Tooth Growth Basic Inferential Analysis

Marcus Vinicius Monteiro de Souza

June 19, 2017

Overview

This work explores the ToothGrowth dataset from the Base R datasets package to find relationships, if there are any, between tooth growth and vitamin C consumption. Specifically, we'll try to answer the following questions:

1. Is there a difference between ascorbic acid and orange juice supplements concerning tooth growth?
2. Is tooth growth affected by vitamin C dosage?

A first look at the data

The ToothGrowth dataset description reads:

> The response is 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 or ascorbic acid (a form of vitamin C and coded as VC).

Let's load the required packages:

```
if (!require('pacman')) {
  install.packages('pacman')
}
```

```
## Loading required package: pacman
```

```
pacman::p_load(lsr, tidyverse)
```

and the data:

```
data('ToothGrowth')

ToothGrowth <- as_tibble(ToothGrowth) %>%
  mutate(dose = as.factor(dose))
```

Now, let's have a look at the data:

```
knitr::kable(ToothGrowth)
```

len	supp	dose
4.2	VC	0.5
11.5	VC	0.5
7.3	VC	0.5
5.8	VC	0.5
6.4	VC	0.5
10.0	VC	0.5
11.2	VC	0.5
11.2	VC	0.5
5.2	VC	0.5
7.0	VC	0.5
16.5	VC	1

len	supp	dose
16.5	VC	1
15.2	VC	1
17.3	VC	1
22.5	VC	1
17.3	VC	1
13.6	VC	1
14.5	VC	1
18.8	VC	1
15.5	VC	1
23.6	VC	2
18.5	VC	2
33.9	VC	2
25.5	VC	2
26.4	VC	2
32.5	VC	2
26.7	VC	2
21.5	VC	2
23.3	VC	2
29.5	VC	2
15.2	OJ	0.5
21.5	OJ	0.5
17.6	OJ	0.5
9.7	OJ	0.5
14.5	OJ	0.5
10.0	OJ	0.5
8.2	OJ	0.5
9.4	OJ	0.5
16.5	OJ	0.5
9.7	OJ	0.5
19.7	OJ	1
23.3	OJ	1
23.6	OJ	1
26.4	OJ	1
20.0	OJ	1
25.2	OJ	1
25.8	OJ	1
21.2	OJ	1
14.5	OJ	1
27.3	OJ	1
25.5	OJ	2
26.4	OJ	2
22.4	OJ	2
24.5	OJ	2
24.8	OJ	2
30.9	OJ	2
26.4	OJ	2
27.3	OJ	2
29.4	OJ	2
23.0	OJ	2

