Exploring The ToothGrowth Dataset

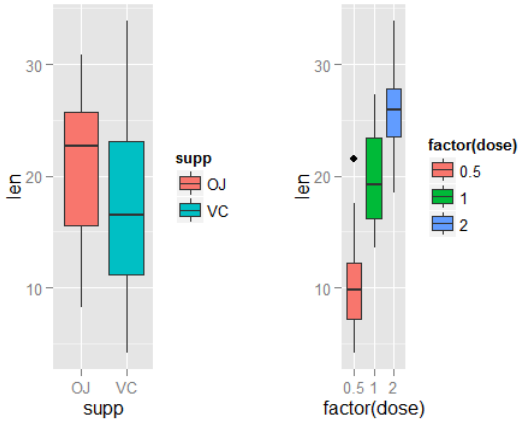
Georges Bodiong

Sunday, February 22, 2015

This document is the second part of the Coursera Statistical Inference Course project. It is an analysis of the ToothGrowth dataset in R. The ToothGrowth dataset explains the relation between the growth of teeth of guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods(orange juice and ascorbic acid).

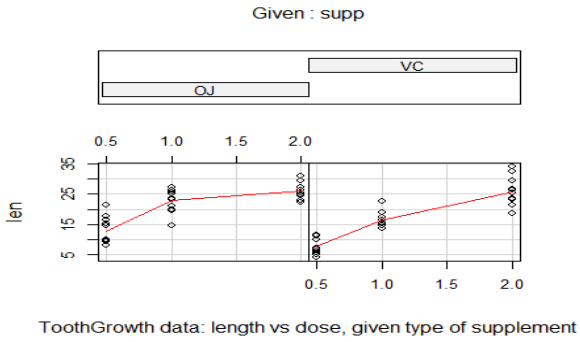
Let's start by loading the data and performing some basic exploratory data analysis

```
library(psych)

## Warning: package 'psych' was built under R version 3.1.2

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.1.2
## Attaching package: 'ggplot2'
## The following object is masked from 'package:psych':
##     %+%

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.1.2

## Loading required package: grid

library(datasets)
data(ToothGrowth)
attach(ToothGrowth)
supp_plot<-ggplot(aes(x = supp, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = supp))
dose_plot<-ggplot(aes(x = factor(dose), y =len), data = ToothGrowth) +
        geom_boxplot(aes(fill = factor(dose)))
grid.arrange(supp_plot, dose_plot, ncol = 2)
```



Summary of data

```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



ToothGrowth data: length vs dose, given type of supplement

```
head(ToothGrowth, 5)

##    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

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

dose<-as.factor(dose)
describe(len)

##   vars  n  mean   sd median trimmed  mad min  max range  skew kurtosis
## 1    1 60 18.81 7.65  19.25   18.95 9.04 4.2 33.9  29.7 -0.14    -1.04
##     se
## 1 0.99

table(supp,dose)

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

round(with(ToothGrowth, sapply(split(len, supp), mean)), 3)
```

```
##     OJ     VC
## 20.663 16.963
```

```
aggregate(len, list(dose), mean)
```

```
##   Group.1      x
## 1     0.5 10.605
## 2       1 19.735
## 3       2 26.100
```

```
aggregate(len, list(supp, dose), mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ       1 22.70
## 4      VC       1 16.77
## 5      OJ       2 26.06
## 6      VC       2 26.14
```

```
aggregate(len, list(supp, dose), sd)
```

```
##   Group.1 Group.2        x
## 1      OJ     0.5 4.459709
## 2      VC     0.5 2.746634
## 3      OJ       1 3.910953
## 4      VC       1 2.515309
## 5      OJ       2 2.655058
## 6      VC       2 4.797731
```

Confidence intervals and hypothesis test to compare tooth growth by supp and dose

```
# T Test by supplemant type
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
```

```
# T test by dose level
Tooth.dose0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Tooth.dose0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Tooth.dose1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))

t.test(len ~ dose, data = Tooth.dose0.5_1.0)
```

```
##  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 = Tooth.dose0.5_2.0)
```

```
##  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 = Tooth.dose1.0_2.0)
```

```
##  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
```

```
# T test for supplement by dose level
Tooth.dose0.5 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose1.0 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose2.0 <- subset(ToothGrowth, dose == 2.0)
```

```
t.test(len ~ supp, data = Tooth.dose0.5)
```

```
##  Welch Two Sample t-test
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data = Tooth.dose1.0)
##  Welch Two Sample t-test
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77

t.test(len ~ supp, data = Tooth.dose2.0)
##  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:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

For dose 0.5, the p-value of OJ in comparison to VC is 0.0064. Since it is less than 0.05 (strong presumption against null hypothesis), it means that there is a difference between both methods.

For dose 1.0, the p-value of OJ in comparison to VC is 0.001. Since it is less than 0.05 (strong presumption against null hypothesis), it means that there is a difference between both methods.

For dose 2.0, the p-value of OJ in comparison to VC is 0.064. Since it is greater than 0.05 (low presumption against null hypothesis), it means that there is a no that much of a difference between both methods.

conclusions and underlying assumptions.

From the values, we can assume that there is an increase in tooth growth depending on the doses. There seems to be no other factor affecting the growth pattern. The delivery methods are independent of the dose size.