STT 810

ICA 18

Aditya Jain

2022-11-08

Contents

1. Read the data into R.

```
setwd("C:\\Users\\Adi\\Desktop\\Fall-22 Study Material\\STT 810\\ICA")
data <- read.csv('hypodata.csv', stringsAsFactors = FALSE, header = TRUE)
data <- data$x
```

2. Find the mean, standard deviation, and standard error.

```
mn <-  mean(data)
sd <- sd(data)
serror  <- sd/sqrt(length(data))
paste(mn, sd, serror)
```

```
## [1] "11.8139804213029 35.3175404078189 1.1168386904374"
```

3. Assuming the data is normal, build a 99% (2-sided) confidence interval with the t distribution.

```
a3 <- mn + serror*qt(c(0.005, 0.995), length(data) - 1)
paste(a3)
```

```
## [1] "8.93168820768327" "14.6962726349226"
```

4. Next, use standard bootstrapping to build the 99% confidence interval.

```
datamean1 <- rep(0,10000)
datasd1 <- rep(0,10000)
for(i in 1:10000){
  datasamp <- sample(data, length(data), replace = TRUE)
  datamean1[i] <- mean(datasamp)
  datasd1[i] <- sd(datasamp)
}
paste(quantile(datamean1, c(0.005, 0.995)))
```

```
## [1] "9.51201233953652" "15.2758318153242"
```

5. Use the Bayesian bootstrapping to build the 99% confidence interval.

```r
library(DirichletReg)
```

```
## Warning: package 'DirichletReg' was built under R version 4.2.2
```

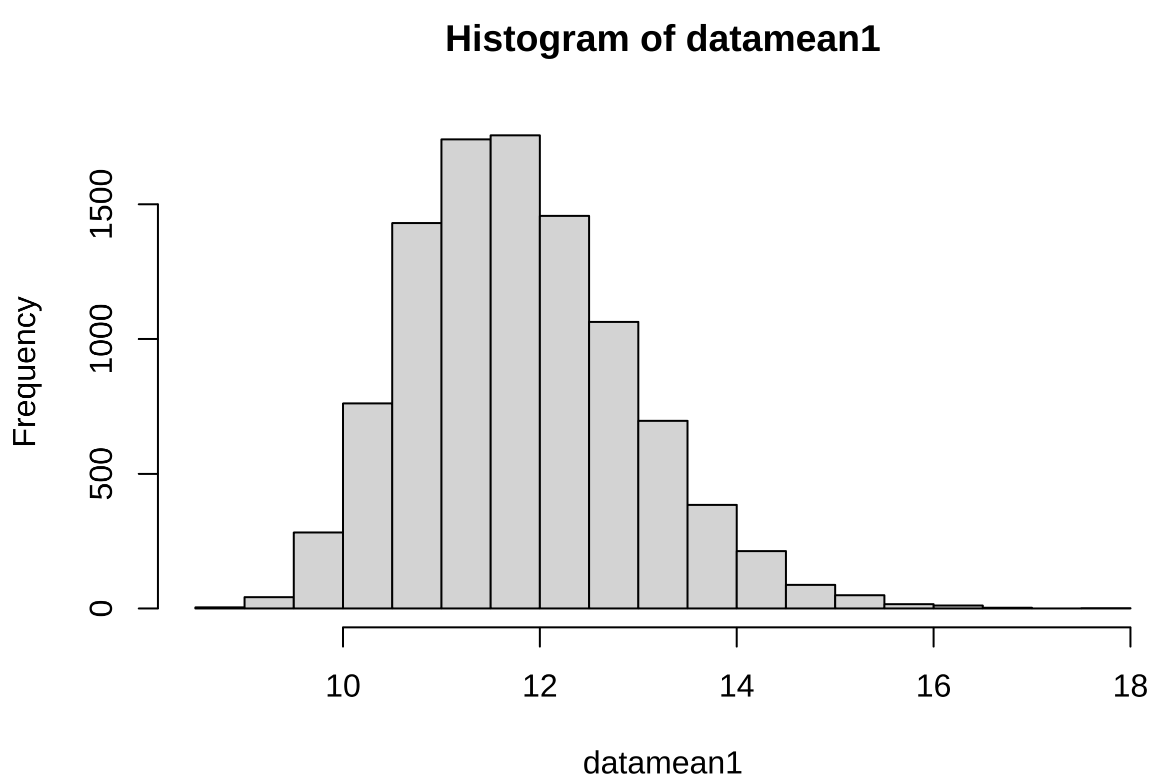
```
## Loading required package: Formula
```

```r
weight <- rep(0,length(data))
datamean2 <- rep(0,10000)
datasd2 <- rep(0,10000)
for(i in 1:10000){
  weight <- rdirichlet(1, rep(1,length(data)))
  datasamp <- sample(data, length(data), prob = weight, replace = TRUE)
  datamean2[i] <- mean(datasamp)
  datasd2[i] <- sd(datasamp)
}
paste(quantile(datamean2, c(0.005, 0.995)))
```

