Statistical Interference - Tooth Growth

Synopsis

ToothGrowth dataset is provided in library. In the dataset, it shows how the consumption of vitamin C by guinea pigs affect its tooth growth. Each one consume vitamin C with three different dose levels, half, one, and two miligrams per day. Based on the analysis result, growth effect by consumption of orange juice is only significant compared to ascorbic acid when the vitamin C dosage is below two miligrams.

Work with the data

```r
# load library
library(datasets)
# load data
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 ...
```

"len" is the tooth length. "supp" is delivery method where OJ is orange juice and VC is ascorbic acid. "dose" is the dosage.

plot data

```r
# use library "ggplot2"
library(ggplot2)
# use ggplot function
ggplot(ToothGrowth, aes(x=factor(dose), y=len)) + geom_point() + stat_summary(fun="mean", col="blue", s:
```

```
## Warning: Removed 3 rows containing missing values (`geom_segment()`).
## Removed 3 rows containing missing values (`geom_segment()`).
```
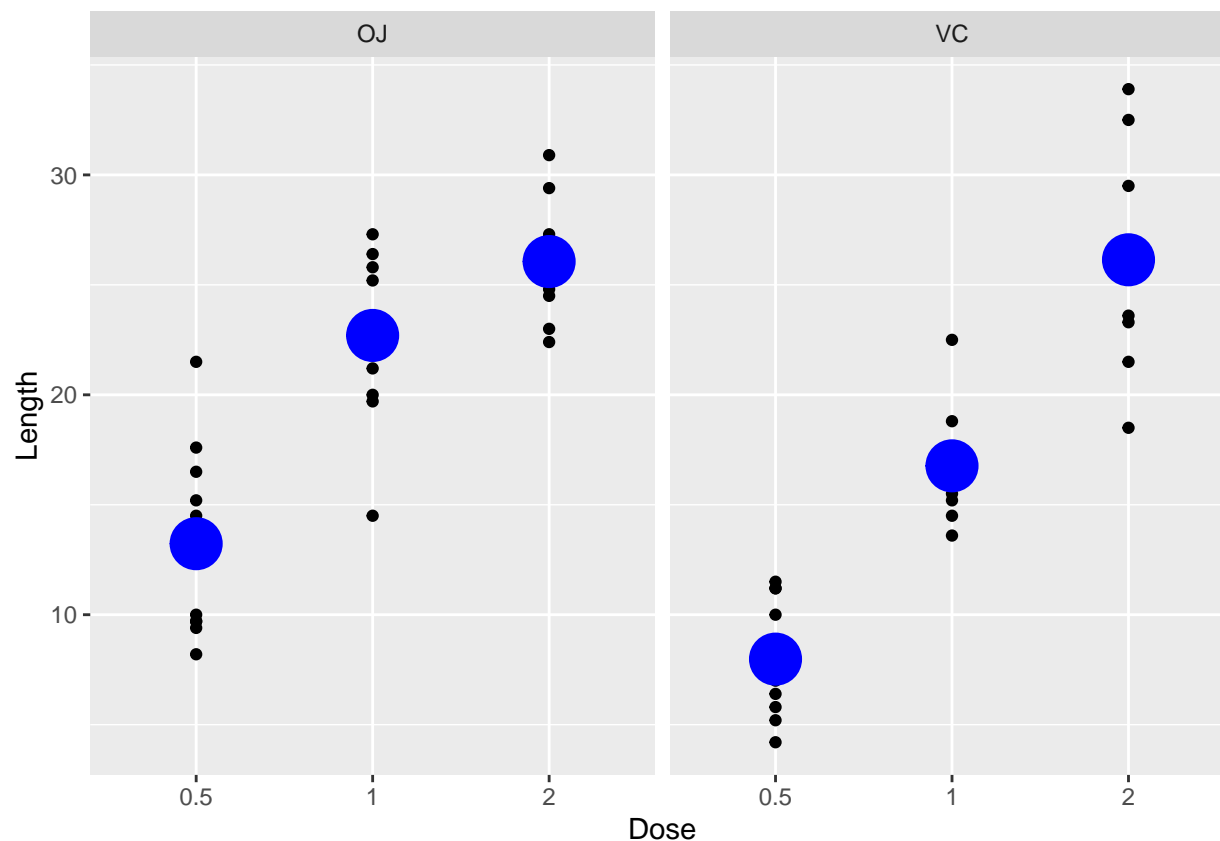
Blue circles are the average length of each dose.

Based on the graph above, it can be seen that overall orange juice is somehow more effective since it has higher mean values, but for two miligrams dosage per day, it does not show significant improvement compared to ascorbic acid. Now, if we re compare them without considering the dose, let's see if orange juice is still more effective.

```
orange_juice <- subset(ToothGrowth, supp=="OJ")
ascorbic_acid <- subset(ToothGrowth, supp=="VC")

t.test(orange_juice$len, ascorbic_acid$len)
```

```
##
##   Welch Two Sample t-test
##
## data:  orange_juice$len and ascorbic_acid$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

what if we compare only two miligrams of dosage

```
orange_juice <- subset(ToothGrowth, supp=="OJ" & ToothGrowth$dose=="2")
ascorbic_acid <- subset(ToothGrowth, supp=="VC" & ToothGrowth$dose=="2")

t.test(orange_juice$len, ascorbic_acid$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  orange_juice$len and ascorbic_acid$len
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

Based on both test above, overall the orange juice eficacy is higher than ascorbic acid. However, for dosage of two miligrams vitamin C, the efficacy are the same considering the p-value on the second test is significantly high.

Conclusion

Hence, the conclusion is that orange juice is only better than ascorbic acid when the dosage is lower than two miligrams.

Assumption

- independent samples and identically distributed
- variance of groups are unequal
- normal distribution (approx.)
- random samples of population