Statistical Inference - Course Project

Part 2

This is the project part 2 of the statistical inference class. In this project, I will use the ToothGrowth dataset available in the R Dataset Package to compare tooth growth by supp and dose.

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

The ToothGrowth dataset provides data about the effect of Vitamin C on Tooth Growth in Guinea Pigs. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
library(datasets)
data(ToothGrowth)
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 ...
```

The data frame has 60 observations of 3 variables

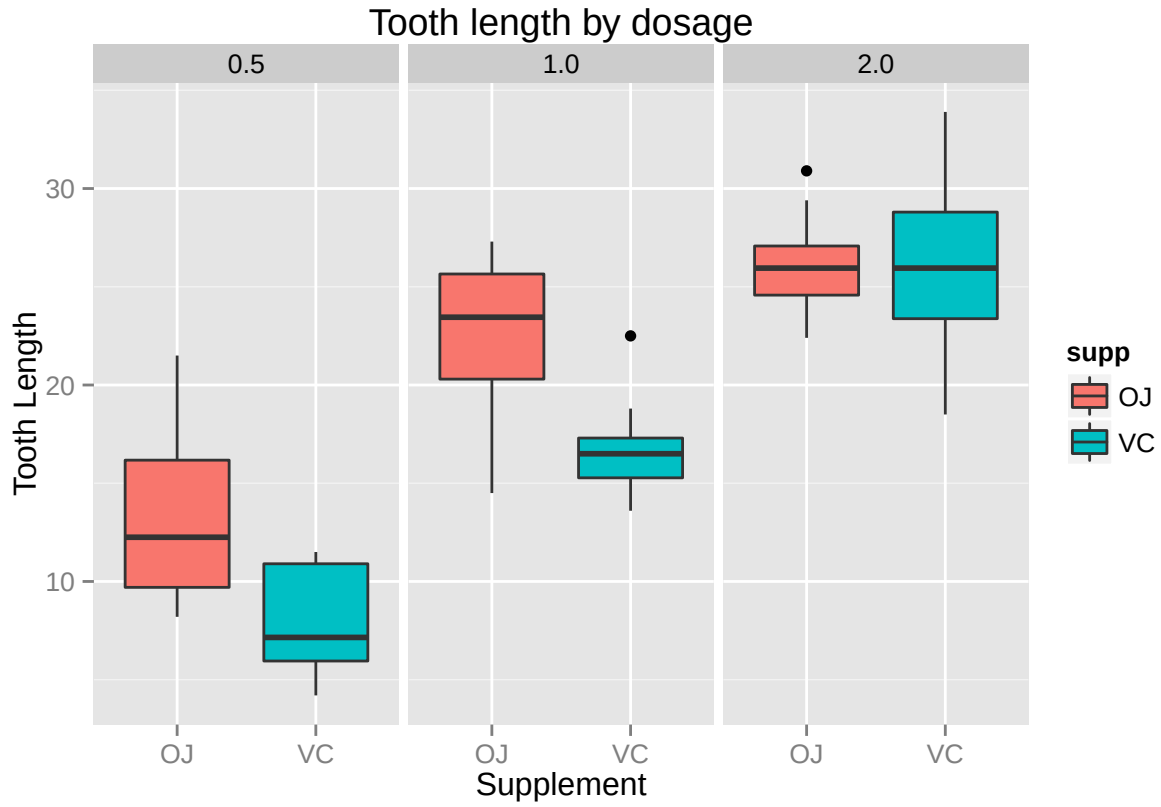
- len - Tooth length
- supp - The Supplement type (Vitamin C (VC) or Orange Juice (OJ))
- dose - The dose in milligrams

```
head(ToothGrowth,5)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
```

2. Provide a basic summary of the data.

Each combinaison of supplements and doses has 10 measures (10 * 3 doses * 2 types = 60 observations). Here is an overview of the distribution:



As we can see, the length has more extrem values when the supplement comes from Vitamin C. However, if we look at the mean, the Orange Juice supplement seems to provide a better tooth growth with smaller doses.

```
aggregate(ToothGrowth$len, by=list('Dose'=ToothGrowth$dose,'Supp'=ToothGrowth$supp), FUN = mean)
```

```
##   Dose Supp     x
## 1  0.5   OJ 13.23
## 2  1.0   OJ 22.70
## 3  2.0   OJ 26.06
## 4  0.5   VC  7.98
## 5  1.0   VC 16.77
## 6  2.0   VC 26.14
```

3. Compare tooth growth by supp and dose

I will first test if the Orange Juice is better than Vitamin C in term of tooth growth. The null hypothesis is that Orange Juice and Vitamin C perform the same.

H_0: μ_{OJ} - $\mu_{VC} = 0$

H_1: μ_{OJ} - $\mu_{VC} > 0$

As the number of observation is not high, I'll use t-test and assume the populations are independent, are not paired and have unequal variances:

```
t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

The 95% confidence interval is [-0.171, 7.571] for μ_{OJ} - μ_{VC} and it contains 0. So I cannot reject the null hypothesis, in other words, Orange Juice is not better than Vitamin C.

The previous plot shows a significant mean difference with the 3 differente dosages. I will repeat the same test using the mean difference with dosage 0.5 ans dosage 2.

```
t <- subset(ToothGrowth, dose %in% c(0.5,2))
t.test(len~dose, paired=F, var.equal=F, data=t)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

With a 95% confidence interval of [-18.16, -12.83] for $\mu_{dose0.5}$ - μ_{dose2} (p-value = 4.398e-14), I reject the null hypothesis that there is no difference in tooth length between dosage 0,5 and 2. According to this test, the higher the dosage is, the longer the teeth are.

Now let's compare the effect of dosage depending on the supplement type.

```
t05 <- subset(ToothGrowth, dose == 0.5)
t1 <- subset(ToothGrowth, dose == 1)
t2 <- subset(ToothGrowth, dose == 2)
t.test(len~supp, paired=F, var.equal=F, data=t05)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len~supp, paired=F, var.equal=F, data=t1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len~supp, paired=F, var.equal=F, data=t2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

- For a dosage of 0.5mg, the 95% confidence interval of [1.719, 8.781] for μ_{OJ} - μ_{VC} (p-value = 0.006359), so I reject the null hypothesis. At this dosage, orange Juice is better than Vitamin C.
- For a dosage of 1g, the 95% confidence interval of [2.802, 9.058] for μ_{OJ} - μ_{VC} (p-value = 0.001038), so I reject the null hypothesis. At this dosage, orange Juice is better than Vitamin C.
- For a dosage of 1g, the 95% confidence interval of [-3.798, 3.638] for μ_{OJ} - μ_{VC} (p-value = 0.9639), so I cannot reject the null hypothesis. At this dosage, orange Juice is as good as Vitamin C.

4. Conclusion

As we saw, if dosage is ignored, there is no significant difference in the tooth length between Orange Juice and Vitamin C. However, if the supplement type is ignored, the plot and the second t.test show that a higher dose provides a significant difference.

One interesting conclusion is that Orange Juice do have a larger impact than Vitamin C on tooth growth with a smaller dosage. At a dosage of 2 mg, the impact became the same.

All tests assume the populations are independent, are not paired and have unequal variances. A pig cannot be "reused" to measure the their tooth growth twice.