Exponential Distribution in R vs Central Limit Theorem (Part 2)

Jon Ting

19/08/2020

```
knitr::opts_chunk$set(warning=FALSE, fig.height=2, fig.width=5)
set.seed(77)
library(ggplot2)
```

Part 2

Overview

This project explores the exponential distribution in R with a lambda of 0.2 and compare it with the Central Limit Theorem. Part 2 involves the analysis of the ToothGrowth data in the R datasets package, which includes:

1. loading the ToothGrowth data and perform some basic exploratory data analyses.
2. providing a basic summary of the data.
3. using confidence intervals and hypothesis tests to compare tooth growth by supp and dose.
4. stating conclusions and the assumptions needed for the conclusions.

Load, Summarize and Preprocess Data

```
# Load the data
data(ToothGrowth)

# Understand structure of data
str(ToothGrowth)
```

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

```
# Summarize the data
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
```

Since the dose variable has only 2 unique values, it is convertd to a categorical variable

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

Exploratory Data Analysis

The following relationships were investigated during exploratory data analysis: - Tooth length vs supplement type - Tooth length vs dosage - Tooth length vs dosage grouped by delivery method

```
# Relationship 1
ggplot(data=ToothGrowth, mapping=aes(x=supp, y=len)) +
  geom_boxplot(mapping=aes(fill=supp)) +
  labs(title="Tooth Length vs Supplement Type", x="Supplement Type", y="Tooth Length")
```
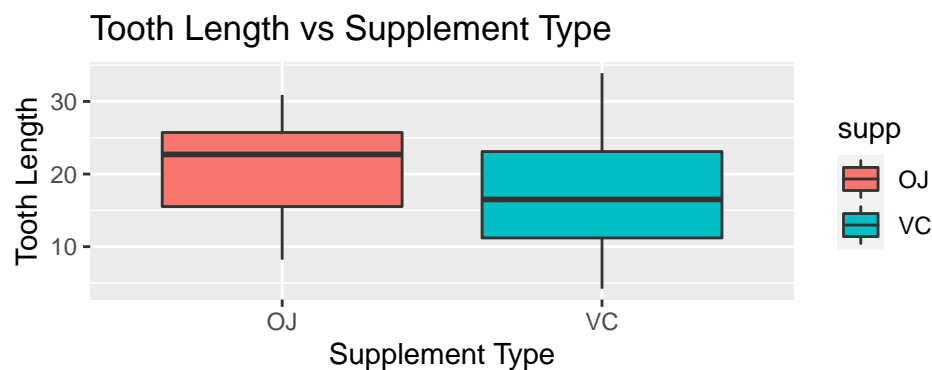

The boxplots seem to indicate that consumption of OJ generally leads to longer teeth.

```
# Relationship 2
ggplot(data=ToothGrowth, mapping=aes(x=dose, y=len)) +
  geom_boxplot(mapping=aes(fill=dose)) +
  labs(title="Tooth Length vs Dosage", x="Dosage", y="Tooth Length")
```
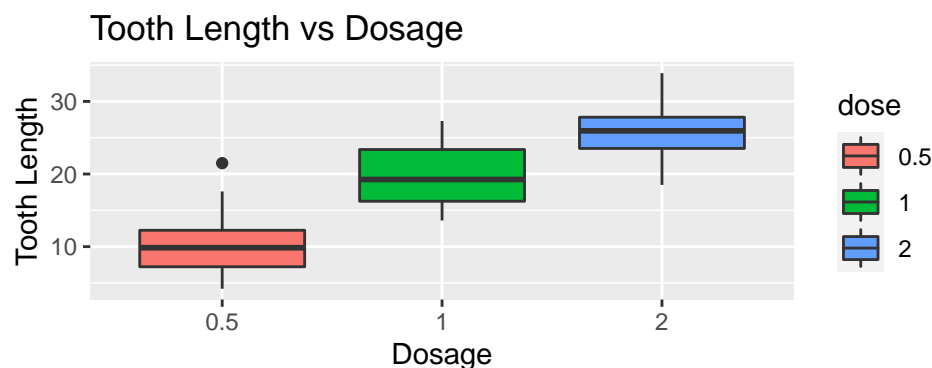

These plots point to a proportional relationship between dosage and tooth length.

```
# Relationship 3
ggplot(data=ToothGrowth, mapping=aes(x=dose, y=len)) +
  geom_boxplot(mapping=aes(fill=dose)) +
  labs(title="Tooth Length vs Dosage by Delivery Method", x="Dosage", y="Tooth Length") +
  facet_grid(~supp)
```


Tooth Length vs Dosage by Delivery Method
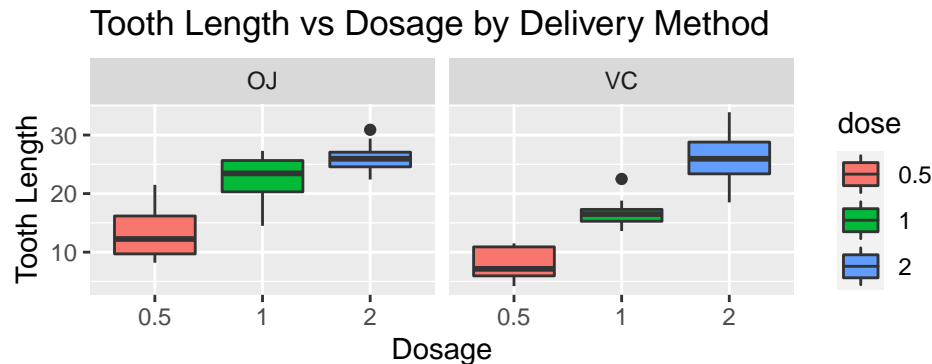
These graphs support the previous indications (i.e. OJ is more effective than VC and that greater dosage leads to longer teeth) in general. Hypothesis tests will be conducted next to compare the correlation between tooth growth by supplement type and dosage.

Compare Tooth Growth by Supp and Dose

A Welch 2 sample T-test is carried out for tooth growth by supplement type:

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

```
##
##  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
```

A p-value of 0.06 is obtained, which indicates that at the 0.05 significance level, there is lack of evidence to reject the null hypothesis that supplement type has no impact on tooth length. Zero is contained within the 95% confidence interval, which supports the conclusion that there is no relationship between the variables.

Another similar test is carried out for tooth growth by dosage as below:

```
# Dose amounts 0.5 and 1.0
t.test(formula=len~dose, data=subset(ToothGrowth, ToothGrowth$dose %in% c(0.5, 1.0)))
```

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

```
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
# Dose amounts 0.5 and 2.0
t.test(formula=len~dose, data=subset(ToothGrowth, ToothGrowth$dose %in% c(0.5, 2.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```r
# Dose amounts 1.0 and 2.0
t.test(formula=len~dose, data=subset(ToothGrowth, ToothGrowth$dose %in% c(1.0, 2.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##           19.735            26.100
```

All of the tests returned extremely small p-values, meaning that there is evidence at the 0.05 significance level to reject the null hypothesis that changing dosage does not correlate with changes in tooth length. None of the confidence intervals contain zero and all upper bounds are negative values, suggesting that increase in dosage leads to longer teeth.

Conclusions and Assumptions

The following conclusions were made according to the results of the t-test analyses:

1. the supplement type has no effect on tooth growth.
2. increased dosages leads to longer teeth.

Under the assumptions that:

1. The samples are representative of the guinea pig population.
2. The observations are independent from each other.
3. Each observation is randomly sampled.