Statistical Inference Report

2023-11-30

Part 1: Simulation Exercise

For the simulation we are using a lambda of 0.2, we are doing 1000 simulations each consisting of 40 exponentials. To achieve this we populate a matrix where each row is a simulation of 40 columns.

```
lambda <- 0.2
simulations <- 1000
n <- 40

set.seed(1)
data <- matrix(rexp(n*simulations, lambda), nrow = simulations, ncol = n)
```

Means

The mean of an exponential distribution is given by $1/\lambda$, in this case $1/0.2 = 5$.

```
1/lambda
```

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

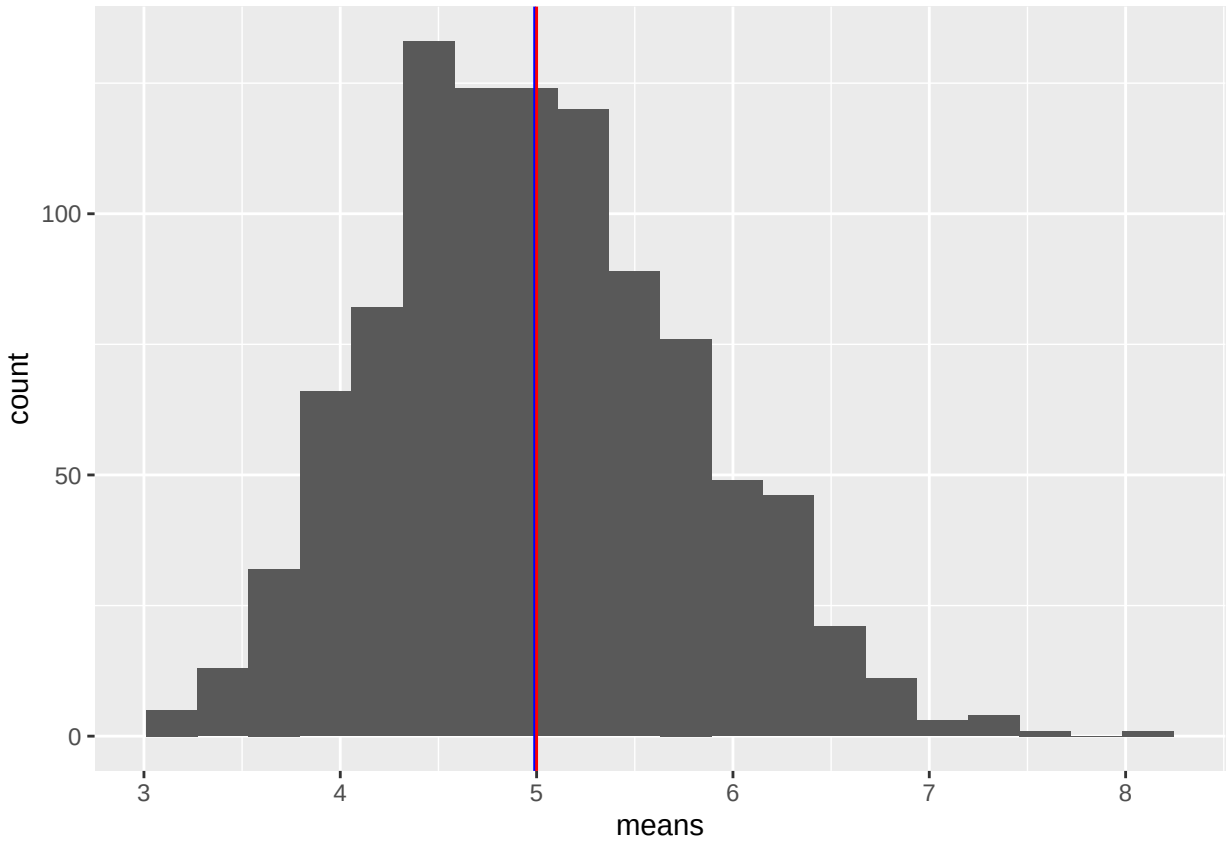
The sample mean of each simulation can be found using apply, we can then find the average of these to compare with the theoretical mean.

```
means <- apply(data, 1, mean)
mean(means)
```

```
## [1] 4.990025
```

Below is a histogram of all of the means of each simulation, the theoretical mean is shown by the red line, and the sample mean is shown in blue. We can see that despite some simulations having a mean far from the theoretical, overall the sample mean averages out to being almost identical to the theoretical mean.

```
ggplot(mapping = aes(means)) +
    geom_histogram(bins=20) +
    geom_vline(xintercept=mean(data), color="blue") +
    geom_vline(xintercept=1/lambda, color="red")
```