See a compact display of what our R object looks like:

```
str(ToothGrowth)
```

```
## Classes 'tbl_df', 'tbl' and '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

and a summary of some of the data's statistics:

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

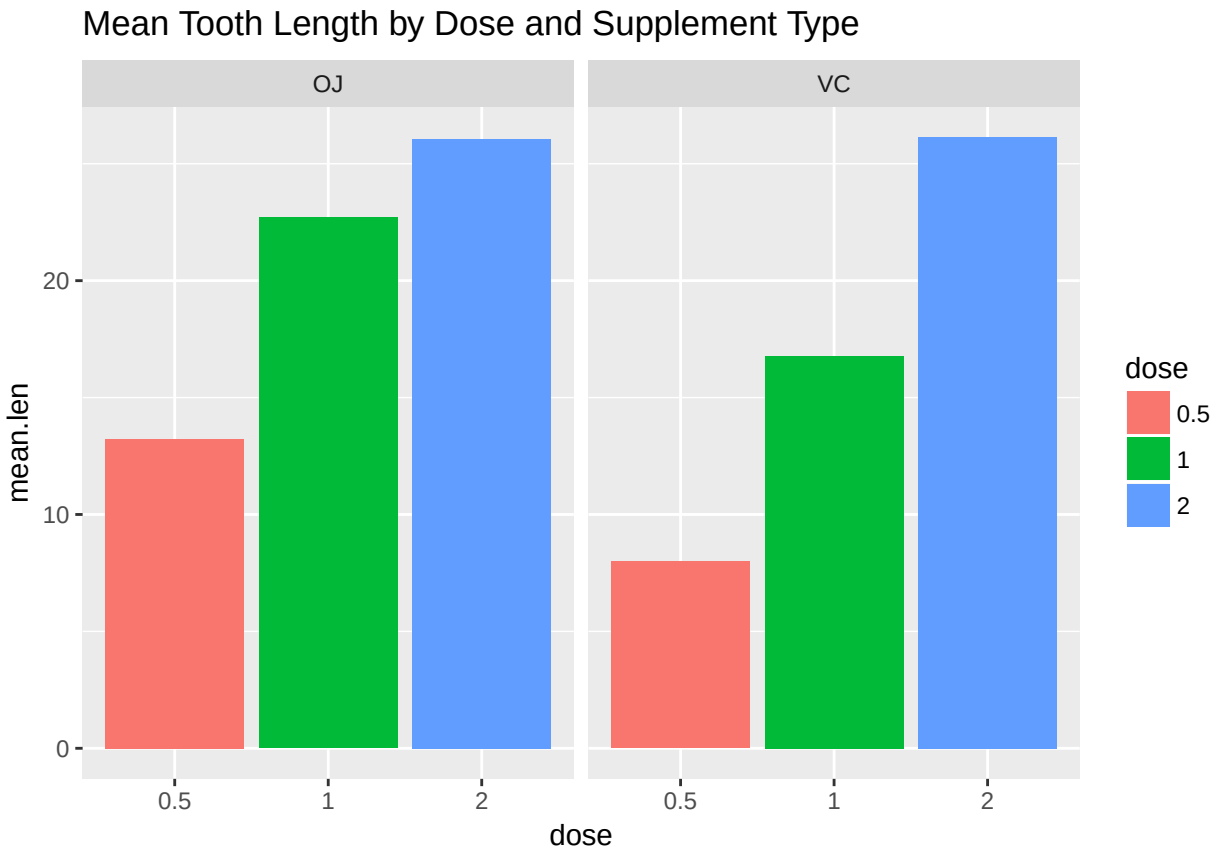
To clarify the meaning of the variables:

Variable Name	Description
len	Tooth length
supp	Supplement type: VC (ascorbic acid) or OJ (orange juice)
dose	Dose in milligrams/day

Is there a difference between ascorbic acid and orange juice supplements concerning tooth growth?

If there was no difference between ascorbic acid and orange juice supplements concerning tooth growth, we would expect that the average tooth length would be the same, plus or minus chance variation. Let's have another look at the data:

