Statistical Inference Course Project Part 2

Chenchen

May 21, 2015

Overview

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

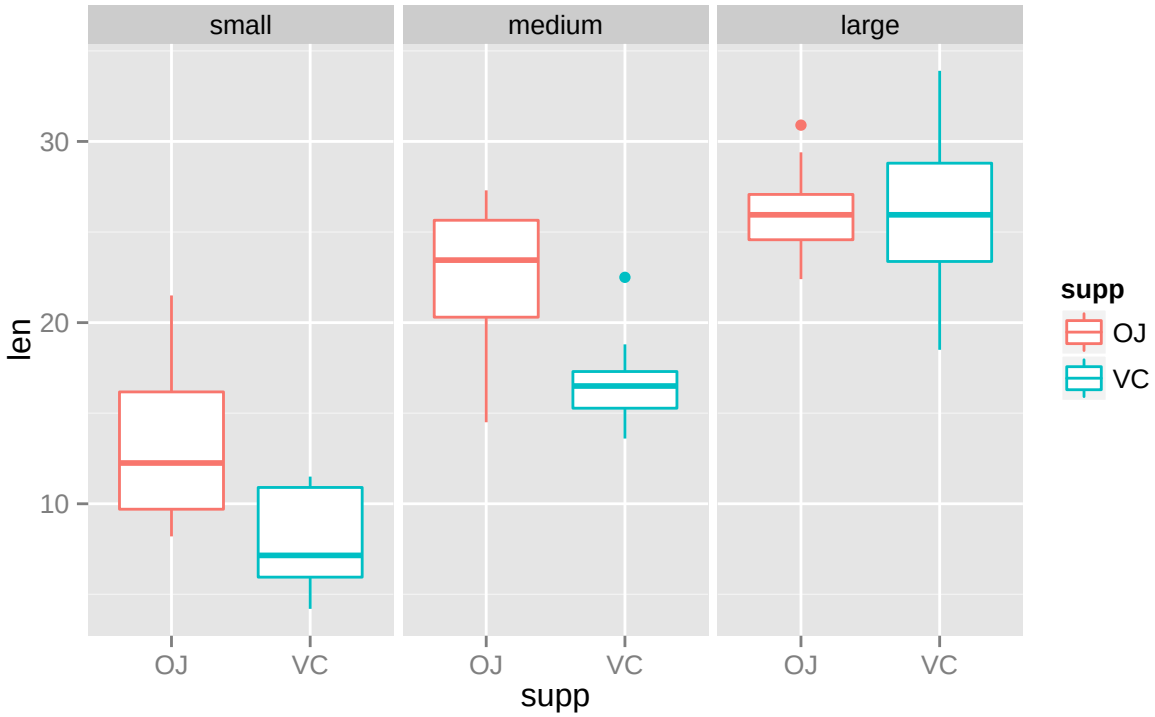
```
# load data
data(ToothGrowth)
head(ToothGrowth)
```

```
##    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
```

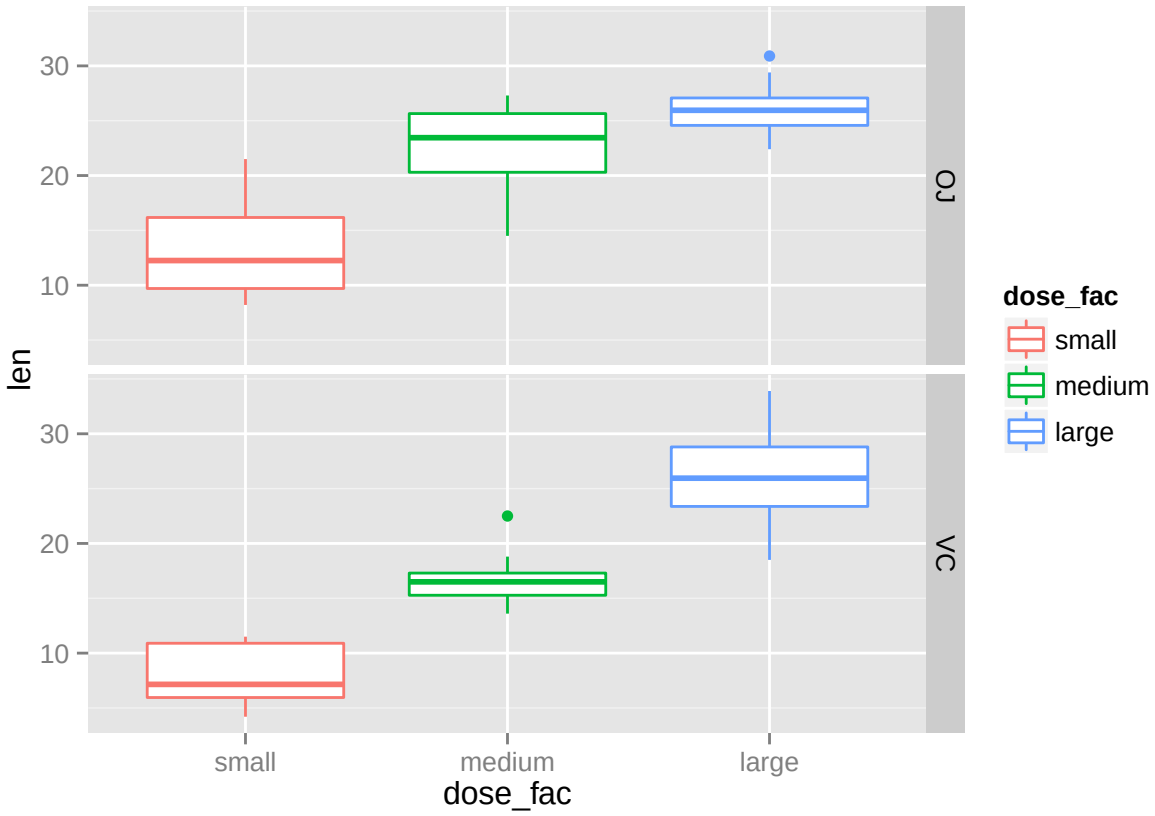
```
# change dose to factors
ToothGrowth$dose_fac = factor(ToothGrowth$dose, levels = c(0.5, 1, 2), labels = c("small", "medium", "la
head(ToothGrowth)
```

```
##    len supp dose dose_fac
## 1  4.2   VC  0.5    small
## 2 11.5   VC  0.5    small
## 3  7.3   VC  0.5    small
## 4  5.8   VC  0.5    small
## 5  6.4   VC  0.5    small
## 6 10.0   VC  0.5    small
```

