Statistical Inference Project - Part 2 Inferential Data Analysis

ALIU

September 18, 2016

Load the ToothGrowth data and Basic exploratory data analyses

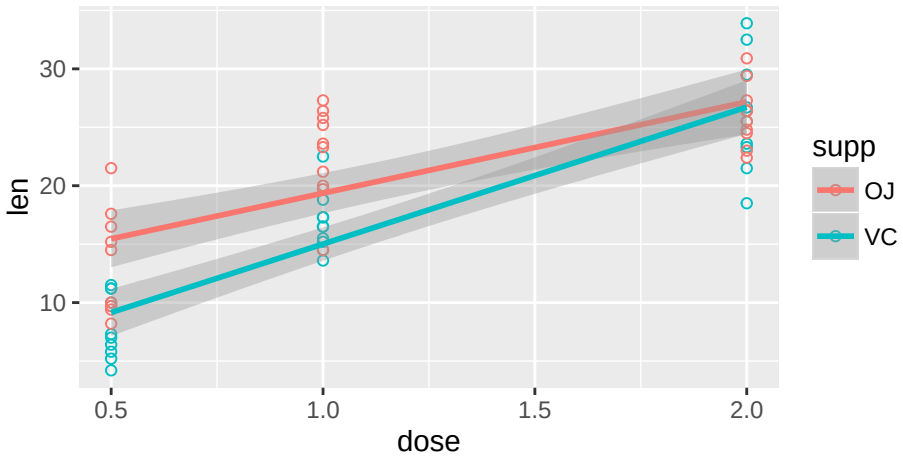
Load the data "ToothGrowth" into R and take a look at the structure using the "str" function.

```
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 ...
```

To explore the relationship between "len" and "dose", we can plot the two variables in an x-y plot, given the type of "supp".

```
g <- ggplot(ToothGrowth, aes(x=dose, y=len, colour=supp)) +
  geom_point(shape=1) + geom_smooth(method=lm)
g
```



Basic summary of the data

From the outputs, we can see that there are 3 variables and 60 observations in the data set, with the variables "len" and "dose" as numeric variables and "supp" as a factor variable of two levels. From

the plot, we can see that as the dose increases, the lent of the tooth tend to be longer as well and this trend is true for both supplement categories. At the same dose level, the supplement OJ seems to have a greater effect on tooth length, although, as the dose increases, the difference between OJ and VC decreases

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Tooth growth by supplement

To compare the tooth growth by supplement, we can Perform a hypothesis test with the null hypothesis: OJ and VC do not have a statistically significant difference in their effect on tooth growth. The results are as follows:

```
test1 <- t.test(len ~ supp,data=ToothGrowth)
test1
```

```
##
##  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
```

The p-value of 0.0606345 is suggests that we do not have sufficient evidence to reject the null hypothesis that OJ and VC do not have a statistically significant effect on tooth growth. The same conclusion can also be drawn by interpreting the confidence interval of -0.1710156, 7.5710156 for mean(OJ)-mean(VC), which contains zero.

Tooth growth by dose

Similiarly, to compare the tooth growth by dose, we can Perform a hypothesis test with the null hypothesis: dose do not have a statistically significant difference in their effect on tooth growth. For this analysis, we will ignore the type of supplement and subset the three dosage levels into groups for the hypothesis test. For simplicity, only the p-values and confidence intervals of each test are shown. The results are as follows:

```
dose12 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose23 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
dose13 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))

test2a <- t.test(len ~ dose, data=dose12)
test2b <- t.test(len ~ dose, data=dose13)
test2c <- t.test(len ~ dose, data=dose23)

test2 <- data.frame(p.value = c(test2a$p.value, test2b$p.value, test2c$p.value),
                    conf.int.lower = c(test2a$conf.int[1],
                                       test2b$conf.int[1],
                                       test2c$conf.int[1]),
                    conf.int.upper = c(test2a$conf.int[2],
                                       test2b$conf.int[2],
                                       test2c$conf.int[2]),
                    row.names = c("Test 2A", "Test 2B", "Test 2C"))

test2
```

```
##              p.value conf.int.lower conf.int.upper
## Test 2A 1.268301e-07     -11.983781      -6.276219
## Test 2B 4.397525e-14     -18.156167     -12.833833
## Test 2C 1.906430e-05      -8.996481      -3.733519
```

With p-values all below 0.05, and the confidence intervals not containing zero, we can reject the null hypotheses and conclude that there is a significant difference in tooth length between dosages among all three levels.

Comparison of tooth growth considering both factors

To compare tooth growth considring both dose and supplement, we subset the data into groups by the dosage levels and perform hyphothesis tests for the supplement types at each dosage level.

```
dose05 <- subset(ToothGrowth, dose == 0.5); test3a <- t.test(len ~ supp, data=dose05)
dose10 <- subset(ToothGrowth, dose == 1.0); test3b <- t.test(len ~ supp, data=dose10)
dose20 <- subset(ToothGrowth, dose == 2.0); test3c <- t.test(len ~ supp, data=dose20)

test3 <- data.frame(p.value = c(test3a$p.value, test3b$p.value, test3c$p.value),
                    conf.int.lower = c(test3a$conf.int[1],
                                       test3b$conf.int[1],
                                       test3c$conf.int[1]),
                    conf.int.upper = c(test3a$conf.int[2],
                                       test3b$conf.int[2],
                                       test3c$conf.int[2]),
                    row.names = c("Test 3A", "Test 3B", "Test 3C"))

test3
```

```
##             p.value conf.int.lower conf.int.upper
## Test 3A 0.006358607       1.719057       8.780943
## Test 3B 0.001038376       2.802148       9.057852
## Test 3C 0.963851589      -3.798070       3.638070
```

Based on the p-values and confidence intervals, we can reject the null hypotheses and conclude that OJ is significantly more effective than VC for tooth growth at dose levels 0.5 and 1.0. However, when dose level reaches 2.0, we cannot reject the null hypothesis and therefore have to conclude that the difference in effect on tooth growth between OJ and VC is not statistically significant.

Conclusions and the assumptions needed

Conclusions

- When ignoring dose levels, there is no statistically significant difference in the tooth growth between the supplement types.
- When ignoring the supplement types, there is a statistically significant difference in tooth length between the dose levels, with higher doses resulting in longer teeth.
- When considering both factors:
  - At the dosage levels 0.5 and 1.0, OJ has a statistically significant effect on tooth length comparing to VC
  - However, as the dosage increases to 2.0, the difference is no longer significant

Assumptions: In the tests and analysis performed, we assume that the variances between the separate populations tested are different (used the default var.equal = FALSE for all the t tests). Also, we assume that the populations are independent and random.