Course 6 - Statistical Inference Week 4

Stephan Leijtens

Part 1 of Peer Assignment

1. Show the sample mean and compare it to the theoretical mean of the distribution.

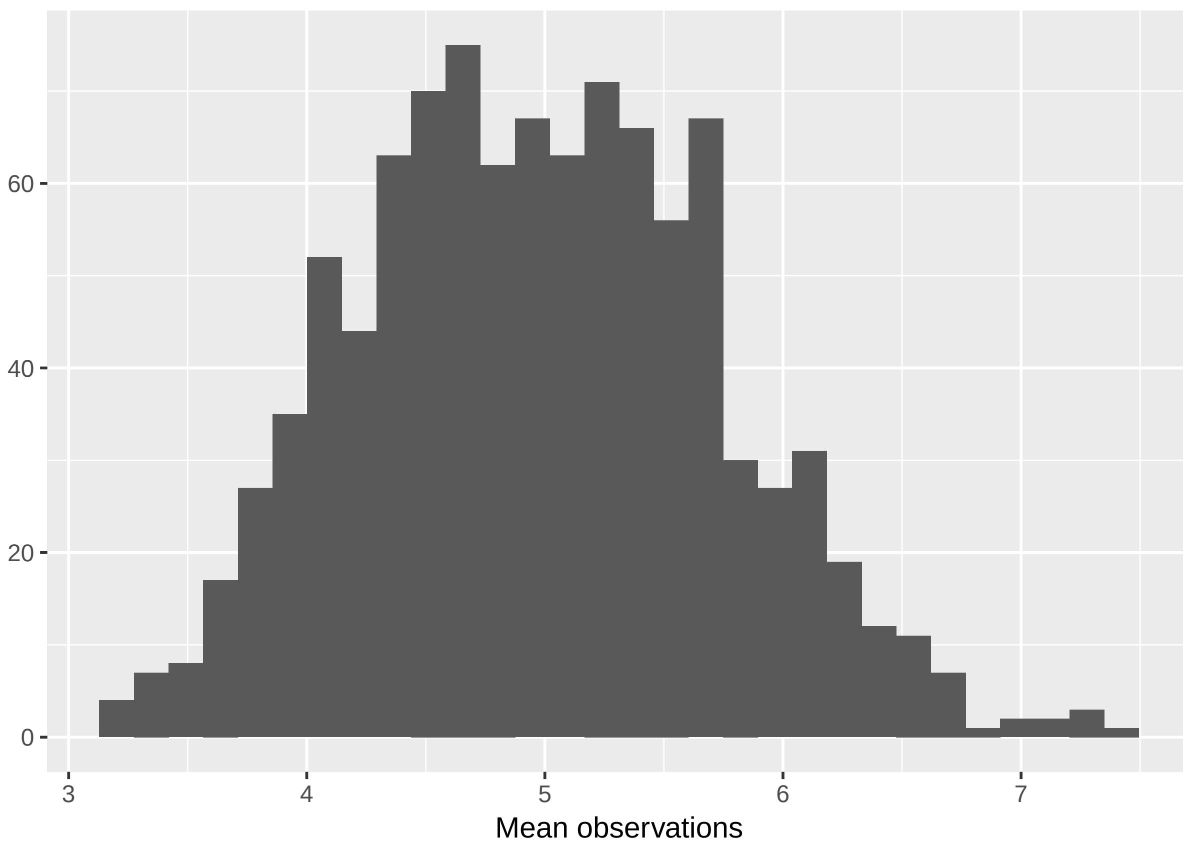
First we'll simulate the numbers and make a histogram to get a feeling of what kind of numbers we're working with.

```r
library(ggplot2)
set.seed(1234)

n <- 40
lambda <- 0.2
sim <- replicate(1000, rexp(n, .2))
sim_mean <- apply(sim, 2, mean)

qplot(sim_mean, geom = "histogram",
      xlab = "Mean observations")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Now we'll show the sample mean.

```
sample_mean <- mean(sim_mean)
print(sample_mean)
```

```
## [1] 4.974239
```

Let's look at the theoretical mean too.