```
ToothGrowth %>%
  group_by(supp, dose) %>%
  summarise(mean.len = mean(len)) %>%
  ggplot(aes(dose, mean.len)) +
  geom_col(aes(fill = dose)) +
  facet_grid(. ~ supp) +
  ggtitle('Mean Tooth Length by Dose and Supplement Type')
```



It shows that the mean tooth length is higher for orange juice at doses 0.5 and 1, and equal at dose 2. But is that difference significant? Or it can be better explained by chance?

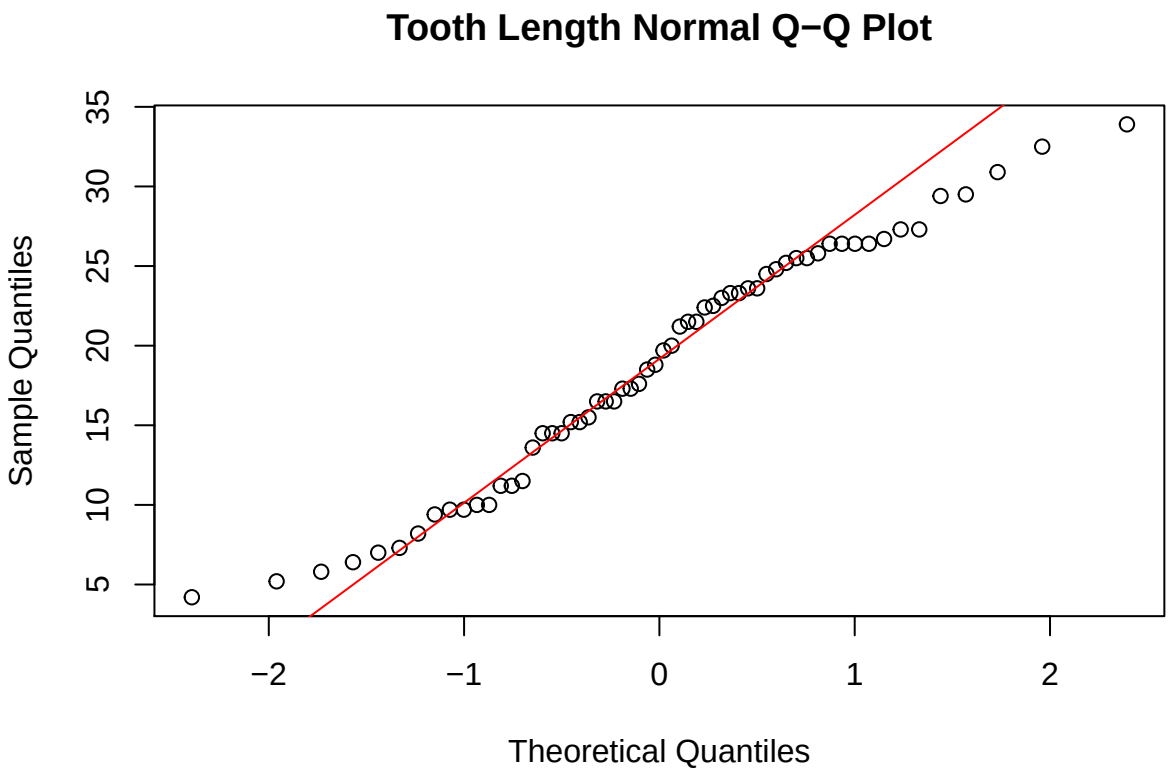
Alternative Hypothesis Test: Orange Juice Supplements Lead to More Tooth Growth Than Ascorbic Acid Ones

To compare the two means statistically, we can perform a t-test, providing the following assumptions about data hold true see:

1. Normality.

First, let's check if the sample can be considered as drawn from a normal distribution:

```
qqnorm(ToothGrowth$len, main = 'Tooth Length Normal Q-Q Plot')
qqline(ToothGrowth$len, col = 2)
```



The majority of the sample's observations fit the theoretical normal line.

2. Independence

The observations were drawn independently from 60 guinea pigs.

Also very importantly, we'll do an unpaired, or independent samples test, because there is no good reason to assume that the guinea pigs were somehow paired, much less that the same pigs were tested with both supplements.

So, finally, let's do the test:

