Statistical inference assignment: part 1

MJM Beuken

5 november 2017

Preparing data, setting parameters and loading packages.

```
## load package
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
## set constants
labda <- 0.2
n <- 40
N.sim <- 1000

## set seed for reproducible results
set.seed(12345)


## run simulation in matrix
Exp.dist <- matrix(data=rexp(n*N.sim, labda), nrow=N.sim)
Exp.dist.mean <- data.frame(means=apply(Exp.dist, 1, mean))


## show result
head(Exp.dist, 5)
```

```
##            [,1]       [,2]     [,3]      [,4]      [,5]       [,6]
## [1,] 2.20903896  5.3732210 1.853568  5.248332 5.6777243  0.9139268
## [2,] 2.63736378 19.8099725 2.882970  5.511062 0.8340732 10.8226244
## [3,] 4.04233657  1.8314817 7.740680  4.445362 4.3855814  0.3543665
## [4,] 0.09224336 20.6291484 7.318112  2.706309 3.5726144  2.5923590
## [5,] 2.27705254  0.5725662 4.830426 14.312186 9.3860453 10.9072053
##           [,7]     [,8]       [,9]     [,10]      [,11]      [,12]
## [1,]  1.545202 3.844758 5.48546895 9.2284197  0.2272441  0.4383067
## [2,] 13.181986 4.876838 0.08253044 5.0027082  0.4681773  2.2016474
## [3,] 16.462073 5.921136 5.52945061 0.8884328  6.0283744  1.9596278
## [4,]  4.902970 1.679661 2.10463492 4.6469601 12.9116764  6.1761876
## [5,]  2.345180 8.542360 1.32601085 7.2669705  0.5884633 13.5684019
##          [,13]     [,14]       [,15]      [,16]        [,17]      [,18]
## [1,] 0.9279281  2.188139 0.963783227 11.8165028  4.150079773  1.5548021
## [2,] 6.4638531  8.767063 4.037945743  3.8683176  0.437579614 10.1253599
## [3,] 1.9286241  4.026319 0.009459571  0.7316528 13.398965140  0.2183817
## [4,] 7.2196560  1.019608 1.601369432  4.6552851  0.006633197  4.4851056
## [5,] 5.5088495 13.451511 3.422279239  1.7348434  1.737818259  1.3803048
##          [,19]      [,20]     [,21]        [,22]     [,23]     [,24]
## [1,] 4.1913093 17.2310240 20.511713  0.006410727 0.4990566 8.8103824
## [2,] 0.1299604  0.1884597  2.936695  3.889874336 2.4158663 0.2574604
## [3,] 3.1411451  9.8430821  3.317262 11.455022608 1.0897307 1.7799943
## [4,] 6.9202768 14.9090158 11.490750  1.139647070 5.0222555 0.5780644
```

```
## [5,] 5.4945848  4.2836313 11.843621  1.318925077 4.3050928 4.5825775
##           [,25]     [,26]    [,27]     [,28]     [,29]     [,30]    [,31]
## [1,]  1.7672793  1.788854 0.491512 7.0480526 6.2092529 3.6903036 8.834342
## [2,]  5.6026637 23.297202 2.561051 6.4519832 0.0274466 4.4535518 2.219804
## [3,]  4.3083472  3.076907 3.230082 4.6728406 9.0935352 4.4165963 0.809141
## [4,] 11.7108194  5.290598 6.837817 1.0919372 2.7117139 6.9822197 1.975513
## [5,]  0.9140074  2.283614 7.287269 0.7714448 6.0938054 0.5581431 8.129194
##          [,32]    [,33]     [,34]      [,35]       [,36]     [,37]
## [1,] 3.4015275 5.912150 0.1290731  5.4557445 5.140924253 0.2914339
## [2,] 4.1631424 2.216768 0.7357926  1.0890007 2.510215573 0.4944616
## [3,] 2.6404487 4.227883 1.0269441  0.3023570 1.230894695 2.1068146
## [4,] 2.1559663 0.173913 1.8172622  0.6863949 6.647168533 0.1804041
## [5,] 0.9452109 2.485991 4.0623063 12.5543250 0.008484203 0.3968482
##          [,38]     [,39]      [,40]
## [1,] 9.0624603  1.454240 13.8079820
## [2,] 2.3054006  1.462731  4.1114190
## [3,] 1.7465557 14.271670 10.0655435
## [4,] 2.5677886 16.868310  0.1983288
## [5,] 0.9870758  7.665424  1.3625874
```

```
head(Exp.dist.mean, 5)
```

