Program_assignment2 - Part 1

Overview

In this study we will test if three dose levels of vitamin C by two delivery methods have effect on odontoblasts (cells responsible for tooth growth).

To do that we use the `ToothGrowth` R dataset in the package `datasets`. In the dataset we have data for 60 guinea pigs and each one received a treatment of vitamin C to measure tooth length according to his treatment.

Data Summary

```
ToothGrowth %>%
  head(5)
```

```
##    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
```

```
ToothGrowth %>%
  group_by(supp, dose) %>%
  summarise(n = n()) %>%
  as.data.frame()
```

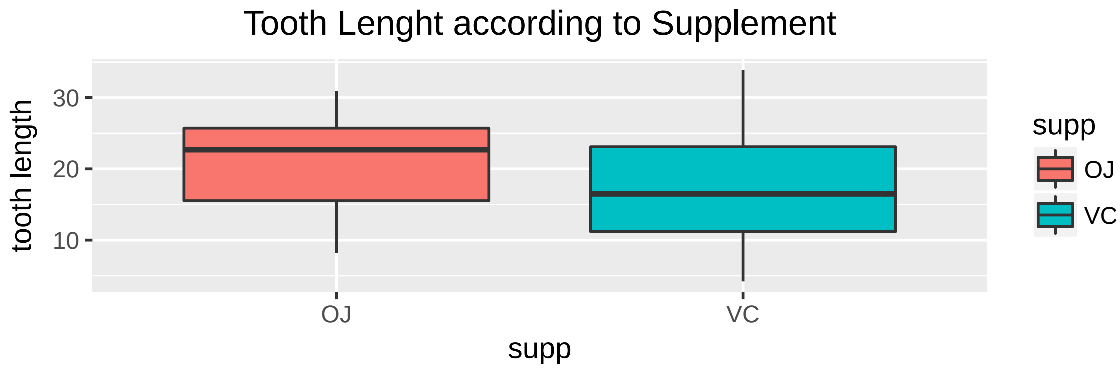
```
##   supp dose  n
## 1   OJ  0.5 10
## 2   OJ  1.0 10
## 3   OJ  2.0 10
## 4   VC  0.5 10
## 5   VC  1.0 10
## 6   VC  2.0 10
```

So in the dataset we have 60 rows and 3 columns. The first column is len that reprensents the tooth length, the second one is supp and represents the supplement, that is the delivery method of vitamin C (`VC` is for ascorbic acid and `OJ` for orange juice). The last column dose is the dose of vitamin C in milligrams/day. The dose can be equal to 0.5, 1 or 2.

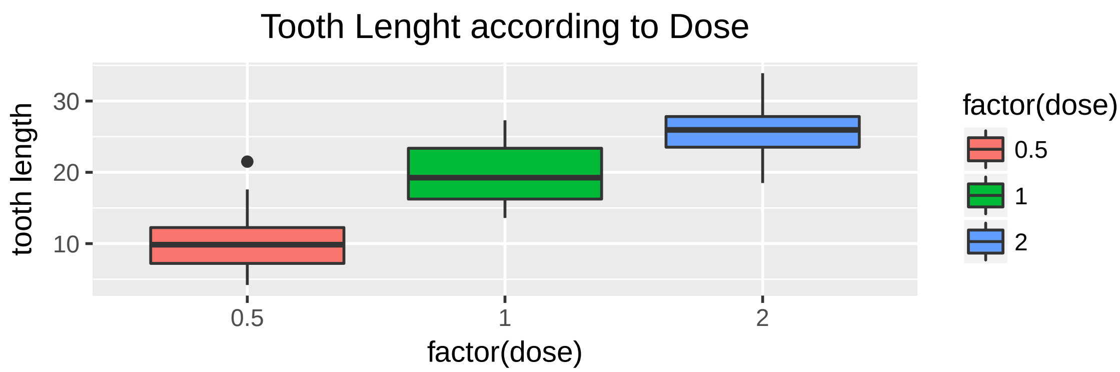
The experimental design is balanced: there is 6 different treatments (combination of dose and delivery method) and each treatment is delivered to 10 pigs.

