Effects of Vitamin C in Guinea Pig Tooth Growth

Peter Pih
Statistical Inference, Corsera(statinference-015)
June 17th, 2015

This assignment uses the **ToothGrowth** data set which is built into R It is based on a study to see the effects of vitamin C intake for guinea pig tooth growth Guinea Pigs, like humans, are dependent on consuming vitamin C

Additional information on the data set is available here: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html

The ToothGrowth Data Set

What does the Tooth Growth data set look like?

```r
class(ToothGrowth)
```

```
## [1] "data.frame"
```

```r
str(ToothGrowth)
```

```
## '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 ...
```

```r
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

The data columns in the data frame are:

- len - tooth length
- supp - kind of supplement used: orange juice or asorbic acid
- dose - dosage in milligrams (0.5, 1.0, 2.0)

Each supplement type and dosage has 10 observations, giving a total of 60 observations/rows in the data set.

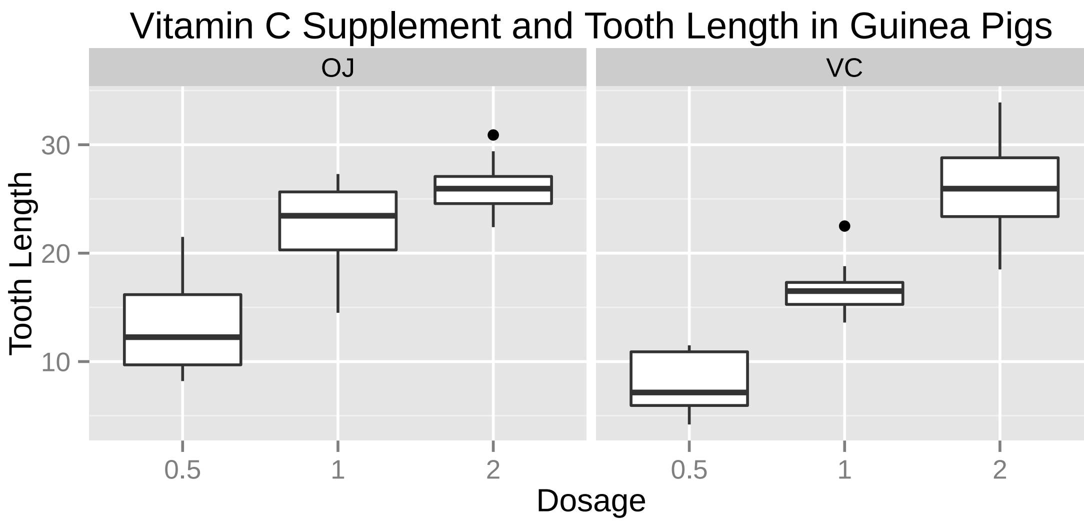
Of Interest

The two things to investigate with this data set:

1) Effect of dosage on tooth growth
2) Effect of supplement type on tooth growth

An initial plot of the data shows that there is a strong trend in dosage level regardless of type of supplement:

```r
require(ggplot2, quietly=TRUE)
ggplot(ToothGrowth, aes(factor(dose), len)) + geom_boxplot() + facet_grid(.~supp) +
  xlab("Dosage") + ylab("Tooth Length") + ggtitle("Vitamin C Supplement and Tooth Length in Guinea Pigs
```



As dosage increases, tooth length increases, regardless of the type of supplement given. However there seems to be differences in rates of effectiveness at different dosage levels depending on the type of supplement given.

To further understand the interaction between supplement and dosage we will calculate p-values based on subsets of supplement type and dosage level.

Summary Data By Supplement / Dosage

Calculate some summary statistics for the data set and put into a data.frame **summ_data** and also create a matrix which will be used to compute the p-values. The new matrix, ***data***, will use `paste(supp, dose)` as column headers for subsetting the tooth length data.