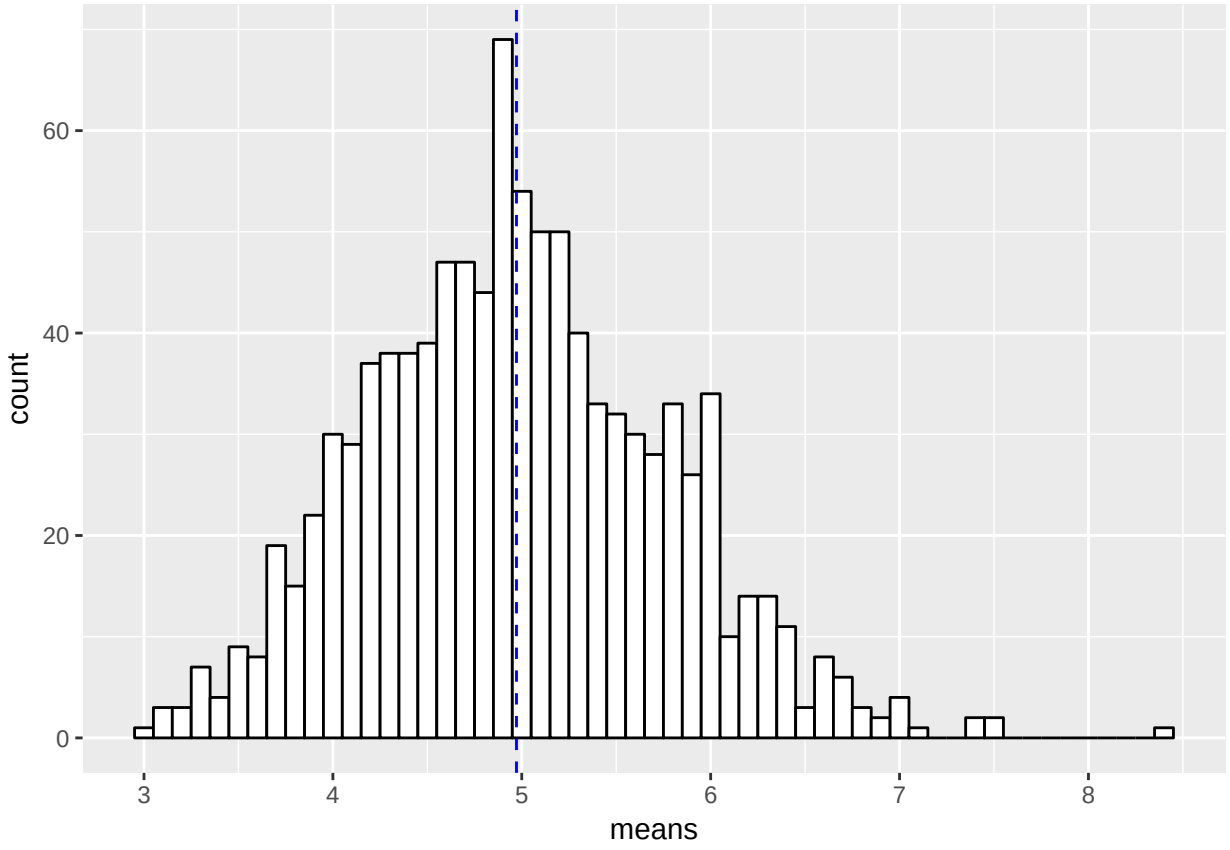
```
##      means
## 1 4.734537
## 2 4.388326
## 3 4.443878
## 4 4.906917
## 5 4.787316
```

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

plot histogram with a mean line.

```
p <- ggplot(Exp.dist.mean, aes(x=means)) +
        geom_histogram(binwidth=0.1, color="black", fill="white")


## Add mean line
p + geom_vline(aes(xintercept=mean(means)), color="blue", linetype="dashed", size=0.5)
```



Compare sample mean to theoretical mean

Setting theoretical parameters and calculating sample parameters.

```
mu <- 1/labda
var.theoretical <- (1/labda^2)/n

## calculating sample mean (x (bar))
x.bar <- mean(Exp.dist.mean$means)

## calculating sample variance s²
var.sample <- var(Exp.dist.mean$means)

## show parameters
mu
```

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

```
var.theoretical
```

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

```
x.bar
```

```
## [1] 4.971972
```

```
var.sample
```

```
## [1] 0.6157926
```

