Part 2: Inferential Data Analysis

Lee Bunce

1 October 2017

In this report we look at the ToothGrowth data in the datasets package. After first performing some exploratory analysis we perform hypothesis tests to compare tooth growth by supplement type and and dose.

Exploratory Analysis

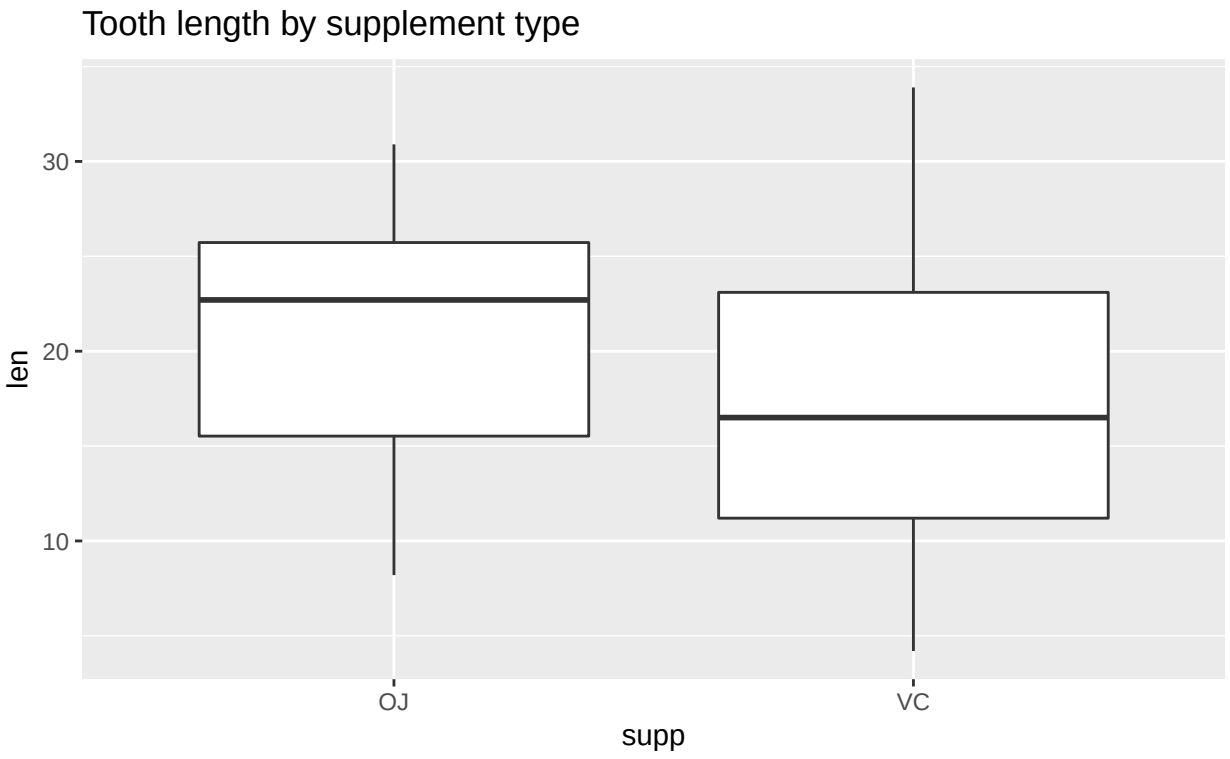
First we persent some summary statistics for the data, alongside some plots.

```
data("ToothGrowth")

summary(ToothGrowth)
```