Variances

The variance of an exponential distribution is given by $1/\lambda^2$, in this case $1/0.2^2 = 25$.

```
1/lambda^2
```

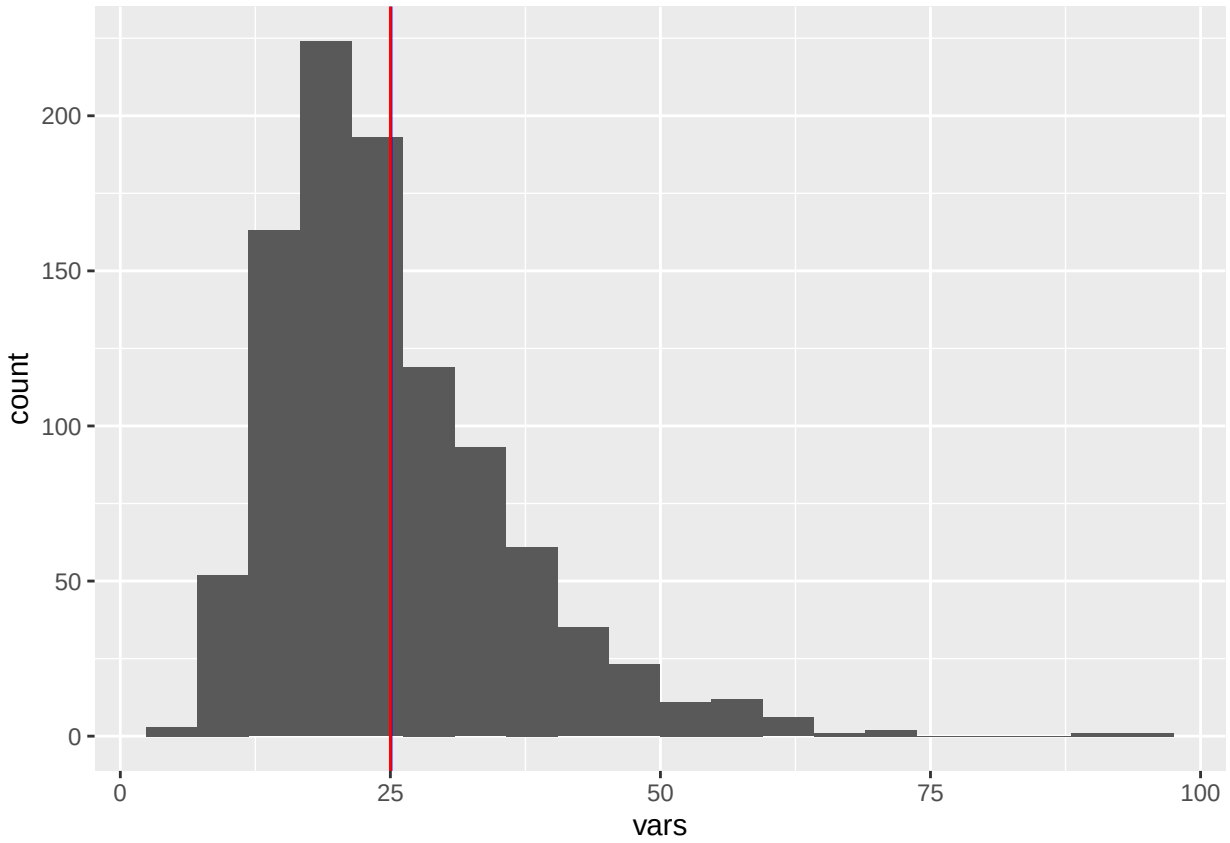
```
## [1] 25
```

We can find the variance of each simulation using apply, then we average this to find the average sample variance across simulations.

```
vars <- apply(data, 1, var)
mean(vars)
```

```
## [1] 25.05783
```

```
ggplot(mapping = aes(vars)) +
    geom_histogram(bins=20) +
    geom_vline(xintercept=mean(vars), color="blue") +
    geom_vline(xintercept=1/lambda^2, color="red")
```



Is the distribution normal

Part 2: Basic Inferential Data Analysis

The ToothGrowth dataset contains data that aims to show the effect of vitamin C on tooth growth in guinea pigs through observations of types of supplements and their dosage. It can be loaded into the ToothGrowth variable with data("ToothGrowth").

```
data("ToothGrowth")
```

It is a data frame containing 60 observations with 3 variables. len contains the length of the tooth, supp contains the suppliment type, and dose contains the dose in milligrams/day.

```
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 ...
```