Programming Assigment 1-2

Analysis of the Effect of Vitamin C on Tooth Growth

Author: Zdravka Cankova

Data: ToothGrowth dataset built into R

Source: C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

Overview

This project makes use of the ToothGrowth dataset that comes with R. This dataset contains the results of a study that examined the effect of vitamin C on guinea pig tooth growth. The independent variables are the dose of vitamin C (0.5, 1, and 2 mg) and the delivery method (orange juice or vitamin C supplement). The dependent variable is the tooth length. The purpose of this project is to provide a basic summary of the data, as well as to determine if there are any changes in tooth growth based on delivery method or dose. Our null hypothesis is that there are no differences in tooth growth as a result of delivery method or dose. The alternative hypothesis is that there are changes in tooth growth associated with supplement and dose (two-sided test).

Data Summary

First, we look at a summary of the data:

```
ToothGrowth <- 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 ...
```

There are three columns:
* len: tooth length, numeric format
* supp: delivery method, factor with levels OJ (orange juice) and VC (vitamin C)
* dose: dose, numeric format

Even though the dose is given in numeric format, it's not a continuous variable, since only 3 types of doses were given out. It would be more appropriate to convert it to a factor with levels 0.5, 1 and 2 mg:

```
ToothGrowth$dose <- factor(ToothGrowth$dose, levels = c("0.5", "1", "2"))
```

Next we can summarize the average tooth length by dose and delivery method in table format:

```
xtabs(len ~ supp + dose, data = ToothGrowth)
```

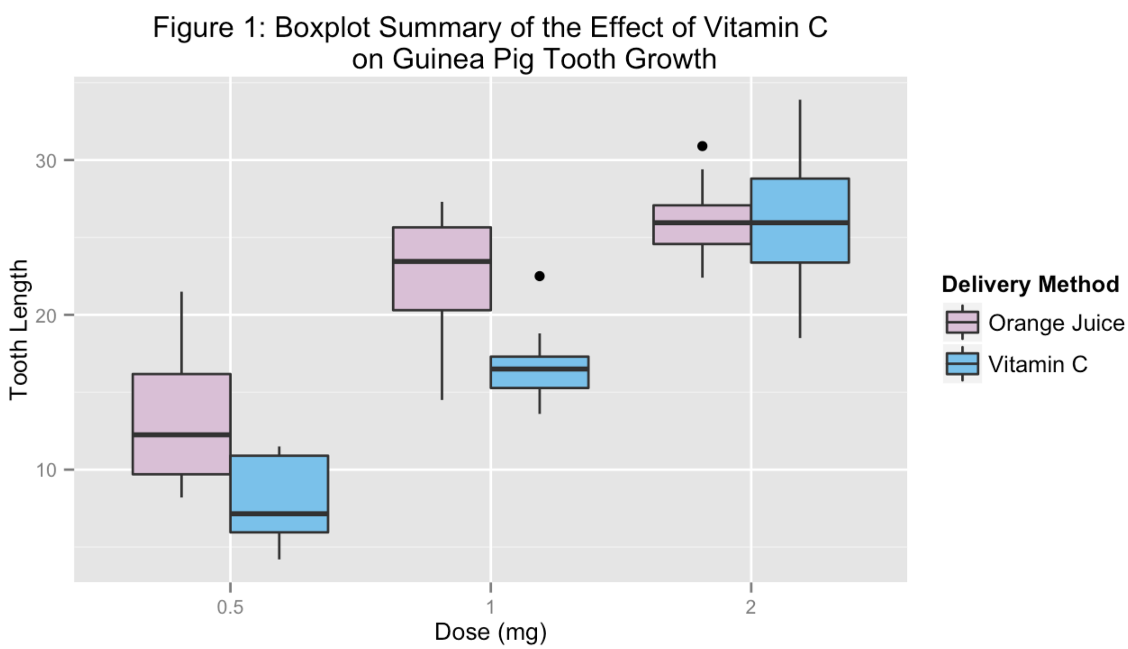
```
##     dose
## supp   0.5     1     2
##   OJ 132.3 227.0 260.6
##   VC  79.8 167.7 261.4
```

After looking at the table, we can see that the average tooth length appears to increase with increasing doses of vitamin C. Also, it seems that the average tooth length is higher when supplementing with orange juice for the two lower doses, but at the highest dose (2mg), the tooth length is the same for both delivery methods.

Finally, we can construct boxplots of tooth length as a function of vitamin C dose for each delivery method:

```
library(ggplot2)

boxplot <- ggplot(ToothGrowth, aes(x = dose, y = len, fill = supp)) +
      geom_boxplot() +
      labs(x = "Dose (mg)",
           y = "Tooth Length",
           title = "Figure 1: Boxplot Summary of the Effect of Vitamin C
           on Guinea Pig Tooth Growth") +
      scale_fill_manual(values=c("thistle", "skyblue2"),
                        name="Delivery Method",
                        labels=c("Orange Juice", "Vitamin C")) +
      theme(axis.text = element_text(size = 8),
            axis.title = element_text(size = 10),
            plot.title = element_text(size = 12))
boxplot
```



The boxplot also confirms the trends seen in the summary table:
1. Tooth length appears to increase with increasing vitamin C dosage.
2. Supplementing with orange juice seems to result in longer teeth for the two lower doses (0.5 and 1 mg), but at the highest dose (2 mg) tooth length seems the same for both delivery methods.

