Statistical Inference report 2

C. Deramond

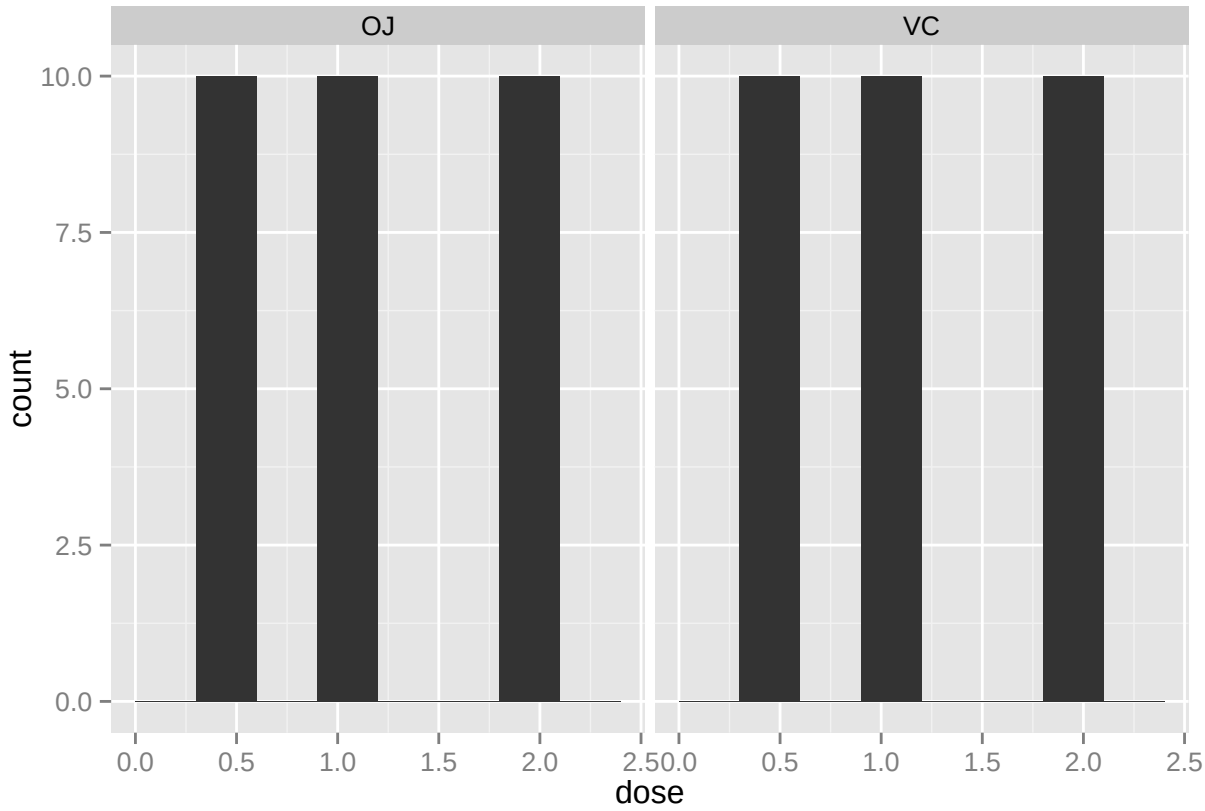
Sunday, December 21, 2014

For this report we'll review results for analysis on the ToothGrowth dataset.

