Statistical Inference

Kaveri Chatra

Part 1: Investigating exponetial distribution and comparing with Central Limit Theorem

Loading Libraries

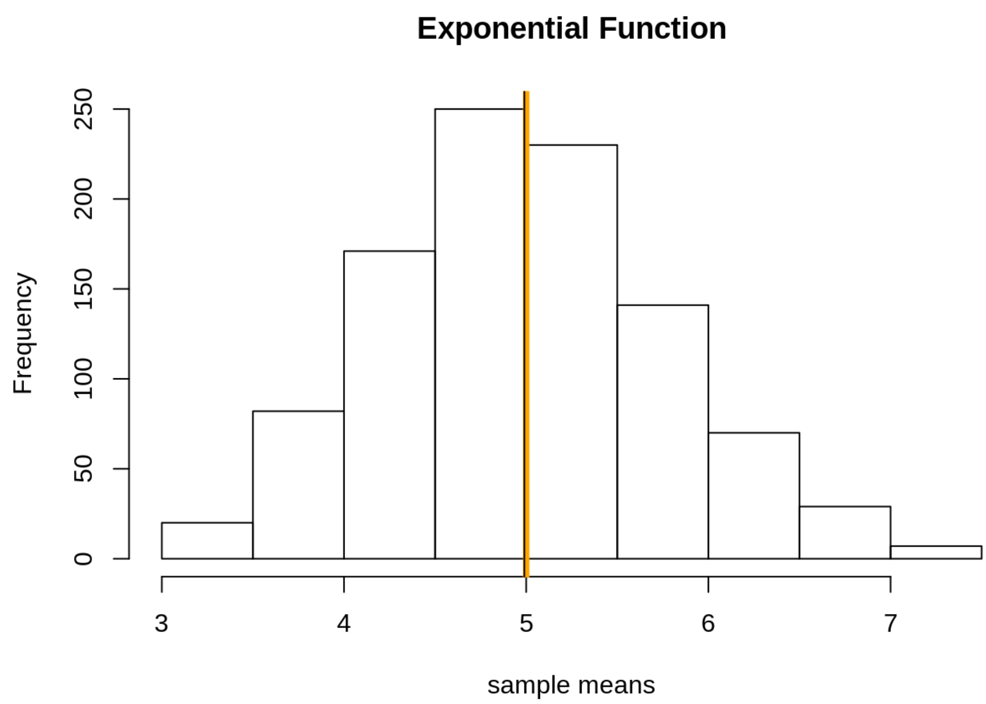
```
library("data.table")
library("ggplot2")
library("datasets")
library("rcompanion")
```

```
set.seed(1)
lambda <- 0.2
sim<- 1000
n <- 40


exp_sim <- replicate(sim, rexp(n, lambda))
means_exp <- apply(exp_sim, 2, mean)
```

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```
sample_mean <- mean(exp_sim)
theoretical_mean <-1/lambda
hist(means_exp, xlab = "sample means", main = "Exponential Function")
abline(v = theoretical_mean, col = "orange", lwd = 4)
abline(v = sample_mean, col = "black")
```



From the graph it can be noticed that sample mean is very close to theoretical mean

2. Show how variable it is and compare it to the theoretical variance of the distribution..

Standard deviation

```
sample_std <- sd(means_exp)
#For exp. distribution mean = std
theoretical_std <- 1/lambda/sqrt(n)
theoretical_std
```

```
## [1] 0.7905694
```

```
sample_std
```

```
## [1] 0.7817394
```

Variance

```
sample_var <- var(means_exp)
theoretical_var <- 1/lambda^2/n
theoretical_var
```

```
## [1] 0.625
```

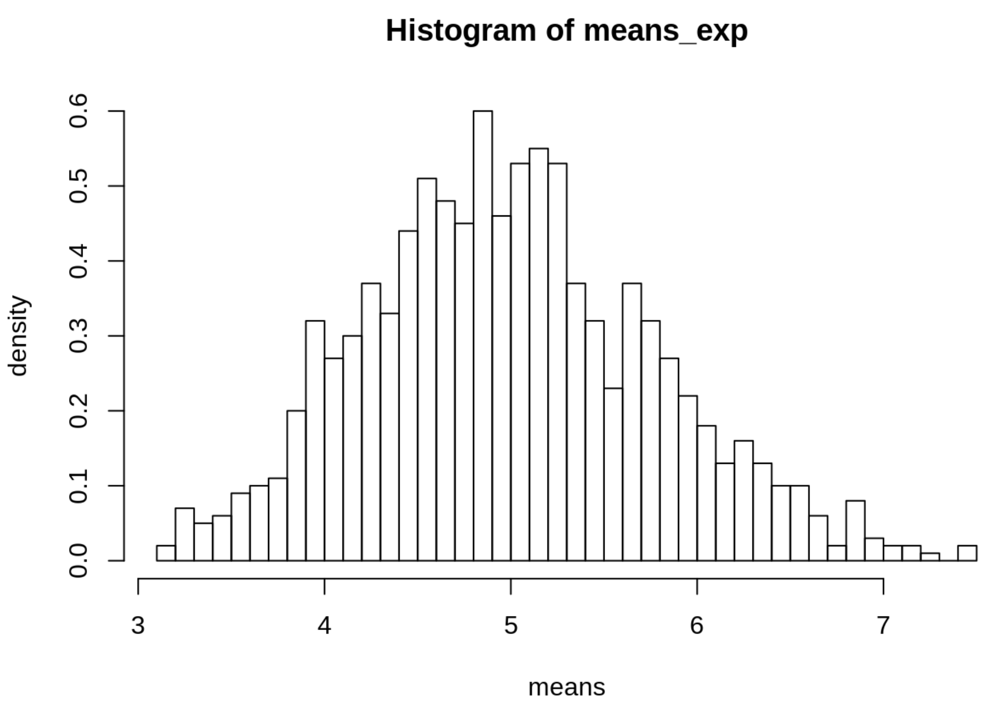
```
sample_var
```

