Tooth Growth Analysis

Kent Mok

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```
library(datasets)  # Load datasets library
data(ToothGrowth)  # Load ToothGrowth data
str(ToothGrowth)  # Show structure of ToothGrowth data
```

The ToothGrowth data frame consists of 60 observations on 3 variables.

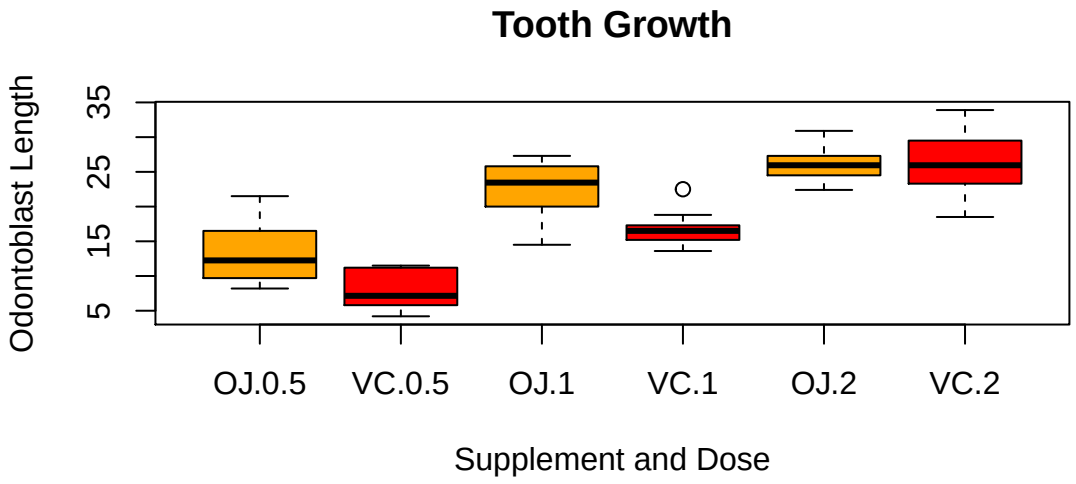
- The **len** variable is a numeric that appears to describe the length of the odontoblasts.
- The **supp** variable is a factor with two levels that indicates which supplement was used, OJ or VC.
- The **dose** variable is a numeric that appears to describe the dosage of the supplement used.

The first 30 observations of the dataframe consists of observations where the supplement VC was used. The latter 30 observations of the dataframe consists of observations where the supplement OJ was used.

The dataframe can be further split down by the dosage of the supplement used. For each supplement, the dosage observations appear to be 10 each at 0.5, 1.0, and 2.0. We can explore these six subsets by looking at the lengths of the odontoblasts based on the supplement used.

2. Provide a basic summary of the data.

```
boxplot(len ~ supp * dose, data = ToothGrowth,
        col = c("orange","red"), main = "Tooth Growth",
        xlab = "Supplement and Dose", ylab = "Odontoblast Length")
```



The mean values of the lengths at 0.5 and 1.0 mg of orange juice were higher than the lengths for the same dosages of ascorbic acid. At 2.0 mg, however, the mean of the lengths for the two supplements were close in value.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

The data of odontoblast length can be considered by grouping based on the supplement or by grouping based on the dosage.

Test by supplement Based on grouping by the supplement, it can be hypothesized that OJ yields higher length values than VC. The null hypothesis for this would be $H_o : \mu_{OJ} - \mu_{VC} > 0$. If the difference between the odontoblast lengths for supplements orange juice (OJ) and ascorbic acid (VC) is reliably greater than 0 (95% confidence), then the null hypothesis can be accepted and the conclusion is that orange juice creates a significant difference in odontoblast length over ascorbic acid.

```
t.test(len ~ supp, data = ToothGrowth)
```

The t-test result shows that the 95% confidence interval contains zero, so the null hypothesis is rejected. (Code output in Appendix.)

Test by dosage Based on grouping by the dosage, it can be hypothesized that dosages of 1.0 mg yield a higher length values than dosages of 0.5 mg. In addition, dosages of 2.0 mg yield higher length values than dosages of 1.0 mg as well as dosages of 0.5 mg. The null hypotheses for these would be

- $H_o : \mu_{1.0mg} - \mu_{0.5mg} > 0$
- $H_o : \mu_{2.0mg} - \mu_{1.0mg} > 0$
- $H_o : \mu_{2.0mg} - \mu_{0.5mg} > 0$

```
t.test(ToothGrowth$len[ToothGrowth$dose == 1],
       ToothGrowth$len[ToothGrowth$dose == 0.5])
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{1.0mg} - \mu_{0.5mg} > 0$ is accepted. (Code output in Appendix.)

```
t.test(ToothGrowth$len[ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$dose == 1])
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{2.0mg} - \mu_{1.0mg} > 0$ is accepted. (Code output in Appendix.)

```
t.test(ToothGrowth$len[ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$dose == 0.5])
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{2.0mg} - \mu_{0.5mg} > 0$ is accepted. (Code output in Appendix.)

Test by supplement and dosage The following t-tests analyze the length data based on the supplements for a given dosage level.

