Peer-graded Assignment: Statistical Inference Course Project | BASIC INFERENTIAL DATA ANALYSIS

Roxana Trejos Ramirez

11/7/2020

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data. Use 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) State your conclusions and the assumptions needed for your conclusions.

##The Effect of Vitamin C on Tooth Growth in Guinea Pigs

R: ToothGrowth data

Description

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Usage

ToothGrowth

Format

A data frame with 60 observations on 3 variables.

- **len** which is the numeric tooth length measurement (unspecified units)
- **supp** which is the delivery method of Vitamin C - either by orange juice (OJ) or by ascorbic acid (VC)
- **dose** which is the dosage of the supplement - 0.5 ml/day, 1.0 ml/day or 2.0 ml/day.

Source

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

References

McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.

```
# Libraries
    library(datasets)
    library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
    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
```

a. Load the Tooth Growth Data.

```
# load the sample dataset containing ToothGrowth data
    data(ToothGrowth)

# workable variable
    dataTG <-ToothGrowth
```

b. Exploratory Data Analysis

```
#   Summary of the type and the values of the features
    str(dataTG)
```

```
## '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 ...
```

```
#   Statistical values of the whole ToothGrowth.
    summary(dataTG)
```

```
##       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
```

```
#   First few rows
    head(dataTG)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
#   Unique dose
    unique(dataTG$dose)
```

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

```
#   Unique supplement
    unique(dataTG$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

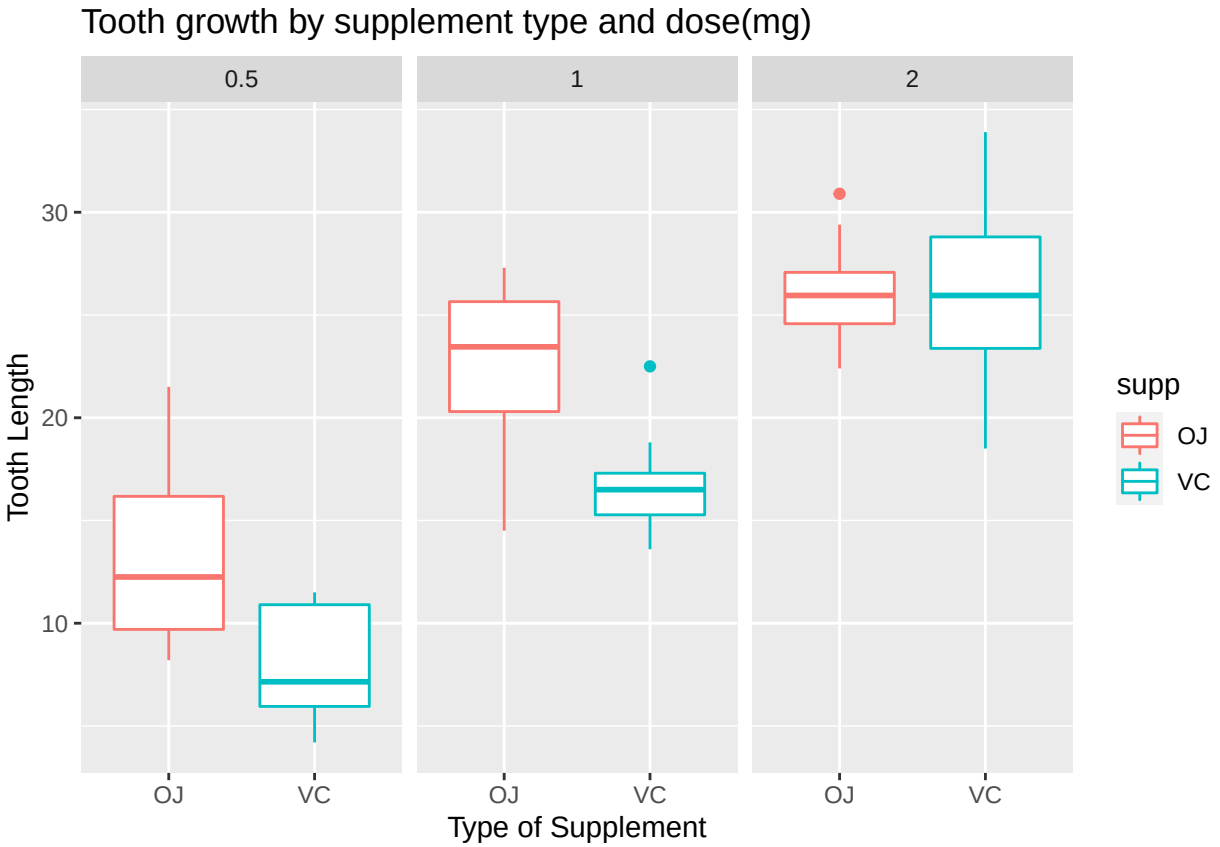
```
# Summarize by dose and supp, the mean length of growth.
    a <- dataTG %>%
        group_by(supp,dose) %>%
        summarize(lenmean=mean(len), lensd=sd(len), count = n())
```

