Statistical Inference Course Project_Part2.Rmd

Mohit Shukla

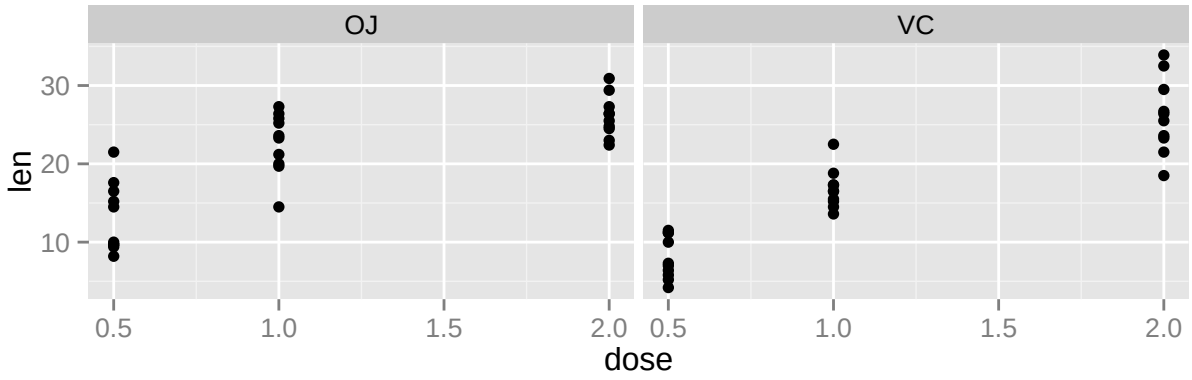
Thursday, May 21, 2015

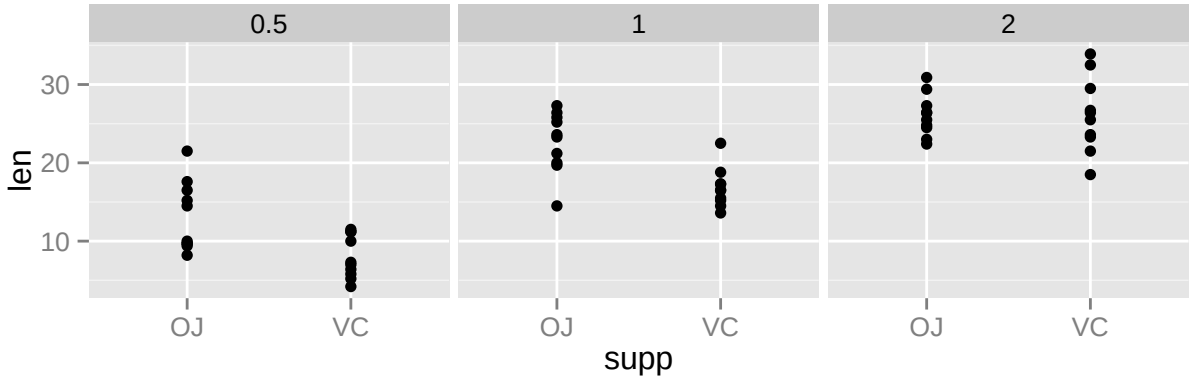
Overview ToothGrowth data set has observations recorded for 10 guinea pigs for 3 different doses of Vitamin and two delivery Methods. In this report hypothesis tests have been performed on tooth growth by dose and supp.

Summary of the data As can be inferred from the plots that OJ is the clear winner for all doses except in the last case where VC seems to dominate for 2 milligrams case.

```
## Loading required package: grid
```

```
##       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
```





```
set1 <- subset(ToothGrowth,supp=="VC")
set2 <- subset(ToothGrowth,supp=="OJ")
g1 <- set1$len  # VC
g2 <- set2$len  #OJ
t.test(g2,g1,paired=TRUE)$conf
```

