Statistical Inference Project

Guy Cole

October 2, 2016

Simulation Exercise

Overview

Investigate the exponential distribution in R and compare it w/the Central Limit Theorem.

As mandated by assignment, exponential distribution will be simulated in R using rexp(n, lambda) where lambda is the rate parameter. Distribution mean and standard deviation are 1/lambda and lambda = 0.2 for all simulations.

Assigment mandates investigation of "distribution of averages of 40 exponentials" w/1000 simulations.

Simulations

Initialization.

```
set.seed(54321)
lambda <- 0.2
n <- 40
simulationLimit <- 1000
standardDeviation <- 1/lambda
```

Sample Mean vs Theoretical Mean

```
simulation <- rexp(simulationLimit * n, rate = lambda)

simulatedMean <- mean(simulation)
simulatedMean
```

```
## [1] 4.990132
```

```
theoreticalMean <- 1/lambda
theoreticalMean
```

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

Sample Variance vs Theoretical Variance

```
datum <- NULL
for (ii in 1:simulationLimit) {
  datum <- c(datum, mean(rexp(n, lambda)))
}
```

Illustrate sample mean vs theoretical mean.

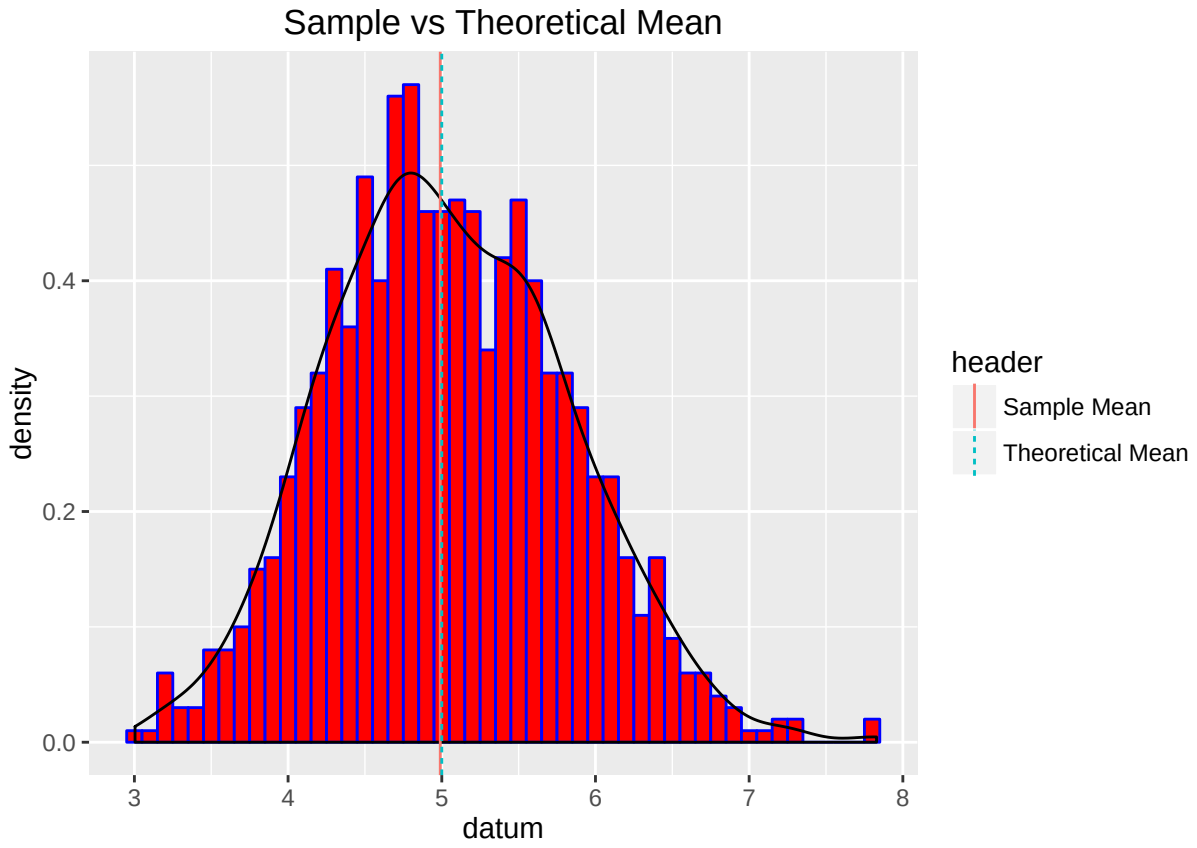
```
library(ggplot2)

dataFrame1 <- data.frame(header = c("Sample Mean", "Theoretical Mean"), values = c(simulatedMean, theor

gg <- ggplot(NULL, aes(x=datum))

plot1 <- gg + geom_histogram(aes(y=..density..), color="blue", fill="red", binwidth=0.1) + labs(title =

plot1
```



