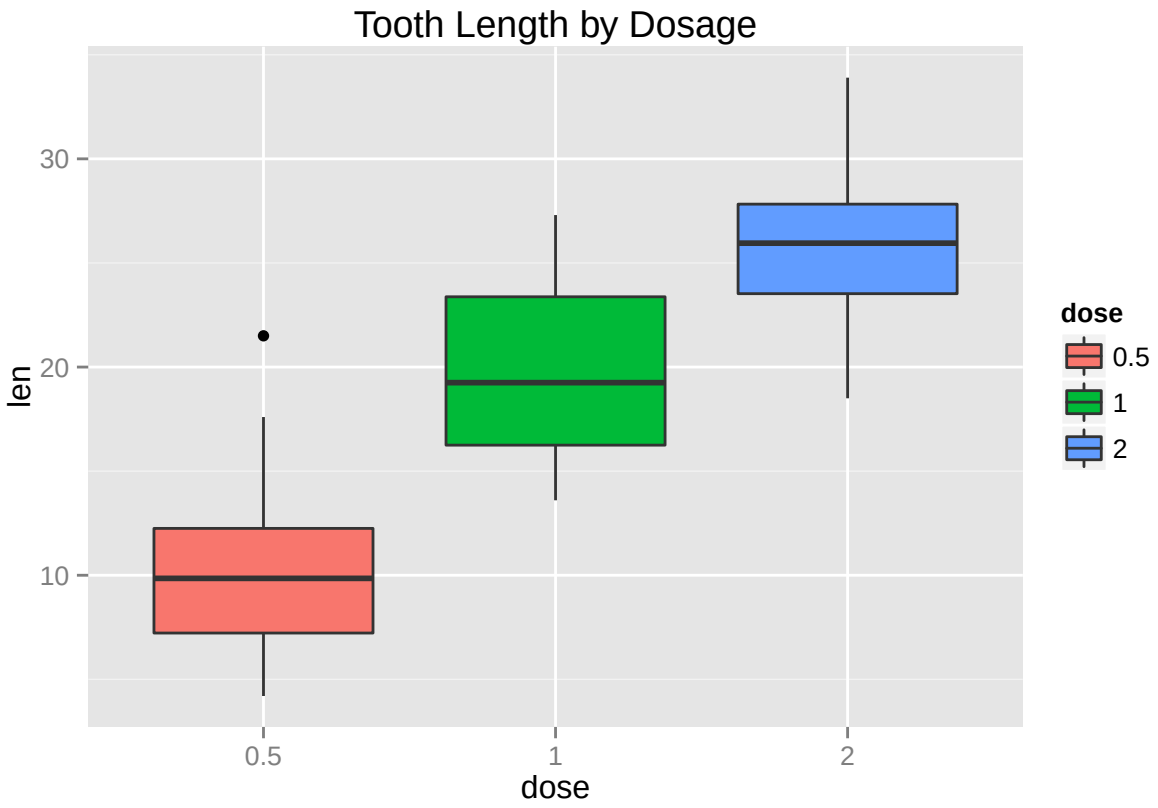
Question 2: Provide a baisc summary of the data

Using the `summary()` R function on the ToothGrowth dataset yields:

