Statistical Inference Project Part 2

Azmi Mohamed Ridwan

March 27, 2016

The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
data("ToothGrowth")
```

From the help description, the data description:

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

The dataset is a dataframe with 60 observations and 3 variables

- [,1] len numeric Tooth length
- [,2] supp factor Supplement type (VC or OJ).
- [,3] dose numeric Dose in milligrams/day

```
#  Look at the class & structure
class(ToothGrowth)
```

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

```
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 ...
```

```
# Print the first 10 rows
head(ToothGrowth)
```

```
##    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
## 6 10.0   VC  0.5
```

```
# Since the dose has discrete values, we can convert the dose as factors
unique(ToothGrowth$dose)
```

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

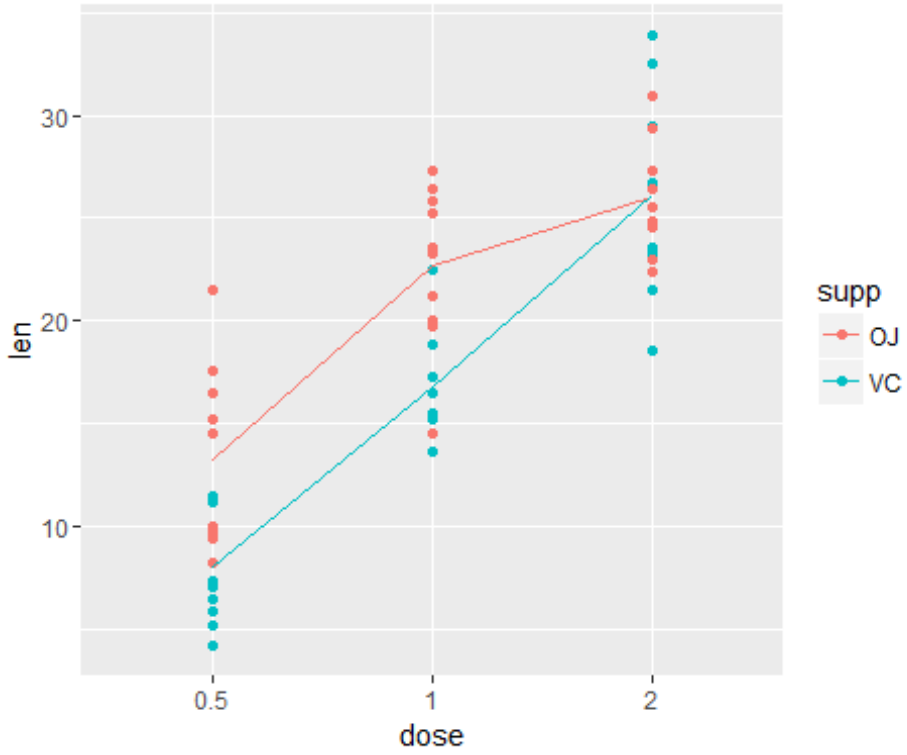
```
ToothGrowth$dose <-  as.factor(ToothGrowth$dose)

# Summary of values
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

To do some visualization, we first do a scatterplot of tooth length against the dosage overlaid with the mean values for each dosage per supplement type.