Standard deviation

```
sd(datum)
```

```
## [1] 0.7773352
```

Variance

```
varianceDatum <- sd(datum)^2
varianceDatum
```

```
## [1] 0.60425
```

Theoretical variance

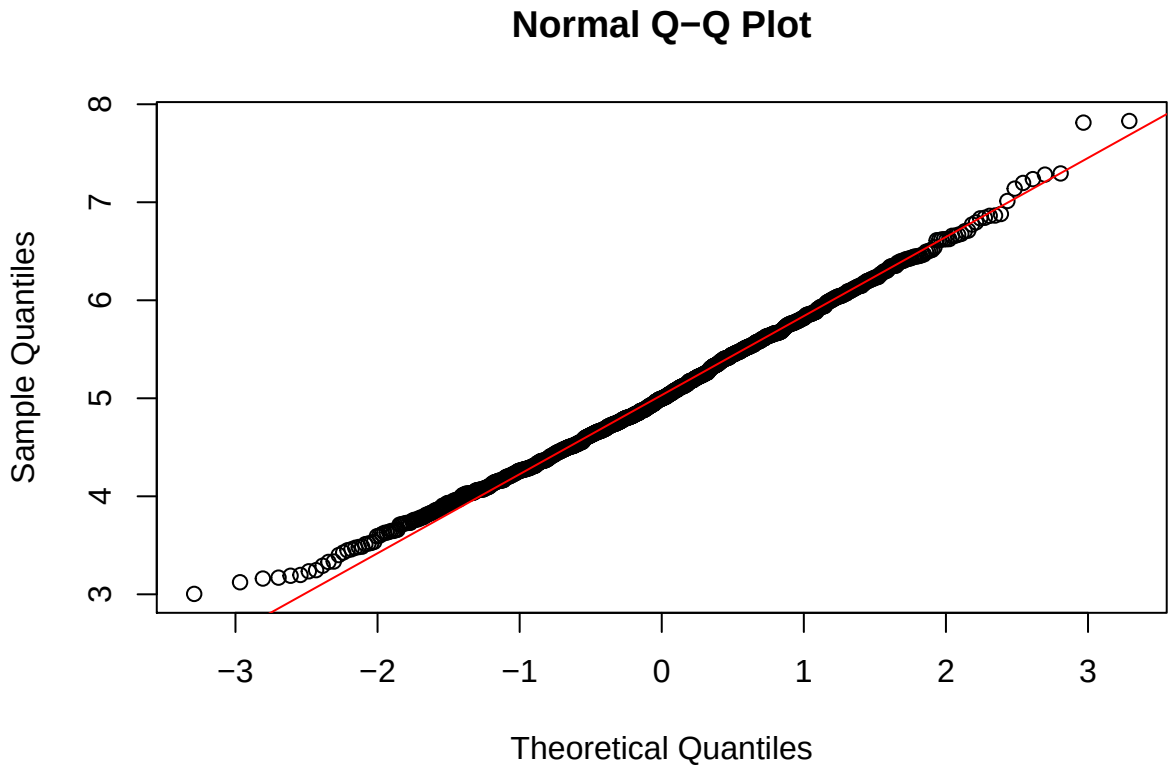
```
theoreticalDatum <- (1/lambda)^2/n
theoreticalDatum
```

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

Normal Distribution

The distribution plot (above) clearly exhibits gaussian bell curve shape, indicated normal distribution. Plot below uses qqnorm() to compare simulation to normal, indicating close approximation.

```
qqnorm(datum)
qqline(datum, col=10)
```



Tooth Growth

Overview

Explore the R datasets "ToothGrowth" data.

"ToothGrowth" captures the length of teeth in ten guinea pigs w/three dose levels of Vitamin C (0.5, 1 and 2 mg), delivered via orange juice or ascorbic acid.

```
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(knitr)
library(datasets)

data("ToothGrowth")
```

Data Dictionary

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

```
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 ...
```

```
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
```

```
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
```