Exploratory Analysis

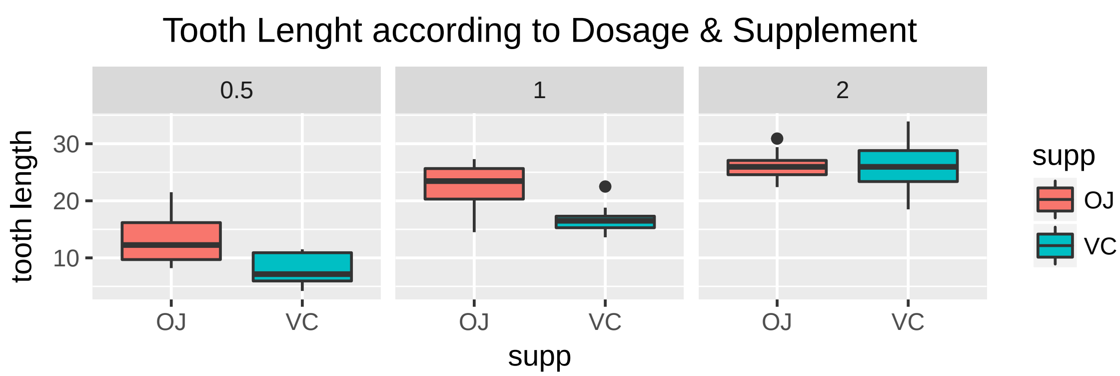
(The code to display the three next boxplots is in the appendix part)



This first boxplot, show difference between OJ and VC methods. The range of values are [8.2 ; 30.9] and [4.2 ; 33.9] respectively. It shows that variance are not the same if you change the delivery method.



Similarly, with this second boxplot, we see an increase of averages for the 0.5, 1.0, & 2.0 equal to 10.61, 19.73, 26.10 respectively And we also notice a increasing range levels according to dose level. The variance is also increasing.



In summation, we observe that dose level have a more important impact on teeth growth, but the last graphic show that for a dose of 2 the delivery method don't seem to affect the result of length tooth. To test this theory, hypothesis tests will be conducted in next section.

Hypothesis tests

We will make some Student tests to determine if dose level and delivery method have effects on tooth growth. The result of each test is summary in one sentence after the test and the conclusion is drawing in the last section.

```r
test_len.vs.supp <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)
print(paste("t-statistic=", test_len.vs.supp$statistic))
```

```
## [1] "t-statistic= 1.91526826869527"
```

```r
print(paste("p-value=", test_len.vs.supp$p.value))
```

```
## [1] "p-value= 0.0606345078809341"
```

Fail to reject the null hypothesis: supplement delivery type have no effect on tooth growth.

```r
dose12 <- ToothGrowth %>%
  filter(dose %in% c(0.5, 1.0))
dose13 <- ToothGrowth %>%
  filter(dose %in% c(0.5, 2.0))
dose23 <- ToothGrowth %>%
  filter(dose %in% c(1.0, 2.0))
```

```r
test_len.vs.dose12 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose12)
print(paste("t-statistic=", test_len.vs.dose12$statistic))
```

```
## [1] "t-statistic= -6.4766477265891"
```

```r
print(paste("p-value=", test_len.vs.dose12$p.value))
```

```
## [1] "p-value= 1.26830072017385e-07"
```

Reject the null hypothesis: dose level is correlated to tooth growth (restricted to dose level of 0.5 or 1).

```r
test_len.vs.dose13 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose13)
print(paste("t-statistic=", test_len.vs.dose13$statistic))
```

```
## [1] "t-statistic= -11.7990459563103"
```

```r
print(paste("p-value=", test_len.vs.dose13$p.value))
```

```
## [1] "p-value= 4.39752495936323e-14"
```

