Basic inferential data analysis

pierre attey

Sunday, January 25, 2015

Background

- The goal of this second part of project is to analyze the ToothGrowth data in the R datasets package through the following steps: `Load the ToothGrowth data and perform some basic exploratory data analyses` `Provide a basic summary of the data.` `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) `State your conclusions and the assumptions needed for your conclusions`

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
 - Using summary statistics to spot problems by using DescTools Packages
tools for descriptive statistics
```

The variable dose should be a factor, not numeric we have to assign it to be a factor

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

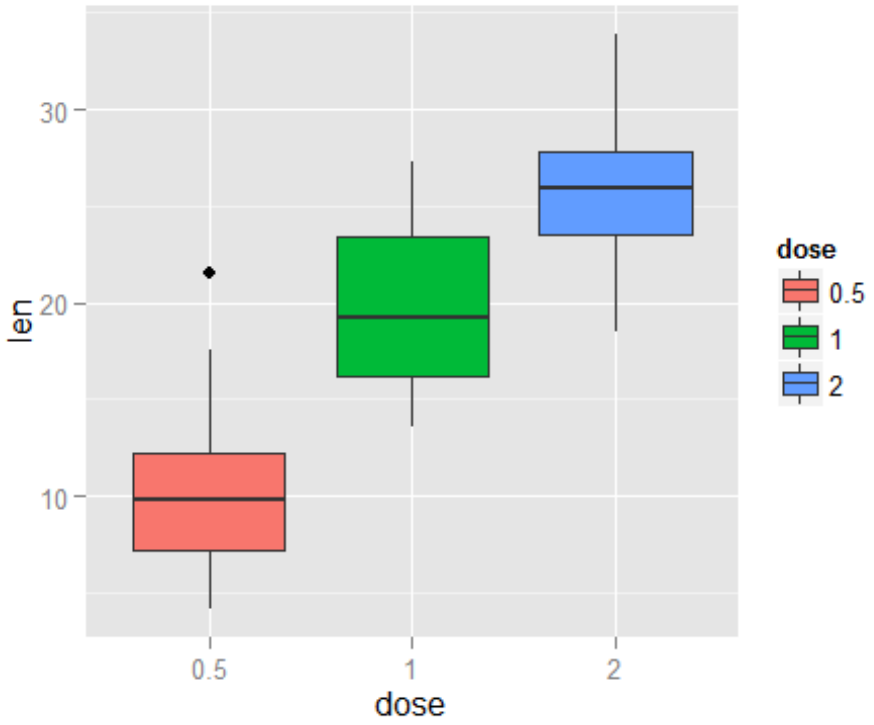
- look at the dataset variables after conversion

2. Provide a basic summary of the data

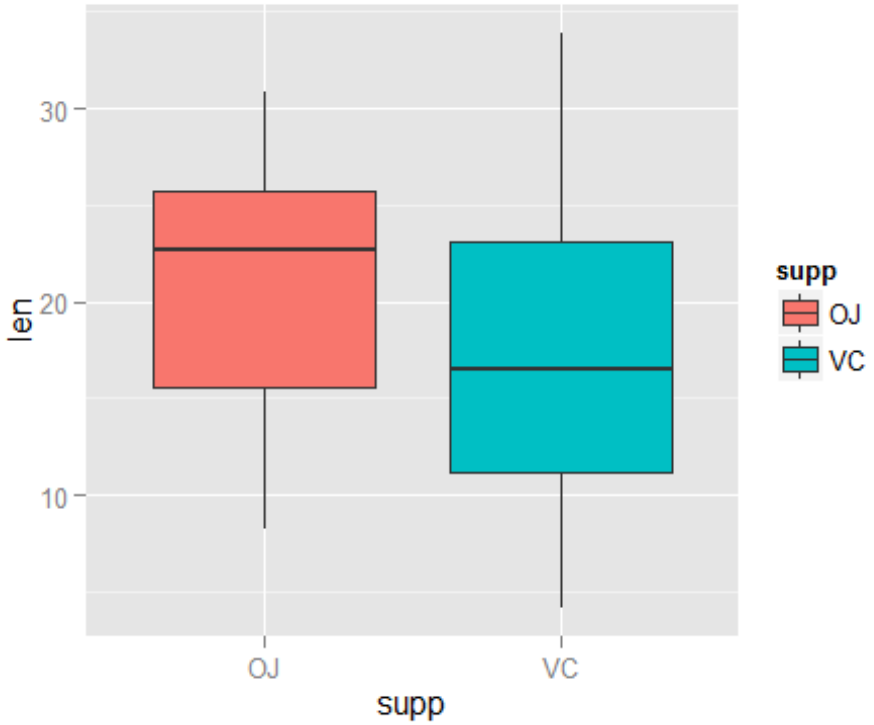
```
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

```
library(ggplot2)
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
```




```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```



3. Use confidence intervals and/or hypothesis tests

- control of the differences between groups assuming unequal inter groups variances

```
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 confidence interval contains zero because of The p-value. This indicates that we can not reject the null hypothesis that the different supplement types have no effect on tooth length

- Subsetting three sub-groups as per dose level pairs
- control of the differences between all subset groups

```
t.test(len ~ dose, data = ToothGrowth.doses1)
```

```
##
##  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.doses2)
```

```
##
##  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.doses3)
```

```
##
##  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 p-value for all this test is less than 0.05, and the confidence interval does not contain zero. The mean tooth length increases on raising the dose level. This indicates that we can not accept the null hypothesis, and establish that increasing the dose level leads to an increase in tooth length.

4. State your conclusions and the assumptions needed for your conclusions

Supplement type has no effect on tooth growth and Inreasing the dose level leads to increased tooth growth.