Exploratory plot 1



Exploratory plot 2



2. Provide a basic summary of the data.

```
summary(ToothGrowth)
```

```
##       len        supp         dose         dose_fac
##  Min.   : 4.20   OJ:30   Min.   :0.500   small :20
##  1st Qu.:13.07   VC:30   1st Qu.:0.500   medium:20
##  Median :19.25           Median :1.000   large :20
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

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

3.1 The distribution of treatment groups

```
tapply(len, list(dose_fac, supp),function(x) length(x))
```

```
##        OJ VC
## small  10 10
## medium 10 10
## large  10 10
```

3.2 perform t-test of OJ vs VC

```
# Subset data by dosage treatment
ToothGrowth_sm = filter(ToothGrowth, dose_fac == "small")
ToothGrowth_md = filter(ToothGrowth, dose_fac == "medium")
ToothGrowth_lg = filter(ToothGrowth, dose_fac == "large")
```

3.2.1 perform t-test of OJ vs VC in small dose group only

```
t_test_sm = t.test(ToothGrowth_sm$len ~ ToothGrowth_sm$supp)
# confidence interval
t_test_sm$conf.int[1:2]
```

```
## [1] 1.719057 8.780943
```

```
# p value
t_test_sm$p.value
```

```
## [1] 0.006358607
```

3.2.2 perform t-test of OJ vs VC in medium dose group only

```
t_test_md = t.test(ToothGrowth_md$len ~ ToothGrowth_md$supp)
# confidence interval
t_test_md$conf.int[1:2]
```

```
## [1] 2.802148 9.057852
```

```
# p value
t_test_md$p.value
```

```
## [1] 0.001038376
```

3.2.3 perform t-test of OJ vs VC in large dose group only

```
t_test_lg = t.test(ToothGrowth_lg$len ~ ToothGrowth_lg$supp)
# confidence interval
t_test_lg$conf.int[1:2]
```

```
## [1] -3.79807  3.63807
```

```
# p value
t_test_lg$p.value
```

```
## [1] 0.9638516
```

3.3 perform t-test of small dosage (small) vs large dosage (large))

```
# For convenience, we only compare small vs large, while drop medium dosages (medium)
ToothGrowth = filter(ToothGrowth, dose_fac != "medium")
ToothGrowth$dose_fac = factor(ToothGrowth$dose_fac)
# Subset data by delivery method
ToothGrowth_OJ = filter(ToothGrowth, supp == "OJ")
ToothGrowth_VC = filter(ToothGrowth, supp == "VC")
```

3.3.1 perform t-test of dose_sm vs dose_lg in OJ group only

```
t_test_OJ = t.test(ToothGrowth_OJ$len ~ ToothGrowth_OJ$dose_fac)
# confidence interval
t_test_OJ$conf.int[1:2]
```

```
## [1] -16.335241  -9.324759
```

```
# p value
t_test_OJ$p.value
```

```
## [1] 1.323784e-06
```

3.3.2 perform t-test of dose_sm vs dose_lg in VC group only

```
t_test_VC = t.test(ToothGrowth_VC$len ~ ToothGrowth_VC$dose_fac)
# confidence interval
t_test_VC$conf.int[1:2]
```

```
## [1] -21.90151 -14.41849
```

```
# p value
t_test_VC$p.value
```

```
## [1] 4.681577e-08
```

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

4.1 conclusions

4.1.1 comparing delivery methods

- When treated with small and medium amount of Vitamin C, orange juice increases more tooth length than ascorbic acid.
- When treated with large amount of Vitamin C, there's no significant difference between the two delivery methods in influencing tooth length.

4.1.2 comparing dose amount

- For both methods, large amount treatment will increase tooth length than small amount

4.2 Assumptions

- These tooth length are approximately normally distributed
- These samples are independent (unpaired)