Exploration of Tooth Growth Data

Matt Dancho

June 9, 2016

Overview

This report analyzes the effect of Vitamin C on tooth growth in guinea pigs. The data used comes from the `ToothGrowth` data set in the `R datasets` package. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice (OJ) or ascorbic acid, a form of vitamin C (VC). The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. The analysis contains an exploratory evaluation and a statistical analysis of the effect of supplement and dose level on tooth growth.

Setup

First, the needed libraries and data are loaded.

```
# Load libraries
library(ggplot2)        # Used for plotting
library(pander)         # Used for table formatting

# Load data
data("ToothGrowth")
tg <- ToothGrowth
```

Exploratory Analysis

Next, an exploratory analysis is performed. The `ToothGrowth` data set contains three features:

1. **len:** Tooth length
2. **supp:** Supplement type (VC or OJ).
3. **dose:** Dose in milligrams/day

A summary of the data is shown in **Table 1**.

```
pander(summary(tg))
```

Table 1: Summary of Data

len	supp	dose
Min. : 4.20	OJ:30	Min. :0.500
1st Qu.:13.07	VC:30	1st Qu.:0.500
Median :19.25	NA	Median :1.000
Mean :18.81	NA	Mean :1.167
3rd Qu.:25.27	NA	3rd Qu.:2.000
Max. :33.90	NA	Max. :2.000

As shown in **Table 2**, a total of ten doses were given by each dosage level and supplement type.

```
pander(table(tg$dose, tg$supp))
```

Table 2: Number of Doses Given by Dose Level and Supplement Type

	OJ	VC
0.5	10	10
1	10	10
2	10	10

Figure 1 shows a box plot of dosage versus length by supplement type. Dose levels appear to have a positive relationship with tooth length. The relationship between supplement type and tooth length is more difficult to determine.

```
ggplot(data=tg, aes(x=factor(dose), y=len)) +
        geom_boxplot(aes(fill=supp)) +
        facet_grid(.~supp) +
        ylab("Length") +
        xlab("Dosage") +
        ggtitle("Dosage versus Tooth Length by Supplement Type")
```

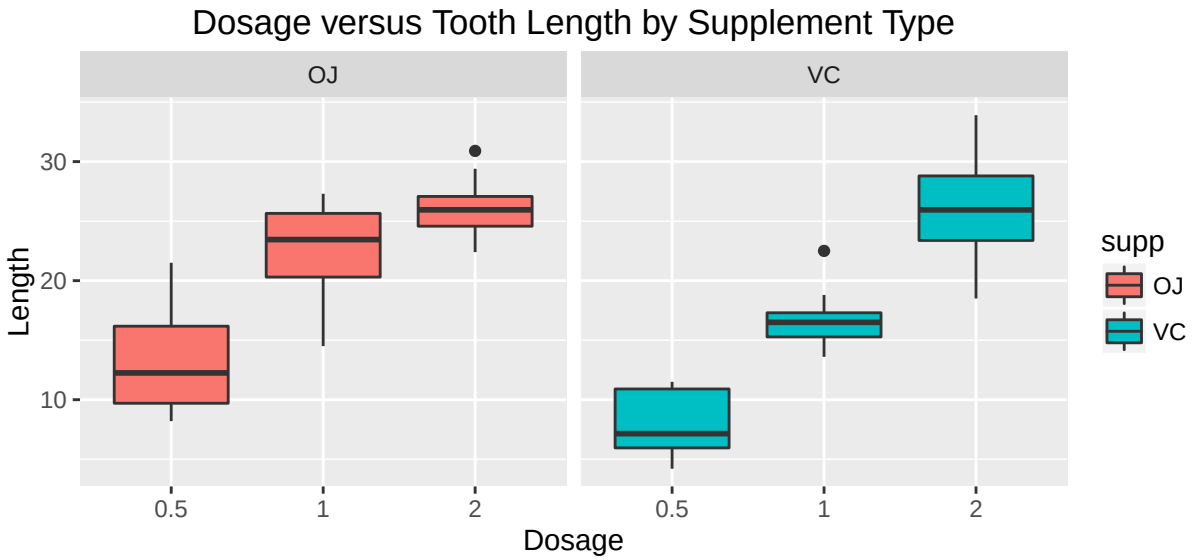

Figure 1: Dosage versus Tooth Length by Supplement Type

Statistical Analysis of Tooth Growth by Supplement and Dose

From the exploratory analysis, there are possibly relationships between:

1. Tooth length and supplement type
2. Tooth length and dosage

Two-sample t-tests are performed to determine if the relationships are statistically significant.

Assumptions

The underlying assumptions are:

1. The samples were selected at random
2. The samples are representative of the entire population
3. The distribution of the means is normal
4. The populations have unequal variance (conservative assumption)
5. 95% confidence is needed to reject the hypothesis

Alternative Hypothesis 1: Supplement type has an effect on tooth length

A t-test is performed on data subset by supplement type with an alternative hypothesis that the supplement type has an effect on tooth length. The t-test is set to compare supplement type VC to OJ. The result is negative indicating VC typically results in a lower tooth length than OC. However, from the t-test results, we can see that the 95% confidence interval contains zero indicating that we should not reject the null hypothesis at a 95% confidence interval. Therefore, we cannot conclude that supplement type has an effect on tooth length.

```
supp_vc <- subset(tg, supp=="VC")
supp_oj <- subset(tg, supp=="OJ")
t.test(supp_vc$len, supp_oj$len, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  supp_vc$len and supp_oj$len
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Alternative Hypothesis 2: Dose level has an effect on tooth length

The alternative hypothesis that the dose level has an effect on tooth length. The t-test is set to compare supplement each dose level. Three dosage levels are present, which results in a total of three t-tests to compare each dosage:

1. Dose level 0.5 to 1.0
2. Dose level 0.5 to 2.0
3. Dose level 1.0 to 2.0

The data is subset to perform the t-tests.

```
# Subset data by dose level
dose_0.5 <- subset(tg, dose==0.5)
dose_1.0 <- subset(tg, dose==1.0)
dose_2.0 <- subset(tg, dose==2.0)
```

Compare Dose Level 0.5 to 1.0

The first t-test concludes the 95% confidence interval is greater than zero indicating the null hypothesis that dose level has no effect can be rejected. Therefore, a dose level of 1.0 has a positive effect on tooth length versus a dose level of 0.5.

```
# Compare dose level 0.5 to 1.0
t.test(dose_1.0$len, dose_0.5$len, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_1.0$len and dose_0.5$len
## 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
```

Compare Dose Level 0.5 to 2.0

The second t-test concludes the 95% confidence interval is greater than zero indicating the null hypothesis that dose level has no effect can be rejected. Therefore, a dose level of 2.0 has a positive effect on tooth length versus a dose level of 0.5.

```
# Compare dose level 0.5 to 2.0
t.test(dose_2.0$len, dose_0.5$len, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_2.0$len and dose_0.5$len
## 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
```

Compare Dose Level 1.0 to 2.0

The third t-test concludes the 95% confidence interval is greater than zero indicating the null hypothesis that dose level has no effect can be rejected. Therefore, a dose level of 2.0 has a positive effect on tooth length versus a dose level of 1.0.

```
# Compare dose level 1.0 to 2.0
t.test(dose_2.0$len, dose_1.0$len, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_2.0$len and dose_1.0$len
## 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
```

Conclusions

1. Supplement type was not proven with 95% confidence to have an effect on tooth length.
2. Dose level was proven with 95% confidence to have an effect on tooth length. All three comparisons of dose level concluded that tooth length increased with increasing dose level.