Compare theoretical mean and variance to sample mean and variance in table

```
Comparing <- matrix(c(mu, var.theoretical, x.bar, var.sample), ncol=2, byrow=TRUE)
colnames(Comparing) <- c("Mean", "Variance")
rownames(Comparing) <- c("Theoretical", "Sample")
Comparing <- as.table(Comparing)
Comparing
```

```
##                  Mean  Variance
## Theoretical 5.0000000 0.6250000
## Sample      4.9719720 0.6157926
```

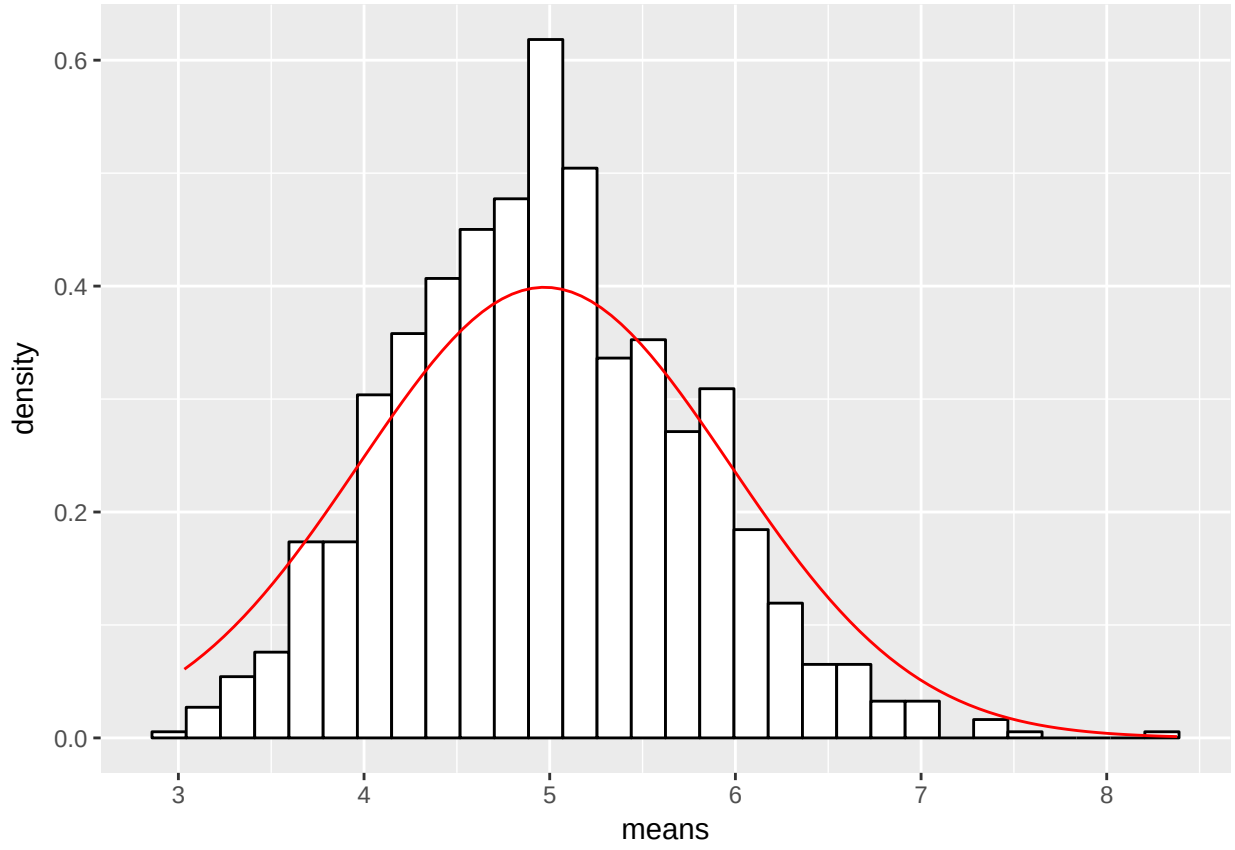
Show that the distribution is approximately normal.

