ToothGrowth Data Analysis

The dataset includes measurements of the tooth length of guinea pigs (len) which have been fed three different levels of Vitamin C, 0.5, 1 or 2 milligrams (dose) using two different delivery methods VC and OJ (supp). The research question is whether a different dose and/or a different delivery method of Vitamin C has an impact on tooth length.

Exploratory Data Analysis

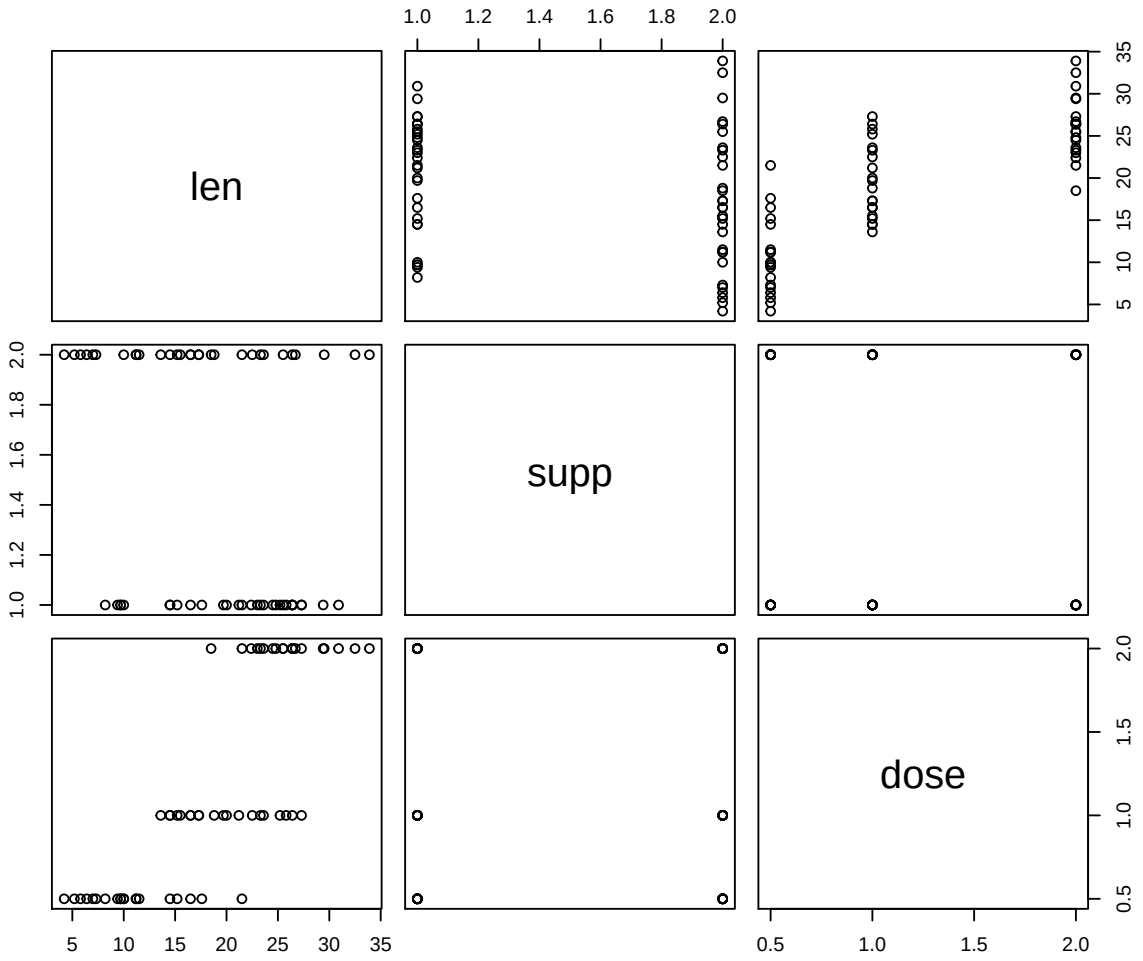
```
library(datasets)
library(dplyr)

options(scipen=100)
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
```

```
pairs(ToothGrowth)
```



The pairs plot indicates a positive linear relationship between tooth length and dose. On the other hand, the relationship between tooth length and delivery method is less obvious.