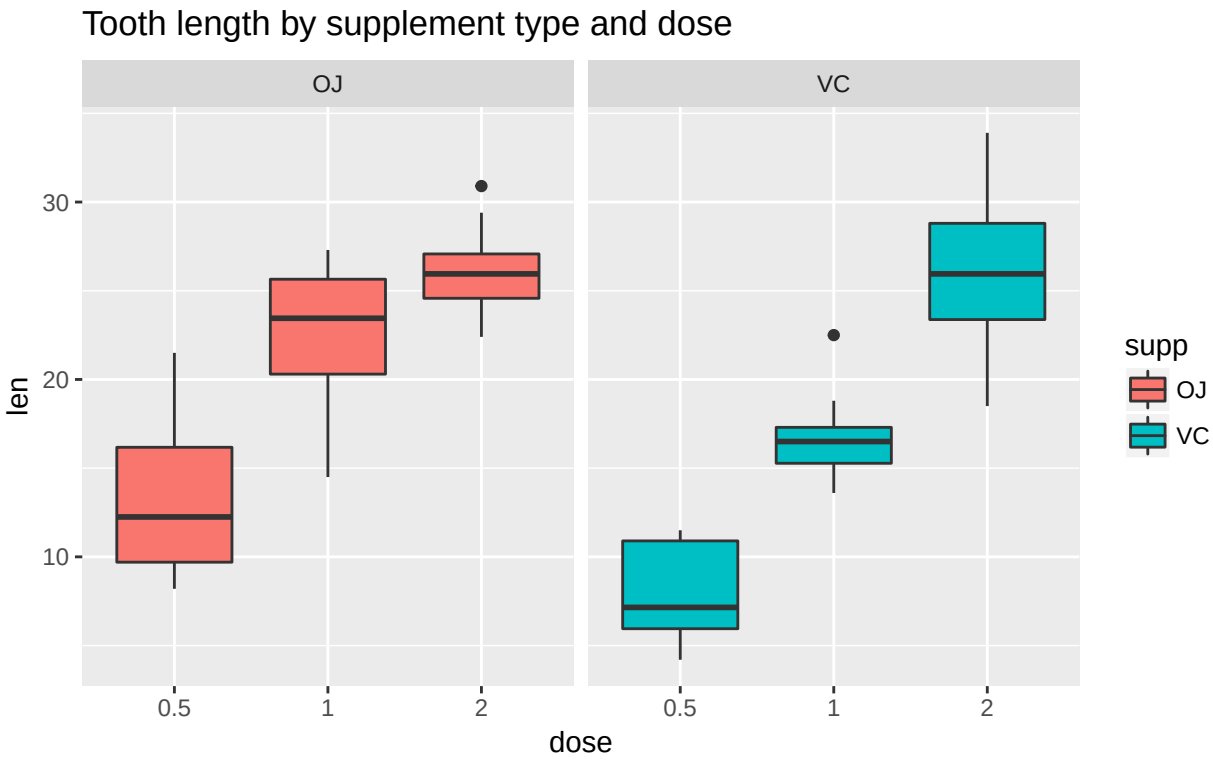
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
ggplot(ToothGrowth, aes(supp, len)) +
  geom_boxplot() +
  labs(title = 'Tooth length by supplement type')
```



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

ggplot(ToothGrowth, aes(dose, len, fill = supp)) +
  geom_boxplot() +
  facet_grid(~supp) +
  labs(title = 'Tooth length by supplement type and dose')
```



Hypothesis Test

One hypothesis suggested by our exploratory data analysis is that orange juice has a stonger effect on tooth growth than Vitamin C. To test that we perform a test test between the two groups.

```r
t.test(len~supp, data = ToothGrowth)
```

```
##
##  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 test gives a p-value of 0.06, meaning that under a 95% confidence test we fail to reject the null hypothesis that the two groups have the same population mean.

A second suggested hypothesis is that increased dose has a positive impact on tooth length. We again use a t-test to test this hypothesis, focusing on the difference between dosages of 0.5 and 2mg per day.

```
t.test(len~dose, data = filter(ToothGrowth, dose %in% c(0.5, 2)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

The p-value here us much less than 0.05, so we can reject the null hypothesis that there is no difference between the two groups.

Assumptions

The conclusions above relies on the assumptions that the sample is representative or the larger population.