Statistical Inference Course Project part 2

Naveen Venkat Raj Jaladi

Overview

This report aims to analyze the ToothGrowth data in the R datasets package. Per the course project instructions, the following items should occur:

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

2.Provide a basic summary of the data.

3.Using confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose(Only use the techniques from class, even if there's other approaches worth considering).

4.Stating conclusions and the assumptions needed from conclusions.

Exploratory Data Analysis

Dataset description:

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).

Codebook:

- len - numeric, Tooth length

- supp - factor, Supplement type (VC or OJ).

- dose - numeric, Dose in milligrams/day

Load data and required packages:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```r
library(purrr)
```

```
## Warning: package 'purrr' was built under R version 3.6.3
```

```r
library(broom)
```

```
## Warning: package 'broom' was built under R version 3.6.3
```

```r
options(tibble.width = Inf)
data("ToothGrowth")
```

Explore the dataset:

```r
glimpse(ToothGrowth)
```

```
## Observations: 60
## Variables: 3
## $ len  <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16.5, 16...
## $ supp <fct> VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC...
## $ dose <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 1....
```

```r
summary(ToothGrowth)
```

```
##       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
```

```r
ToothGrowth$dose <- factor(ToothGrowth$dose)
```

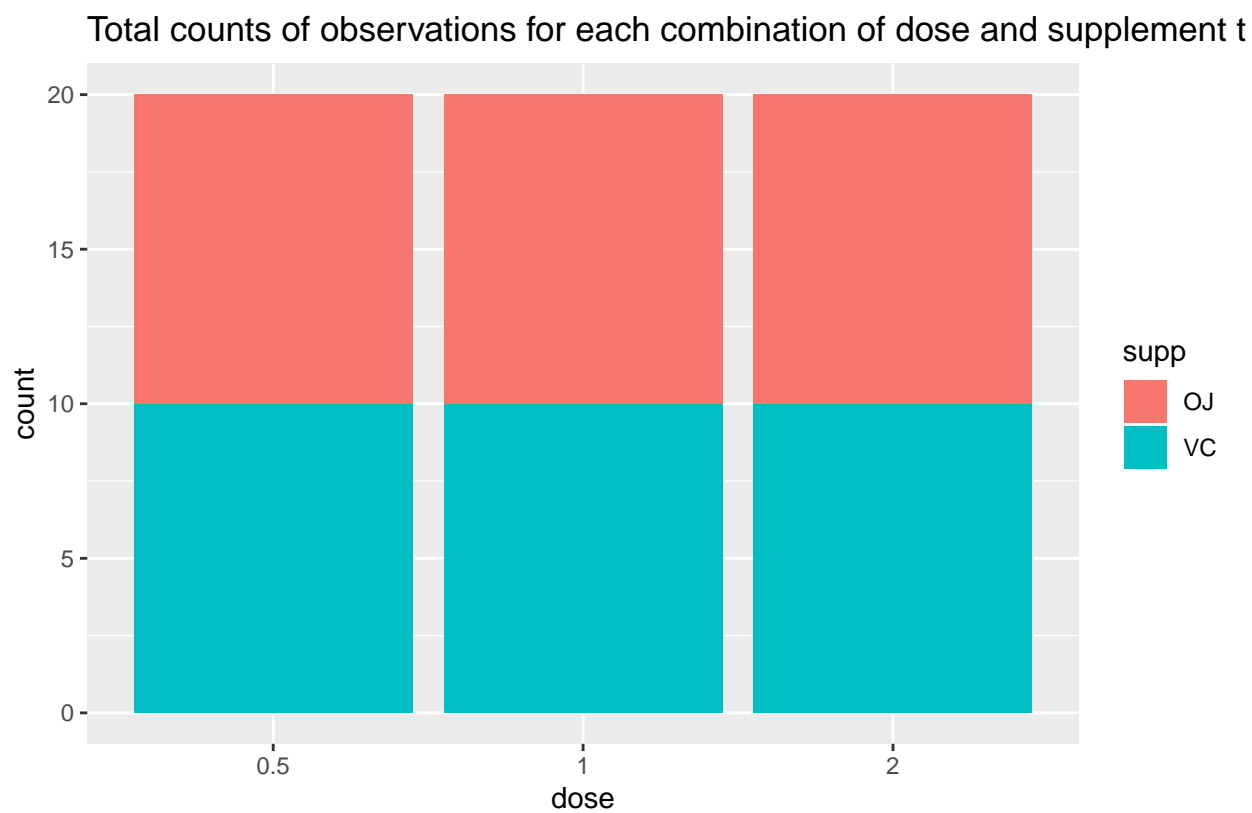Exploring variables via visualization:
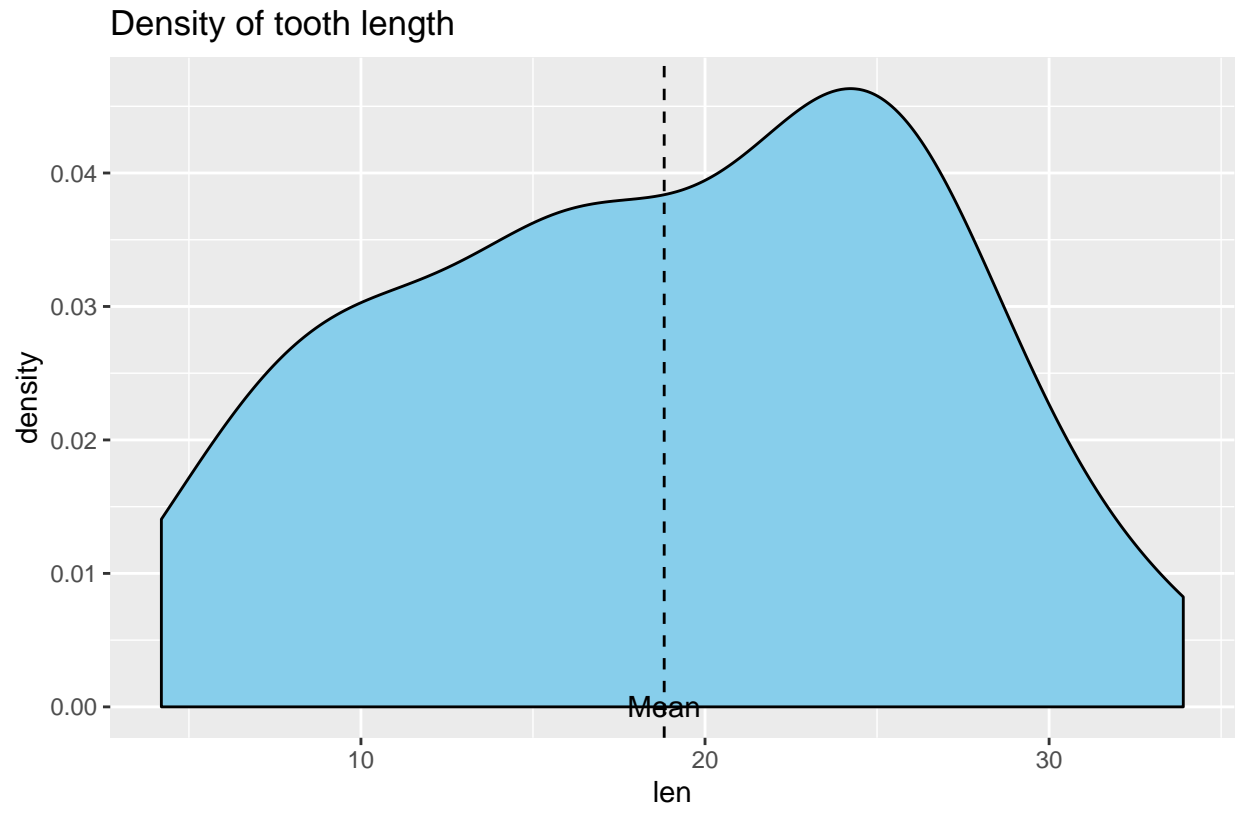
Total counts of observations for each combination of dose and supplement t

Figure 1

Density of tooth length

Mean

Figure 2

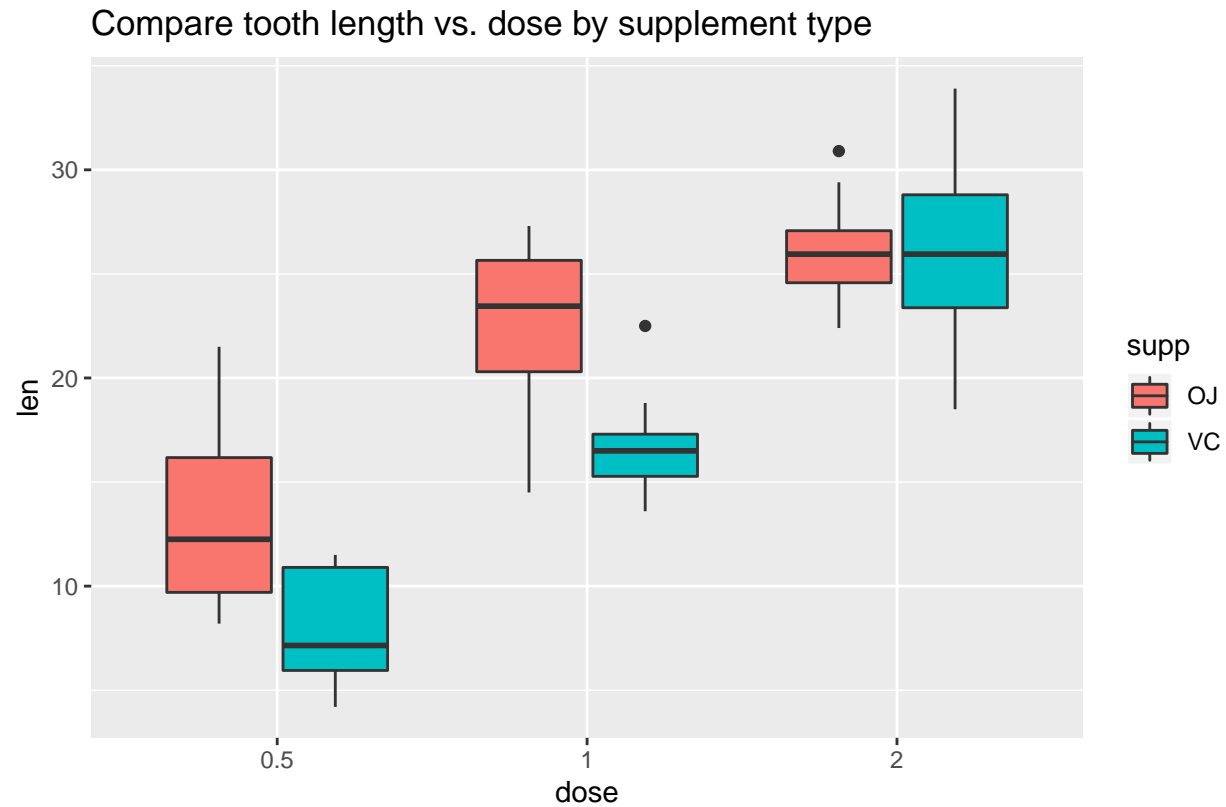## Compare tooth length vs. dose by supplement type

Figure 3

Hypothesis test

In the previous step of exploratory data analysis we see that there is a obvious relationship between the dose levels of vitamin and tooth growth. Therefore we can say the dosage affects the tooth length. Also visualizations show us that supplement delivery method has no robust impacts on tooth length.

Let's calculate t-test for both groups.

1. Testing tooth growth by supplement delivery method:

Null Hypothesis : different supplement types have no effect on tooth length

