Statistical Inference Project - Part Two

Brandon Robinson

2024-04-13

Overview

The purpose of the project is to investigate two parts relating statistical analysis of data. Part 1 is to perform a simulation on random data and analyze following the Central Limit Theorem. Part 2 uses a dataset from the R dataset library to perform an analysis and derive a conclusion.

Part 2 - Basic Inferential Data Analysis

Instructions for Part 2 are as follows: 1. Load the ToothGrowth data and perform some basic exploratory data analyses 2. Provide a basic summary of the data. 3. 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) 4. State your conclusions and the assumptions needed for your conclusions.

Question 1 - Load the ToothGrowth data and perform some basic exploratory data analyses
First, the required packages will be loaded in order to run the analysis and create the plots

```
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)
```

Next, we load the data from the dataset library

```
data("ToothGrowth")
```

The dataset shows the effect of vitamin c on the tooth growth in 60 guinea pigs. Each animal received dose levels of 0.5, 1, and 2 mg/day by two delivery methods, orange juice (OJ) and vitamin c (VC)

Question 2 - Perform some basic exploratory data analyses Then, we perform a basic summary of the dataset

```
# Overview of each variable in the dataset
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
```

```
head(ToothGrowth)
```

```
##    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(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6 14.5
## [16] 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6  9.7  8.2
## [31]  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4 23.0
```

```
unique(ToothGrowth$supp)
```

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

```
unique(ToothGrowth$dose)
```

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

```
# View the first 6 occurrences
head(ToothGrowth)
```

```
##    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
```

```r
# Identify the unique values for each variable
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6 14.5
## [16] 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6  9.7  8.2
## [31]  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4 23.0
```

```r
unique(ToothGrowth$supp)
```

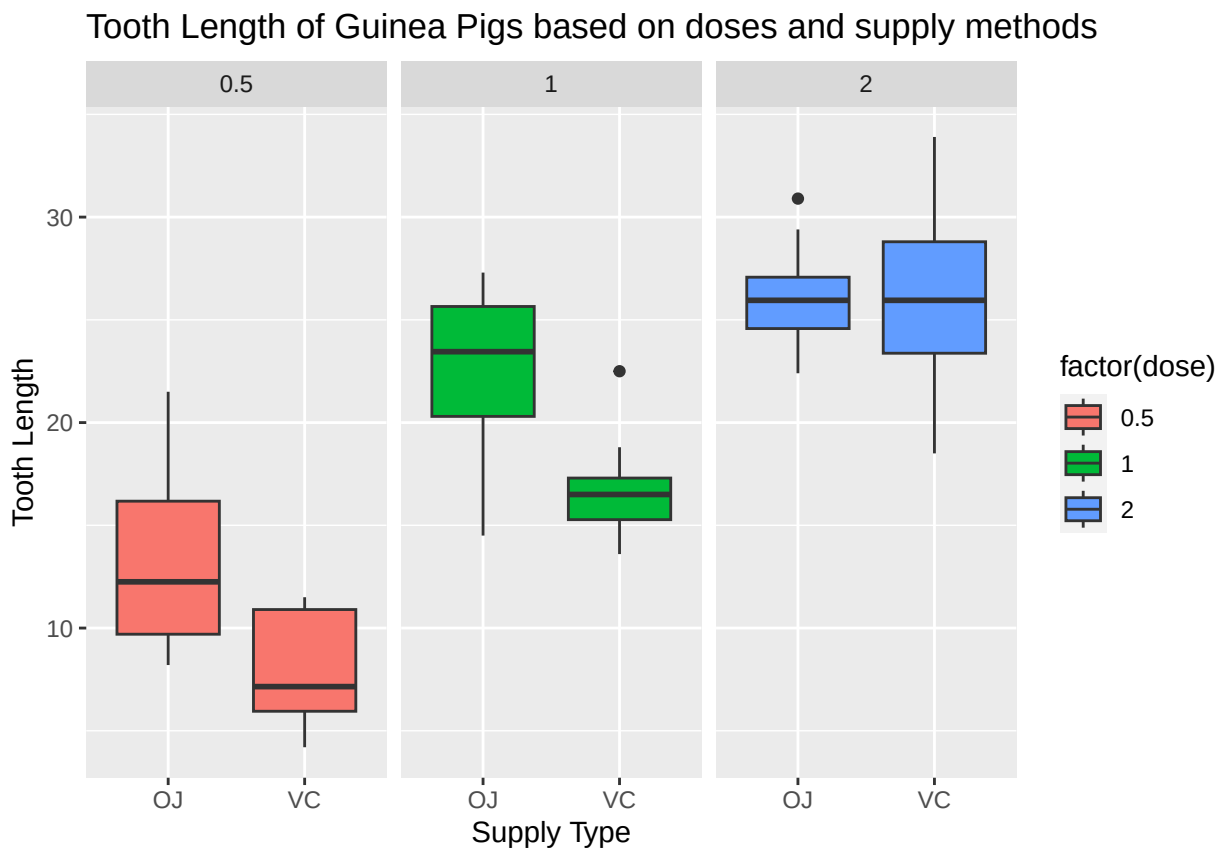
```
## [1] VC OJ
## Levels: OJ VC
```

```r
unique(ToothGrowth$dose)
```

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

Next, we plot the initial findings from the dataset

```r
ggplot(ToothGrowth, aes(x=factor(supp), y=len)) +
        facet_grid(.~dose) +
        geom_boxplot(aes(fill = factor(dose)), show.legend = TRUE) +
        labs(title = "Tooth Length of Guinea Pigs based on doses and supply methods ", x="Supply Type",
```



From the plot we can see as we approach a dosage of 2mg it cannot be concluded if one supplement is more effective than another. We will test this in the next section.

Question 3: Use confidence intervals or hypothesis test to compare the tooth growth by supp and dose. The first test will look at tooth growth by supplement

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  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 between group OJ and group VC 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
```

Next, we will look at the dosage level of each supplement compared to tooth growth

```
t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Between 0.5 to 1mg the hypothesis that OJ delivers more than growth can be accepted since the p-value is < 0.05 and the confidence interval does not include 0 At 2mg the null hypothesis can be rejected since the p-value is > 0.05 and the confidence interval includes 0

Conclusion Based on the full dataset it cannot be concluded that one supplement is not more effective than another for tooth growth