Statistical Inference Course Project

Paul Martinez

2023-03-02

PART 1: Project Instructions

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations. :

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
# set simulation
lambda<- 0.2
n <- 40
n_simulations <- 1000

# setting seed to reproduce in future
set.seed(2)

# run matrix
ex_dis <- matrix(data=rexp(n * n_simulations, lambda), nrow=n_simulations)
ex_dis_m <- data.frame(means=apply(ex_dis, 1, mean))
```

Sample mean compared to the theoretical mean of the distribution

The expected mean of a exponential distribution of rate lambda is 1/lambda.Therefore, the theoretical mean is as follows:

```
theoretical_m <- 1/lambda
theoretical_m
```

```
## [1] 5
```

The average sample mean of the 1000 simulations of 40 randomly sampled exponential distributions is as follows:

1

```
#meansofmeans
m_m <- mean(ex_dis_m$means)
m_m
```

```
## [1] 5.016356
```

As demonstrated above, one can see that both means are very close to each other.

Sample Variance compared to Theoretical Variance of distribution:

The theoretical variance is as follows:

```
variance_t <- 1/lambda/sqrt(n)
variance_t
```
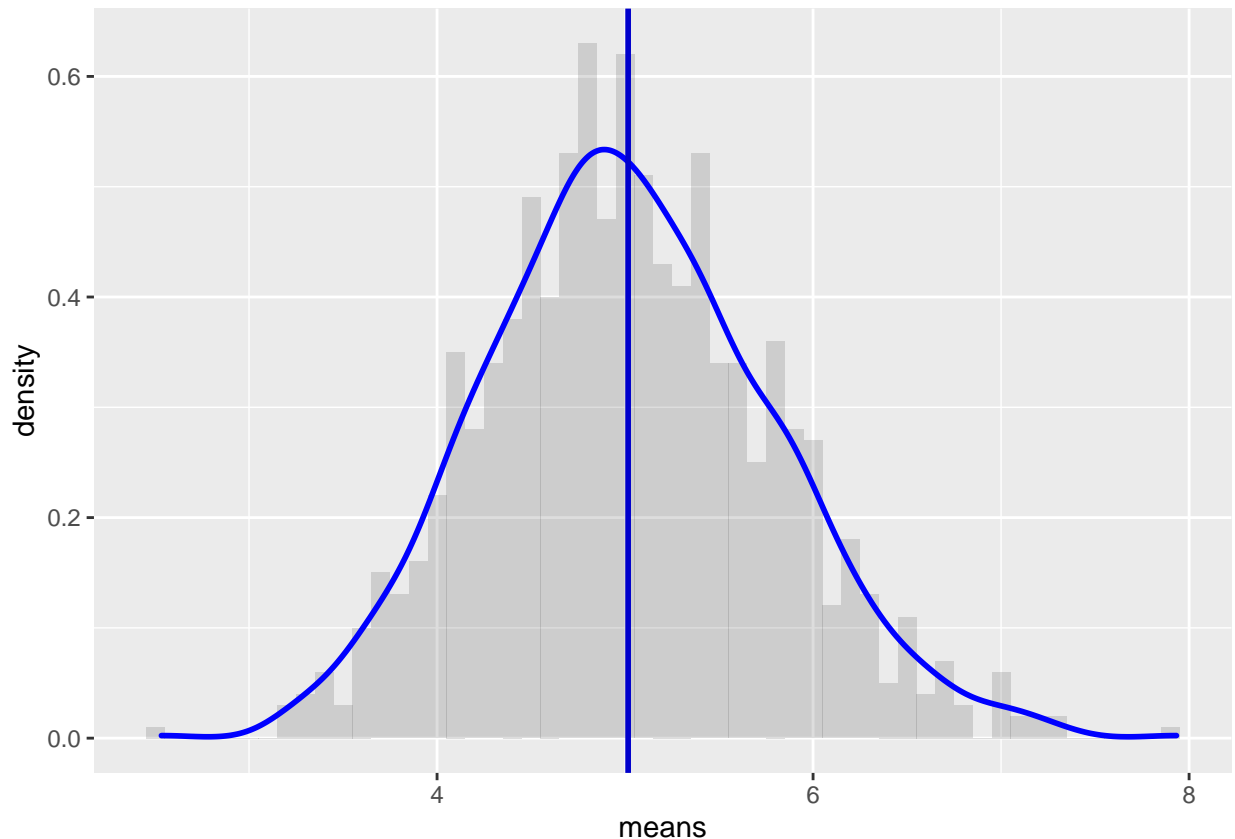
```
## [1] 0.7905694
```

The sample variance of distribution is as follows:

```
variance_S <- sd(ex_dis_m$means)
variance_S
```

```
## [1] 0.7600087
```

Similar to the means, the variance of distribution between both the sample and the theoritical are very similar.

##Distribution is Approximately Normal.

As shown in the graph above, the distribution of the sample mean shows a normal distribution. Note the line shows the exact mean of the sample data.

PART 2: Project Instructions:

Now in the second portion of the project, 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.