```
theor_mean <- 1/0.2
print(theor_mean)
```

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

As you can see, the two are very similar, nearly identical.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

First we'll calculate the variance from the simulation means with give the sample variance.

```
var(sim_mean)
```

```
## [1] 0.5706551
```

Then the theoretical variance which is (lambda * sqrt(n))^-2.

```
(lambda * sqrt(n))^-2
```

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

Just a small difference between the two

```
abs(var(sim_mean)-(lambda * sqrt(n))^-2)
```

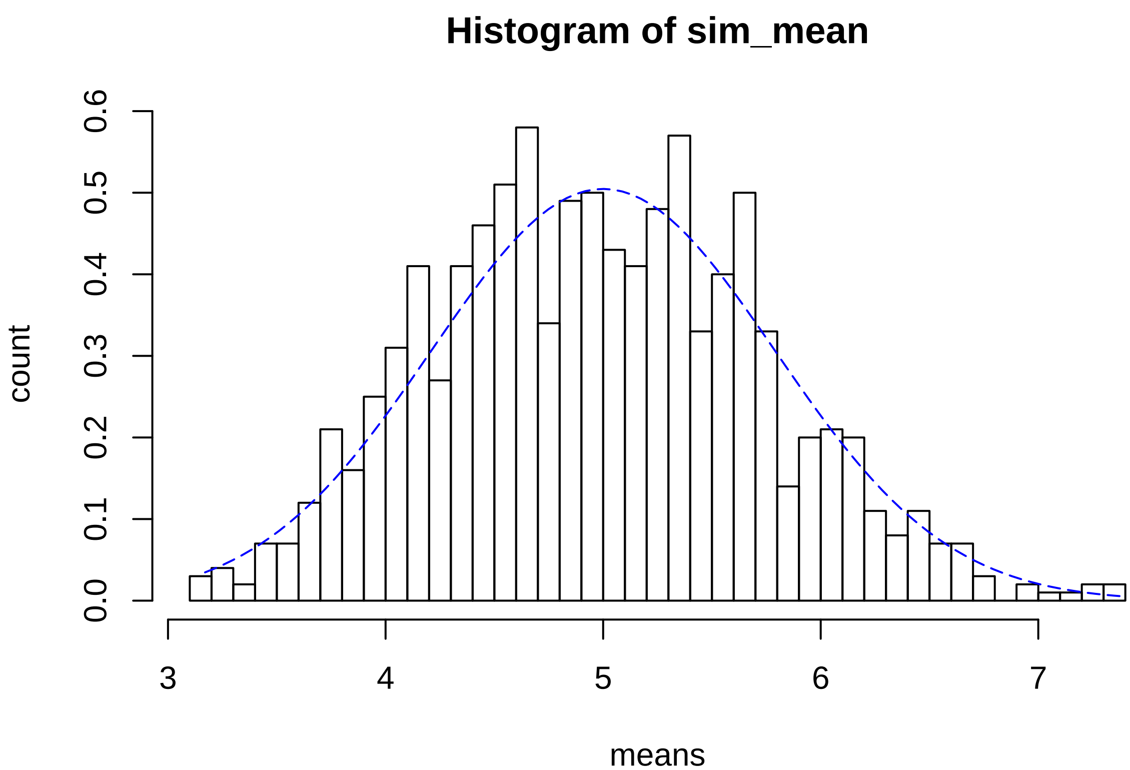
```
## [1] 0.05434495
```

3. Show that the distribution is approximately normal.

Let's plot the distribution of the simulation means and compare to the normal distribution (overlay, with a mean of lambda^-1 and standard deviation of (lambda*sqrt(n))^-1)

```
fit <- seq(min(sim_mean), max(sim_mean), length=100)
standard_fit <- dnorm(fit, mean=theor_mean, sd=(1/.2)/sqrt(n))

hist(sim_mean, breaks = n, prob=T, xlab = "means", ylab = "count")
lines(fit, standard_fit, pch=2, col="blue", lty=5)
```



As you can see, the means follow the normal distribution quite well.