Statistical Inference Course Project - Part 2

Background

The aim of this project is to analyse the `ToothGrowth` data in the R datasets package.

Extract from **R Help**:

This `ToothGrowth` dataset is used to analyse **The Effect of Vitamin C on Tooth Growth in Guinea Pigs**. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

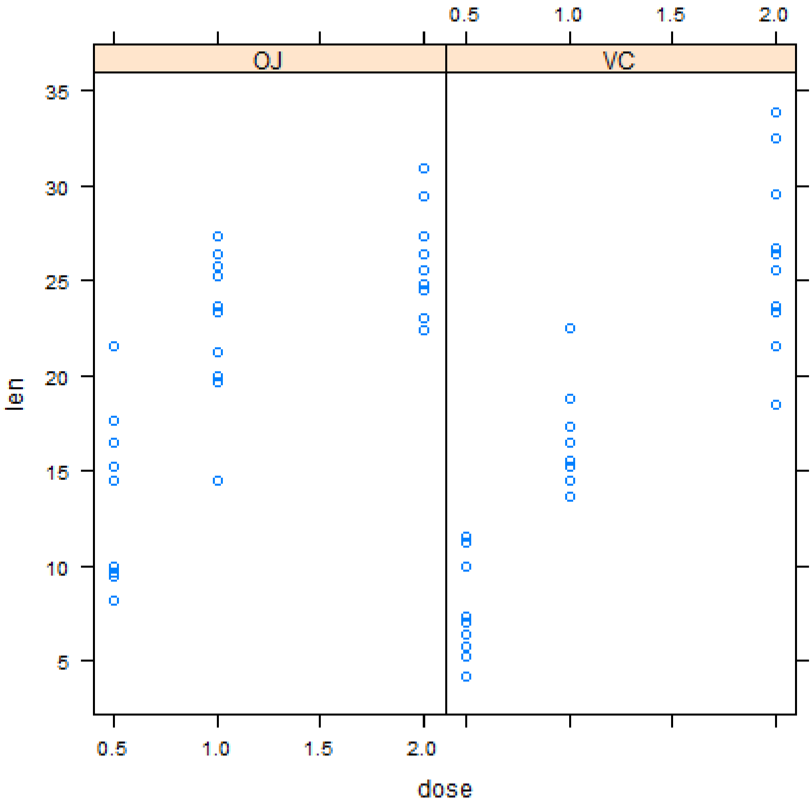
No.1 - Load the ToothGrowth data and perform some basic exploratory data analysis

```
data(ToothGrowth)
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Exploratory plot to highlight the basic features of the data:

```
library(lattice)
xyplot(len ~ dose | supp, data = ToothGrowth)
```



No.2 - Provide a basic summary of the data

We note the following points about the dataset:

- There are 60 observations (rows) and 3 variables (columns) in the data frame
- There is no missing value
- The `len` variable indicates the tooth length
- The `supp` variable indicates the supplement type: VC (for ascorbic acid) or OJ (for orange juice)
- The `dose` variable indicates the dose in milligrams: 0.5, 1.0 and 2.0
- There were 60 independent guinea pigs in total (i.e. 10 for each 'supplement-dose' combination group)

From the plot, we can make the following hypothesis:

- OJ has more positive effect on tooth growth compared to VC
- Higher dose level has more positive effect on tooth growth

No.3 - Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

We add a new variable to the dataset to indicate the 6 unique 'supplement-dose' combination groups:

```
library(dplyr)
my_df <- mutate(ToothGrowth, grouping = paste(supp, dose, sep = "_"))
head(my_df)
```

```
##    len supp dose grouping
## 1  4.2   VC  0.5   VC_0.5
## 2 11.5   VC  0.5   VC_0.5
## 3  7.3   VC  0.5   VC_0.5
## 4  5.8   VC  0.5   VC_0.5
## 5  6.4   VC  0.5   VC_0.5
## 6 10.0   VC  0.5   VC_0.5
```

To compare the effect of Vitamin C on tooth growth across the two delivery methods (VC and OJ) within the same dose level

Hence, we can perform t-Test for the following groups to compute the 95% confidence interval:

- (i) VC_0.5 against OJ_0.5
- (ii) VC_1 against OJ_1
- (iii) VC_2 against OJ_2

t-Test (i):

```
g1 <- my_df$len[my_df$grouping == "VC_0.5"]
g2 <- my_df$len[my_df$grouping == "OJ_0.5"]
t.test(g2, g1, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

t-Test (ii):

```
g3 <- my_df$len[my_df$grouping == "VC_1"]
g4 <- my_df$len[my_df$grouping == "OJ_1"]
t.test(g4, g3, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

t-Test (iii):

```
g5 <- my_df$len[my_df$grouping == "VC_2"]
g6 <- my_df$len[my_df$grouping == "OJ_2"]
t.test(g6, g5, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

To compare the effect of Vitamin C on tooth growth across the different dose levels within the same delivery method

Hence, we can perform t-Test for the following groups to compute the 95% confidence interval:

- (iv) VC_0.5 against VC_1
- (v) VC_1 against VC_2
- (vi) OJ_0.5 against OJ_1
- (vii) OJ_1 against OJ_2

t-Test (iv):

```
h1 <- my_df$len[my_df$grouping == "VC_0.5"]
h2 <- my_df$len[my_df$grouping == "VC_1"]
t.test(h2, h1, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  6.314288 11.265712
## attr(,"conf.level")
## [1] 0.95
```

t-Test (v):

```
h3 <- my_df$len[my_df$grouping == "VC_1"]
h4 <- my_df$len[my_df$grouping == "VC_2"]
t.test(h4, h3, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  5.685733 13.054267
## attr(,"conf.level")
## [1] 0.95
```

t-Test (vi):

```
h5 <- my_df$len[my_df$grouping == "OJ_0.5"]
h6 <- my_df$len[my_df$grouping == "OJ_1"]
t.test(h6, h5, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  5.524366 13.415634
## attr(,"conf.level")
## [1] 0.95
```

t-Test (vii):

```
h7 <- my_df$len[my_df$grouping == "OJ_1"]
h8 <- my_df$len[my_df$grouping == "OJ_2"]
t.test(h8, h7, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 0.1885575 6.5314425
## attr(,"conf.level")
## [1] 0.95
```

No.4 - State the conclusions and the assumptions needed for the conclusions

In all the t-Tests above, we made the following assumptions:

- The groups of guinea pigs are independent, hence we set `paired` = False in the argument.
- The variances are unequal, hence we set `var.equal` = False in the argument. (Note: This is the recommended assumption when in doubt)

Hypothesis - OJ has more positive effect on tooth growth compared to VC:

Based on the t-Test results from (i), (ii) and (iii), we can see that the 95% confidence interval is above zero for (i) and (ii), but not (iii). This suggests that OJ has a more positive effect on tooth growth compared to VC at dose level 0.5 and 1.0, but not at dose level 2.0.

Hypothesis - Higher dose level has more positive effect on tooth growth:

Based on the t-Test results from (iv) to (vi), we can see that the 95% confidence interval is above zero for all 4 cases. This suggests that higher dose level has more positive effect on tooth growth for both delivery methods.