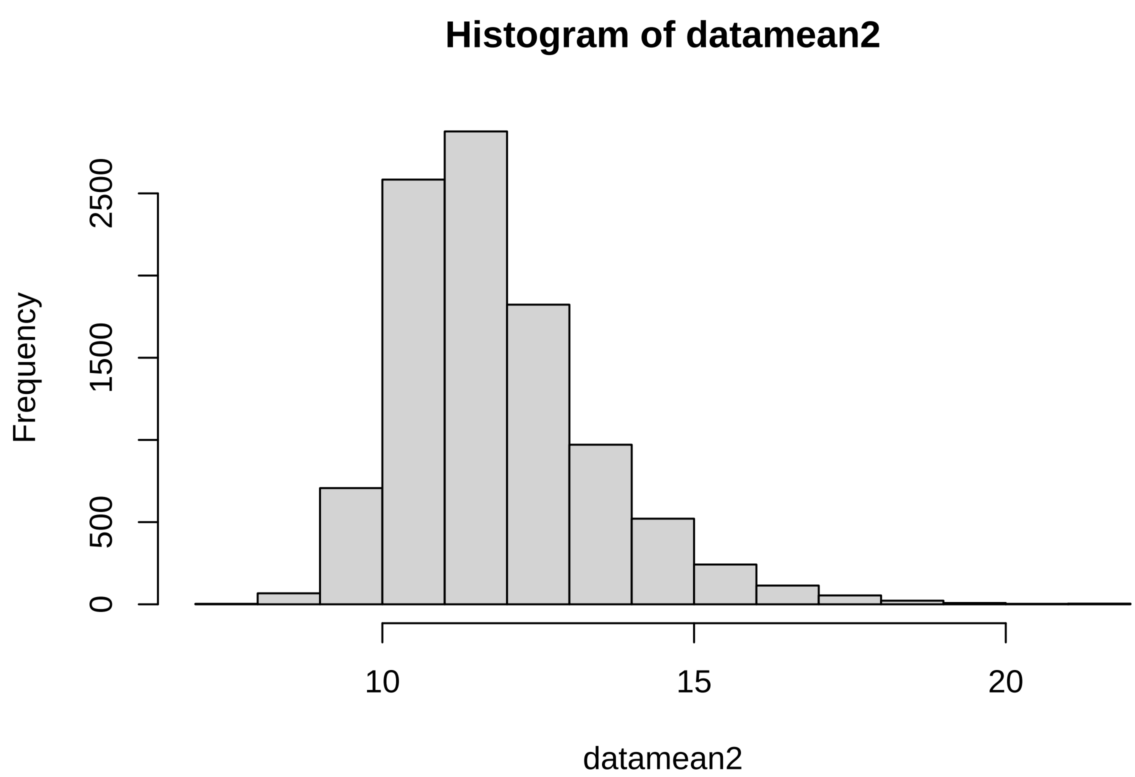
```
## [1] "8.89235027982231" "17.6013217873393"
```

6. Plot a histogram of the simulations in (4) and (5). Does it look like the simulated means fit a normal distribution?

```r
hist(datamean1, breaks = 20)
```



```r
hist(datamean2, breaks = 20)
```



The simulated mean does not fit the normal distribution.

7. How do the 3 confidence intervals compare?

The confidence interval increases from t distribution to normal bootstrapping to bayesian bootstrapping. Thus, we can say that bayesian method gives us more broder estimate of our confidence interval.

8. Next, use both standard and Bayesian bootstrapping to build a 99% CI for the standard deviation. Plot a histogram of each. Is the graph symmetric or skewed?