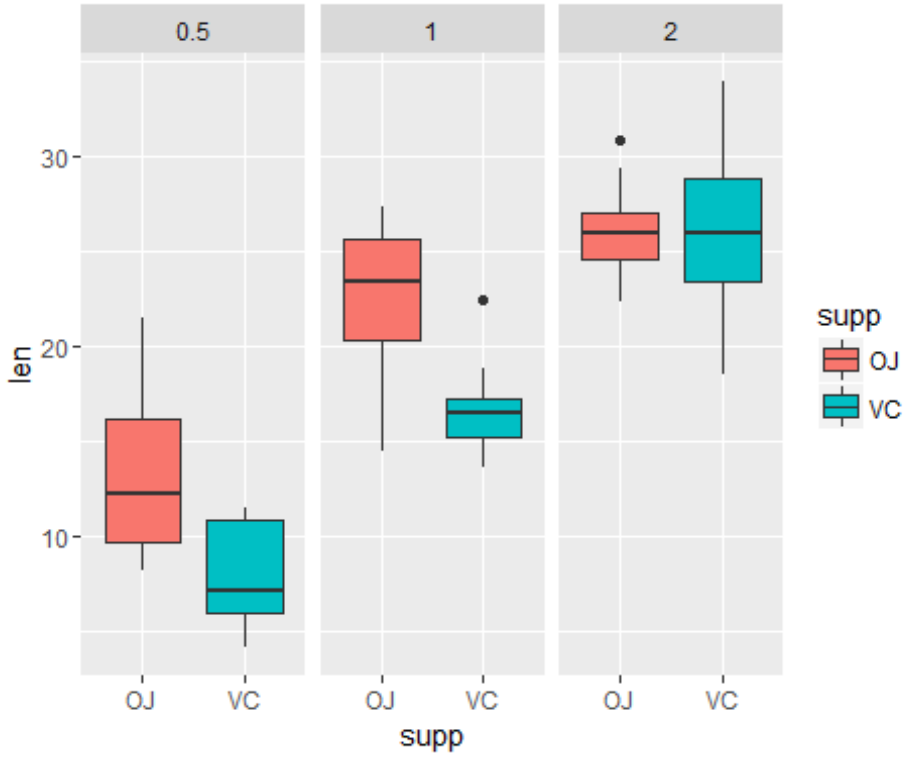
```
library(ggplot2)
g <- ggplot(aes(x=dose, y = len), data = ToothGrowth)
g <- g + geom_point(aes(color = supp))
avg <- aggregate(len~.,data=ToothGrowth,mean)
g <- g + geom_line(data=avg,aes(group=supp,colour=supp))
g
```



As an initial assessment, we can see that in both supplements, the increase in dosage results in the increase of tooth length with OJ giving better results.

As an alternative visualization, we can use box plots instead which leads to the same initial assessment.

```
ggplot(aes(x=supp,y=len),data=ToothGrowth)+geom_boxplot(aes(fill=supp)) +
    facet_wrap(~ dose)
```



Provide a basic summary of the data.

We can summarize the data by looking at the tooth length by supplement and dosage amount.

```
by(ToothGrowth$len, INDICES = list(ToothGrowth$supp, ToothGrowth$dose),
summary)

## : OJ
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
## ------------------------------------------------------------
## : VC
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
## ------------------------------------------------------------
## : OJ
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## ------------------------------------------------------------
## : VC
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.27   16.50   16.77   17.30   22.50
## ------------------------------------------------------------
```

```
## : OJ
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.58   25.95   26.06   27.08   30.90
## ------------------------------------------------------------
## : VC
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

First we compare between the 2 supplements in the whole dataset by doing an unpaired t-test.

```
result1 <- t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
print(result1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

Since the p value is **0.0606345** and the confidence interval includes 0, we cannot reject the null hypotheses that there is no difference due to the 2 supplements between the 2 groups. However this is not the best way to conduct this test.

Next, we subset the data by dosage and look at the results then.

```
# subset the data by dosage
dose0_5 <- subset(ToothGrowth, dose == 0.5)
dose1 <- subset(ToothGrowth, dose == 1.0)
dose2 <- subset(ToothGrowth, dose == 2.0)

# Run the t-test for each subset data
result2 <- t.test(len ~ supp, paired = F, var.equal = F, data = dose0_5)
result3 <- t.test(len ~ supp, paired = F, var.equal = F, data = dose1)
result4 <- t.test(len ~ supp, paired = F, var.equal = F, data = dose2)
print(result2)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
print(result3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
print(result4)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            26.06            26.14
```

The confidence intervals for dose levels 0.5mg **[1.7190573, 8.7809427]** and 1.0mg **[2.8021482, 9.0578518]** allow for the rejection of the null hypothesis i.e. there is a significant evidence that the supplement effects the tooth growth differently. However, the confidence interval for dose level 2.0mg **[-3.7980705, 3.6380705]** is not enough to reject the null hypothesis.

State your conclusions and the assumptions needed for your conclusions.

Assumptions

- The populations between the 2 supplement groups are independent
- The variances between populations are different
- A random population was used
- Double blind research methods were used.

Conclusions

- When dose equals 0.5mg or 1.0mg, the effect of Orange Juice (OJ) on Tooth Growth is better than the effect of ascorbic acid (VC).
- When dose equals 2.0, there is no significant difference between the 2 supplements.