Statistical Inference Course Project Part2

Atsuko Yamamoto

January 24, 2015

Part2 Basic inferential data analysis.

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

```
setwd("~/Rdata/Inference")
library(datasets)
data(ToothGrowth)
```

2. Provide a basic summary of the data.

```
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 ...
```

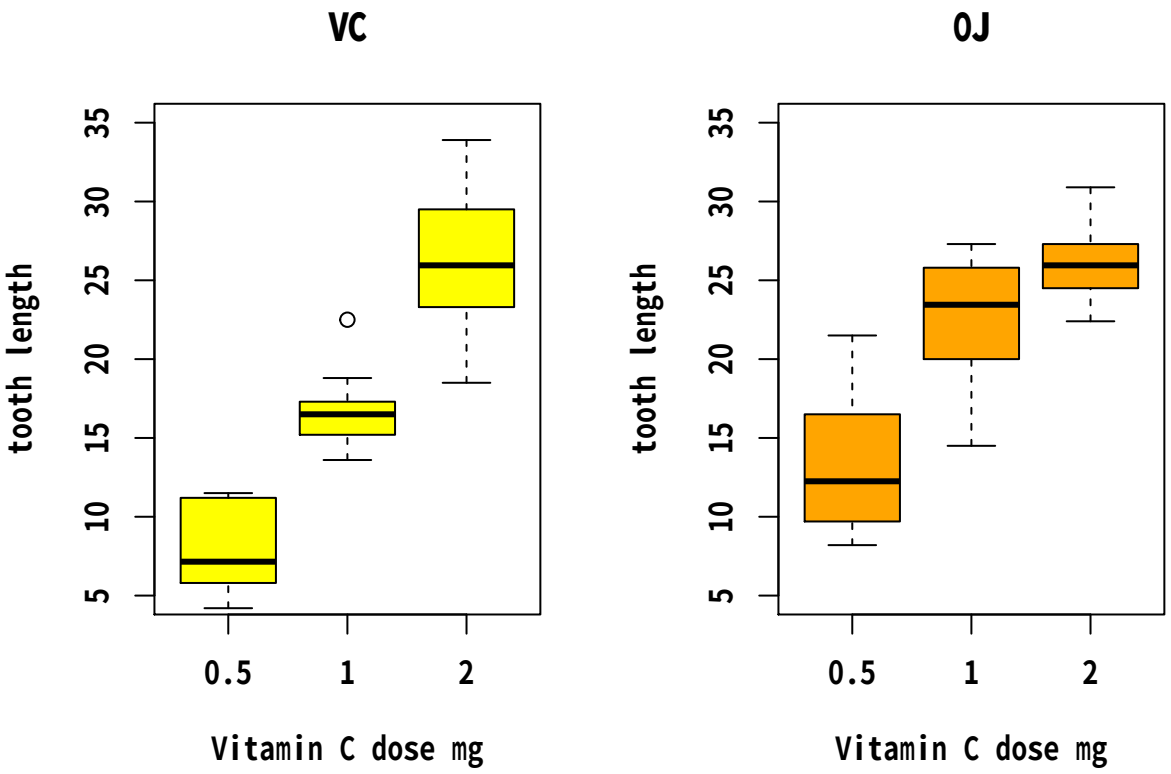
```
summary(ToothGrowth)
```

```
##       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
```

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

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

```
# barplot
library(ggplot2)
par(mfrow = c(1, 2))
boxplot(len ~ dose, data=ToothGrowth,subset = supp == "VC", col = "yellow", main="VC", xlab = "Vitamin C
boxplot(len ~ dose, data=ToothGrowth,subset = supp == "OJ", col = "orange", main="OJ", xlab= "Vitamin C
```



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)

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

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
t.test(ToothGrowth$len[ToothGrowth$dose==0.5], ToothGrowth$len[ToothGrowth$dose==1], var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
t.test(ToothGrowth$len[ToothGrowth$dose==0.5], ToothGrowth$len[ToothGrowth$dose==2], var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5] and ToothGrowth$len[ToothGrowth$dose == 2]
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15352 -12.83648
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```
t.test(ToothGrowth$len[ToothGrowth$dose==1], ToothGrowth$len[ToothGrowth$dose==2], var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

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