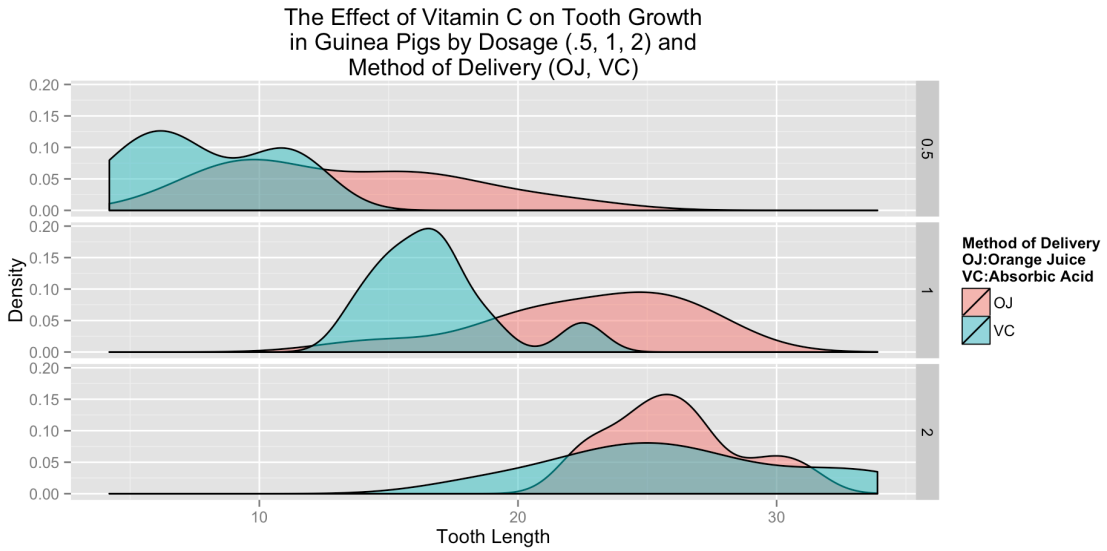
Statistical Inference Programming Assignment Part 2 Author: "Hamel Husain"

This is the report for part 2 (inference) of the Coursera statistical inference class. The full code for this report can be viewed at: **https://github.com/hamelsmu/datasciencecoursera/tree/master/StatInference (https://github.com/hamelsmu/datasciencecoursera/tree/master/StatInference)**

1. Load the ToothGrowth data and provide basic summary of data.

```
## '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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

2. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. I will first explore the data via density plot:



Based on the the density plot, I have several hypothesis that I would like to test. **(1)** For a medium dose (dose 1), Vitaman C adminstered through OJ is associated with longer tooth length than VC. **(2)** Controlling for delivery method, higher doses lead to higher tooth length.

Test Hypothesis 1

```
OJ = ToothGrowth[ToothGrowth$dose ==1 & ToothGrowth$supp == 'OJ', ]$len
VC = ToothGrowth[ToothGrowth$dose ==1 & ToothGrowth$supp == 'VC', ]$len
t.test(OJ, VC)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean of x mean of y
```

```
##     22.70     16.77
```

Conclusion: Based on these results, I reject the null hypothesis that the mean difference between OJ and VC where dose = 1 is zero. I feel confident in saying that OJ is associated with longer teeth length than VC for medium dosage. I am assuming that the variances of these two groups are equal.

Test Hypothesis 2

```
low_dose_OJ = ToothGrowth[ToothGrowth$dose ==.5 & ToothGrowth$supp == 'OJ', ]$len
high_dose_OJ = ToothGrowth[ToothGrowth$dose ==2 & ToothGrowth$supp == 'OJ', ]$len
low_dose_VC = ToothGrowth[ToothGrowth$dose ==.5 & ToothGrowth$supp == 'VC', ]$len
high_dose_VC = ToothGrowth[ToothGrowth$dose ==2 & ToothGrowth$supp == 'VC', ]$len
t.test(high_dose_OJ, low_dose_OJ)
```

```
##
##  Welch Two Sample t-test
##
## data:  high_dose_OJ and low_dose_OJ
## t = 7.817, df = 14.67, p-value = 1.324e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   9.325 16.335
## sample estimates:
## mean of x mean of y
##     26.06     13.23
```

```
t.test(high_dose_VC, low_dose_VC)
```

```
##
##  Welch Two Sample t-test
##
## data:  high_dose_VC and low_dose_VC
## t = 10.39, df = 14.33, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  14.42 21.90
## sample estimates:
## mean of x mean of y
##     26.14      7.98
```

Conclusion: I reject the null hypothesis that high dosages have the same mean tooth length as low dosages, when controlling for delivery method. Furthermore, I noticed that the effect of dosage is more prounounced when the delivery method is VC. I again assumed the variance of the two groups were equal when performing the t-test.