Statistical Analysis

Next, we will use hypothesis testing to detemine whether these trends are statistically significant at a 95% confidence level (alpha = 0.05). We will assume that each animal is receiving only one treatment, so we can perform t-test to compare the means of two independent samples. The null and alternative hypotheses are as follows:

H_0: mu1 = mu2
H_A: mu1 != mu2

We can perform several tests on groups of two from the data to determine if any of the tends in the boxplot are significant. First, let us subset the data:

```
OJ0.5 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == "0.5"]
OJ1 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == "1"]
OJ2 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == "2"]
VC0.5 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == "0.5"]
VC1 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == "1"]
VC2 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == "2"]
length <- data.frame(cbind(OJ0.5, OJ1, OJ2, VC0.5, VC1, VC2))
```

Now, we can do the t-test for each two independent groups:

Changes with Increasing Dose

- **Orange juice: 0.5 vs. 1 mg**

```
t.test(length$OJ1 - length$OJ0.5)
```

```
##
##  One Sample t-test
##
## data:  length$OJ1 - length$OJ0.5
## t = 4.1635, df = 9, p-value = 0.002435
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   4.324616 14.615384
## sample estimates:
## mean of x
##      9.47
```

- **Orange juice: 0.5 vs. 2 mg**

```
t.test(length$OJ2 - length$OJ0.5)
```

```
##
##  One Sample t-test
##
## data:  length$OJ2 - length$OJ0.5
## t = 7.4919, df = 9, p-value = 3.724e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   8.956042 16.703958
## sample estimates:
## mean of x
##     12.83
```

- **Orange juice: 1 vs. 2 mg**

```
t.test(length$OJ2 - length$OJ1)
```

```
##
##  One Sample t-test
##
## data:  length$OJ2 - length$OJ1
## t = 1.9435, df = 9, p-value = 0.08384
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.5509376  7.2709376
## sample estimates:
## mean of x
##      3.36
```

- **Vitamin C: 0.5 vs. 1 mg**

```
t.test(length$VC1 - length$VC0.5)
```

```
##
##  One Sample t-test
##
## data:  length$VC1 - length$VC0.5
## t = 6.1364, df = 9, p-value = 0.0001715
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   5.549601 12.030399
## sample estimates:
## mean of x
##      8.79
```

- **Vitamin C: 0.5 vs. 2 mg**

```
t.test(length$VC2 - length$VC0.5)
```

```
##
##  One Sample t-test
##
## data:  length$VC2 - length$VC0.5
## t = 9.7912, df = 9, p-value = 4.264e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  13.9643 22.3557
## sample estimates:
## mean of x
##     18.16
```

- **Vitamin C: 1 vs. 2 mg**

```
t.test(length$VC2 - length$VC1)
```

```
##
##  One Sample t-test
##
## data:  length$VC2 - length$VC1
## t = 5.346, df = 9, p-value = 0.0004648
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   5.405082 13.334918
## sample estimates:
## mean of x
##      9.37
```

To summarize, tooth length significantly increases with increasing vitamin C dosage for both delivery methods. For orange juice delivery, there were significant differences beween all groups ($p < 0.05$), except 1 mg vs. 2 mg ($p = 0.08384$). For the vitamin C supplement, all groups were significantly different ($p < 0.05$).

Changes with Delivery Method

- **0.5 mg: Orange juice vs. vitamin C**

```
t.test(length$OJ0.5 - length$VC0.5)
```

```
##
##  One Sample t-test
##
## data:  length$OJ0.5 - length$VC0.5
## t = 2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.263458 9.236542
## sample estimates:
## mean of x
##      5.25
```

- **1 mg: Orange juice vs. vitamin C**

```
t.test(length$OJ1 - length$VC1)
```

```
##
##  One Sample t-test
##
## data:  length$OJ1 - length$VC1
## t = 3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.951911 9.908089
## sample estimates:
## mean of x
##      5.93
```

- **2 mg: Orange juice vs. vitamin C**

```
t.test(length$OJ2 - length$VC2)
```

```
##
##  One Sample t-test
##
## data:  length$OJ2 - length$VC2
## t = -0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.328976  4.168976
## sample estimates:
## mean of x
##     -0.08
```

As expected based on the boxplots, the orange juice supplement resulted in significantly longer teeth at dosages of 0.5 and 1 mg ($p < 0.05$). However, at the highest dosage of 2 mg, tooth length was about the same for both orange juice and vitamin C supplements ($p = 0.967$).

Conclusions

After performing two-sided t-tests with aplha = 0.05, we can conclude that:

1. Tooth length increased significantly with increasing vitamin C dosage, regardless of delivery method.
2. Supplementing with orange juice resulted in significantly longer teeth for the two lower doses (0.5 and 1 mg), when compared to the vitamin C supplement. However, at the highest dose (2 mg), there were no significant differences in tooth length between the delivery methods.

These results are based on the assumption that the distribution of average lengths is approximately normal. However, since the sample size is somewhat small (10 animals per treatment per dose), this may not be a reasonable assumption. Increasing the number of animals would help improve the significance of these resuls. Also, we've assumed that the groups are independent, which is not quite the case, since some of them are receiving either the same dose or use the same delivery method. Doing multiple comparisons instead of independent t-tests might also improve the validity of these results.