Statistical Inference report 2

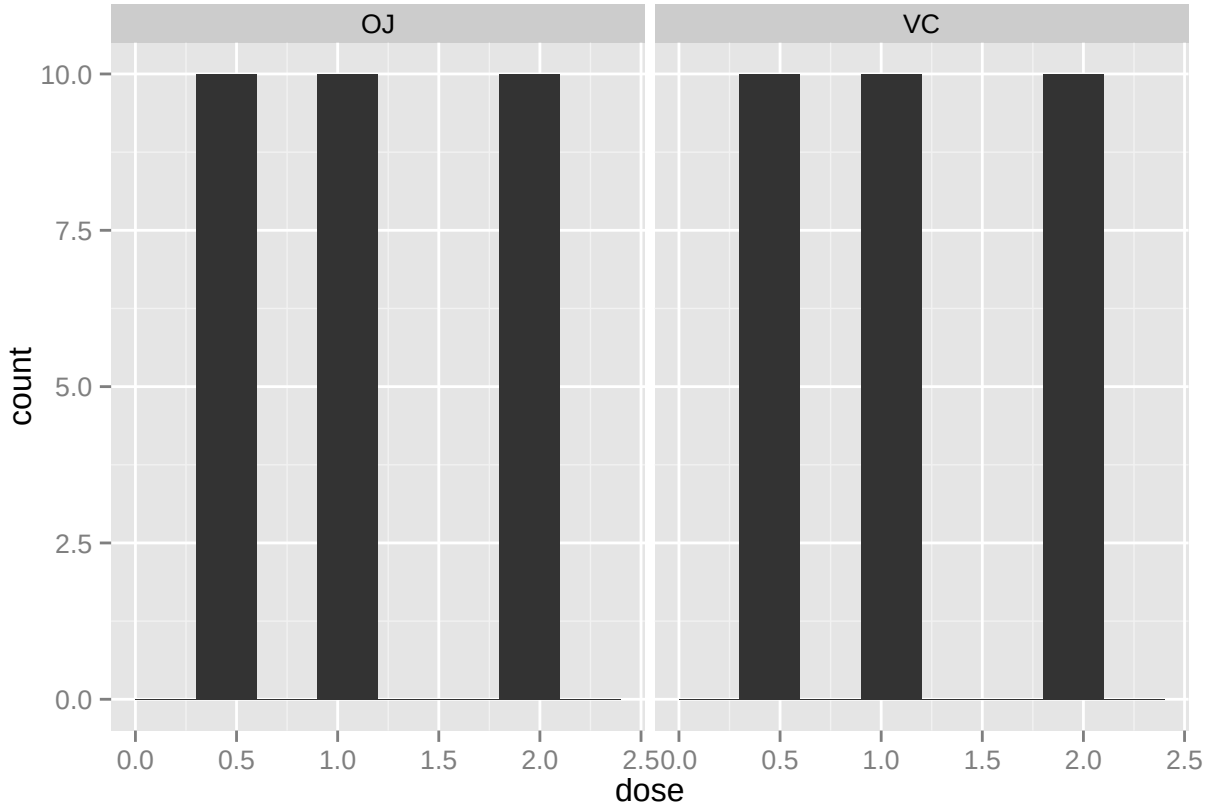
C. Deramond

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

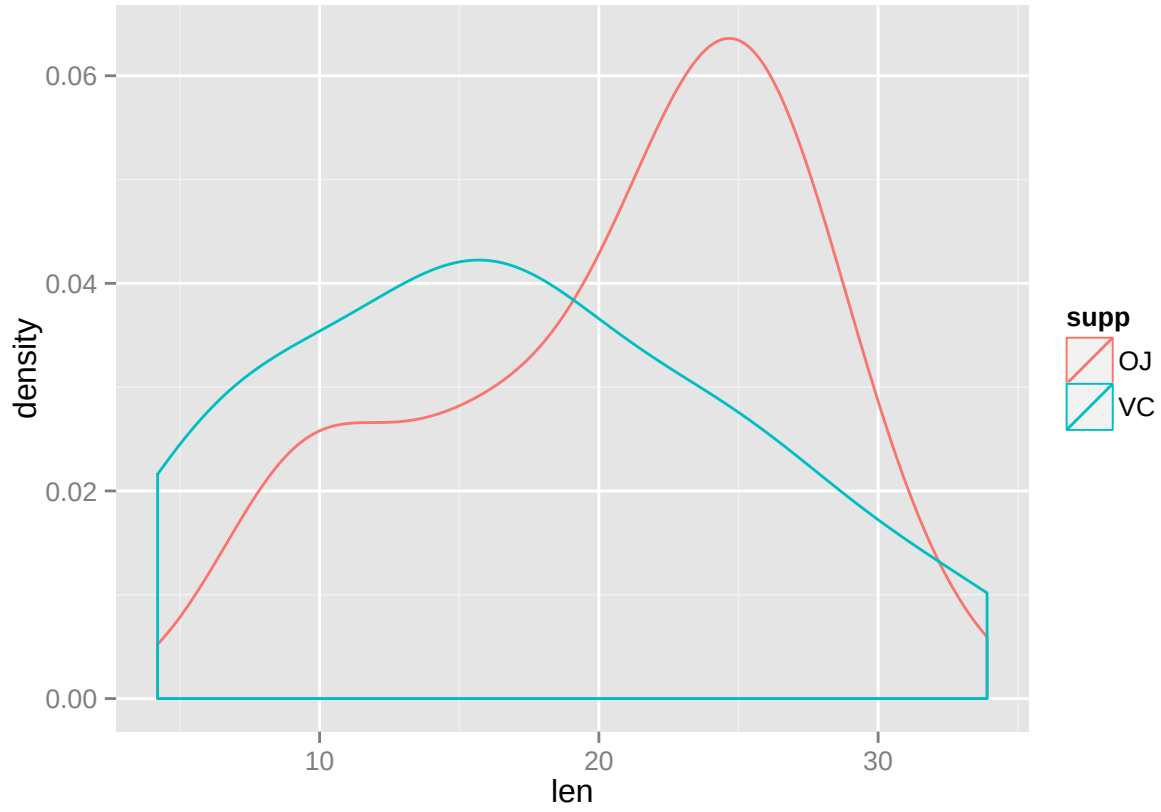
1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
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:



at first glance: it might not be a bad hypothesis that there is significant difference in average lenght gained 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
```

Analogous process, so to divide by the supplement:

```
## 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)

As shown by the summary table, there's not much sense in comparing dosage among supplements, as they are exactly equal. What might be of interest is to review whether is there a difference among the lengths obtained by different supplements:

- note the t-confidence interval for the lenght non discriminating by supplement: 16.8373057, 20.789361

so, it's safe to assume there is no difference among supplements as the mean is contained in the interval.

Digging deeper: we'll try the 2 sample non-paired t-test for H_0: $\mu_1 = \mu_2$ under different variances:

```
# for safety we assume different variances for non-paired t-test
t.test(len ~ supp, paired = F, data = data, var.equal = F)
```

```
##
##  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
```

so we fail to reject null hypothesis.

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

- There's no significant difference in length gained between the 2 supplements.
- Note that the only assumptions we are taking are the standard considerations for a Welch's test.