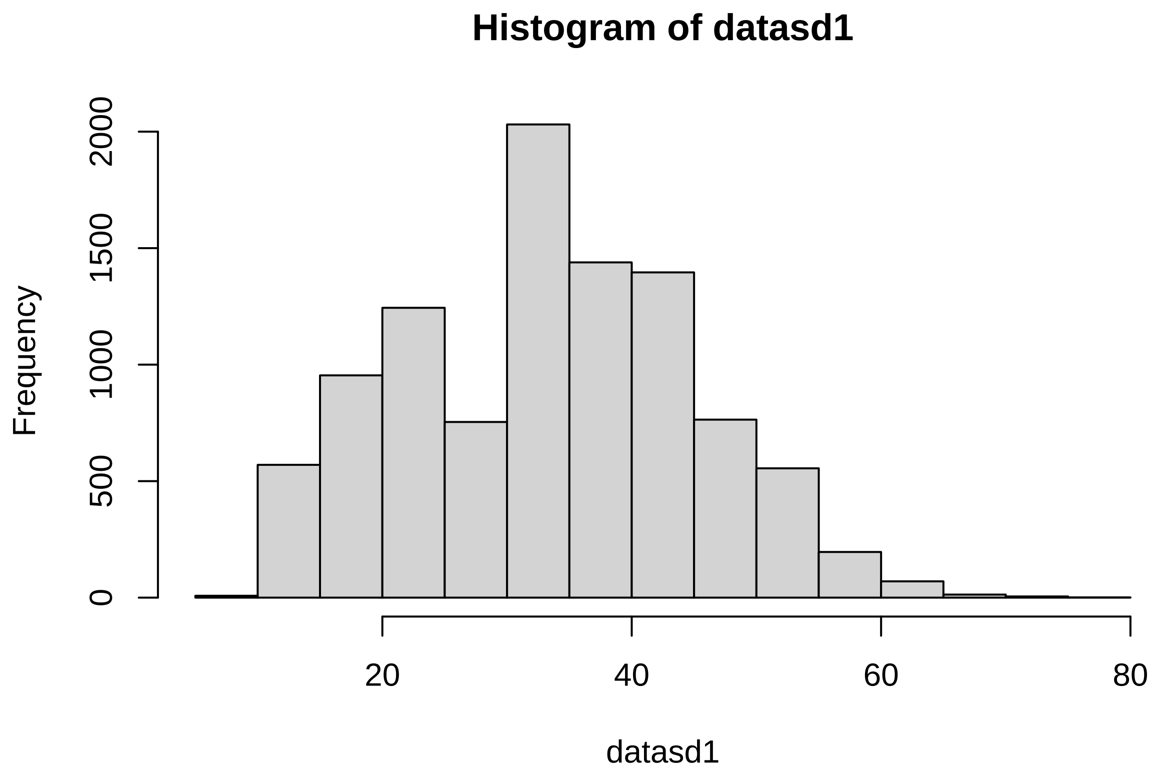
```r
paste(quantile(datasd1, c(0.005, 0.995)))
```

```
## [1] "11.3467440376771" "61.6527065931835"
```

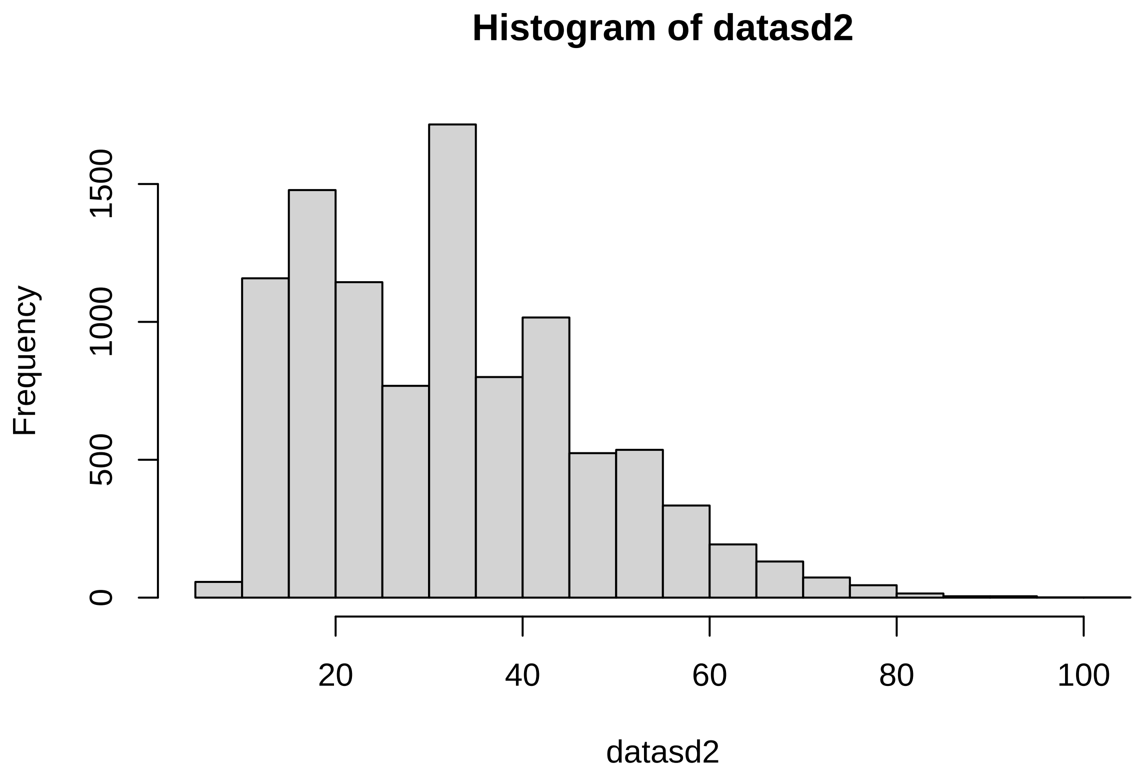
```r
paste(quantile(datasd2, c(0.005, 0.995)))
```

```
## [1] "9.85814139204787" "76.662711757753"
```

```r
hist(datasd1, breaks = 20)
```




```
hist(datasd2, breaks = 20)
```



The standard deviation is not normal as well.