

Statistical Inference Course Project Part 2

Mark Nicholls

24/04/2021

Basic Inferential Data Analysis

Overview

Here I will load up the ToothGrowth data, then perform some basic exploratory data analyses. I'll then show an appropriate plot to summarize the data. I'll then use hypothesis tests to compare tooth growth by supplement and dose, and then state my assumptions and conclusions

Basic summary of the data

```
#load the data  
data("ToothGrowth")  
#number of rows and columns  
dim(ToothGrowth)
```

```
## [1] 60 3
```

```
#first few 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
```

```
#last few rows  
tail(ToothGrowth)
```

```
##      len supp dose  
## 55 24.8   OJ  2  
## 56 30.9   OJ  2  
## 57 26.4   OJ  2  
## 58 27.3   OJ  2  
## 59 29.4   OJ  2  
## 60 23.0   OJ  2
```

```
# how many NAs?  
sum(apply(ToothGrowth, 2, is.na))
```

```
## [1] 0
```

```
# how many of each
table(ToothGrowth$supp, ToothGrowth$dose)
```

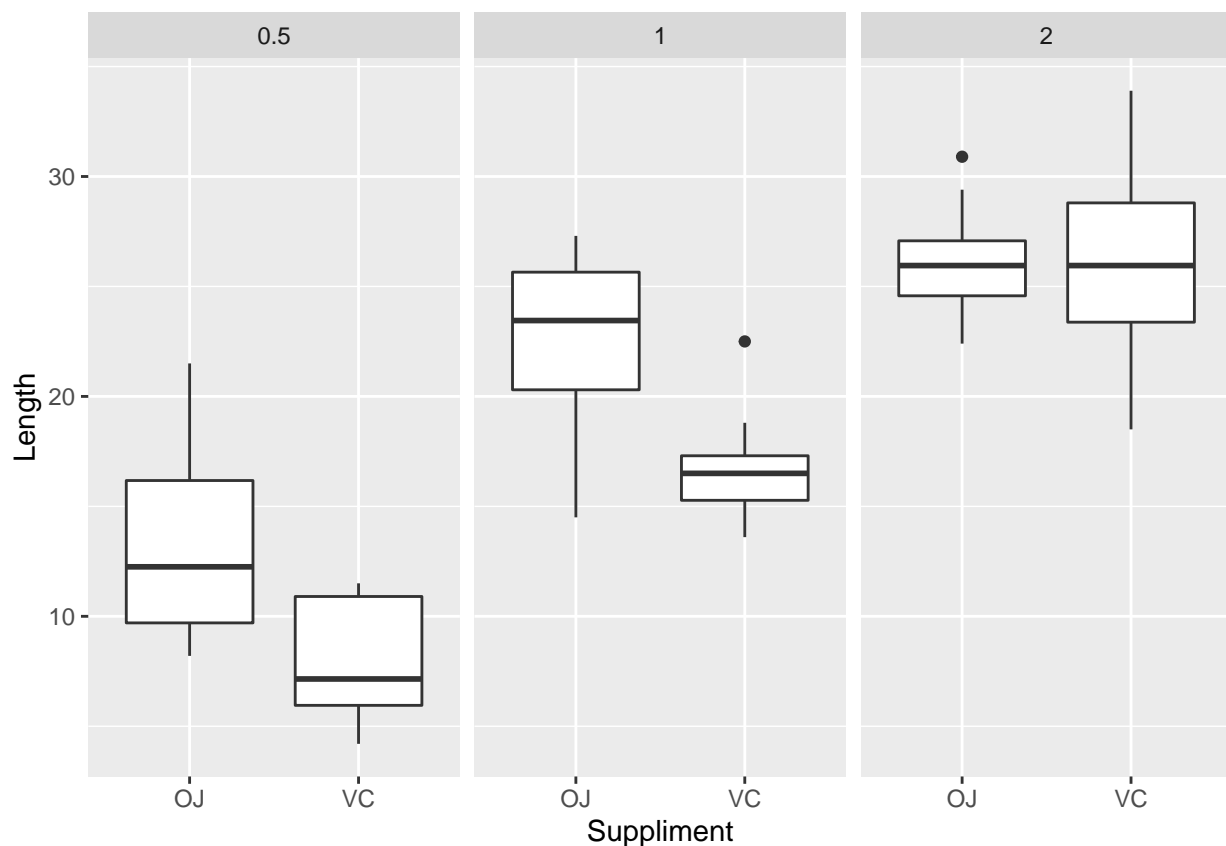
```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Data Summary

Ok, it looks like we have 60 hamsters split into 2 groups, one for each supplement and each supplement has 3 doses, (0.5, 1 & 2). I'll use a box & whisker plot to summarize the data.

```
#load library
library(ggplot2)

ggplot(data = ToothGrowth) +
  geom_boxplot(aes(x = supp, y = len)) +
  facet_grid(. ~ dose) +
  labs(x = "Suppliment", y = "Length")
```



It appears from the box & whisker plot that for lower doses OJ results in greater average tooth length, and at the greatest dose the average tooth length appears the same, lets test those differences to a significance level of 0.05 (95% confidence)

Hypothesis tests

For all these tests I'll use H_0 : no difference in the mean, and H_a : the means are not equal, that is to say that any increase is meaningful. My significance level or alpha is 0.05, so any p value below that I will consider sufficient evidence to reject H_0 .

I'll separate the relevant values and then do use a t.test to compare the means, default values all apply here so I'll use those.

First, for dose = 0.5

```
vc_05 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5,1]
oj_05 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5,1]

t.test(oj_05, vc_05)
```

```
##
## Welch Two Sample t-test
##
## data:  oj_05 and vc_05
## 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 of x mean of y
##      13.23      7.98
```

The p value is below my chosen significance level, so I reject H_0 in favour of the alternative; dose 0.5 OJ has a significantly greater tooth length.

Second, do the same analysis for dose = 1

```
vc_1 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1,1]
oj_1 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1,1]

t.test(oj_1, vc_1)
```

```
##
## Welch Two Sample t-test
##
## data:  oj_1 and vc_1
## 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 of x mean of y
##      22.70      16.77
```

Again the p value is far less than my chosen significance level so I reject H_0 in favour of the alternative: at dose = 1 OJ has a significantly greater tooth length.

Last, for dose = 2

```
vc_2 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2,1]
oj_2 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2,1]

t.test(oj_2, vc_2)
```

```
##
## Welch Two Sample t-test
##
## data:  oj_2 and vc_2
## 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
```

Here the p value is above my chosen significance level, so I fail to reject H_0 : the means are equal, on there is no significant difference in average tooth growth between the supplements.

Conclusion

For this analysis I assume that that the trials were properly randomized and the data has no underlying confounding factors which affected the tooth growth rate.

For dose equal to 2 there was not a significant difference in the mean of the tooth growth rate for the supplements.

However at both doses lower than 2 the OJ supplement resulted in a greater tooth length when tested at a significance level of 0.05 (95%)