Project 2

Data

```r
data(ToothGrowth)
```

Analysis

Setup the environment

```r
library(ggplot2)
library(dplyr)
library(datasets)
data(ToothGrowth)
```

Exploratory data analysis

```r
ToothGrowth %>% group_by(supp, dose) %>%
  summarise(meanLength = mean(len), medianLength = median(len), sdLength = sd(len))
```

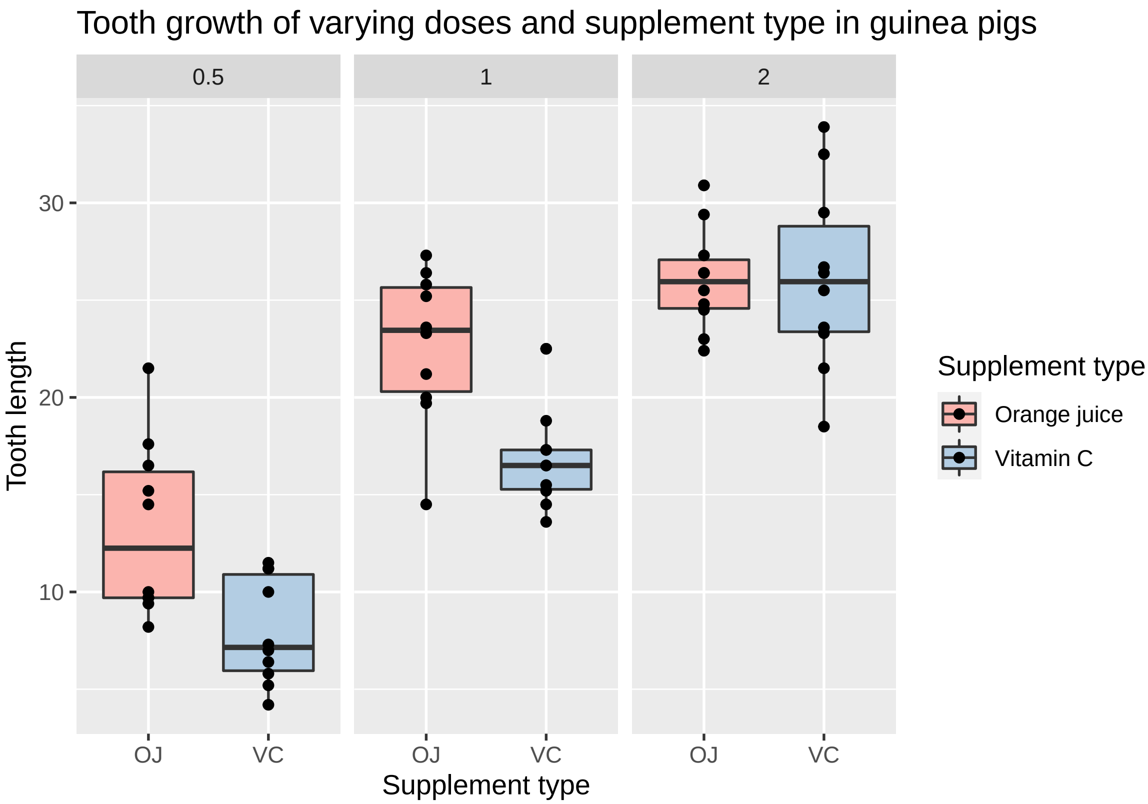
```
## `summarise()` regrouping output by 'supp' (override with `.groups` argument)
```

```
## # A tibble: 6 x 5
## # Groups:   supp [2]
##   supp   dose meanLength medianLength sdLength
##   <fct> <dbl>      <dbl>        <dbl>    <dbl>
## 1 OJ      0.5      13.2         12.2      4.46
## 2 OJ      1        22.7         23.5      3.91
## 3 OJ      2        26.1         26.0      2.66
## 4 VC      0.5       7.98         7.15     2.75
## 5 VC      1        16.8         16.5      2.52
## 6 VC      2        26.1         26.0      4.80
```

Observations:

- Guinea pigs treated with 0.5 and 1mg/day of orange juice have on average longer tooth length than those treated with vitamin C. However, this trend is not obvious in guinea pigs treated with 2mg/day of orange juice vs vitamin C

2. Plot the distribution of the raw data



Hypothesis testing: difference in tooth growth treated with orange juice and vitamin C

- Null hypothesis: there is no difference in tooth growth in guinea pigs treated with the same dose of orange juice and vitamin C
- Alternative hypothesis: there is a difference in tooth growth in guinea pigs treated with the same dose of orange juice and vitamin C
- Control type I error at less than 0.05 ($p<0.05$)

