Statistical Inference - Project, Part 2

N. Lakhani

17 January 2018

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

Synopsis

This report is part of the Coursera project on Statistical Inference. It provides Basic inferential analysis using *ToothGrowth* data. The data is explored for impact of 2 different supplements and dosage on tooth growth length. A hypothesis test is conducted to explore if the difference in mean growth is statistically different given the 2 supplements and their dosage

Analyze ToothGrowth

This section does an analysis of toothgrowth data providing

- A basic summary of data
- Use of hypothesis testing to compare tooth growth by supplement of 'OJ' - Orange Juice or 'VC' - Ascorbic acid and by 'dose'.

I use a null hypothesis: stating that there is no difference (at a .05 - alpha significance level) of the mean tooth growth between the supplement methods at various doses

Assumption

I have assumed that the difference in growths between the 2 supplements is normal, the variance are equal, the data is not paired and the observations are independent

Basic analysis of data

The code below extracts the basis structure of the data file and statistics of the tooth growth data for each of the 2 supplement methods.

- There are 30 data points for each of the 2 supplement methods
- The mean growth of the 'OJ' supplement is 20.66, median at 22.70
- The mean growth of the 'VC' supplement is 18.81 with median at 19.25

From the boxplot below by supplement & dose, it does appear that the means & median of the growth between the 2 methods seem different across different doses. The ranges also appear to be different. As an example for 0.5 dose the mean for Oranje juice is about 12.5 vs 8 for Ascorbic acid.

I use a hypothesis test to explore if the mean growth corresponding to the 2 supplement methods are statistically different for the 3 different dose levels

```
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 ...
```

```
summary(ToothGrowth[ToothGrowth$supp == 'OJ',])
```

```
##       len        supp         dose
##  Min.   : 8.20   OJ:30   Min.   :0.500
##  1st Qu.:15.53   VC: 0   1st Qu.:0.500
##  Median :22.70           Median :1.000
##  Mean   :20.66           Mean   :1.167
##  3rd Qu.:25.73           3rd Qu.:2.000
##  Max.   :30.90           Max.   :2.000
```

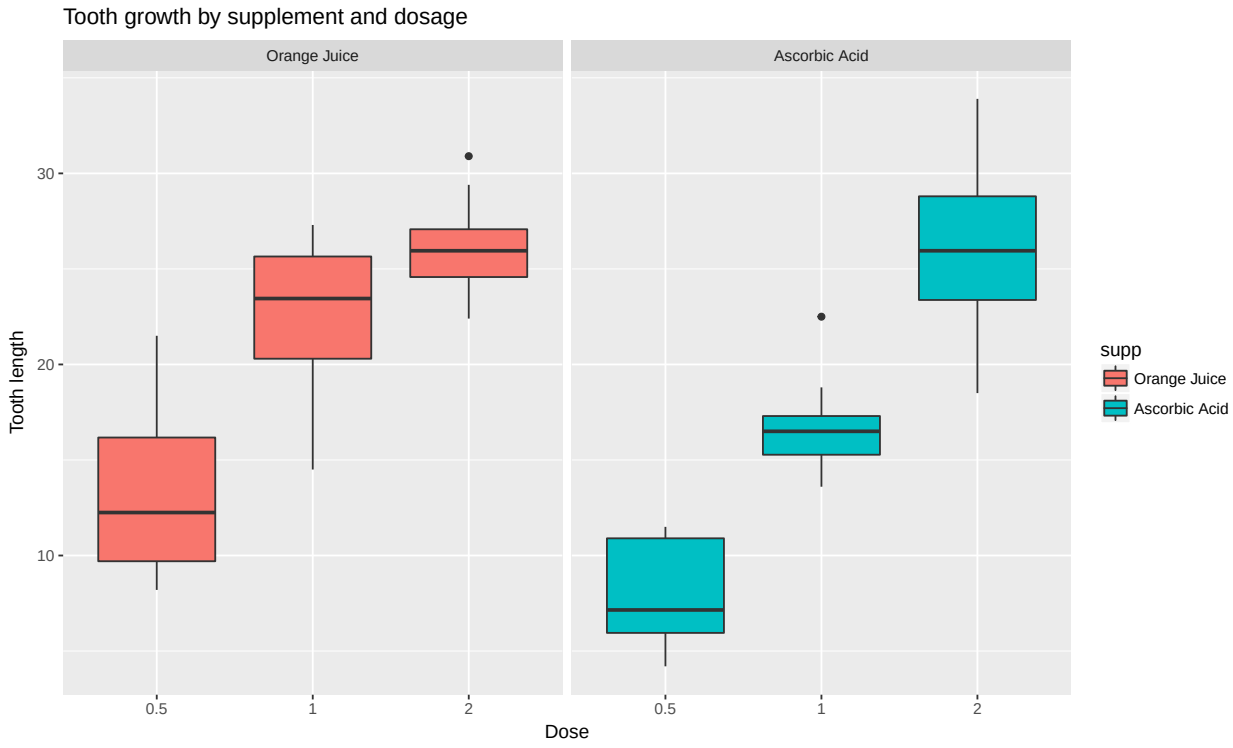
```
summary(ToothGrowth[ToothGrowth$supp == 'VC',])
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ: 0   Min.   :0.500
##  1st Qu.:11.20   VC:30   1st Qu.:0.500
##  Median :16.50           Median :1.000
##  Mean   :16.96           Mean   :1.167
##  3rd Qu.:23.10           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
levels(ToothGrowth$supp) <- c('Orange Juice','Ascorbic Acid')
xtabs(ToothGrowth$len ~ ToothGrowth$dose + ToothGrowth$supp)/c(10,10,10)
```

```
##                 ToothGrowth$supp
## ToothGrowth$dose Orange Juice Ascorbic Acid
##              0.5         13.23          7.98
##              1           22.70         16.77
##              2           26.06         26.14
```

```
ggplot(ToothGrowth,aes(x=factor(dose),y=len)) + geom_boxplot(aes(fill=supp)) +labs(title = 'Tooth growth
```