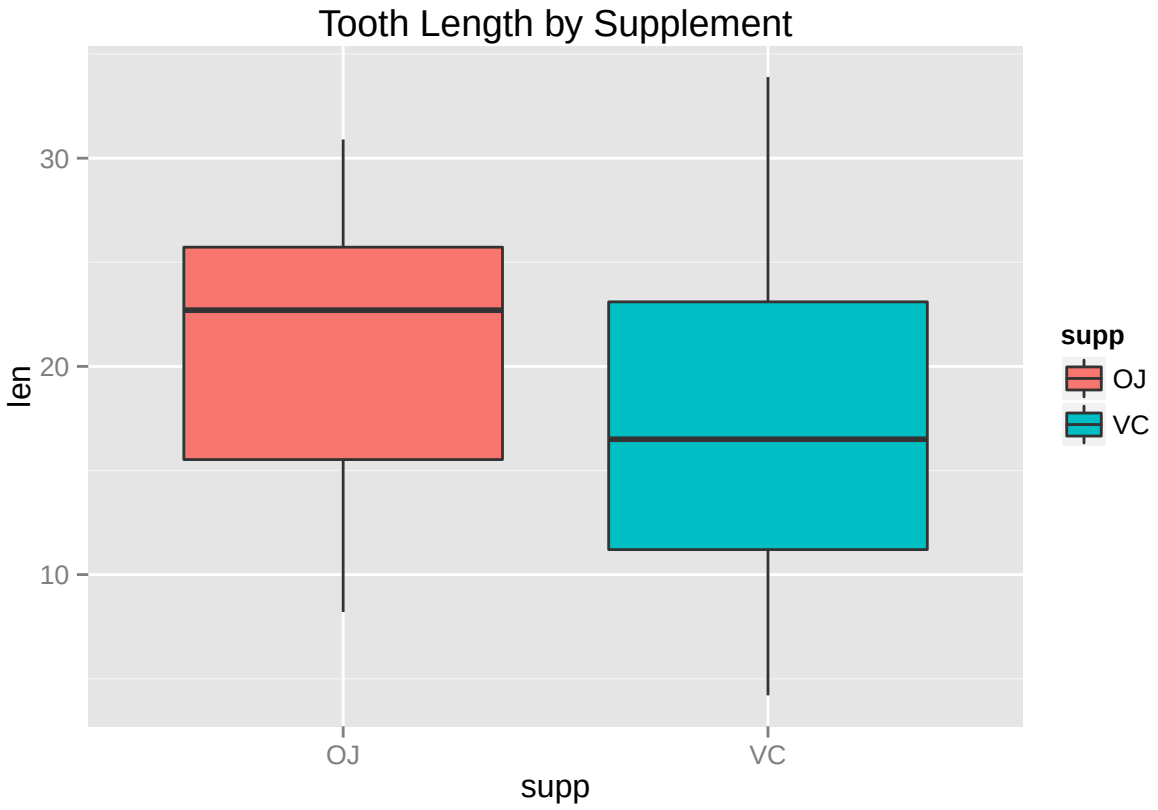
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Let's generate some basic box plots to gain a better understand of how the data behaves:

```
library(ggplot2)
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) +
        ggtitle("Tooth Length by Dosage")
```




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



Question 3: Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Consider the following hypothesis:

H_0: Supplements have no effect on tooth length

H_A: Supplements have an effect on tooth length

We will use a t-test to assess that the H_0 is true.

```
t.test(len~supp, ToothGrowth)
```

```
##
##  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
```

The p-value of the t-test is **0.06**. Therefore, we cannot reject the null hypothesis.

Consider the following hypothesis:

H_0: Increasing the dosage has no effect on tooth length

H_A: Increasing the dosage has an effect on tooth length

Since we have 3 distinct dosages, we have 3 possible dosage pairs to consider. Therefore, we will perform a t-test on each of the dosage pairings. Let's begin by subsetting by dosage pairings: 0.5 & 1.0, 0.5 & 2.0, 1.0 & 2.0:

```
ToothGrowth.doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
ToothGrowth.doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
ToothGrowth.doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

Let's perform the 3 t-tests to validate H_0:

```
t.test(len ~ dose, data = ToothGrowth.doses_0.5_1.0)
```

```
##
##  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
```

```
t.test(len ~ dose, data = ToothGrowth.doses_0.5_2.0)
```

```
##
##  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
```

```
t.test(len ~ dose, data = ToothGrowth.doses_1.0_2.0)
```

```
##
##  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
```

Since all 3 p-values are less than 0.05 and the confidence intervals do not contain 0, we can reject H_0 and state that increasing the dosage leads to increased tooth length.

Question 4: State you conclusions and the assumptions needed for your conclusions.

Conclusions

1. Supplement type does not seem to have an effect on tooth length.
2. Increasing dosage levels does appear to lead to increased tooth length.

Assumptions

1. The sample population of guinea pigs is representative of the entire population of guinea pigs tested.
2. The sample population of guinea pigs were randomly selected.
3. All dosage pair subset considered have the same variance.