Basic Inferential Data Analysis

Oleksandr Fialko

11/28/2016

`ToothGrowth` dataset contains the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice (coded as OJ) or ascorbic acid (a form of vitamin C and coded as VC).

```
data("ToothGrowth")
str(ToothGrowth)
```

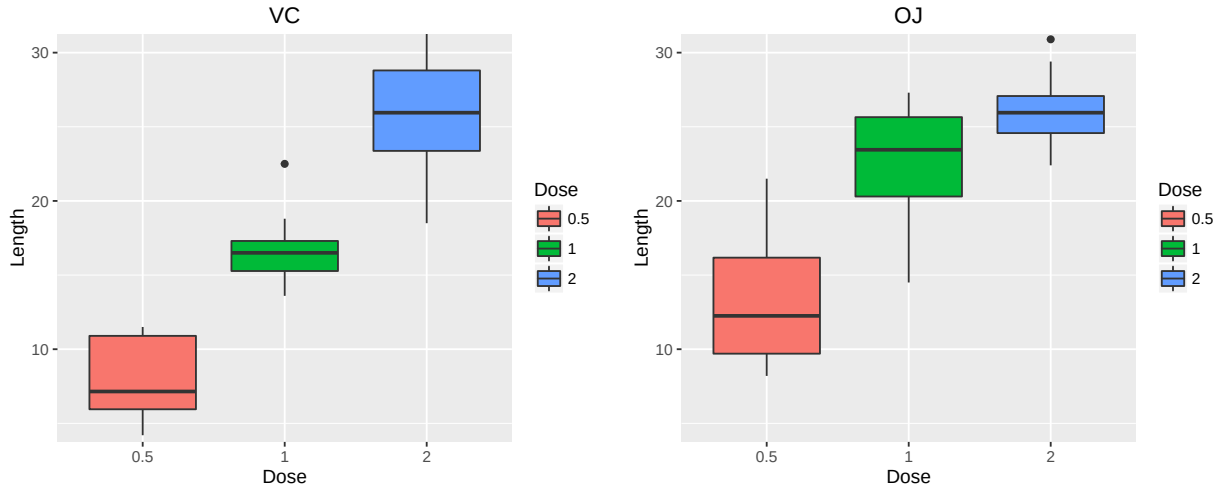
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Subsetting on `OJ` and `CV`:

```
data_OJ <- subset(ToothGrowth,supp=='OJ')
data_VC <- subset(ToothGrowth,supp=='VC')
```

Visualizing the resulting datasets:

```
library(ggplot2)
library(gridExtra)
g1<-ggplot(data_VC,aes(x=as.factor(dose),y=len,fill=as.factor(dose)))
plot1<-g1+geom_boxplot()+labs(fill='Dose',x='Dose',y='Length',title='VC')+
    coord_cartesian(ylim = c(5,30))
g2<-ggplot(data_OJ,aes(x=as.factor(dose),y=len,fill=as.factor(dose)))
plot2<-g2+geom_boxplot()+labs(fill='Dose',x='Dose',y='Length',title='OJ')+
    coord_cartesian(ylim = c(5,30))
grid.arrange(plot1,plot2,ncol=2)
```



Below I run several `t.test` to check different hypothesis derived from the naive visual analysis. I will report p-values. If a p-value is below 0.05, then a hypothesis is correct.

Hypothesis 1: Higher Dose of VC increases the length of odontoblasts

```
d05_VC <- data_VC[data_VC$dose==0.5,]$len
d10_VC <- data_VC[data_VC$dose==1.0,]$len
test<-t.test(d05_VC,d10_VC,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 6.492265e-07
```

```
d20_VC <- data_VC[data_VC$dose==2.0,]$len
test<-t.test(d10_VC,d20_VC,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 3.397578e-05
```

Hypothesis 2: Higher Dose of OJ increases the length of odontoblasts

```
d05_OJ <- data_OJ[data_OJ$dose==0.5,]$len
d10_OJ <- data_OJ[data_OJ$dose==1.0,]$len
test<-t.test(d05_OJ,d10_OJ,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 8.357559e-05
```

```
d20_OJ <- data_OJ[data_OJ$dose==2.0,]$len
test<-t.test(d10_OJ,d20_OJ,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 0.0373628
```

Hypothesis 3: OJ is more efficient than VC in growing odontoblasts

```
test<-t.test(d05_OJ,d05_VC,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 0.005303661
```

```
test<-t.test(d10_OJ,d10_VC,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 0.0007807262
```

```
test<-t.test(d20_OJ,d20_VC,var.equal = TRUE,paired = FALSE)
test$p.value
```

```
## [1] 0.9637098
```

Results:

All three hypothesis seem to be correct except one statement: when the dose is 2 mg/day, the length of odontoblasts seem do not depend on the method of delivery.