We will use hypothesis test to test the difference in mean growth between the 2 supplements.

- The null hypothesis H0: mu1 - mu2 = 0, that is growths for both supplements at same dosage level is identical at a 0.05 signficance level
- The alternative hypothesis Ha: mu1 - mu2 < or > 0, that is the growths are not identical. We use a 2-sided test.

The hypothesis test is conducted using a t.test using R below. Four tests are conducted for a) all dose combined and b) each of the 3 doses, resulting in **p-values** for each test as below:

```
t.test(data=ToothGrowth,len ~ supp,alternative='two.sided',conf.int=.95)
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                    20.66333                    16.96333
```

```
t.test(len ~ supp,data=subset(ToothGrowth,dose==0.5),alternative='two.sided')
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                       13.23                       7.98
```

```
t.test(len ~ supp,data=subset(ToothGrowth,dose==1),alternative='two.sided')
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                       22.70                       16.77
```

```
t.test(len ~ supp,data=subset(ToothGrowth,dose==2),alternative='two.sided')
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                      26.06                       26.14
```

**The t-test results show*:

- for all doses combined, the p value is 0.06 (> 0.05), hence we cannot reject the null hypothesis, the conclusion is that the supplement across doses do not make a difference in growth. The mean growth for orange juice is 20.66 vs 16.96 for Ascorbic acid.
- for dose of 0.5, p value is 0.006 (< 0.05), hence we reject the null hypothesis; the conclusion is that the supplement across dose of 0.5 makes a difference and the mean growth for orange juice is 13.23 vs 7.98 for Ascorbic acid
- for dose of 1, p value is 0.001 (< 0.05), hence we reject the null hypothesis; the conclusion is that the supplement across dose of 1 makes a difference and the mean growth for orange juice is 22.7 vs 16.77 for Ascorbic acid
- for dose of 2, p value is 0.96 (> 0.05), hence we cannot reject the null hypothesis, the conclusion is that the supplement at dose of 2 makes no difference and the mean growth for orange juice is 26.06 vs 26.147 for Ascorbic acid. The CI for difference in means is [-3.97,3.63]

Conclusion

Conducted hypothesis tests to examine if the growth supplement made an impact on tooth growth across various doses. The tests show that a difference in growth is statistically valid at dose of 0.5 and 1.0, but does not hold at dose of 2, for a significance level of 0.5