Statistical Inference Project: Part 2

Angie Marchany-Rivera

7/22/2021

Overview

This project consists of 2 parts: a simulation excercise and basic inferential data analysis. Part 1 shows a comparison between the exponential distribution and the Central Limit Theorem (CLT). Part 2 shows the results of confidence intervals and hypothesis tests of the ToothGrowth dataset in R.

Part 2: Basic Inferential Data Analysis

Now in the second portion of the project the ToothGrowth data in the R datasets package will be analyzed through confidence intervals and hypothesis testing.

Data Variables

The ToothGrowth dataset shows "The Effect of Vitamin C on Tooth Growth in Guinea Pigs". 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 (coded as OJ) or ascorbic acid (a form of vitamin C and coded as VC).

Loading the ToothGrowth data and performing some basic exploratory data analyses

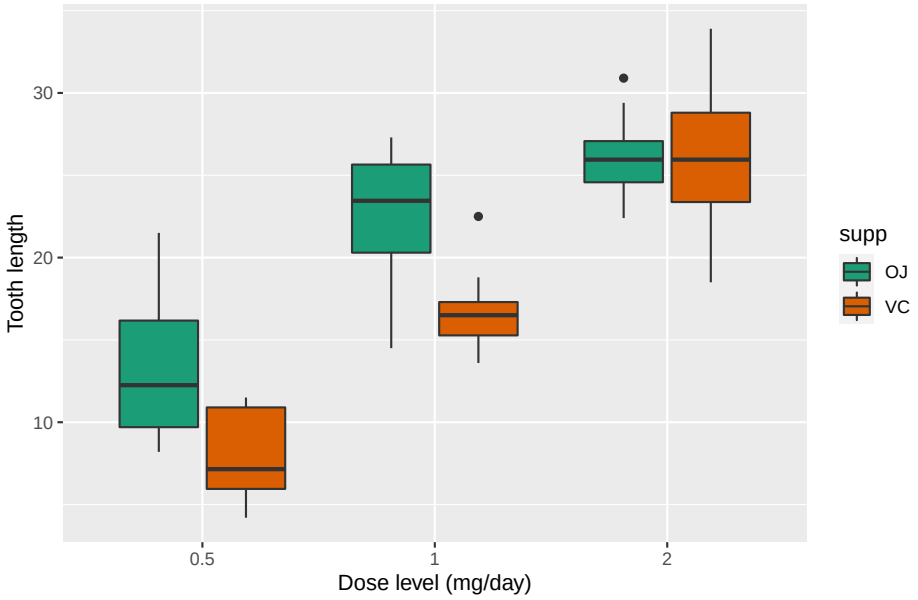
```
library(dplyr)
library(gtsummary)
library(ggplot2)
data<-ToothGrowth
str(data)
```

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

Where "len" represents the tooth length, "supp" represents the supplement type, and "dose" represents the dose level in mg/day.

```
ggplot(data, aes(x=factor(dose), y=len, fill=supp)) + geom_boxplot()+
        labs(title="Figure 1: Exploratory analysis of the ToothGrowth Data",
     x="Dose level (mg/day)", y = "Tooth length")+
        scale_fill_brewer(palette="Dark2")
```

Figure 1: Exploratory analysis of the ToothGrowth Data



Summary of the ToothGrowth data

The objective of this data analysis is to evaluate the impact of the variables "supp" and "dose" on the variable "len". Table 1 shows the relationship between the supplement type and the tooth length while Table 2 shows the relationship between the dose level and the tooth length.

```
data %>% tbl_summary(by=supp,  type = all_continuous() ~ "continuous2",
                statistic = all_continuous() ~ c("{median} ({p25}, {p75})",
                                                 "{mean} ({sd})",
                                                 "{min}, {max}")) %>%
        modify_header(label ~ "**Variable**") %>%
        modify_caption("**Summary of the data by supplement type**")%>%
        bold_labels()
```