```
## `summarise()` regrouping output by 'supp' (override with `.groups` argument)
```

```
    print(a)
```

```
## # A tibble: 6 x 5
## # Groups:   supp [2]
##   supp   dose lenmean lensd count
##   <fct> <dbl>   <dbl> <dbl> <int>
## 1 OJ      0.5   13.2   4.46    10
## 2 OJ      1     22.7   3.91    10
## 3 OJ      2     26.1   2.66    10
## 4 VC      0.5    7.98  2.75    10
## 5 VC      1     16.8   2.52    10
## 6 VC      2     26.1   4.80    10
```

```
g<-ggplot(dataTG,
          aes(x=supp,
              y=len,
              color=supp)) +
    geom_boxplot() +
    facet_grid(facets = ~ dose) +
    labs(title="Tooth growth by supplement type and dose(mg)" ,
         y = "Tooth Length",
         x = "Type of Supplement")
g
```



b.1 Data Analysis Conclusions

1. There are three doses: 0.5, 1.0 and 2.0.
2. The analysis shows only two doses: VC and OJ:
    - OJ: orange juice.
    - VC: ascorbic acid~Vitamin C.
3. In the VC:
    - Tooth length is longer when the doses increase.
    - The tooth growth has a linear relationship with dosage.
4. In the OJ:
    - Higher dossage (2.0mg) has less improvement in tooth growth.
    - This supplement, generally induces more tooth growth than VC except at higher dosage (2.0 mg).

c. Use 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)

c.1 Assumptions:

1. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs.

2. 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).
3. A data frame with 60 observations on 3 variables.
4. 95% Confidence Interval.

c.2 Inferential statistic to determine if there is a significant difference between the means of OJ and VC doses' groups.