```
independentSamplesTTest(len ~ supp, as.data.frame(ToothGrowth))
```

```
##
##    Welch's independent samples t-test
##
## Outcome variable:   len
## Grouping variable:  supp
##
## Descriptive statistics:
##                 OJ     VC
##    mean     20.663 16.963
##    std dev.  6.606  8.266
##
## Hypotheses:
##    null:        population means equal for both groups
##    alternative: different population means in each group
##
```

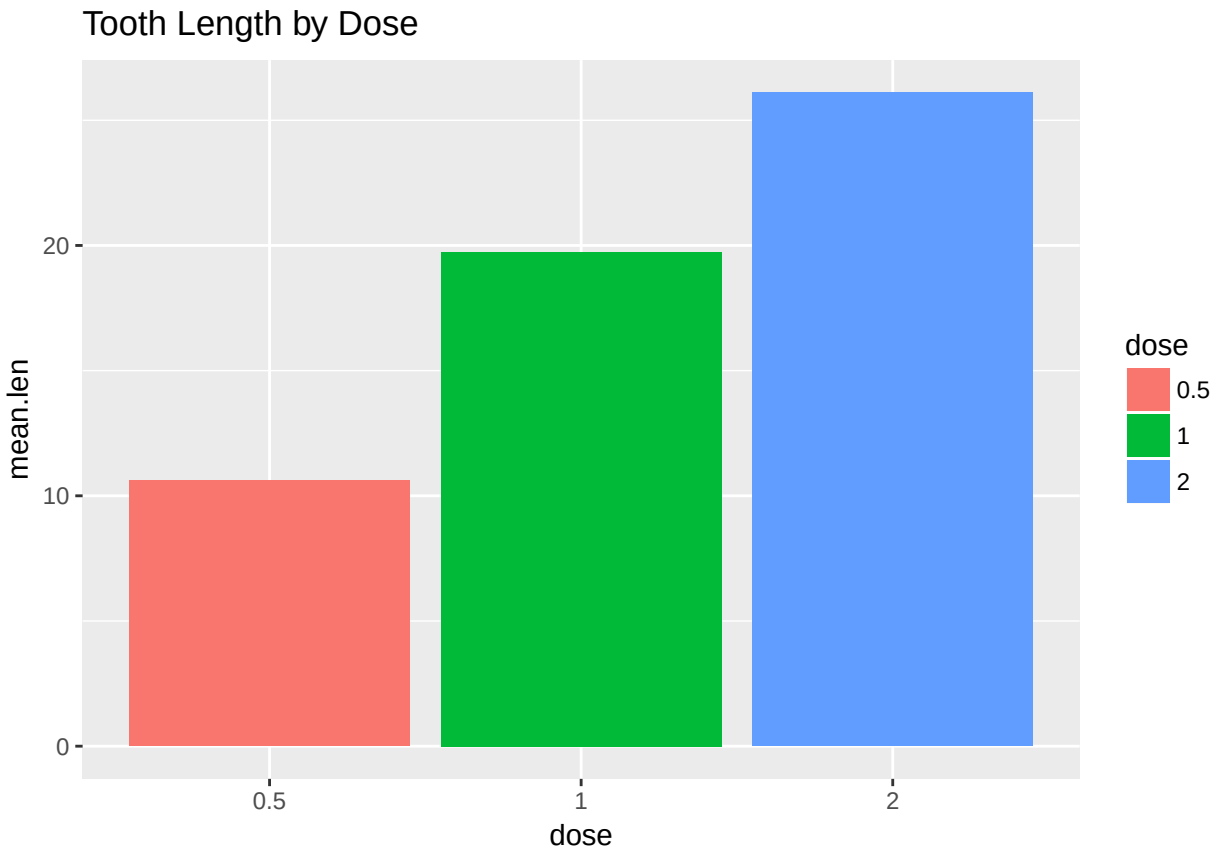
```
## Test results:
##    t-statistic:  1.915
##    degrees of freedom:  55.309
##    p-value:  0.061
##
## Other information:
##    two-sided 95% confidence interval:  [-0.171, 7.571]
##    estimated effect size (Cohen's d):  0.495
```

The conclusion is: the mean tooth length of the pigs given orange juice supplements was *20.66 (std dev = 6.60)* whereas the mean tooth length of the pigs given ascorbic acid supplements was *16.96 (std dev = 8.26)*. A Welch's independent samples t-test showed that this 3.7 unit lengths of difference was **not significant** *(t(55.30) = 1.915, p > .05, CI95 = [-0.171, 7.571], d = 0.495)*, suggesting that **there is not a difference between taking orange juice and ascorbic acid supplements concercing tooth growth**.

Is there a difference between the supplements' dosage concerning tooth growth?

If there was no difference between the supplements dosage concerning tooth growth, we would expect that the average tooth length at every dosage level would be the same, plus or minus chance variation. Let's have a look at the data:

