Final Assignment for the Statistical Inference module: Part 2

In the second part of the assignment, we were tasked to work with the ToothGrowth dataset in the R "datasets" package. It contains the data on the effects of Vitamin C supplements on the growth of odontoblast cells in the teeth of Guinea Pigs.

Lets load it up and take a look at it right now.

```
toof<-ToothGrowth

str(toof)
```

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

The dataset has 3 variables, which are (from left to right) the length of odontoblasts (microns), type of supplement, which is either ascorbic acid or orange juice and the dose of the supplement which was added to the Guinea Pig's diet. There were a total of 60 Guinea Pigs participating in the study and the table has 60 rows of observations, one for each Pig.

The study has a total of 6 groups of 10 Guinea pigs. Each group was given either orange juice or ascorbic acid as a supplement, in the amount of 0.5, 1 or 2 mg/day.

Currently the data has a lot of abbreviations which makes it harder to work with. Lets fix that.

```
names(toof)<- c("cell_length", "supplement", "dose")

toof$supplement<- gsub(pattern = "OJ", replacement = "Orange juice", x = toof$supplement)

toof$supplement<- gsub(pattern = "VC", replacement = "Ascorbic acid", x = toof$supplement)

str(toof)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ cell_length: num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supplement : chr  "Ascorbic acid" "Ascorbic acid" "Ascorbic acid" "Ascorbic acid" ...
##  $ dose       : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Next, lets add a fourth variable, which denotes which group did the Guinea Pig belong to in this study.

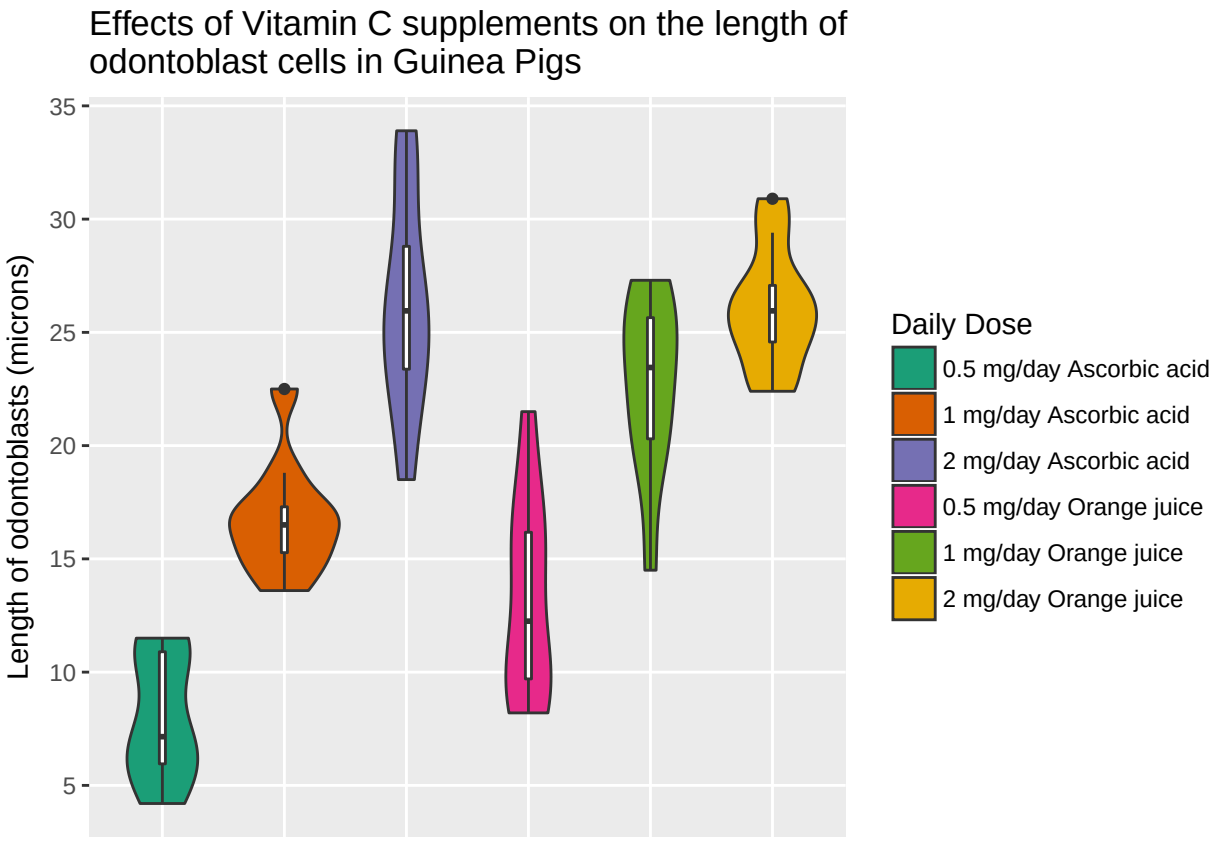
```
list<-rep("mg/day", 60)