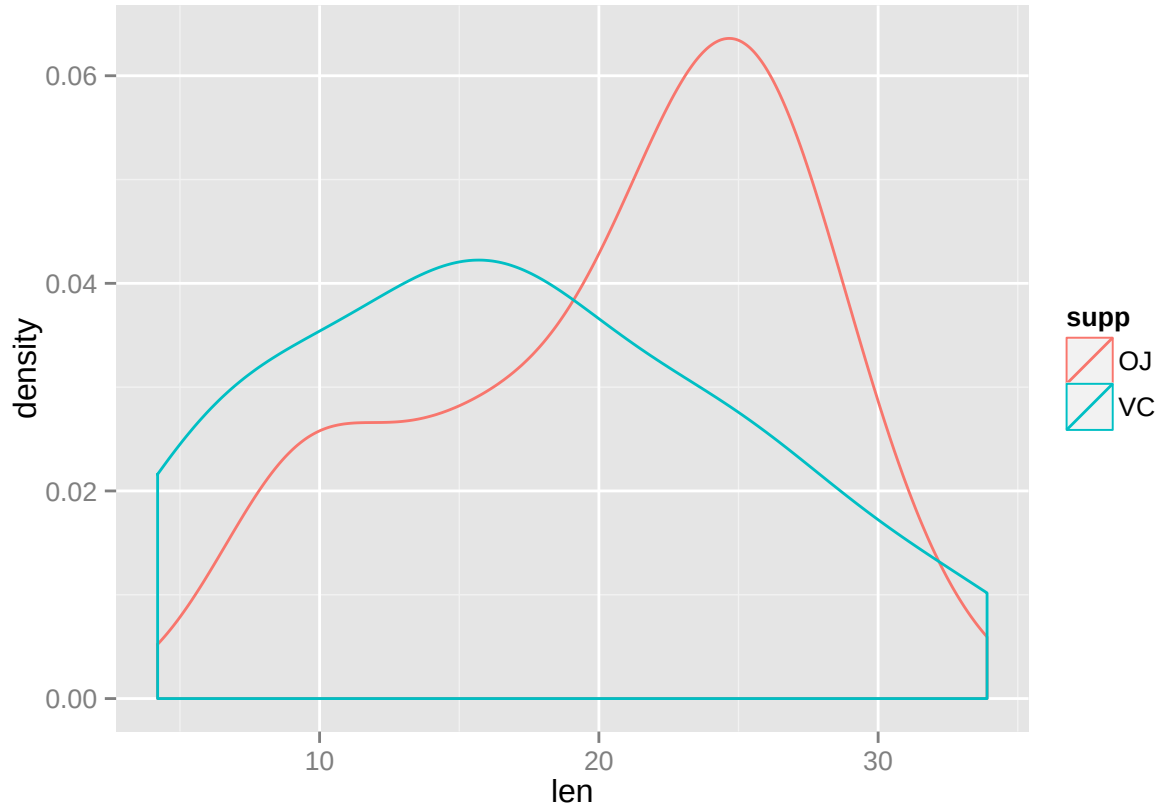
1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
library(reshape2)
library(dplyr, warn.conflicts = F)
library(ggplot2)
data <- ToothGrowth
```

Initially let's see if there's any difference in dosage:



as there is no noticeable difference in dosage, we move on to compare distributions of lenghts:



so it's not a bad hypothesis that there is significant difference between lenght gained in average between supplements

2. Provide a basic summary of the data.

We'll summarise the data for the whole dataset:

```
data %>%
        melt() %>%
        group_by(variable) %>%
        summarise(Mean = mean(value), SD = sd(value), Variance = var(value), n = n() ) %>%
        print
```

```
## Using supp as id variables
```

```
## Source: local data frame [2 x 5]
##
##   variable      Mean        SD   Variance  n
## 1      len 18.813333 7.6493152 58.5120226 60
## 2     dose  1.166667 0.6288722  0.3954802 60
```

Now, divided by the supplement:

```
data %>%
        melt() %>%
        group_by(variable, supp ) %>%
        summarise(Mean = mean(value), SD = sd(value), Variance = var(value), n = n() ) %>%
        print
```

```
## Using supp as id variables

## Source: local data frame [4 x 6]
## Groups: variable
##
##   variable supp      Mean        SD   Variance  n
## 1      len   OJ 20.663333 6.6055610 43.6334368 30
## 2      len   VC 16.963333 8.2660287 68.3272299 30
## 3     dose   OJ  1.166667 0.6342703  0.4022989 30
## 4     dose   VC  1.166667 0.6342703  0.4022989 30
```

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)

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