Statistical Inference Project 2

ToothGrowth Exploratory and Inferential Analysis

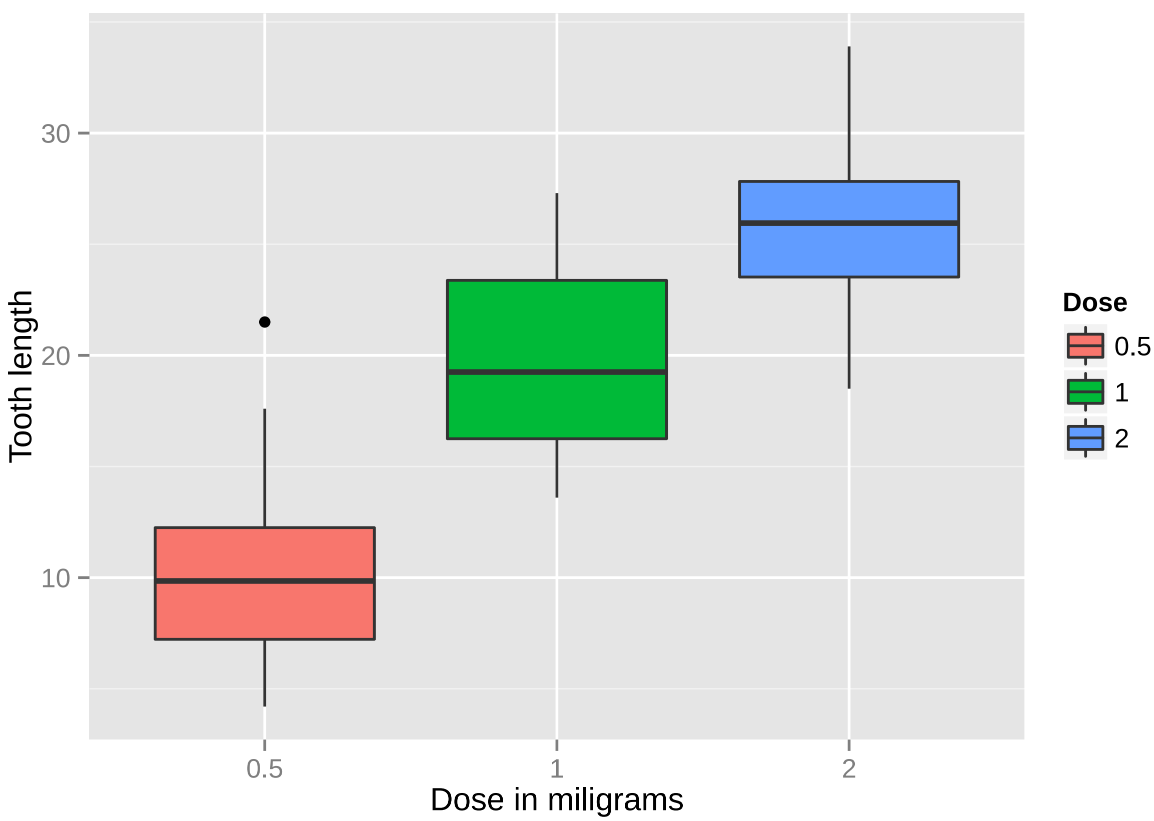
Chris Jones

In this analysis I will provide a basic summary of the ToothGrowth dataset and perform a basic inferential analysis. The ToothGrowth dataset describes the length of teeth in 10 guinea pigs at each of three doses of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
library(ggplot2)
library(datasets)
tg <- ToothGrowth
tg$dose <- as.factor(tg$dose) #convert dose to a factor for easier analysis
```

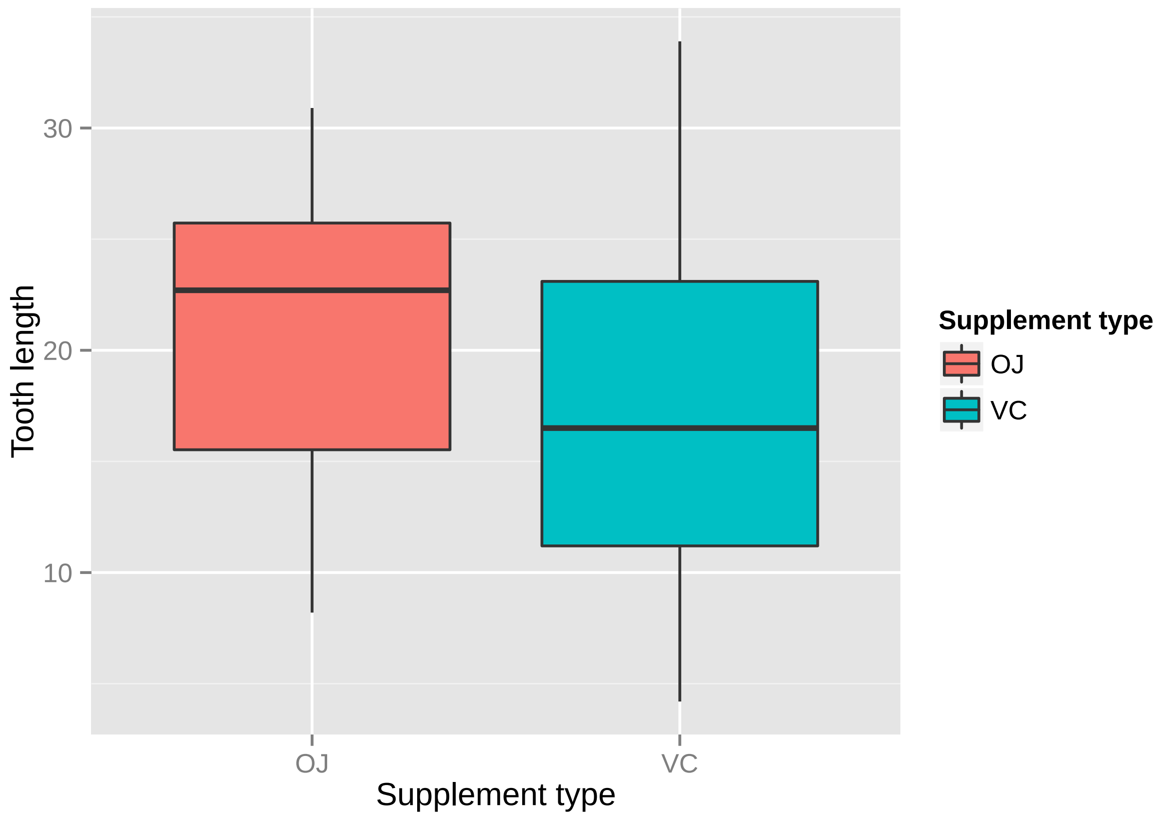
Basic exploratory analysis

```
ggplot(aes(x=dose, y=len), data=tg) +
  geom_boxplot(aes(fill=dose)) +
  xlab("Dose in miligrams") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Dose"))
```



Dosage appears to increase tooth length.

```
ggplot(aes(x=supp, y=len), data=tg) +
  geom_boxplot(aes(fill=supp)) +
  xlab("Supplement type") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```



Overall, orange juice appears to be associated with longer tooth length than ascorbic acid.

Inferential Analysis

Conditions

The analysis below relies on several assumptions:

1. That the guinea pigs under analysis were randomly sampled, and thus representative of the population
2. That dosages and supplement types were administered in a randomised order
3. That the means follow a normal distribution

Effect of dosage on tooth length

