Analysis of the ToothGrowth data in R

Statistical Inference Course project part 2

Ray Qiu

October 22, 2015

Overview

In the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package, and try to draw some conclusions based on statistical inference.

Load required R libraries

```
library(ggplot2)
library(dplyr)
```

Load the ToothGrowth data and perform some basic exploratory data analysis. Provide a basic summary of the data.

```
data("ToothGrowth")
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 ...
```

```
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
```

```
tail(ToothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```
summary(ToothGrowth)
```

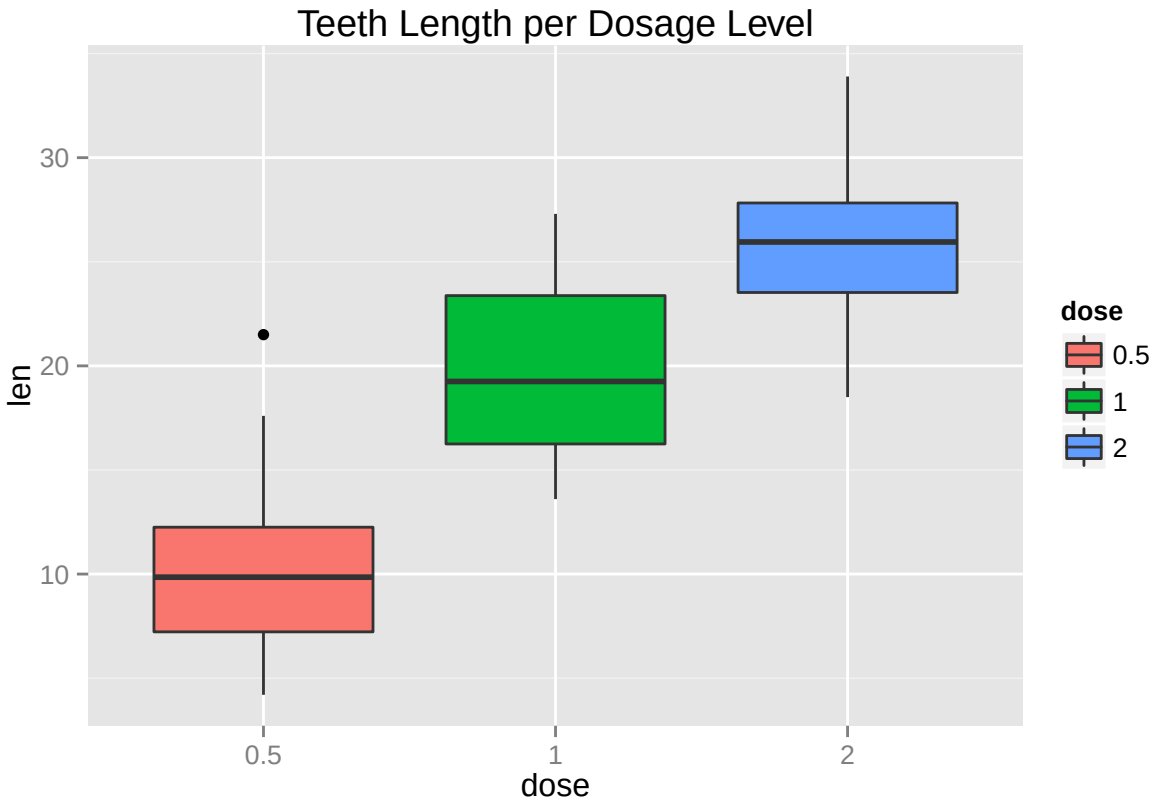
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Convert dose to factors