At dosage levels of 0.5 mg, it can be hypothesized that orange juice yields a higher odontoblast length than ascorbic acid or: $H_o : \mu_{OJ@0.5mg} - \mu_{VC@0.5mg} > 0$

```
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5],
       ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5])
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis is accepted. (Code output in Appendix.)

At dosage levels of 1.0 mg, it can be hypothesized that orange juice yields a higher odontoblast length than ascorbic acid or: $H_o : \mu_{OJ@1.0mg} - \mu_{VC@1.0mg} > 0$

```
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1],
       ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1])
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis is accepted. (Code output in Appendix.)

At dosage levels of 2.0 mg, it can be hypothesized that orange juice yields a higher odontoblast length than ascorbic acid or: $H_o : \mu_{OJ@2.0mg} - \mu_{VC@2.0mg} > 0$

```
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2])
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis is accepted. (Code output in Appendix.)

4. State your conclusions and the assumptions needed for your conclusions.

Conclusions

- When the t-test is performed on the length vs supplement only, the null hypothesis $H_o : \mu_{OJ} - \mu_{VC} > 0$ is rejected. It can be concluded that there is no significant difference in odontoblast lengths across supplement types.
- When the t-test is performed on the length vs dosage only, the null hypotheses for all three dosages were accepted. It can be concluded that higher dosage levels have a significant effect on odontoblast lengths.
- When the t-test is performed on the length vs supplement with the dosages isolated, the null hypotheses at dosage levels of 0.5 mg and 1.0 mg were accepted but the null hypothesis at dosage level of 2.0 mg was rejected. The conclusion is that at the lower dosages of 0.5 mg and 1.0 mg, the orange juice had a significantly higher effect on the odontoblast length than the ascorbic acid, but the significance is lost at the dosage level of 2.0 mg.

Assumptions

- For this analysis, it must be assumed that the samples were selected randomly from the population.
- It must also be assumed that the population follows a normal distribution.
- It must also be assumed that the samples unequal variances.

Appendix

Background Information

The purpose of this experiment is to analyze the ToothGrowth data in the R datasets package. The data contains the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs, each receiving one of three dose levels of Vitamin C (0.5, 1.0, and 2.0 mg) with one of two delivery methods (orange juice or an aqueous solution of ascorbic acid).

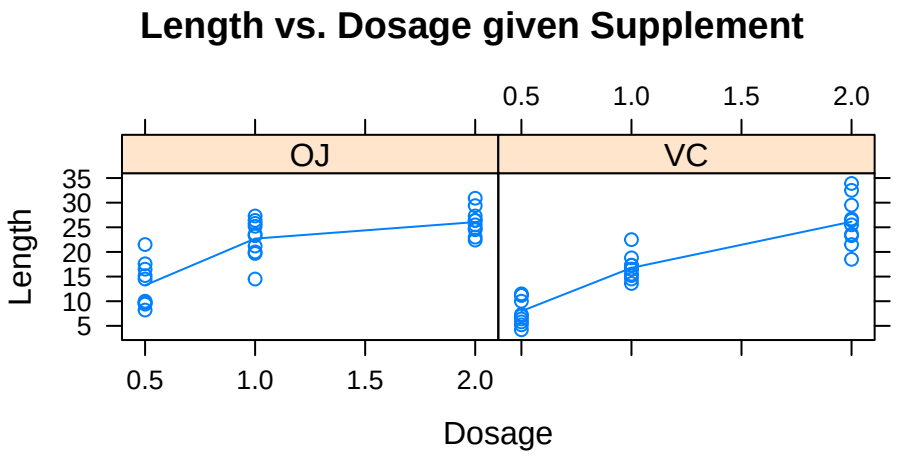
Data Loading

```
## '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 ...
```

Basic Data Summary

A quick scatter plot of the data separated by supplement shows that in general, the average length of odontoblasts were longer for orange juice as a supplement versus ascorbic acid as a supplement.

```
library(lattice)
xyplot(len ~ dose | supp,  # Plot length vs dosage given supplement
       data = ToothGrowth, xlab = "Dosage", ylab = "Length",
       main = "Length vs. Dosage given Supplement",
       type = c("p", "a"))  # Adds line through mean of lengths based on dosage
```



T-test Evaluation - Test by supplement

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

T-test Evaluation - Test by dosage

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{1.0mg} - \mu_{0.5mg} > 0$ is accepted.

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{2.0mg} - \mu_{1.0mg} > 0$ is accepted.

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83383 18.15617
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{2.0mg} - \mu_{0.5mg} > 0$ is accepted.

T-test Evaluation - Test by supplement and dosage

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose ==  and ToothGrowth$len[ToothGrow
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{OJ@0.5mg} - \mu_{VC@0.5mg} > 0$ is accepted.

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose ==  and ToothGrowth$len[ToothGrow
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{OJ@1.0mg} - \mu_{VC@1.0mg} > 0$ is accepted.

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose ==  and ToothGrowth$len[ToothGrow
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

The t-test result shows that the entire 95% confidence interval is greater than zero, so the null hypothesis $H_o : \mu_{OJ@2.0mg} - \mu_{VC@2.0mg} > 0$ is accepted.