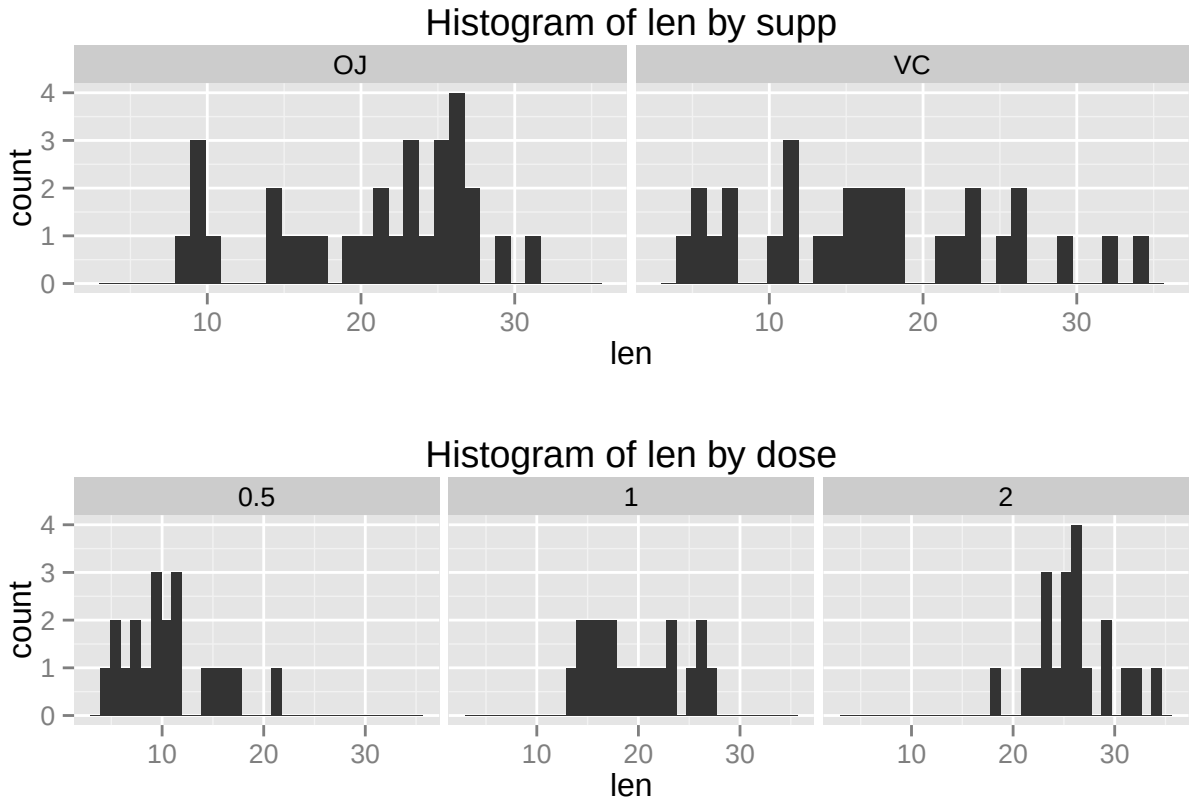
Inferential Data Analysis

Zheng Wu

June 7, 2015

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

```
library(datasets)
library(ggplot2)
library(gridExtra)
data <- ToothGrowth
plot1 <- ggplot(data, aes(x = len)) + geom_histogram() + facet_grid(.~supp) + labs(title = "Histogram o
plot2 <- ggplot(data, aes(x = len)) + geom_histogram() + facet_grid(.~dose) + labs(title = "Histogram o
grid.arrange(plot1, plot2)
```



Provide a basic summary of the data.

```
summary(data)
```

```
##       len        supp         dose
```

```
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

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

```
lOJ <- data$len[data$supp == "OJ"]
lVC <- data$len[data$supp == "VC"]
t.test(lOJ,lVC, alternative = "greater")
```

By supp OJ vs VC

```
##
##  Welch Two Sample t-test
##
## data:  lOJ and lVC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Since the p value of the one sided t-test is smaller that the 0.05 alpha threshould, I would reject the null hypothesis that the true difference in means in the two groups is 0 in favor of the alternative hypothesis that the true difference in means is greater than 0.

```
l0.5 <- data$len[data$dose == 0.5]
l1 <- data$len[data$dose == 1]
t.test(l0.5,l1, alternative = "less")
```

By dose 0.5 vs 1

```
##
##  Welch Two Sample t-test
##
## data:  l0.5 and l1
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Since the p value of the one sided t-test is smaller that the 0.05 alpha threshould, I would reject the null hypothesis that the true difference in means in the two groups is 0 in favor of the alternative hypothesis that the true difference in means is less than 0.

```
l2 <- data$len[data$dose == 2]
t.test(l1,l2, alternative = "less")
```

By dose 1 vs 2

```
##
##  Welch Two Sample t-test
##
## data:  l1 and l2
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Since the p value of the one sided t-test is smaller that the 0.05 alpha threshould, I would reject the null hypothesis that the true difference in means in the two groups is 0 in favor of the alternative hypothesis that the true difference in means is less than 0.

```
t.test(l0.5,l2, alternative = "less")
```

By dose 0.5 vs 2

```
##
##  Welch Two Sample t-test
##
## data:  l0.5 and l2
## t = -11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -13.27926
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

Since the p value of the one sided t-test is smaller that the 0.05 alpha threshould, I would reject the null hypothesis that the true difference in means in the two groups is 0 in favor of the alternative hypothesis that the true difference in means is less than 0.

Assumptions needed for all the above conclusions:

1.Each of the two populations being compared should follow a normal distribution. 2.The data used to carry out the test should be sampled independently and randomly from the two populations being compared.