Coursera, Statistical Inference, Part 2

Heather Dye

3/16/2018

Overview: Basic Inferential Data Analysis

Analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Load Data

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

Basic Summary

```
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(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
```

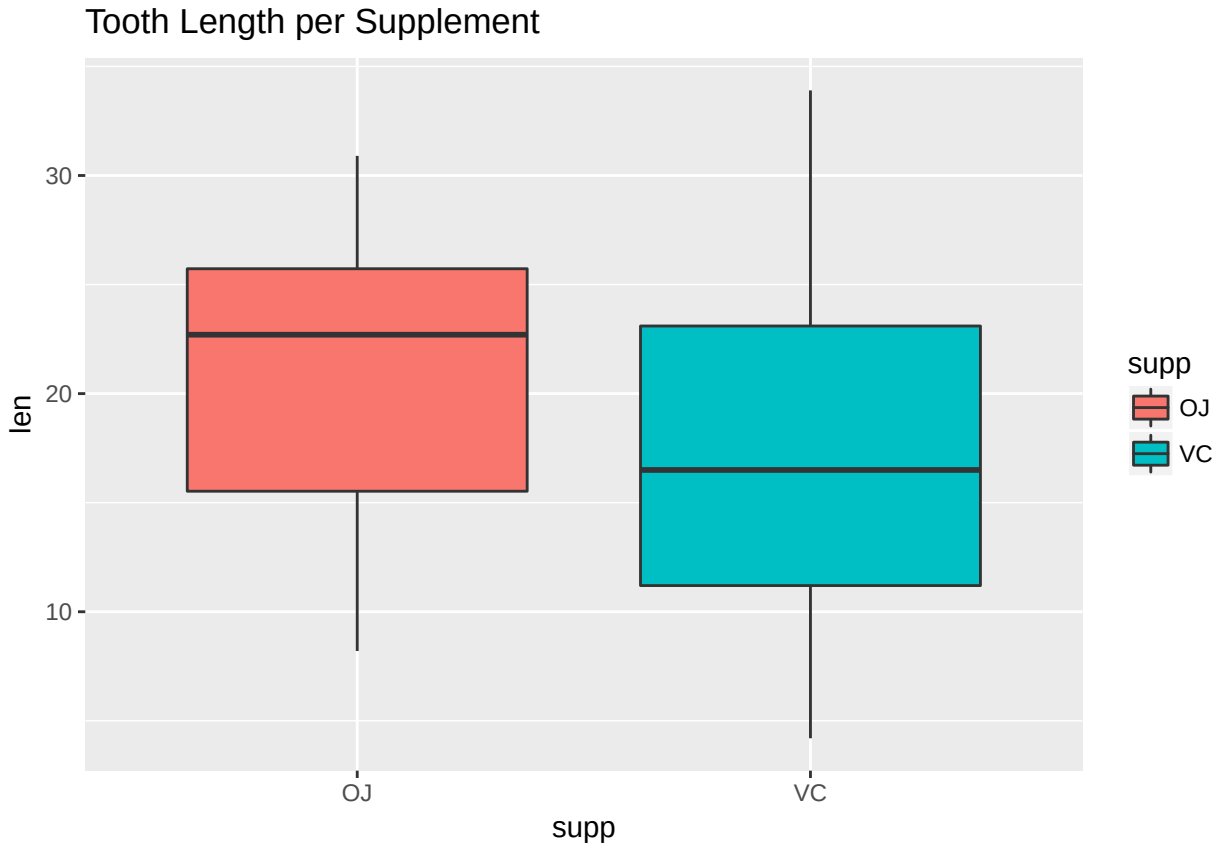
Ten observations per each set of variables are avialble. The variables are dose of 0.5, 1 and 2 and supplement: OJ and VC.