```
## [1] 0.6111165
```

3. Show that the distribution is approximately normal.

From the Central Limit Theorem, the distribution of averages is often normal, even if the distribution that the data is being sampled from is non-normal.

```
par(mfrow = c(1,1))
hist(means_exp,breaks=n,prob=T,xlab = "means",ylab="density")
```




```
#curve(dnorm(x, 0, 1), -3, 3, col = 'blue',add=T)
#lines(density(scale(means_exp)), col = 'red')
```

Part 2: Basic Inferential Data Analysis

Loading and performing some exploratory data analysis

```
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
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
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)
```

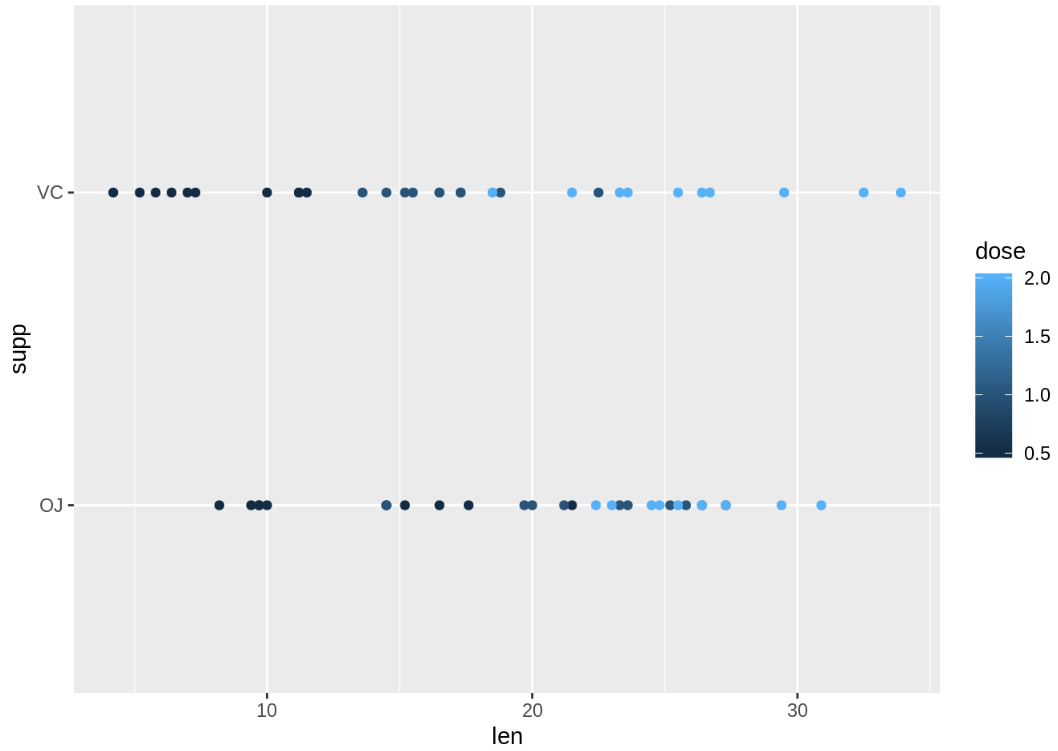
```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000
```

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

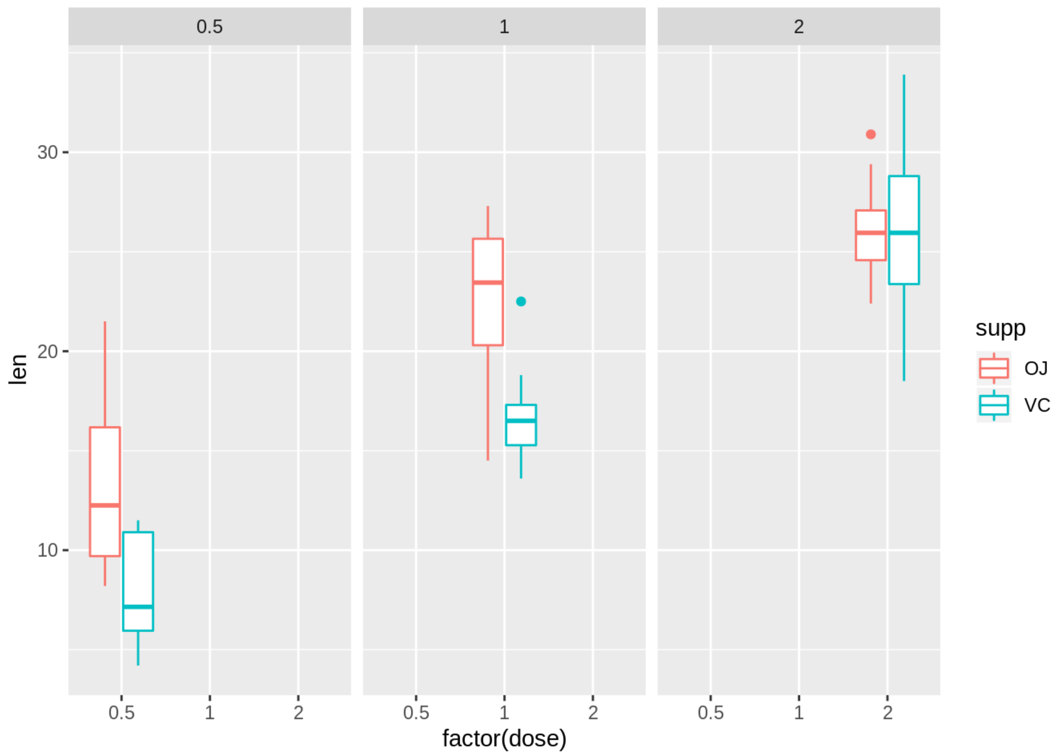
```
## [1] 0.5 1.0 2.0
```

Plots