```
    t.test(len ~ supp, paired=FALSE, var.equal=FALSE, data=dataTG)
```

```
##
##  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
```

Summary:

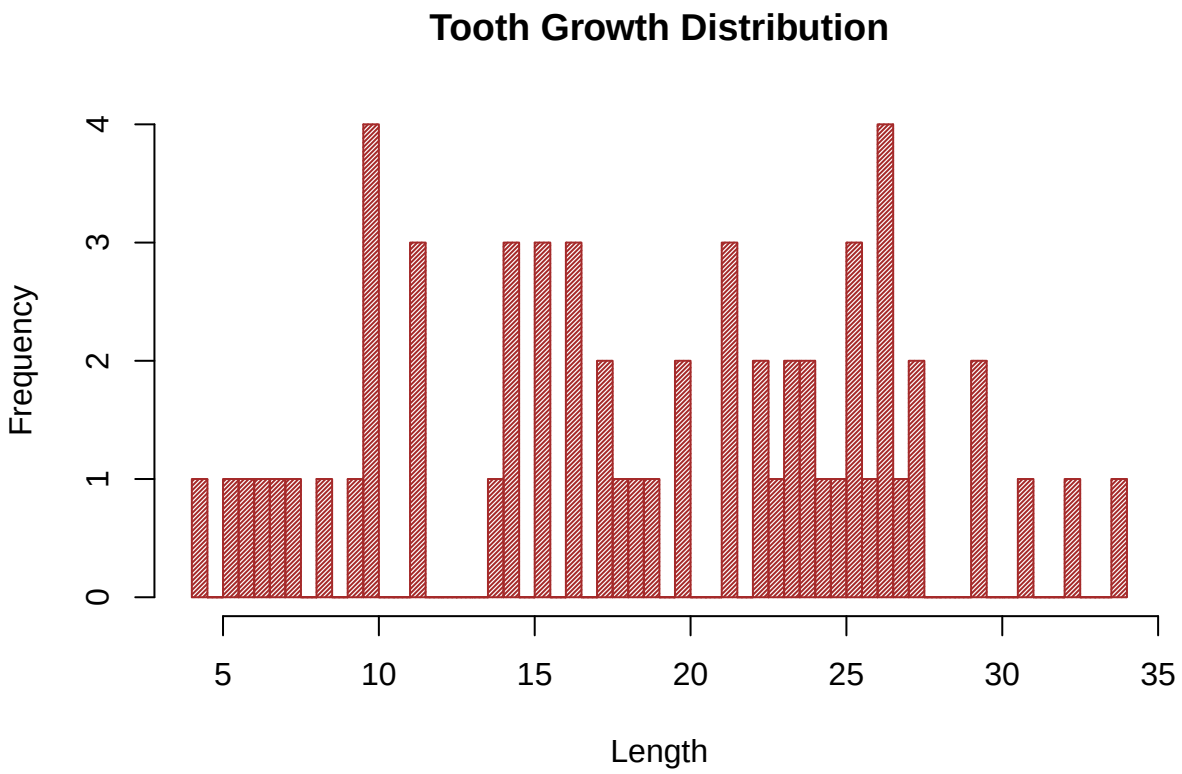
1. The type of supplementation did not matter for the tooth length increase (length is not significantly different for the two supplement types).
2. Comparing a NULL hypothesis (difference of means = 0) against an alternative hypothesis: *we fail to reject the NULL hypothesis*, since the NULL hypothesis value (delta means = 0) is within:
    - confidence interval of 95% confidence.
    - 1.915 t quantile.
    - probability is 0.061 which is greater than our mu (0.05).

c.3 Tooth Length by Delivery Method

```
ggplot(
    data = dataTG,
    mapping = aes(x = supp, y = len, color = supp, fill = supp)
) +
    geom_boxplot(alpha = 0.2) +
    geom_jitter(width = 0.05) +
    scale_x_discrete(labels = c("OJ", "VC")) +
    labs(
        title = "Tooth Length by Delivery Method",
        x = "Delivery Method (Type of the supplement)",
        y = "Length of the tooth"
    ) +
    theme(legend.position = "none")
```

c4. Tooth Growth Distribution

```
hist(dataTG$len,
     col = "brown",
     main = "Tooth Growth Distribution",
     xlab = "Length",
     ylab = "Frequency",
     density=60,
     breaks = 60)
```



c5. Conclusions:

1. **We fail to reject the NULL hypothesis**, since the NULL hypothesis value (delta means = 0) is within:
    - confidence interval of 95% confidence.
    - 1.915 t quantile.
    - probability is 0.061 which is greater than our mu (0.05).
2. The amount of tooth length increase is directly analogous to the vitamin intake, regardless of type of ingestion.
3. The difference of the means of teeth length for the two supplements is not significantly different from zero at the 95% confidence level.