Initial Analysis

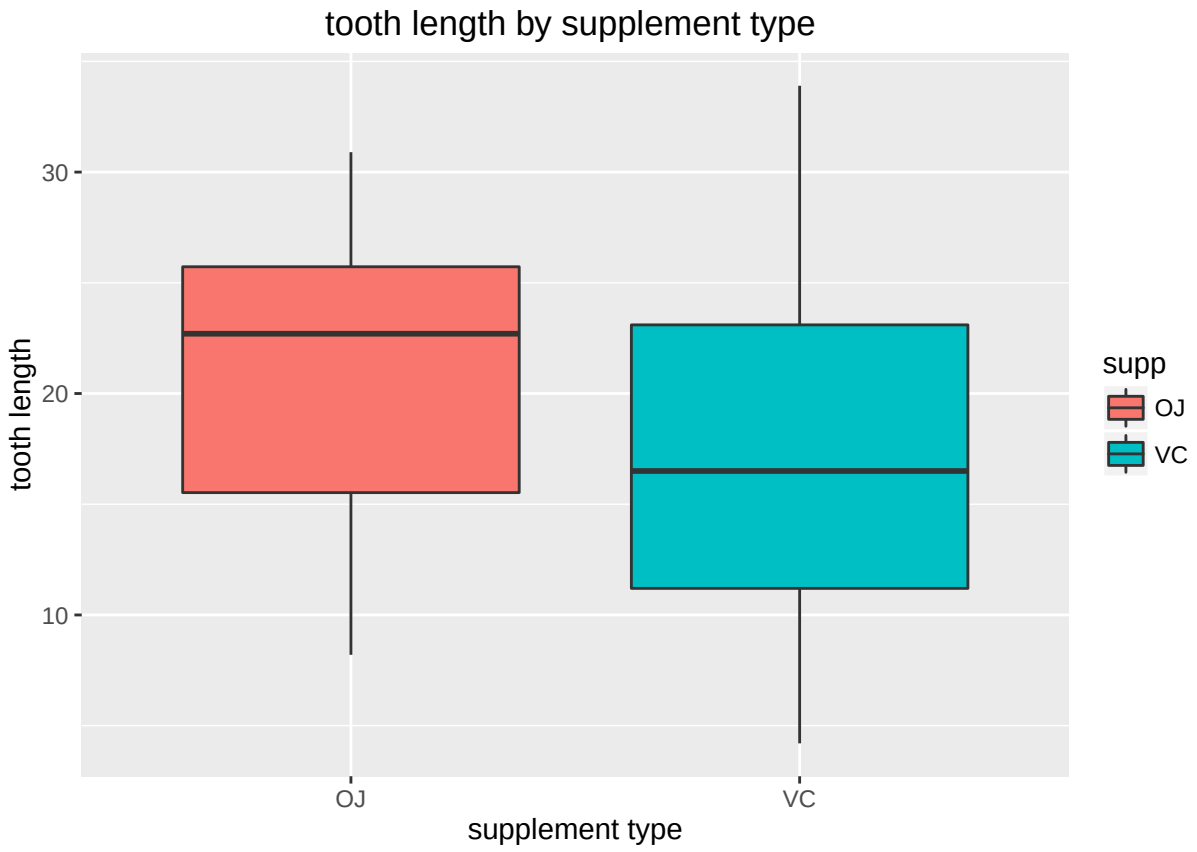
Initial analysis indicates that longer tooth length correlates to higher vitamin C dosage.

```
summaryDataFrame <- ToothGrowth %>% group_by(supp, dose) %>% summarise(count = n(), mean=mean(len), medi

kable (summaryDataFrame ,digits = 3, align = 'c')
```

supp	dose	count	mean	median	standard deviation
OJ	0.5	10	13.23	12.25	4.460
OJ	1.0	10	22.70	23.45	3.911
OJ	2.0	10	26.06	25.95	2.655
VC	0.5	10	7.98	7.15	2.747
VC	1.0	10	16.77	16.50	2.515
VC	2.0	10	26.14	25.95	4.798

```
ggplot(data=ToothGrowth, aes(x=supp, y=len))+
        geom_boxplot(aes(fill=supp))+
        xlab("supplement type") +
        ylab("tooth length")+
        ggtitle("tooth length by supplement type ")
```



Confidence Interval Testing

Employ t test to verify null hypotheses (both supply menthods produce same results).

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"], ToothGrowth$len[ToothGrowth$supp=="VC"], paired=FALSE, v
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

Return p-value of 0.06063 which is not large enough to reject null hypothesis. Conclusion is that tooth length not caused by supply method.

Test increasing dose from 0.5 to 1.

```
t.test(ToothGrowth$len[ToothGrowth$dose==0.5], ToothGrowth$len[ToothGrowth$dose==1], paired=FALSE, var.
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Dramatically lower p value (1.266e-07) implies dosage impacts mean.

```
t.test(ToothGrowth$len[ToothGrowth$dose==1], ToothGrowth$len[ToothGrowth$dose==2], paired=FALSE, var.equ
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Which also yields a very low p value (1.811e-05) which implies that increased vitamin C dosage encourages tooth growth.

Conclusion

Analysis using t-tests indicate that delivery methods do not impact tooth growth. Increased levels of vitamin C dosage encourages increased tooth growth.