Coursera: Statistical Inference Course Project - Tooth Growth analysis

Sunday, September 21, 2014

Description of the assignment

We're going to analyze the ToothGrowth data in the R datasets package. Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.State your conclusions and the assumptions needed for your conclusions.

Data load and exploratory analysis

```
library(datasets)
data(ToothGrowth)
attach(ToothGrowth)

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
```

```
summary(ToothGrowth)
```

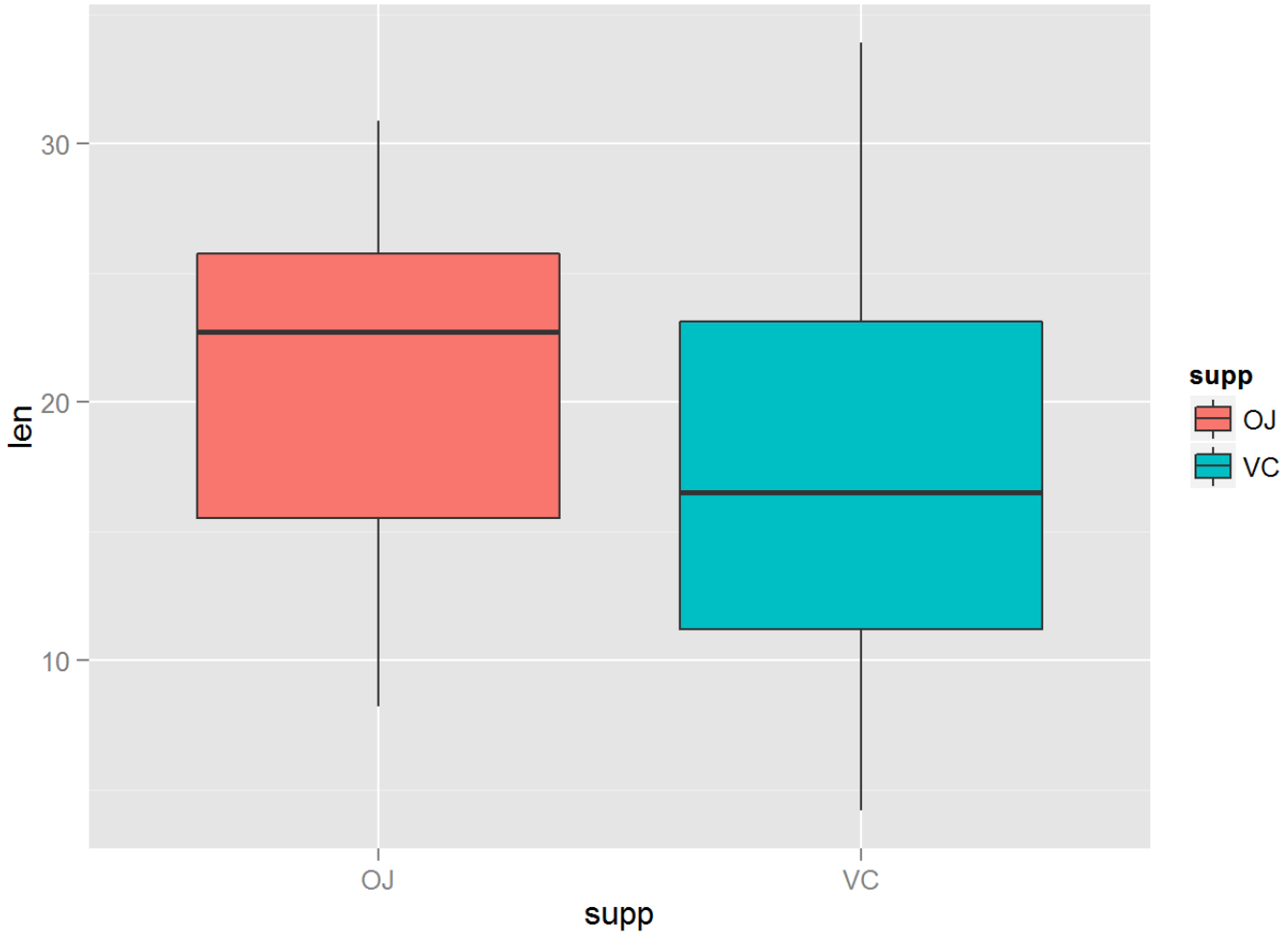
```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

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

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

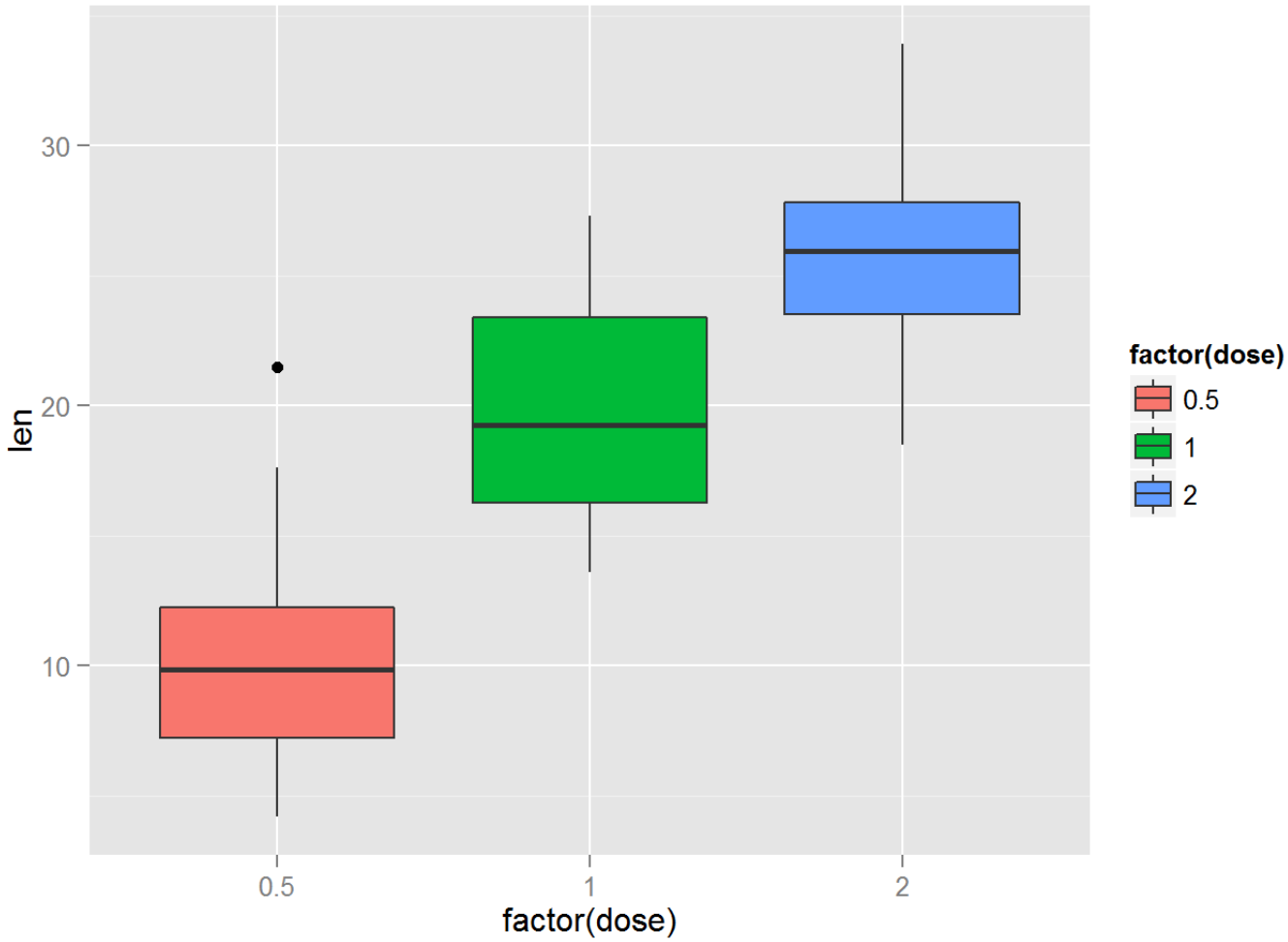
Plots to show relationships between length of tooth, supplement type and dosage levels

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



We see above that tooth growth length is greater for the supplement type OJ vs. VC.

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



We see above that tooth growth length increases with increasing dosage.

T-Test for mean difference by supplement type

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

p-value is not small and the confidence interval contains 0, i.e. we cannot reject null hypothesis. This means that there is no difference by supplement type

T-Test for mean difference by dosage level

Lets take three subsets according to dosage pairs