1. Stratify data by three doses: low dose (0.5mg/day), medium dose (1mg/day), high dose (2mg/day)

```
ld <- ToothGrowth[ToothGrowth$dose == 0.5, ]
md <- ToothGrowth[ToothGrowth$dose == 1, ]
hd <- ToothGrowth[ToothGrowth$dose == 2, ]
```

2. Perform hypothesis testing for low dose (0.5mg/day)

```
t.test(ld$len ~ ld$supp, alternative = 'two.sided', paired = F, var.equal = F)$p.value
```

```
## [1] 0.006358607
```

As p-value is less thean type I error(0.05), we can reject null hypothesis.

3. Perform hypothesis testing for medium dose (1mg/day) - code not shown but similar to above

```
## [1] 0.001038376
```

As p-value is less thean type I error(0.05), we can reject null hypothesis.

4. Perform hypothesis testing for high dose (2mg/day) - code not shown

```
## [1] 0.9638516
```

The mean tooth length when treated at 2mg/day with orange juice is not significantly different than that of vitamin C.

Hypothesis testing: difference in tooth growth treated with different doses of supplements

- Null hypothesis: there is no difference in tooth growth in guinea pigs treated with different doses of orange juice or vitamin C
- Alternative hypothesis: there is a difference in tooth growth in guinea pigs treated with different doses of orange juice or vitamin C
- Control the type I error rate at less than 0.05 ($p<0.05$)

1. Stratify data by two supplement types, orange juice and vitamin C, and then strtify the data to contain only two doses (since t.test can only take conlumns containing two levels) - code not shown

2. Perform hypothesis testing for orange juice at different doses Three different doses are compared to each other. The family-wise error rate will be controlled using the Bonferroni correction to adjust the p-value

```r
oj_lm_p <- t.test(oj_lm$len ~ oj_lm$dose, alternative = 'two.sided', paired = F, var.equal = F)$p.value
oj_lh_p <- t.test(oj_lh$len ~ oj_lh$dose, alternative = 'two.sided', paired = F, var.equal = F)$p.value
oj_mh_p <- t.test(oj_mh$len ~ oj_mh$dose, alternative = 'two.sided', paired = F, var.equal = F)$p.value
data.frame(comparisons = c('0.5 vs 1', '0.5 vs 2', '1 vs 2'),
           pvalue = c(oj_lm_p, oj_lh_p, oj_mh_p),
           bonferroni = p.adjust(c(oj_lm_p, oj_lh_p, oj_mh_p), method = 'bonferroni'))
```

```
##   comparisons       pvalue   bonferroni
## 1    0.5 vs 1 8.784919e-05 2.635476e-04
## 2    0.5 vs 2 1.323784e-06 3.971352e-06
## 3      1 vs 2 3.919514e-02 1.175854e-01
```

Since the Bonferroni-corrected p-values for 0.5 vs 1 and 0.5 vs 2 comparisons are less than 0.05, the null hypotheses are rejected.

3. Perform hypothesis testing for vitamin C at different doses

Three different doses are compared to each other. The family-wise error rate will be controlled using the Bonferroni correction to adjust the p-value - code not shown

```
##   comparisons       pvalue   bonferroni
## 1    0.5 vs 1 6.811018e-07 2.043305e-06
## 2    0.5 vs 2 4.681577e-08 1.404473e-07
## 3      1 vs 2 9.155603e-05 2.746681e-04
```

Since the Bonferroni-corrected p-values for all comparisons are less than 0.05, the null hypotheses are rejected.

Conclusions

1. The assumptions made to perform the statistical tests

- At different dose level there is a different variance in Orange juice and Vitamin-C treated groups
- Variance are different in the orange juice- and vitamin C-treated groups at different dose levels
- The different treatment groups are non-paired - this is not so much an assumption but how the experiment is actually condudcted, as specified in the **?ToothGrowh** description
- Assume all the observations are independent and identically distributed (iid)
- Assume tooth growht follows normal distribution

2. Conculsions

- At doses of 0.5mg/day and 1mg/day, guinea pigs treated with orange juice have longer tooth length compared to those treated with vitamin C; however, the difference is not significant at 2mg/day
- In guinea pigs treated with orange juice, treatment at 1mg/day and 2mg/day results in on average longer tooth length than those treated at 0.5mg/day; however, the difference is not significant in 1mg/day compared to 2mg/day treatment groups
- In guinea pigs treated with vitamin C, there is a does-dependent effect in that the higher the dose is, the longer the tooth length is. Whether this relationship is linear requires further exploration.