toof$group<- paste(toof$dose, list, sep = " ")

toof$group<- paste(toof$group, toof$supplement, sep = " ")

toof$group<-factor(toof$group, ordered = TRUE, levels = unique(toof$group))

str(toof)
```

```
## 'data.frame':    60 obs. of  4 variables:
##  $ cell_length: num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supplement : chr  "Ascorbic acid" "Ascorbic acid" "Ascorbic acid" "Ascorbic acid" ...
```

```
##  $ dose       : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ group      : Ord.factor w/ 6 levels "0.5 mg/day Ascorbic acid"<..: 1 1 1 1 1 1 1 1 1 ...
```

Finally, lets visualize this data with a violin plot.

```
library(ggplot2)

ggplot(toof, aes(toof$group, toof$cell_length)) + geom_violin(aes(fill=toof$group)) +

    scale_fill_brewer(name= "Daily Dose", palette = "Dark2") +

    geom_boxplot(width=0.05) +

    theme(axis.title.x=element_blank(), axis.text.x=element_blank(),

        axis.ticks.x=element_blank()) +

    scale_y_continuous(name = "Length of odontoblasts (microns)",

        breaks = seq(0, 35, 5), minor_breaks = FALSE) +

    ggtitle("Effects of Vitamin C supplements on the length of
odontoblast cells in Guinea Pigs")
```



From the violin plot, we can see that very clearly that higher doses of vitamin C lead to longer odontoblast cells. Thus, it would be more interesting to determine if the choice of supplement had an effect on cell length, for each of the 3 doses.

Our main hypothesis (Ho), will maintain that on average, both supplements had the same effect on cell length for each dosage, i.e. mean(0.5 mg/day Ascorbic acid)-mean(0.5 mg/day Orange Juice)=0.

Conversely, Ha will claim that mean(Ascorbic acid)-mean(Orange juice)!=0.

Since our sample size for each of the 6 groups is 10, which is quite small, we will use a joint t distribution with 18 degrees of freedom to conduct a 95% confidence test. For this test, we will assume that both groups have equal variance.

Firstly, lets separate our data by food supplement and dosage. Groups 1,3,5 will be Ascorbic acid with concentration of 0.5, 1 and 2. Groups 2,4,6 will be Orange juice.

```
list1<-toof[toof$group == "0.5 mg/day Ascorbic acid", 1]

list3<-toof[toof$group == "1 mg/day Ascorbic acid", 1]

list5<-toof[toof$group == "2 mg/day Ascorbic acid", 1]

list2<-toof[toof$group == "0.5 mg/day Orange juice", 1]

list4<-toof[toof$group == "1 mg/day Orange juice", 1]

list6<-toof[toof$group == "2 mg/day Orange juice", 1]
```

Then, we will run a paired t test for groups 1 and 2, 3 and 4, 5 and 6.

```
t1<-t.test(list1, list2, paired = FALSE, var.equal = TRUE)

t2<-t.test(list3, list4, paired = FALSE, var.equal = TRUE)

t3<-t.test(list5, list6, paired = FALSE, var.equal = TRUE)

c(t1$p.value, t2$p.value, t3$p.value)
```

```
## [1] 0.0053036613 0.0007807262 0.9637097790
```

The p values for the first 2 tests are less than 1%, so we can safely reject the null hypothesis. On the other hand, p statistic for the last test is over 96% percent, so in this case we fail to reject the null. To see which supplement yielded higher growth rates, lets look at the confidence intervals.

```
rbind(t1$conf.int, t2$conf.int, t3$conf.int)
```

```
##           [,1]      [,2]
## [1,] -8.729738 -1.770262
## [2,] -9.019308 -2.840692
## [3,] -3.562999  3.722999
```

Thus, for the 0.5 and 1 mg/day doses, Orange juice had a greater effect on the length of odontoblasts, but for 2 mg/day, there was no clear winner. One potential implication of this fact is that Orange juice had concentration of Vitamin C or was more easily digested by the Guinea Pigs, which lead to a higher cell growths. Furthermore, it is possible that 2mg/day of either supplements provides more Vitamin C to the animals than necessary and their bodies filtered out the excess, which would explain equal tooth growth rates in each case.