

Course Project 2

Jeffrey Uslan

Tuesday, July 21, 2015

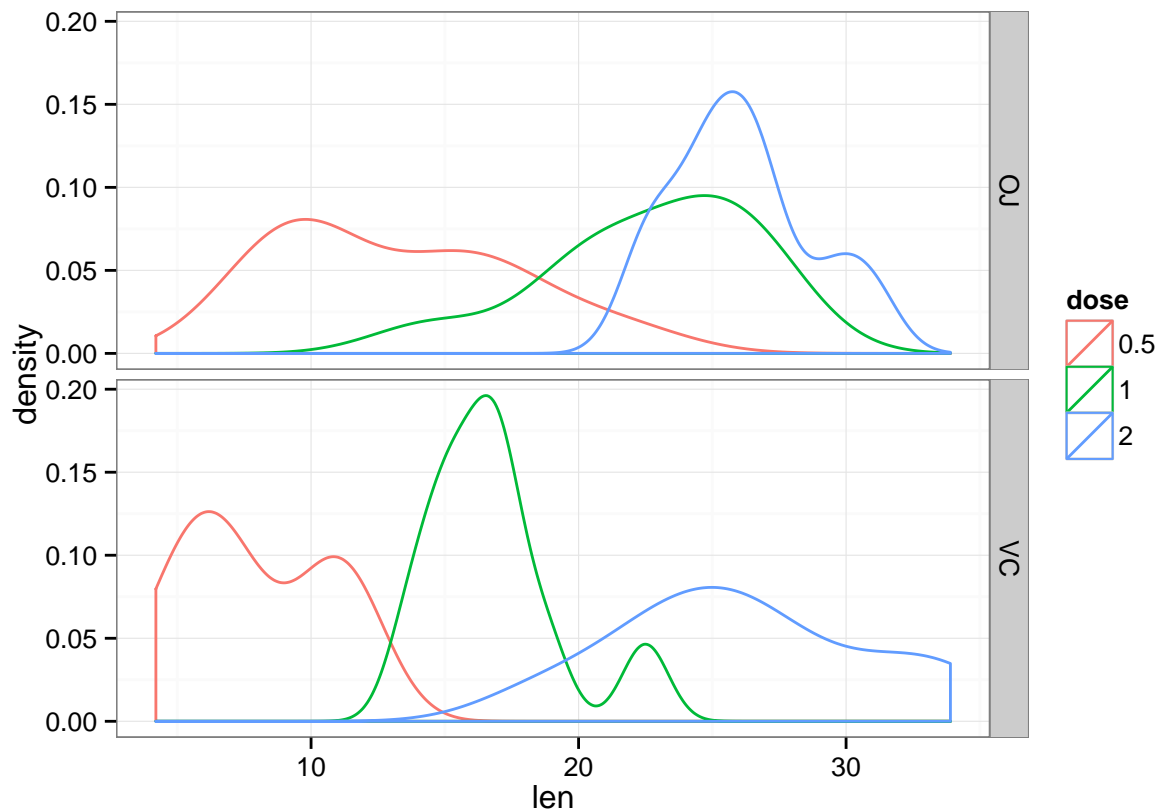
1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
toothGrowth=ToothGrowth
toothGrowth=tbl_df(toothGrowth)
toothGrowth$dose=as.factor(toothGrowth$dose)

str(toothGrowth)
```

```
## Classes 'tbl_df', 'tbl' and '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
ggplot(data=toothGrowth)+theme_bw()+geom_density(aes(x=len,color=dose))+facet_grid(supp~.)
```



2. Provide a basic summary of the data.

```
summary(toothGrowth)
```

```
##      len      supp      dose
## Min.   : 4.20    OJ:30    0.5:20
## 1st Qu.:13.07    VC:30    1 :20
## Median :19.25                2 :20
## Mean   :18.81
## 3rd Qu.:25.27
## Max.   :33.90
```

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

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

```
toothGrowth %>% group_by(supp) %>% summarise(min(len),mean(len),max(len))
```

```
## Source: local data frame [2 x 4]
##
##      supp min(len) mean(len) max(len)
## 1    OJ      8.2  20.66333    30.9
## 2    VC      4.2  16.96333    33.9
```

```
toothGrowth %>% group_by(dose) %>% summarise(min(len),mean(len),max(len))
```

```
## Source: local data frame [3 x 4]
##
##      dose min(len) mean(len) max(len)
## 1  0.5      4.2    10.605    21.5
## 2    1     13.6    19.735    27.3
## 3    2     18.5    26.100    33.9
```

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

Gather relevant indices for comparing dose. Collecting Confidence intervals

```
ind_.5=which(toothGrowth$dose==.5)
conf_.5=t.test(x = toothGrowth$len[ind_.5])$conf.int[1:2]
ind_1=which(toothGrowth$dose==1)
conf_1=t.test(x = toothGrowth$len[ind_1])$conf.int[1:2]
ind_2=which(toothGrowth$dose==2)
conf_2=t.test(x = toothGrowth$len[ind_2])$conf.int[1:2]
```

The 95% confidence intervals for doses of .5, 1, and 2 respectively.

```
conf_.5
```

```
## [1] 8.499046 12.710954
```

```
conf_1
```

```
## [1] 17.66851 21.80149
```

```
conf_2
```

```
## [1] 24.33364 27.86636
```

Gather relevant indices for comparing supp. Collecting Confidence intervals

```
OJ_ind=which(toothGrowth$supp=="OJ")
oj_conf=t.test(x = toothGrowth$len[OJ_ind])$conf.int[1:2]
vc_conf=t.test(x = toothGrowth$len[-OJ_ind])$conf.int[1:2]
```

The 95% confidence intervals for supp of OF and VC respectively.

```
oj_conf
```

```
## [1] 18.19678 23.12989
```

```
vc_conf
```

```
## [1] 13.87675 20.04992
```

A t-test to compare the means of the lengths by supp.

```
OJ_ind=which(toothGrowth$supp=="OJ")
t.test(x = toothGrowth$len[OJ_ind], y = toothGrowth$len[-OJ_ind], alt = "two.sided", paired = FALSE)

##
## Welch Two Sample t-test
##
## data:  toothGrowth$len[OJ_ind] and toothGrowth$len[-OJ_ind]
## 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 of x mean of y
## 20.66333 16.96333
```

The low p-value suggest that we have statistical significant evidence to reject the null hypothesis that the true difference in means is equal to 0 and accept the alternative hypothesis that the true difference is not 0.