output: html_document --- Inference on ToothGrowth ## Author: Edward Lau

## Overview: I am going to perform statistical inference on the ToothGrowth dataset in the R datasets package, to ask the question of whether the dose and delivery methods of vitamin C may have a significant effect on the growth of odontoblasts (teeth) in a population of 10 guinea pigs.

## Part 2: Inference

This will load the ToothGrowth dataset, which examines the effect of Vitamin C on tooth growth in guinea pigs. len = length of teeth for 10 guinea pigs; supp = OC or VC delivery methods (orange juice or ascorbic acid); dose = dose level

## Question 1: Load the ToothGrowth data and perform basic exploratory data analyses

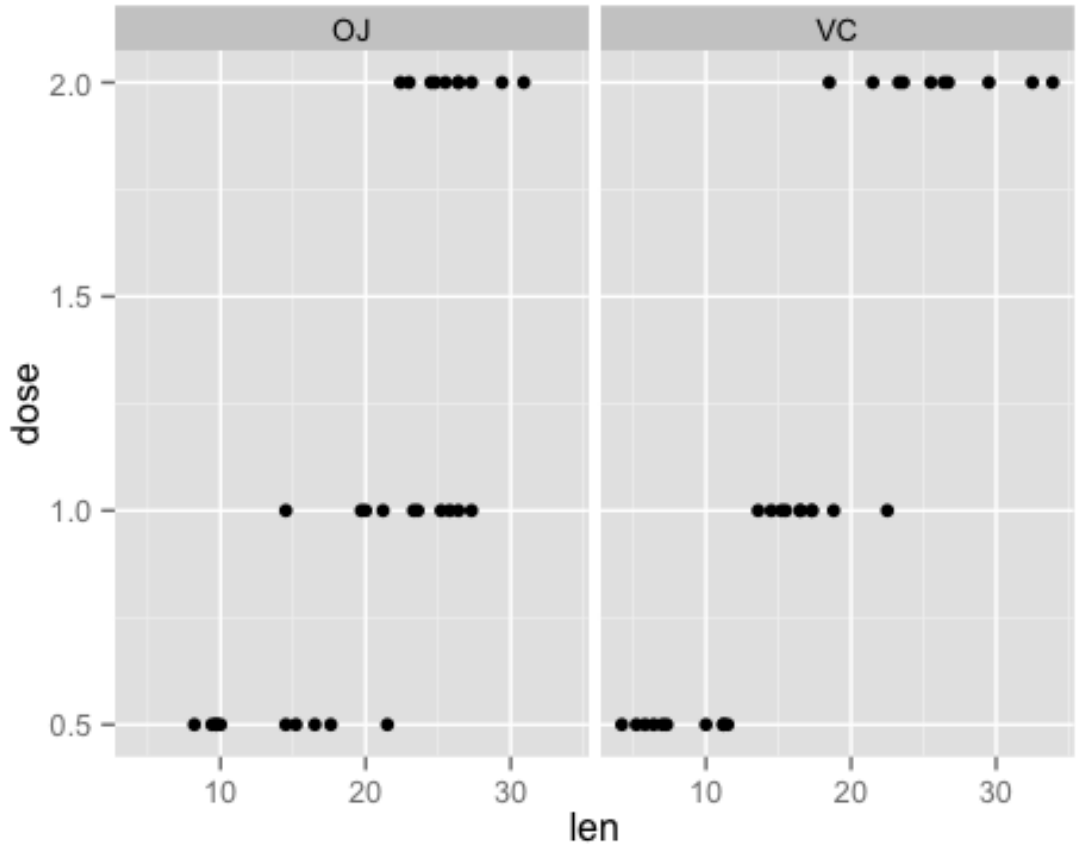
```
library(datasets)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
glimpse(ToothGrowth)
```

```
## Observations: 60
## Variables:
## $ len  (dbl) 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16....
## $ supp (fctr) VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, ...
## $ dose (dbl) 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1....
```

```
qplot(len, dose, data = ToothGrowth, facets = .~supp)
```




Conclusion: it looks like there are increases in tooth length by doses of OJ and VC.

## Question 2: Summarizing the data by dose and supp.

```
group_by(ToothGrowth, supp, dose) %>% summarize(mean = mean(len))
```

```
## Source: local data frame [6 x 3]
## Groups: supp
##
##   supp dose  mean
## 1   OJ  0.5 13.23
## 2   OJ  1.0 22.70
## 3   OJ  2.0 26.06
## 4   VC  0.5  7.98
## 5   VC  1.0 16.77
## 6   VC  2.0 26.14
```

The above table summarizes the mean of tooth length for each dose/ supplement combination.

## Question 3: Compare dose and supplement delivery method using a t test

```
OJdata_0.5 <-filter(ToothGrowth, supp=="OJ", dose == 0.5)
VCdata_0.5 <-filter(ToothGrowth, supp=="VC", dose == 0.5)
t.test(OJdata_0.5$len,VCdata_0.5$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJdata_0.5$len and VCdata_0.5$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```
OJdata_1.0 <-filter(ToothGrowth, supp=="OJ", dose == 1.0)
VCdata_1.0 <-filter(ToothGrowth, supp=="VC", dose == 1.0)
t.test(OJdata_1.0$len,VCdata_1.0$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJdata_1.0$len and VCdata_1.0$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
OJdata_2.0 <-filter(ToothGrowth, supp=="OJ", dose == 2.0)
VCdata_2.0 <-filter(ToothGrowth, supp=="VC", dose == 2.0)
t.test(OJdata_2.0$len,VCdata_2.0$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJdata_2.0$len and VCdata_2.0$len
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

It appears that there is a significant difference in tooth length when given low doses (0.5 and 1.0) of vitamin C through orange juice and through ascorbic acids, but this significance is no longer there at high doses (2.0)

## Question 4: Conclusion

Vitamin C probably promotes tooth growth in a dose-dependent manner between 0.5 and 2.0 milligram. At lower dosages, orange juice is a better delivery method than ascorbic acid. An assumption is that data from the two populations of guinea pig are independent and have identical distribution.