```
g <- ggplot(ToothGrowth, aes(len, supp, color = dose)) + geom_point()
print(g)
```




```
g <- ggplot(ToothGrowth, aes(factor(dose), len, color = supp)) + geom_boxplot() + facet_grid(.~dose)
print(g)
```



Using confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

Confidence interval

```
groupwiseMean(len ~ supp, data = ToothGrowth, conf = 0.95, digits = 3)
```

```
##   supp  n Mean Conf.level Trad.lower Trad.upper
## 1   OJ 30 20.7       0.95       18.2       23.1
## 2   VC 30 17.0       0.95       13.9       20.0
```

```
groupwiseMean(len ~ factor(dose), data = ToothGrowth, conf = 0.95, digits = 3)
```

```
##   dose  n Mean Conf.level Trad.lower Trad.upper
## 1  0.5 20 10.6       0.95        8.5       12.7
## 2  1.0 20 19.7       0.95       17.7       21.8
## 3  2.0 20 26.1       0.95       24.3       27.9
```

```
groupwiseMean(len ~ factor(dose) + factor(supp), data = ToothGrowth, conf = 0.95, digits = 3)
```

```
##   dose supp  n  Mean Conf.level Trad.lower Trad.upper
## 1  0.5   OJ 10 13.20       0.95      10.00      16.40
## 2  0.5   VC 10  7.98       0.95       6.02       9.94
## 3  1.0   OJ 10 22.70       0.95      19.90      25.50
## 4  1.0   VC 10 16.80       0.95      15.00      18.60
## 5  2.0   OJ 10 26.10       0.95      24.20      28.00
## 6  2.0   VC 10 26.10       0.95      22.70      29.60
```

T test

Tooth length corresponding to OJ vs VC supplement. Null hypothesis being means of both are same and Alternate hypothesis being mean of tooth growth corresponding to OJ is greater than that of VC supplements.

```
OJ <- ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC = ToothGrowth$len[ToothGrowth$supp == 'VC']
t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y 
##  20.66333  16.96333
```

Tooth Growth by dose. Dose is divided into 3 groups, group1 = 0.5, group2 = 1.0, group3 = 2.0

```
group1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
group2 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
group3 <- ToothGrowth$len[ToothGrowth$dose == 2.0]

t.test(group1, group2, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  group1 and group2
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y 
##    10.605    19.735
```

```
t.test(group2, group3, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  group2 and group3
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

```
OJtwomg <- ToothGrowth$len[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 2.0]
VCtwomg <- ToothGrowth$len[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 2.0]

t.test(OJtwomg, VCtwomg, alternative = "two.sided", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJtwomg and VCtwomg
## 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
```

Conclusions

From the result of the T tests and confidence intervals, it can be concluded that:

```
1.  The supplement OJ had more impact than supplement VC on tooth growth.
2.  Higher the dose more is the growth.
3.  The impact on tooth growth 2.0 mg dose with VC supplement vs 2.0 mg dose with OJ supplement can't be det
ermined.
```