Reject the null hypothesis: dose level is correlated to tooth growth (restricted to dose level of 0.5 or 2).

```r
test_len.vs.dose23 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose23)
print(paste("t-statistic=", test_len.vs.dose23$statistic))
```

```
## [1] "t-statistic= -4.90048431719355"
```

```r
print(paste("p-value=", test_len.vs.dose23$p.value))
```

```
## [1] "p-value= 1.9064295136718e-05"
```

Reject the null hypothesis: dose level is correlated to tooth growth (restricted to dose level of 1 or 2).

```r
dose.5 <- subset (ToothGrowth, dose == 0.5)
dose1 <- subset (ToothGrowth, dose == 1.0)
dose2 <- subset (ToothGrowth, dose == 2.0)
```

```r
test_len.vs.supp_with0.5 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose.5)
print(paste("t-statistic=", test_len.vs.supp_with0.5$statistic))
```

```
## [1] "t-statistic= 3.16973278367081"
```

```r
print(paste("p-value=", test_len.vs.supp_with0.5$p.value))
```

```
## [1] "p-value= 0.0063586067640968"
```

Reject the null: supplement type, when dosage level equal to 0.5, is correlated to tooth growth.

```r
test_len.vs.supp_with1 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose1)
print(paste("t-statistic=", test_len.vs.supp_with1$statistic))
```

```
## [1] "t-statistic= 4.03276963371935"
```

```r
print(paste("p-value=", test_len.vs.supp_with1$p.value))
```

```
## [1] "p-value= 0.00103837587229988"
```

Reject the null: supplement type, when dosage level equal to 1, is correlated to tooth growth.

```r
test_len.vs.supp_with2 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose2)
print(paste("t-statistic=", test_len.vs.supp_with2$statistic))
```

```
## [1] "t-statistic= -0.0461361049092349"
```

```r
print(paste("p-value=", test_len.vs.supp_with2$p.value))
```

```
## [1] "p-value= 0.963851588723373"
```

Fail to reject the null hypothesis: supplement type, when dosage level equal to 2, is not correlated to tooth growth.

Conclusion

With the following hpothesis:

- The 60 guinea pigs observed were chosen at random and represent a population
- The supplement types and dosage levels were distributed equally and within a controlled environment

Initially, our exploratory data analysis graphically shows that delivery method was not a significant factor in tooth growth. Finally, the first assumption was proven partially false by hypothesis tests. The combined supplement method and dose levels of 0.5 and 1 are significant. However, at the supplement method and dose level of 2, it is not. Supplement delivery method at the 0.5 and 1 dose levels factors significantly impacts tooth growth, while at higher level, it does not. Regarding only dose levels, the exploratory data analysis and hypothesis tests concluded that dose levels are a significant factor in tooth growth: the more dose we provide, the higher growth teeth are.

Appendix

```r
ToothGrowth %>%
  ggplot(aes(x = supp, y = len)) +
  geom_boxplot(aes(fill = supp)) +
  ggtitle("Tooth Lenght according to Supplement") +
  ylab("tooth length") +
  theme(plot.title = element_text(hjust = 0.5))

ToothGrowth %>%
  ggplot(aes(x = factor(dose), y = len)) +
  geom_boxplot(aes(fill = factor(dose))) +
  ggtitle("Tooth Lenght according to Dose") +
```

```
  ylab("tooth length") +
  theme(plot.title = element_text(hjust = 0.5))

TоothGrowth %>%
  ggplot(aes(x = supp)) +
  geom_boxplot(aes(y = len, fill = supp)) +
  facet_wrap(~ dose) +
  ggtitle("Tooth Lenght according to Dosage & Supplement") +
  ylab("tooth length") +
  theme(plot.title = element_text(hjust = 0.5))
```