```
dose1 <- subset(ToothGrowth,dose %in% c(0.5,1.0))
dose2 <- subset(ToothGrowth,dose %in% c(0.5,2.0))
dose3 <- subset(ToothGrowth,dose %in% c(1.0,2.0))
```

Now we use T-tests to determine if there are difference 2 dosage groups.

```
t.test(len ~dose, data = dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

```
t.test(len ~dose, data = dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

```
t.test(len ~dose, data = dose3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

Each of the above have very small p-values which means the null hypothesis can be rejected.

T-Test for mean difference of supplement by dosage level

To test the effect of supplements (OJ & VC) for each dosage type, we will have to subset the data into 3 different sets.

```
supp1 <- subset(ToothGrowth , dose ==0.5)
supp2 <- subset(ToothGrowth , dose ==1.0)
supp3 <- subset(ToothGrowth , dose ==2.0)
```

Now we use T-tests to determine if there are difference 2 dosage groups.

```
t.test(len ~supp, data = supp1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~supp, data = supp2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~supp, data = supp3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

First two above i.e. with dose 0.5 and 1.0 are statistically significant since the p-values are small, hence, for them the null hypothesis can be rejected. However, for the thrid one ie. dose = 2.0, has large p-value which means the null hypothesis will have to be retained.

Assumption

1.Experiment assumed IID samples

2.Sample is representative of the population

Conclusions after performing t-tests

1. No difference by supplement type i.e. OJ or VC do not have any effect independently on tooth length growth

2. Dosage levels on their own have a significant effect on growth of tooth length

3. For dosage levels 0.5 and 1.0, the supplement OJ has a greater effect on tooth length growth than the supplement VC. And for the dosage level 2.0, there is no difference in effect of supplements OJ or VC on tooth length growth.