```
ToothGrowth %>%
  group_by(dose) %>%
  summarise(mean.len = mean(len)) %>%
  ggplot(aes(dose, mean.len)) +
  geom_col(aes(fill = dose)) +
  ggtitle('Tooth Length by Dose')
```



Apparently, the higher the supplement dose, the longer the tooth, but we need to check if this difference is significant:

```
dose.0.5.and.1 <- ToothGrowth %>% filter(dose %in% c(.5, 1))
dose.0.5.and.2 <- ToothGrowth %>% filter(dose %in% c(.5, 2))
dose.1.and.2 <- ToothGrowth %>% filter(dose %in% c(1, 2))

independentSamplesTTest(len ~ dose, as.data.frame(dose.0.5.and.1))
```

```
## Warning in independentSamplesTTest(len ~ dose, as.data.frame(dose.0.5.and.
## 1)): grouping variable has unused factor levels
```

```
##
##    Welch's independent samples t-test
##
## Outcome variable:   len
## Grouping variable:  dose
##
## Descriptive statistics:
##                 0.5      1
##    mean      10.605 19.735
##    std dev.   4.500  4.415
##
## Hypotheses:
##    null:        population means equal for both groups
##    alternative: different population means in each group
##
```

```
## Test results:
##    t-statistic:  -6.477
##    degrees of freedom:  37.986
##    p-value:  <.001
##
## Other information:
##    two-sided 95% confidence interval:  [-11.984, -6.276]
##    estimated effect size (Cohen's d):  2.048
```

```
independentSamplesTTest(len ~ dose, as.data.frame(dose.0.5.and.2))
```

```
## Warning in independentSamplesTTest(len ~ dose, as.data.frame(dose.0.5.and.
## 2)): grouping variable has unused factor levels
```

```
##
##    Welch's independent samples t-test
##
## Outcome variable:   len
## Grouping variable:  dose
##
## Descriptive statistics:
##                0.5      2
##    mean     10.605 26.100
##    std dev.  4.500  3.774
##
## Hypotheses:
##    null:        population means equal for both groups
##    alternative: different population means in each group
##
## Test results:
##    t-statistic:  -11.799
##    degrees of freedom:  36.883
##    p-value:  <.001
##
## Other information:
##    two-sided 95% confidence interval:  [-18.156, -12.834]
##    estimated effect size (Cohen's d):  3.731
```

```
independentSamplesTTest(len ~ dose, as.data.frame(dose.1.and.2))
```

```
## Warning in independentSamplesTTest(len ~ dose, as.data.frame(dose.1.and.
## 2)): grouping variable has unused factor levels
```

```
##
##    Welch's independent samples t-test
##
## Outcome variable:   len
## Grouping variable:  dose
##
## Descriptive statistics:
##                  1      2
##    mean     19.735 26.100
##    std dev.  4.415  3.774
##
## Hypotheses:
##    null:        population means equal for both groups
```

```
##    alternative: different population means in each group
##
## Test results:
##    t-statistic:  -4.9
##    degrees of freedom:  37.101
##    p-value:  <.001
##
## Other information:
##    two-sided 95% confidence interval:  [-8.996, -3.734]
##    estimated effect size (Cohen's d):  1.55
```

The conclusion is: the mean tooth length of the pigs given a dosage of 0.5mm/day of vitaming C supplements was *10.60 (std dev = 4.41)*, mean tooth length of the pigs given a dosage of 1mm/day was *19.735 (std dev = 4.41)*, and mean tooth length of the pigs given a dosage of 2mm/day was *26.100 (std dev = 3.77)*. A series of Welch's independent samples t-tests showed that the difference of unit lengths between each level was **highly significant**:

1. *(t(37.98) = -6.477, p < .001, CI95 = [-11.984, -6.276], d = 2.048)*
2. *(t(36.88) = -11.799, p < .001, CI95 = [-18.156, -12.834], d = 3.731)*
3. *(t(37.10) = -4.9, p < .001, CI95 = [-8.996, -3.734], d = 1.55)*

suggesting that **there is a difference between each dosage level concerning tooth growth**.