Histogram with normal distribution to get a first graphical indication of normality of sample distribution.

```
ggplot(Exp.dist.mean, aes(x=means)) +
        geom_histogram(aes(y=..density..), colour="black", fill="white")+
        stat_function(fun=dnorm, color="red", args=list(mean=x.bar), sd=sqrt(var.sample))
```

```
## Warning: Ignoring unknown parameters: sd
```

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

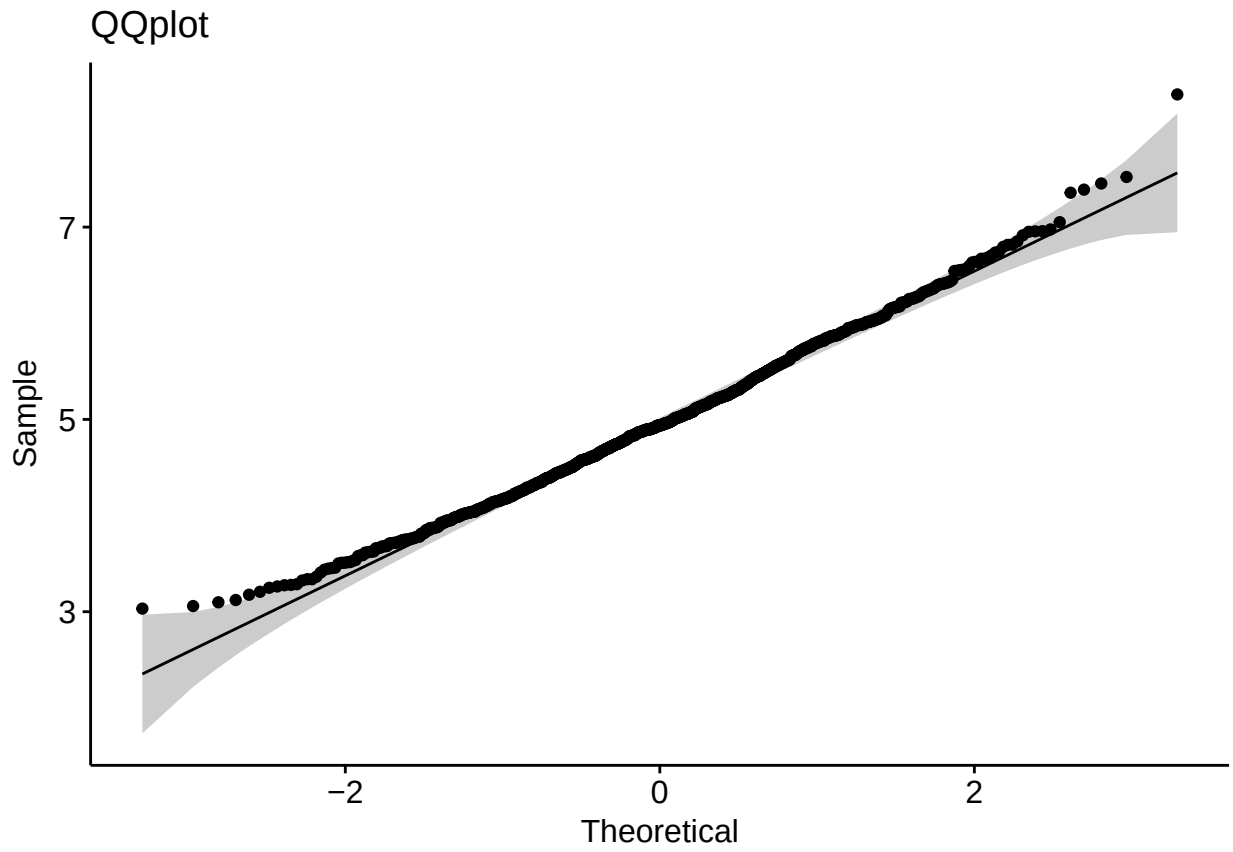



```
## Checking also with qqplot
library("ggpubr")
```

```
## Warning: package 'ggpubr' was built under R version 3.4.2
```

```
## Loading required package: magrittr
```

```
ggqqplot(Exp.dist.mean$means) + ggtitle("QQplot")
```



###As all the points lay approximately along the reference line, we can assume normality.

Using Kolmogorov-Smirnov goodness of fit test to compare the sample distribution to the normal distribution based on theoretical parameters.

Hypothesis under alpha = 5%:

Ho: The sample data are not significantly different than a normal population.

Ha: The sample data are significantly different than a normal population.

```
ks.test(Exp.dist.mean, pnorm, mu, sqrt(var.theoretical))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  Exp.dist.mean
## D = 0.042807, p-value = 0.05122
## alternative hypothesis: two-sided
```

p-value = 0.05122 > 0.025; which means that Ho is failed to be rejected under a 5% probability of a type 1 error.

The sample distribution isn't significantly different than a normal distribution.