Inferential Analysis of Tooth Growth Data

Hamed Bastan-Hagh

22 July 2016

Synopsis

This is an analysis of the ToothGrowth data in R. This records the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. 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, coded as OJ; or ascorbic acid, a form of vitamin C, coded as VC.

This analysis concludes that there is a statistically significant effect of dose amount on tooth growth, but no such effect for supplement type.

1. Load Data and Perform Basic Exploratory Data Analysis

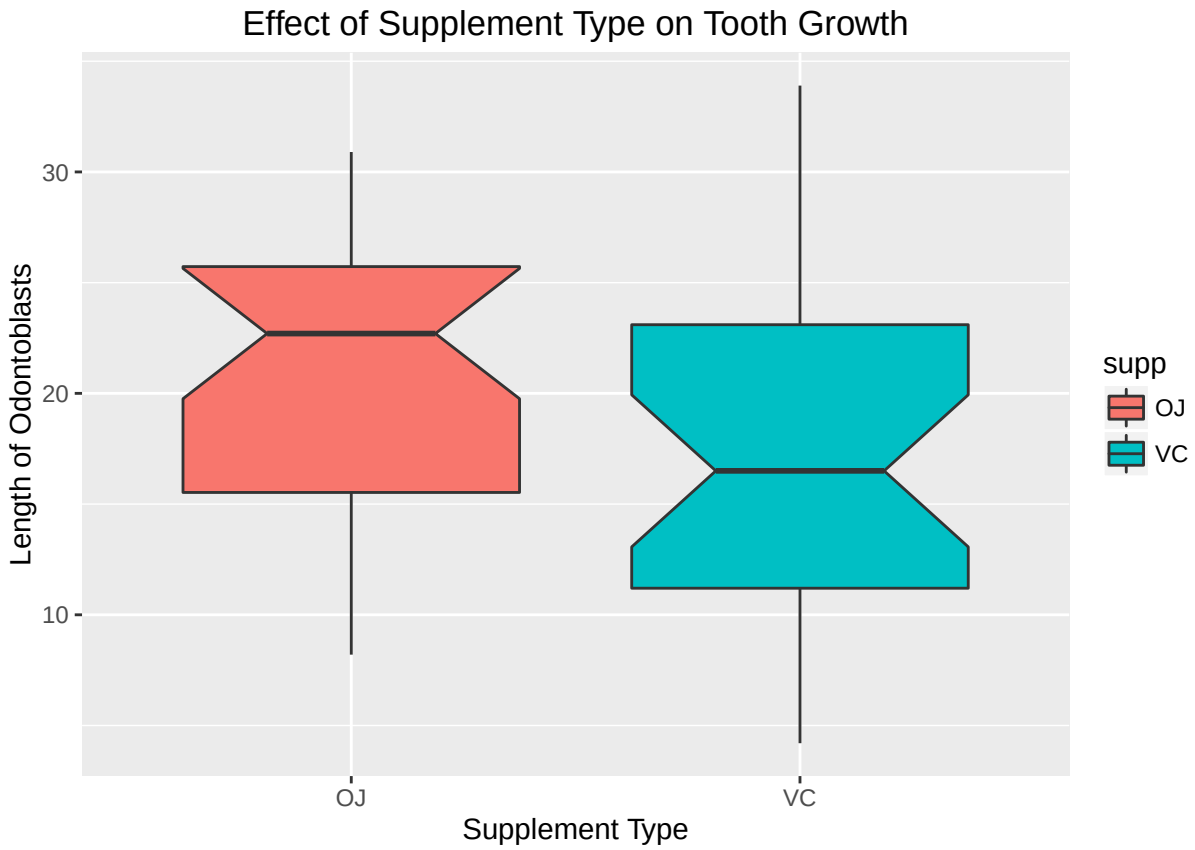
```
# Load required libraries
library(datasets)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
# Assign ToothGrowth data to variable df in 'tibble' form for easier printing
df <- tbl_df(ToothGrowth)
```