```
library(datasets)

data(ToothGrowth)
toothGrowth <- ToothGrowth
```

Summary of the data

Data Description: 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).

```
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 ...
```

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
```
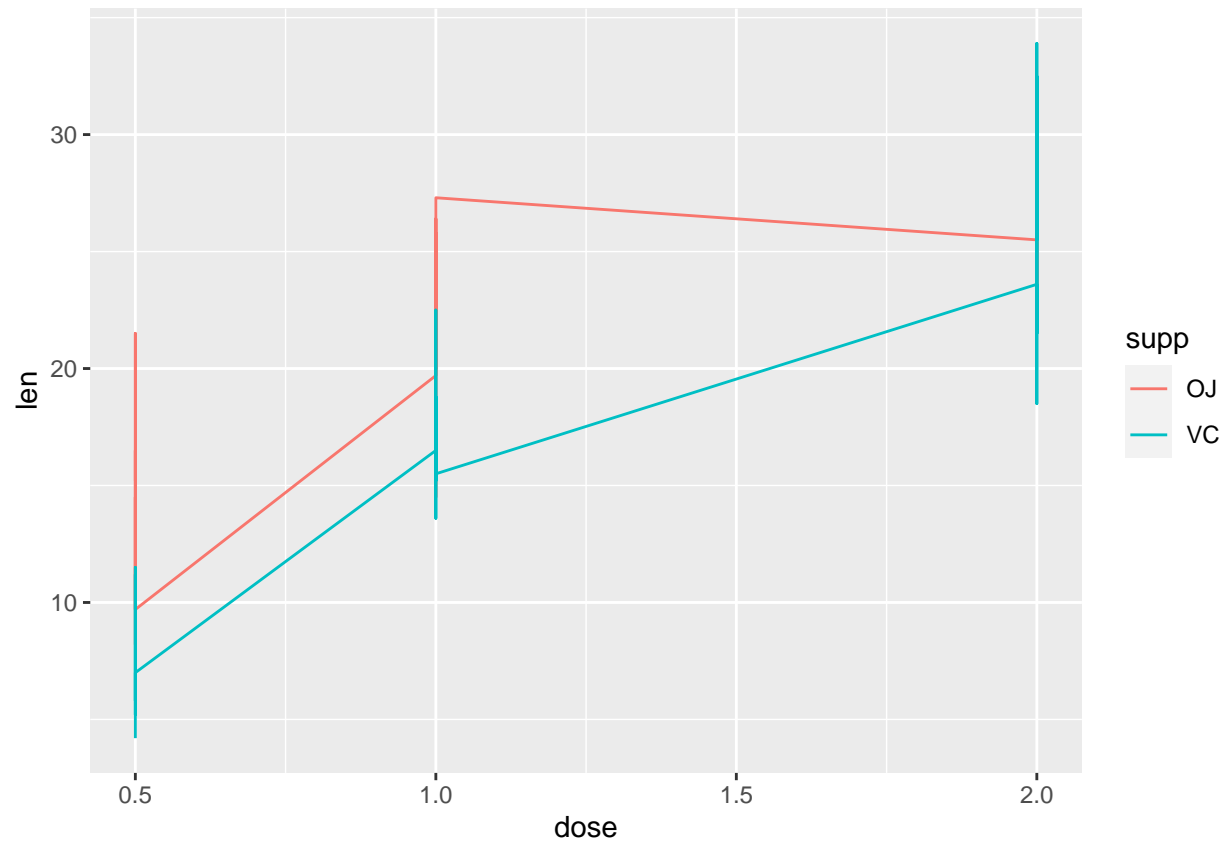
head(toothGrowth)

```
##    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
```

table(toothGrowth$supp, toothGrowth$dose)

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## supp         1  205.4   205.4  12.317 0.000894 ***
## dose         1 2224.3  2224.3 133.415  < 2e-16 ***
## supp:dose    1   88.9    88.9   5.333 0.024631 *
## Residuals   56  933.6    16.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova results indicate that there is a statistical significant association in the interaction betweeen orange juiCe (OJ) and ascorbic acid (VC). The linear graph showing the relationship between dosage amount and the response to legth of odontoblasts also further clarifies the relationship between OJ and VC. As shown in the graph we can see they are very similar while OJ being slightly more effective on guinea pigs.