I have chosen the t-test because of the small sample size and paired nature of the data. I have divided the dataset into three pairs of dosages, so a t-test can be performed on each pair.

```r
dose_0.5_10 <- subset (tg, dose %in% c(0.5, 1.0))  #divide the dosages into testable pairs
dose_0.5_2.0 <- subset (tg, dose %in% c(0.5, 2.0))
dose_1.0_2.0 <- subset (tg, dose %in% c(1.0, 2.0))
```

```r
t.test(len ~ dose, data = dose_0.5_2.0, paired = FALSE)$conf.int #0.5g vs 2g
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(len ~ dose, data = dose_0.5_10, paired = FALSE)$conf.int #0.5g vs 1g
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(len ~ dose, data = dose_1.0_2.0, paired = FALSE)$conf.int #1g vs 2g
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Note that in each t-test, the 95 per cent confidence interval does not contain zero. This means that each increase in dosage results in a statistically significant increase in tooth length.

Effect of delivery method on tooth length

```r
t.test(len ~ supp, data = tg, paired = FALSE)$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

The 95 per cent confidence interval includes zero, meaning that we retain the null hypothesis that supplement type has no effect on tooth length.

Note that the t-tests examine the effect of the explanatory variables in isolation. Further analysis may wish to use multiple regression to examine the effects of both dosage and supplement type while holding the other constant.