Data Summary

The following table shows the mean tooth length for every combination of dose and delivery method.

```
by_supp_dose <- group_by(ToothGrowth, supp, dose)

sum_tooth <- summarise(by_supp_dose, mean_len = mean(len))

sum_tooth
```

```
## Source: local data frame [6 x 3]
## Groups: supp
##
##   supp dose mean_len
## 1   OJ  0.5    13.23
```

```
## 2   OJ  1.0    22.70
## 3   OJ  2.0    26.06
## 4   VC  0.5     7.98
## 5   VC  1.0    16.77
## 6   VC  2.0    26.14
```

The different means lead to the same conclusion as the pairs plot. The dose appears to have a larger impact on the tooth length than the delivery method.

Hypothesis Tests

Delivery Method

- Hypothesis *H0*: The delivery method does not have an impact on tooth length. The means for both groups VC and OJ are equal.
- Hypothesis *H1*: The delivery method does have an impact on tooth length. The means are different.

```
t.supp <- t.test(len ~ supp, data=ToothGrowth, var.equal = T)
t.supp
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

At a 5% level of confidence we fail to reject *H0*, because the p-Value `0.0603934` exceeds 5%. The 95% confidence level `-0.1670064, 7.5670064` of the difference between the group means includes 0, hence no difference.

Dose

Since we have three different types of dose with different means, 0.5 mg (10.605), 1 mg (19.735) and 2 mg (26.1) we need to compare each group to each other. We conduct three test for the same question (influence of dose on tooth length), therefore we need to compensate for the multiple testing and expect the p-Value to be below ~0.017 (0.05/3) for a significance level of 5%.

- Hypothesis *H0.05.1*: The group means of tooth length of 0.5mg dose and 1mg dose are equal.
- Hypothesis *H0.05.2*: The group means of tooth length of 0.5mg dose and 2mg dose are equal.
- Hypothesis *H0.1.2*: The group means of tooth length of 1mg dose and 2mg dose are equal.
- Hypothesis *H1*: The respective group means are not equal.

```r
Tooth_05_1 <- subset(ToothGrowth, dose=="0.5"| dose=="1")
Tooth_05_2 <- subset(ToothGrowth, dose=="0.5"| dose=="2")
Tooth_1_2 <- subset(ToothGrowth, dose=="1"| dose=="2")

t.05.1 <- t.test(len ~ dose, data=Tooth_05_1, var.equal = T)
t.05.1
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 0.0000001266
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
t.05.2 <- t.test(len ~ dose, data=Tooth_05_2, var.equal = T)
t.05.2
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 38, p-value = 0.00000000000002838
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15352 -12.83648
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```r
t.1.2 <- t.test(len ~ dose, data=Tooth_1_2, var.equal = T)
t.1.2
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 0.00001811
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

All three p-Values are way below 0.017, so we can reject all three null hypothesis that two of the respective group means are equal.

Conclusion

The hypothesis tests indicate that there is an effect of the dose of Vitamin C on the tooth length of guinea pigs. There is also an effect of the delivery method on the tooth length, but it's not significant on a 5% level. Whether a p-Value of around 6% is enough to reject the null hypothesis ultimately depends on the context.

Assumptions of Analysis

- Normality Assumption
  Using the t - distribution assumes that the population is normally distributed. This makes sense for the size of body parts of mammals.
- Equal Variances
  The t-tests were conducted under the assumptions of equal variances. This makes sense for the tooth length of guinea pigs, unless we assume that guinea pigs react differently to delivery method and dose of Vitamin C.