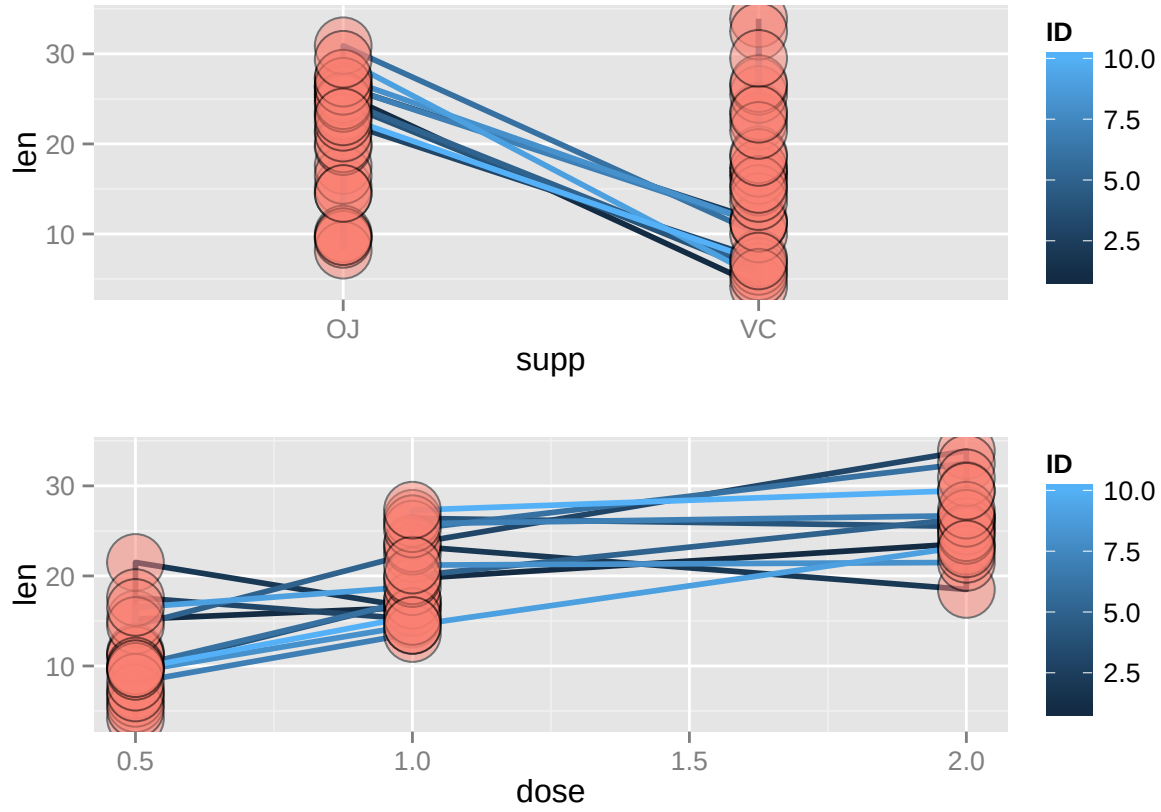
Hypothesis testing based on supp

```
## [1] 1.408659 5.991341
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len~I(relevel(supp,"OJ")),paired=TRUE, data=ToothGrowth)$conf
```

```
## [1] 1.408659 5.991341
## attr(,"conf.level")
## [1] 0.95
```

```
ToothGrowth[,"ID"] <- rep(1:10,6)
plot1 <- ggplot(ToothGrowth,aes(x=supp, y=len, group=factor(ID)))
plot1 <- plot1 + geom_line(size=1,aes(color=ID)) + geom_point(size=10, pch=21, fill="salmon", alpha=.5)
plot2 <- ggplot(ToothGrowth,aes(x=dose, y=len, group=factor(ID)))
plot2 <- plot2 + geom_line(size=1,aes(color=ID)) + geom_point(size=10, pch=21, fill="salmon", alpha=.5)
grid.arrange(plot1,plot2, nrow=2, ncol=1)
```



It turns out that entire confidence interval is above zero. So OJ has more effect on Tooth Growth as compared to VC which was expected from exploratory analysis and plots showing paired variance also tell the same story.

```
set1 <- subset(ToothGrowth, dose==.5)
set2 <- subset(ToothGrowth, dose==1.0)
set3 <- subset(ToothGrowth, dose==2.0)
g1 <- set1$len #.5
g2 <- set2$len #1.0
g3 <- set3$len #2.0
t1 <- t.test(g2,g1,paired=TRUE)$conf
t2 <- t.test(g3,g2,paired=TRUE)$conf
t3 <- t.test(g3,g1,paired=TRUE)$conf
names <- c("upper(g2-g1)","lower(g2-g1)","upper(g3-g2)","lower(g3-g2)","upper(g3-g1)","lower(g3-g1)")
values <- c(t1,t2,t3)
cbind(names,values)
```

Hypothesis testing based on dose

```
##      names           values
## [1,] "upper(g2-g1)" "6.38712117680639"
## [2,] "lower(g2-g1)" "11.8728788231936"
## [3,] "upper(g3-g2)" "3.47181434422758"
## [4,] "lower(g3-g2)" "9.25818565577242"
```

```
## [5,] "upper(g3-g1)" "12.6228018311841"
## [6,] "lower(g3-g1)" "18.3671981688159"
```

Here three tests have been performed for each set of groups and all the resulting confidence intervals are above zero. So more the amount of dose more will be the tooth growth. This was also expected from the exploratory analysis. So 2 milligrams of dose results in maximum Tooth Growth whereas .5 results in minimum.