Table 1: **Summary of the data by supplement type**

Variable	**OJ**, N = 30	**VC**, N = 30
len		
Median (IQR)	23 (16, 26)	16 (11, 23)
Mean (SD)	21 (7)	17 (8)
Range	8, 31	4, 34
dose		
0.5	10 (33%)	10 (33%)
1	10 (33%)	10 (33%)
2	10 (33%)	10 (33%)

```
data %>% tbl_summary(by=dose,  type = all_continuous() ~ "continuous2",
                statistic = all_continuous() ~ c("{median} ({p25}, {p75})",
                                                 "{mean} ({sd})",
                                                 "{min}, {max}")) %>%
```

```
        modify_header(label ~ "**Variable**") %>%
        modify_caption("**Summary of the data by dose level**") %>%
        bold_labels()
```

Table 2: **Summary of the data by dose level**

Variable	**0.5**, N = 20	**1**, N = 20	**2**, N = 20
len			
Median (IQR)	10 (7, 12)	19 (16, 23)	26 (24, 28)
Mean (SD)	11 (4)	20 (4)	26 (4)
Range	4, 22	14, 27	18, 34
supp			
OJ	10 (50%)	10 (50%)	10 (50%)
VC	10 (50%)	10 (50%)	10 (50%)

Based on the information presented in the exploratory analysis and the summary tables, the following hypotheses are worth evaluating:

1. The delivery of a low (0.5 mg/day) to medium (1 mg/day) dosage is more effective (i.e., longer tooth length) when ingesting orange juice versus the absorbic acid method.
2. There is no significant difference in the tooth length when ingesting either of the two supplement types at a high dose level (2 mg/day).
3. Overall, a high dose level yields a longer tooth length than a low dose level, regardless of the supplement type.

T-tests will be used to perform these hypothesis tests since the total amount of ToothGrowth data is small.

Using confidence intervals and hypothesis tests to compare tooth growth length by supplement type and dose level

Testing Null Hypothesis 1: The supplement type has no effect on tooth growth length for a low to medium dose level.

```
lowMedDose <- data[data$dose %in% c(0.5,1),c("len","supp")] # Dose level = 0.5 and 1
t.test(lowMedDose$len[lowMedDose$supp=="OJ"],lowMedDose$len[lowMedDose$supp=="VC"],
       var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  lowMedDose$len[lowMedDose$supp == "OJ"] and lowMedDose$len[lowMedDose$supp == "VC"]
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean of x mean of y
##    17.965    12.375
```

The p-value is lower than 0.05 and the 95% confidence interval does not contain 0 hence we can reject the NULL hypothesis and conclude that the supplement type does have an impact in tooth growth length at low to medium dose levels.

Testing NULL Hypothesis 2: The supplement type has no effect on tooth growth length for a high dose level.

```
highDose <- data[data$dose == 2.0,c("len","supp")] # Dose level = 2.0
t.test(highDose$len[highDose$supp=="OJ"],highDose$len[highDose$supp=="VC"],
       var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  highDose$len[highDose$supp == "OJ"] and highDose$len[highDose$supp == "VC"]
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

The p-value here is larger than 0.05 and the 95% confidence interval includes 0 hence we failed to reject the NULL hypothesis and conclude that the supplement type has no significant effect on tooth growth length at the high dose level.

Testing NULL Hypothesis 3: The dose level has no effect on tooth growth length regardless of the supplement type.

```
t.test(lowMedDose$len,highDose$len,var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  lowMedDose$len and highDose$len
## t = -8.3085, df = 56.202, p-value = 2.347e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.565108  -8.294892
## sample estimates:
## mean of x mean of y
##     15.17     26.10
```

The p-value here is much lower than 0.05 and the 95% confidence interval does not include 0 hence we reject the NULL hypothesis and conclude that an increase in dose level has a significant effect on tooth growth length regardless of the supplement type.

Conclusions:

1. The supplement type does have an impact in tooth growth length at low to medium dose levels.
2. The supplement type has no significant effect on tooth growth length at the high dose level.
3. An increase in dose level has a significant effect on tooth growth length regardless of the supplement type.

Assumptions:

- Variances are not equal.
- Independent and identically distributed (iid) variables.
- Significance level alpha = 0.05.