The Exponential Distribution in R *versus* the Central Limit Theorem (CLT) — Part 2

Assignment: Statistical Inference Course Project

Sample Project Report Structure

Author: calaca

Course available at Coursera

Overview

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

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.

About the data

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).

A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length
[,2] supp factor Supplement type (VC or OJ)
[,3] dose numeric Dose in milligrams/day

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

Analysis

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

```
# Loading the dataset
library(datasets)
data <- ToothGrowth

# Taking a look at the first parts of the dataset
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
```

```
# Taking a look at how the data object is structured
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 ...
```

```
# Converting dose to factor instead of numeric
data$dose <- as.factor(data$dose)

# Checking the conversion
str(data)
```

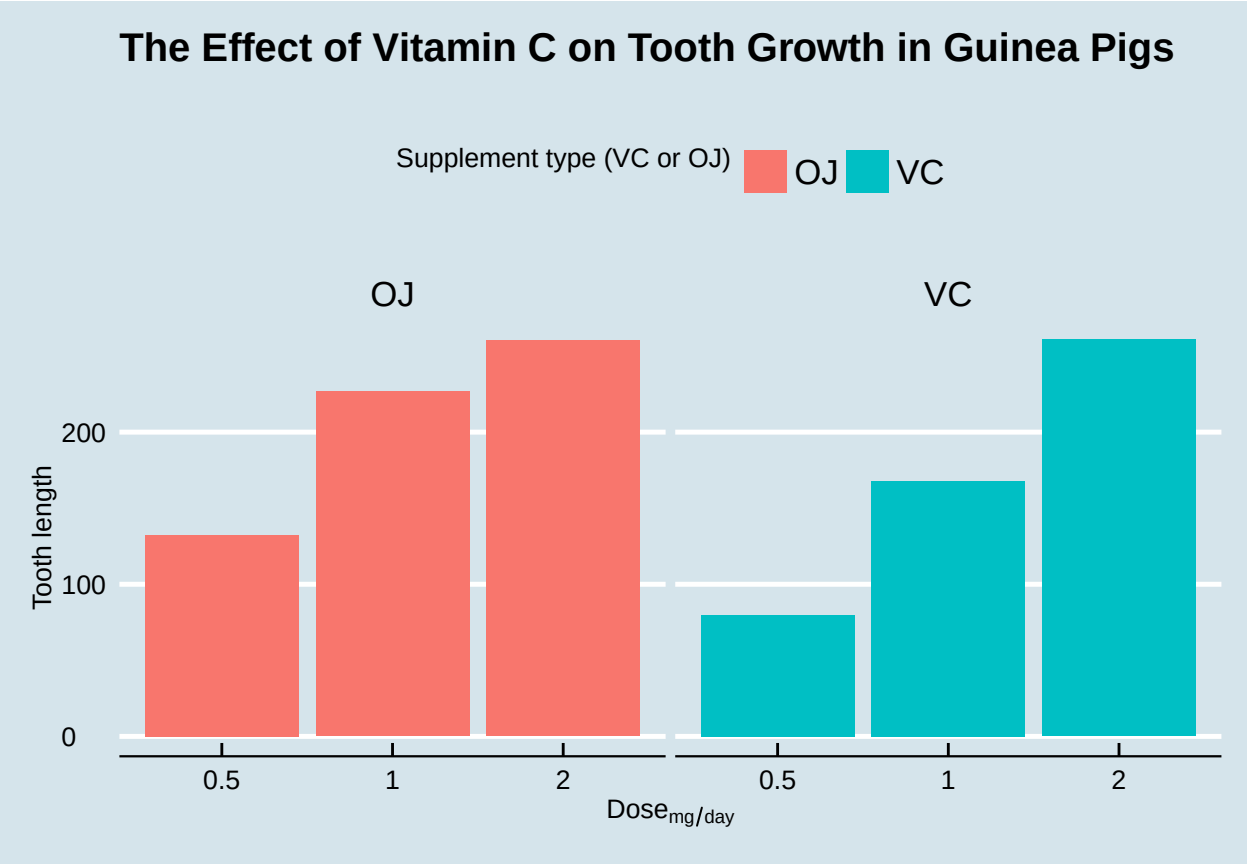
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

2. Provide a basic summary of the data

```
# Summary of the dataset
summary(data)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
library(ggplot2)
library(ggthemes)
ggplot(data, aes(x = dose, y = len, fill = supp)) + geom_bar(stat = "identity") +
  facet_grid(. ~ supp) + labs(x = expression("Dose"[mg/day])) + ylab("Tooth length") +
  guides(fill = guide_legend(title = "Supplement type (VC or OJ)\n")) +
  ggtitle("The Effect of Vitamin C on Tooth Growth in Guinea Pigs") +
  theme_economist()
```



Observing the plot, we see that the amount of supplement (OJ or VC) given to a guinea pig seems to make their teeth grow bigger.

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)

```
# H0: does supplement type affects tooth growth?
t.test(len ~ supp, data)
```

```
##
##  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
```

P-value (0.06063) is greater than 0.05, therefore we cannot reject the null hypothesis (H_0).

```r
# Fetching the data again (now we need dosage as numeric)
data <- ToothGrowth
# Ha: does dosage affects tooth growth?
t.test(data$len, data$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$len and data$dose
## t = 17.81, df = 59.798, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  15.66453 19.62881
## sample estimates:
## mean of x mean of y
## 18.813333  1.166667
```

P-value ($<$ 2.2e-16 or $<$ 0.00000000000000022) is less than 0.05, therefore we can reject the alternative hypothesis (H_a).

4. State your conclusions and the assumptions needed for your conclusions

H_0 = supplement type affects tooth growth: This hypothesis cannot be rejected. Which means that there is not enough evidence to affirm that the a type of supplement is better than the other.
H_a = dosage affects tooth growth: This hypothesis can be rejected. Which means that the amount of supplement affects the tooth growth.Thus, when dosage is increased the guinea pig's tooth grows bigger.