```r
supp_types <- unique(ToothGrowth$supp)             # supplementary types
dosage_type <- unique(ToothGrowth$dose)            # dosages

summ_data <- data.frame()                          # summary data frame
data <- numeric()                                  # new matrix to be used to compute p-values
for (i in supp_types){                             # for each supplementary type
    supp_subset <- subset(ToothGrowth, ToothGrowth$supp == i)

    for (k in dosage_type){                        # for each dosage
        dosage_subset <- subset(supp_subset, supp_subset$dose == k)

        t_mean <- mean(dosage_subset$len)
        t_std <- sd(dosage_subset$len)
        t_count <- nrow(dosage_subset)

        summ_data <- rbind(summ_data, data.frame(supp=i, dose=k, mean=t_mean, std=t_std, count=t_count)

        data <- cbind(data, dosage_subset$len)     # save the subset data as a new column for use later
```

```
        colnames(data)[ncol(data)] <- paste(i, k) # create the category names to be used as column names
    }
}
```

The summary data (shown in the Appendix)confirms the upward slopes of means in the graphs as dosage increases and the differing variances in each of the subset categories.

Compute The Pair-wise p-value Table

Initiailize the statistic matrix, **p__vals**, where the pairwise p-values will be stored. We will use the column factors (supplement/dosage pair) as column and row names in the output p-value matrix.

```
p_vals<-matrix(nrow=ncol(data), ncol=ncol(data))
colnames(p_vals) <- colnames(data)      # make the p-value matrix symmetric
rownames(p_vals) <- colnames(data)      # by using the same column and row names

colnames(data)    # 6 categories total
```

```
## [1] "VC 0.5" "VC 1"   "VC 2"   "OJ 0.5" "OJ 1"   "OJ 2"
```

Now we calculate the pairwise p-values across all supplementary types and dosages. Since our level of significance is alpha=0.05, we look for p-values less than 0.05 to show the means are different.

```
categories <- colnames(data)            # categories are supplement/dosage pairs
for (i in categories){
    for (k in categories){
        p_vals[i,k] <- t.test(data[, i], data[,k], conf.level=0.95, paired=FALSE)$p.value
    }
}
```

Results

The resulting p-value matrix is symmetric around the upper-left to lower-right diagonal Where each category is run against itself and the p-value is 1.0 as expected.

The p-values for changing dosages show that increasing dosages genrally have significant effects. For example, reading down the first column, the dosage is **VC 0.5** (asorbic acid, 0.5mg), there are no values greater than 0.05 (expect for the very first against itself) showing that there were significant changes in the observed mean for other treatments.

It is interesting that **OJ 2** and **VC 2** (p-value=0.9639), the higher dosages of both supplements are not significantly different and **OJ 1** and **VC 2** (p-value=0.0965), show less difference than other pairings where dosage levles are similar, indicating orange juice effectiveness to be higher at lower dosages.

Possible explainations for the apparent effectiveness of orange juice over asorbic acid might be other trace elements and minerals which would present in orange juice but not necessarily in pharmalogical grade asorbic acid. The contributions of trace elements and minerals in nutrition have been widely known and reported on thoght their interactions are not completely understood.

Appendix

Summary Data Table

```r
summ_data
```

```
##   supp dose  mean      std count
## 1   VC  0.5  7.98 2.746634    10
## 2   VC  1.0 16.77 2.515309    10
## 3   VC  2.0 26.14 4.797731    10
## 4   OJ  0.5 13.23 4.459709    10
## 5   OJ  1.0 22.70 3.910953    10
## 6   OJ  2.0 26.06 2.655058    10
```

Pairwise p-value Table

```r
round(p_vals, 4)
```

```
##        VC 0.5   VC 1   VC 2 OJ 0.5   OJ 1   OJ 2
## VC 0.5 1.0000 0.0000 0.0000 0.0064 0.0000 0.0000
## VC 1   0.0000 1.0000 0.0001 0.0460 0.0010 0.0000
## VC 2   0.0000 0.0001 1.0000 0.0000 0.0965 0.9639
## OJ 0.5 0.0064 0.0460 0.0000 1.0000 0.0001 0.0000
## OJ 1   0.0000 0.0010 0.0965 0.0001 1.0000 0.0392
## OJ 2   0.0000 0.0000 0.9639 0.0000 0.0392 1.0000
```