Statistical Inference Course Project - Part 2

Arash Farazdaghi

August 2014

Global Directives

Environment setup including the libraries and knitr parameters required.

```
knitr::opts_chunk$set(echo=TRUE); library(reshape2)
```

Data Analysis

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

```
# Load data
data(ToothGrowth); data <- ToothGrowth
# pre-processing for analysis
data.melt <- melt(data, names(data)[2:3], names(data)[1])
# examine length by supplement type
data.meanbysupp <- dcast(data.melt, supp ~ variable, mean)
data.varbysupp <- dcast(data.melt, supp ~ variable, var)
# examine length by dosage
data.meanbydose <- dcast(data.melt, dose ~ variable, mean)
data.varbydose <- dcast(data.melt, dose ~ variable, var)
# examine length by supplement type and dosage
data.meanbyall <- dcast(data.melt, supp + dose ~ variable, mean)
data.varbyall <- dcast(data.melt, supp + dose ~ variable, var)
```

Question 2 Provide a basic summary of the data.

There are a total of 60 unique records in the data set with 3 variables. By examining the mean and variance of the tooth lengths, we can see that orange juice (OJ) is a more effective supplement at low dosages than Vitamin C, while both OJ and VC have nearly the same effect with the 2mg dosage. Please see the table below for the details.

```
names(data.meanbyall) <- c("Supplement","Dosage","ToothLength"); data.meanbyall
```

```
##   Supplement Dosage ToothLength
## 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
```

Question 3 Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Gosset's t-Test is used to examine the tooth length by supplement type and dosage.

```r
# t-Test of VC v. OJ supplement for 0.5mg dosage
data.ttest.low <- t.test(data$len[1:10],data$len[31:40],paired=TRUE); data.ttest.low
```

```
##
##  Paired t-test
##
## data:  data$len[1:10] and data$len[31:40]
## t = -2.979, df = 9, p-value = 0.01547
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.237 -1.263
## sample estimates:
## mean of the differences
##                   -5.25
```

```r
# t-Test of VC v. OJ supplement for 1.0mg dosage
data.ttest.mid<- t.test(data$len[11:20],data$len[41:50],paired=TRUE); data.ttest.mid
```

```
##
##  Paired t-test
##
## data:  data$len[11:20] and data$len[41:50]
## t = -3.372, df = 9, p-value = 0.008229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.908 -1.952
## sample estimates:
## mean of the differences
##                   -5.93
```

```r
# t-Test of VC v. OJ supplement for 2.0mg dosage
data.ttest.high <- t.test(data$len[21:30],data$len[51:60],paired=TRUE); data.ttest.high
```

```
##
##  Paired t-test
##
## data:  data$len[21:30] and data$len[51:60]
## t = 0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.169  4.329
## sample estimates:
## mean of the differences
##                    0.08
```

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

As can be seen, the p-values for dosages 0.5mg and 1mg are below 0.5 (0.0155 and 0.0082 respectively), which indicate that there is a significant difference between the two supplements at this dosage. However, the p-value for the 2mg dosage is 0.967, which is above 0.5, with a confidence interval of -4.169, 4.329, and mean of differences 0.08, indicating that both supplements have nearly the same effect. This is based on the assumption that the data for the 10 subjects is paired and there are no significant external factors.