```
t.result1 <-
    tidy(t.test(len~supp, data = ToothGrowth)) %>%
    print()
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##      <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl>
## 1     3.70      20.7      17.0      1.92  0.0606      55.3   -0.171      7.57
##   method                alternative
##   <chr>                 <chr>
## 1 Welch Two Sample t-test two.sided
```

5

Here We can not reject the Null Hypothesis, cause the Confidence interval contains zero and also the p.value (0.0606345) is greater than 0.05.

2. Testing tooth growth by dose:

Null Hypothesis : dosage levels have no effect on tooth length

We will compare doses 0.5 with 1, 1 with 2 and 2 with 0.5

```
t.result2 <-
ToothGrowth %>%
    mutate(dose.group = "0.5-1") %>%
    bind_rows(ToothGrowth) %>%
    mutate(dose.group = if_else(is.na(dose.group), "1-2", dose.group)) %>%
    bind_rows(ToothGrowth) %>%
    mutate(dose.group = if_else(is.na(dose.group), "2-0.5", dose.group)) %>%
    filter((dose.group == "0.5-1" & dose %in% c(.5, 1)) |
           (dose.group == "1-2" & dose %in% c(1, 2)) |
           (dose.group == "2-0.5" & dose %in% c(2, .5))) %>%
    select(-supp) %>%
    group_by(dose.group) %>%
    do(tidy(t.test(len~dose, data = .))) %>%
    print()
```

```
## # A tibble: 3 x 11
## # Groups:   dose.group [3]
##   dose.group estimate estimate1 estimate2 statistic  p.value parameter conf.low
##   <chr>         <dbl>     <dbl>     <dbl>     <dbl>    <dbl>     <dbl>    <dbl>
## 1 0.5-1         -9.13      10.6      19.7     -6.48 1.27e- 7      38.0    -12.0
## 2 1-2           -6.37      19.7      26.1     -4.90 1.91e- 5      37.1    -9.00
## 3 2-0.5        -15.5       10.6      26.1    -11.8  4.40e-14      36.9    -18.2
##   conf.high method                    alternative
##       <dbl> <chr>                     <chr>
## 1     -6.28 Welch Two Sample t-test two.sided
## 2     -3.73 Welch Two Sample t-test two.sided
## 3    -12.8  Welch Two Sample t-test two.sided
```

The confidence intervals for these doses groups allow us to reject the Null Hypothesis.

Conclusion

Given the following assumptions:

- The sample is representative of the population.

- The distribution of the sample means follows the Central Limit Theorem.

In reviewing our t-test analysis from above, we can conclude that

- supplement delivery method has no effect on tooth growth/length.

- however increased dosages do result in increased tooth length.

Appendices

Code to plot Figure 1:

```
ggplot(ToothGrowth, aes(x = dose, fill = supp)) +
    geom_bar() +
    labs(caption = "Figure 1", title = "Total counts of observations for each combination of dose and s
```

Code to plot Figure 2:

```
ggplot(ToothGrowth, aes(x = len)) +
    geom_density(fill = "skyblue") +
    geom_vline(xintercept = mean(ToothGrowth$len), lty = "dashed") +
    annotate(geom = "text", x = mean(ToothGrowth$len), y = 0, label = "Mean") +
    labs(caption = "Figure 2", title = "Density of tooth length")
```

Code to plot Figure 3:

```
ggplot(ToothGrowth, aes(x = dose, y = len, fill = supp)) +
    geom_boxplot() +
    labs(caption = "Figure 3", title = "Compare tooth length vs. dose by supplement type")
```