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

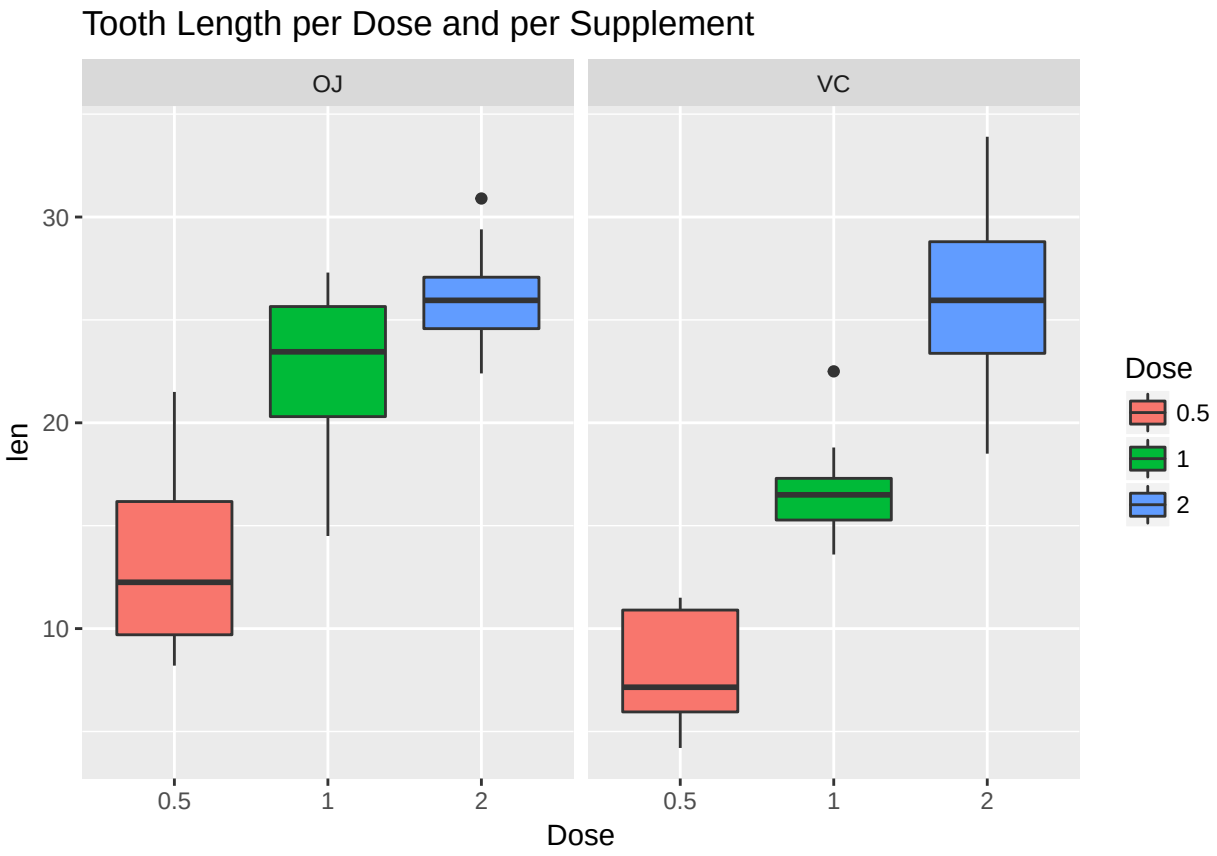
```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Boxplots

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




```
ggplot(ToothGrowth, aes(x=as.factor(dose),y=len,fill = as.factor(dose))) +
  geom_boxplot() +
  facet_wrap( ~ supp) +
  ggtitle("Tooth Length per Dose and per Supplement") +
  labs(x = "Dose", fill = "Dose")
```



Compare tooth growth by supp and dose using hypothesis tests and / or intervals

Comparison is done by performin t test on each set per dosage.

```
dose_0.5 <- ToothGrowth[ToothGrowth$dose == 0.5, ]
dose_1.0 <- ToothGrowth[ToothGrowth$dose == 1.0, ]
dose_2.0 <- ToothGrowth[ToothGrowth$dose == 2.0, ]
```

T test for 0.5 dosage

```
test0.5 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose_0.5)
sprintf("p-value for dose = 0.5: %g", test0.5$p.value)
```

```
## [1] "p-value for dose = 0.5: 0.00635861"
```

```
sprintf("Confidence Interval for Dose = 0.5: %g", test0.5$conf)
```

```
## [1] "Confidence Interval for Dose = 0.5: 1.71906"
## [2] "Confidence Interval for Dose = 0.5: 8.78094"
```

T test for 1.0 dosage

```
test1.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose_1.0)
sprintf("p-value for dose = 1.0: %g", test1.0$p.value)
```

```
## [1] "p-value for dose = 1.0: 0.00103838"
```

```
sprintf("Confidence Interval for Dose = 1.0: %g", test1.0$conf)
```

```
## [1] "Confidence Interval for Dose = 1.0: 2.80215"
```

```
## [2] "Confidence Interval for Dose = 1.0: 9.05785"
```

T test for 2.0 dosage

```
test2.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose_2.0)
sprintf("p-value for dose = 2.0: %g", test2.0$p.value)
```

```
## [1] "p-value for dose = 2.0: 0.963852"
```

```
sprintf("Confidence Interval for Dose = 2.0: %g", test2.0$conf)
```

```
## [1] "Confidence Interval for Dose = 2.0: -3.79807"
## [2] "Confidence Interval for Dose = 2.0: 3.63807"
```

Conclusions

P-values are lower than 0.05 for dose 0.5 and 1.0. Therefore we can reject the null hypothesis and accept the alternative hypothesis that OJ is correlated with the increased rate of tooth growth of guinea pigs.

P-value is above 0.05 for 2.0 dosage therefore we accept the null hypothesis, meaning there is no statistical difference between controlled and test group and the use of supplement cannot be correlated with different rates of tooth growth of guinea pigs.