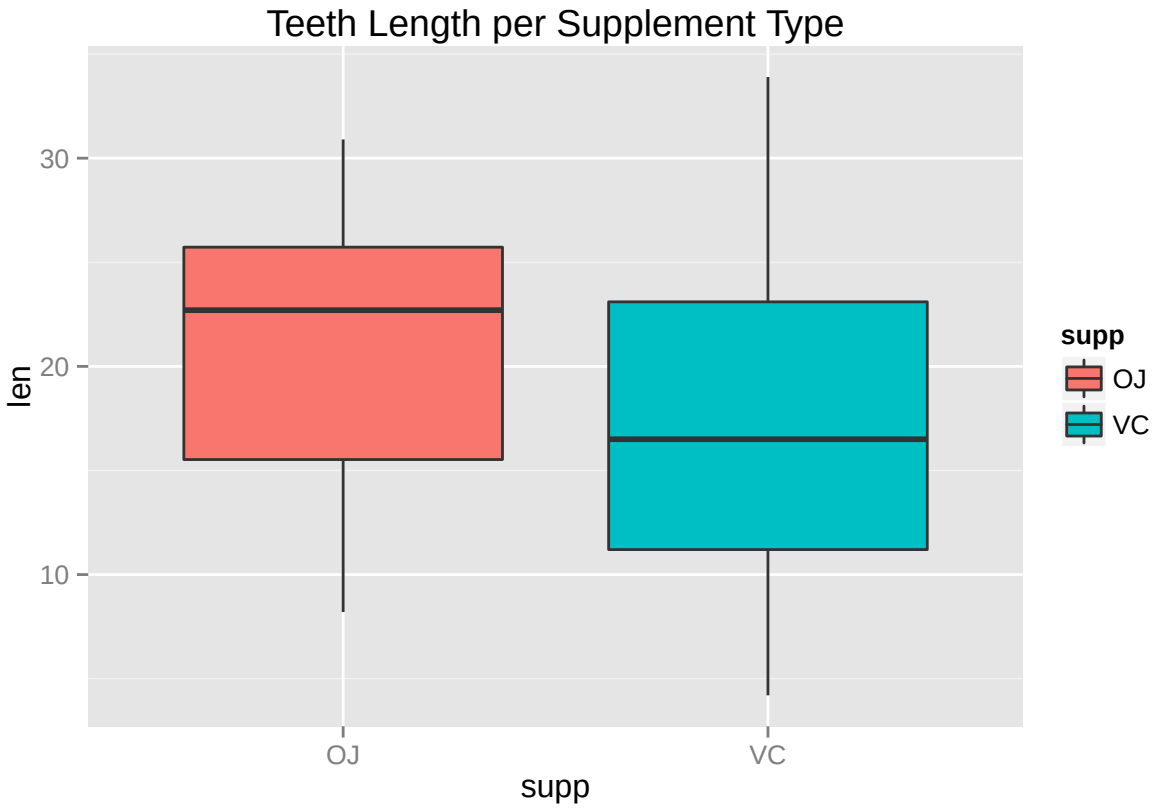
```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

Some exploratory plots

```
ggplot(ToothGrowth, aes(x = dose, y = len)) +
    ggtitle("Teeth Length per Dosage Level") +
    geom_boxplot(aes(fill = dose))
```




```
ggplot(ToothGrowth, aes(x = supp, y = len)) +
    ggtitle("Teeth Length per Supplement Type") +
    geom_boxplot(aes(fill = supp))
```



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

Test whether supplement types have effects on teeth length

```
t.test(len ~ supp, data = 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 test result shows that p-value is 0.06063, which is greater than 0.05; And 95 percent confidence interval values contain 0. Based on this we cannot reject the null hypothesis that different supplement types have no effect on teeth length.

Separate data into three dose groups where each group contains a pair of doses, then test each dose pair to see whether different doses have different effects on teeth length.

```r
doseGroup1 <- ToothGrowth %>% filter(dose != 2.0)
doseGroup2 <- ToothGrowth %>% filter(dose != 1.0)
doseGroup3 <- ToothGrowth %>% filter(dose != 0.5)
```

Test doseGroup1

```r
t.test(len ~ dose, data = doseGroup1)
```

```
##
##  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
```

Test doseGroup2

```r
t.test(len ~ dose, data = doseGroup2)
```

```
##
##  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
```

Test doseGroup3

```r
t.test(len ~ dose, data = doseGroup3)
```

```
##
##  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
```

The test result shows that p-value is close to 0; And 95 percent confidence interval values don't contain 0. Based on this we can reject the null hypothesis, and conclude that increasing dosage will increase teeth growth.

State your conclusions and the assumptions needed for your conclusions.

Conclusions

- Supplement type has no effect on teeth growth.
- Increasing dosage level leads to increased teeth growth.

Assumptions

- The experiment design suggest that each Guinea pig was randomly assigned to a combination of dosage and supplement type, so the tests use the independent samples methodology.
- The sample of 60 Guinea pigs is representative of all Guinea pigs, so we can generalize the test conclusion to the population.
- For t-test regarding tooth length per supplement type, the variances are assumed to be different for the two groups being compared. For t-tests regarding teeth length per dosage level, we assume a constant variance across the groups.