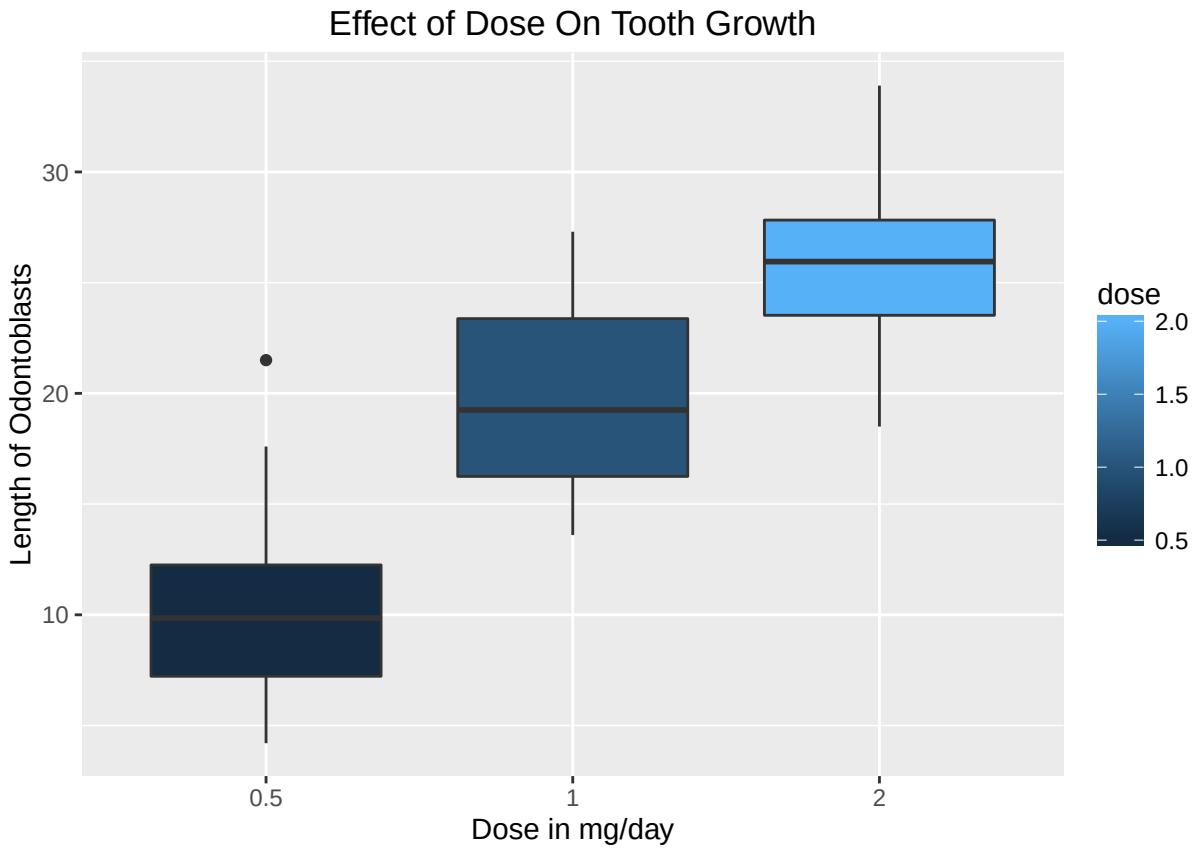
With the data loaded we will do some basic exploratory analysis.

```
# Boxplot of the values separated by supplement type
ggplot(df, aes(supp, len)) +
    geom_boxplot(aes(fill = supp), notch = TRUE) +
    labs(title = "Effect of Supplement Type on Tooth Growth",
         x = "Supplement Type",
         y = "Length of Odontoblasts")
```



The notches for the two groups do not overlap, which is evidence that the median value for Orange Juice is higher than Ascorbic Acid. We will investigate this further in section 3.

```
# Boxplot of the values separated by dose amount
ggplot(df, aes(factor(dose), len)) +
    geom_boxplot(aes(fill = dose)) +
    labs(title = "Effect of Dose On Tooth Growth",
         x = "Dose in mg/day",
         y = "Length of Odontoblasts")
```



This indicates a positive correlation between dose amount and length, which we will investigate further in section 3.

2. Summary of the Data

Here are some basic summaries for the data set.

```
#Print put basic summary of our data
summary(df)
```

```
##       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
```

```
# Print out levels for supplement type 'supp'
levels(df$supp)
```

```
## [1] "OJ" "VC"
```

```r
# Print out unique values for dose amount 'dose'
unique(df$dose)
```

```
## [1] 0.5 1.0 2.0
```

Now we can see a summary of the data grouped by dose amount.

```r
df %>%
    group_by(dose) %>%
        summarise(len = mean(len))
```

```
## # A tibble: 3 x 2
##    dose    len
##   <dbl>  <dbl>
## 1   0.5 10.605
## 2   1.0 19.735
## 3   2.0 26.100
```

And finally a summary of the data grouped by supplement type.

```r
df %>%
    group_by(supp) %>%
        summarise(len = mean(len))
```

```
## # A tibble: 2 x 2
##     supp      len
##   <fctr>    <dbl>
## 1     OJ 20.66333
## 2     VC 16.96333
```

3. Compare tooth growth by "supp" and "dose"

3a. Supplement Type

Firstly we will explore the effect of supplement type on tooth growth with a t-test. Our null hypothesis is:

H_0: mu = mu_0

The alternative hypothesis:

H_a: mu != mu_0

```r
# Use function t.test on the dataset
t.test(len ~ supp, df)
```

```
##
##  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
```

Our p-value is 0.06, which is slightly higher than our threshold for 95% confidence. Therefore we **fail to reject the null hypothesis**, and cannot conclude that supplement type has an effect on tooth growth.

3b. Dose Amount

Now we will test the effect of dose amount on tooth growth in our data. We will test the ToothGrowth data in three sections based on dose amount: testing 0.5 against 1.0; 0.5 against 2.0; and 1.0 against 2.0. For each of these our null hypothesis is that:

`H_0: mu = mu_0`

Alternative hypothesis:

`H_a: mu != mu_0`

```
# Testing 0.5 against 1.0 by removing those rows where dose = 2.0
t.test(len ~ dose, df[df$dose != 2, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
# Testing 0.5 against 2.0 by removing those rows where dose = 1.0
t.test(len ~ dose, df[df$dose != 1, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
# Testing 1.0 against 2.0 by removing those rows where dose = 0.5
t.test(len ~ dose, df[df$dose != 0.5, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

P-values for each of these tests is low enough that we can **reject the null hypothesis**. We can conclude that dose amount does have an impact on tooth growth.

4. Conclusions

Based on our analysis we can conclude that:

- Supplement type does not have a statistically significant effect on tooth growth in our data.
- Dose amount does have a statistically significant effect on tooth growth in our data.