Assignment 0: Practicing R

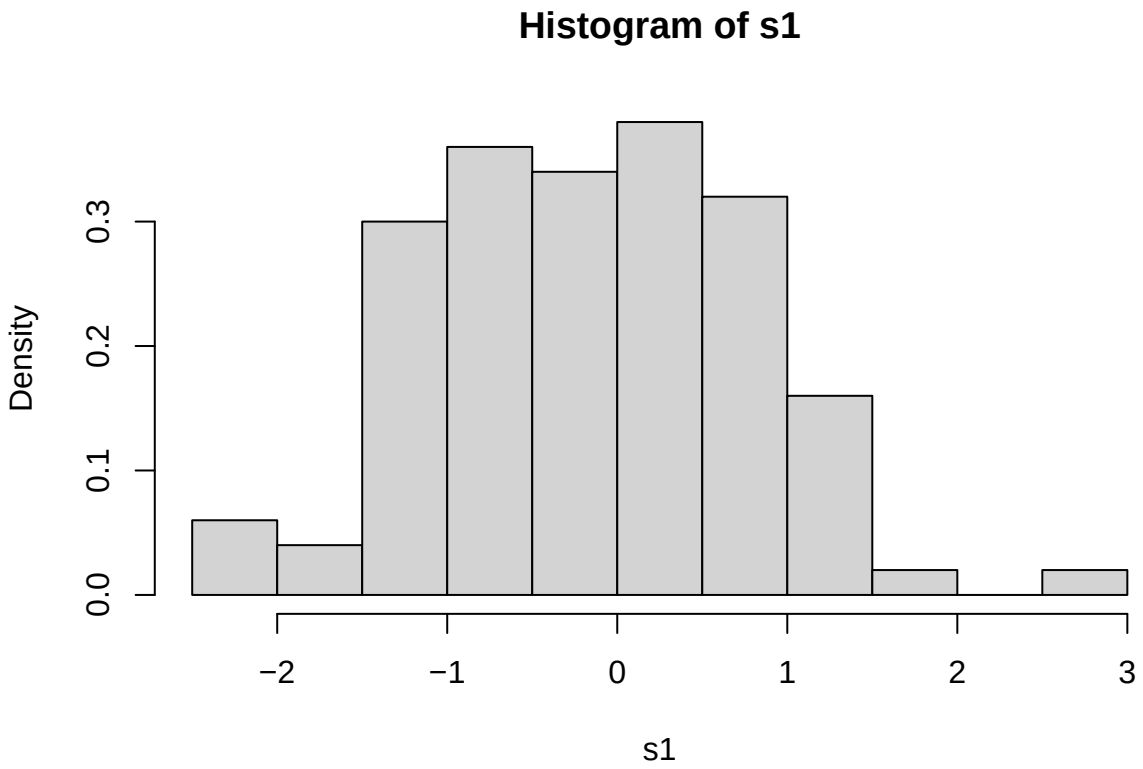
Exercise 1

Practicing working with normal distributions.

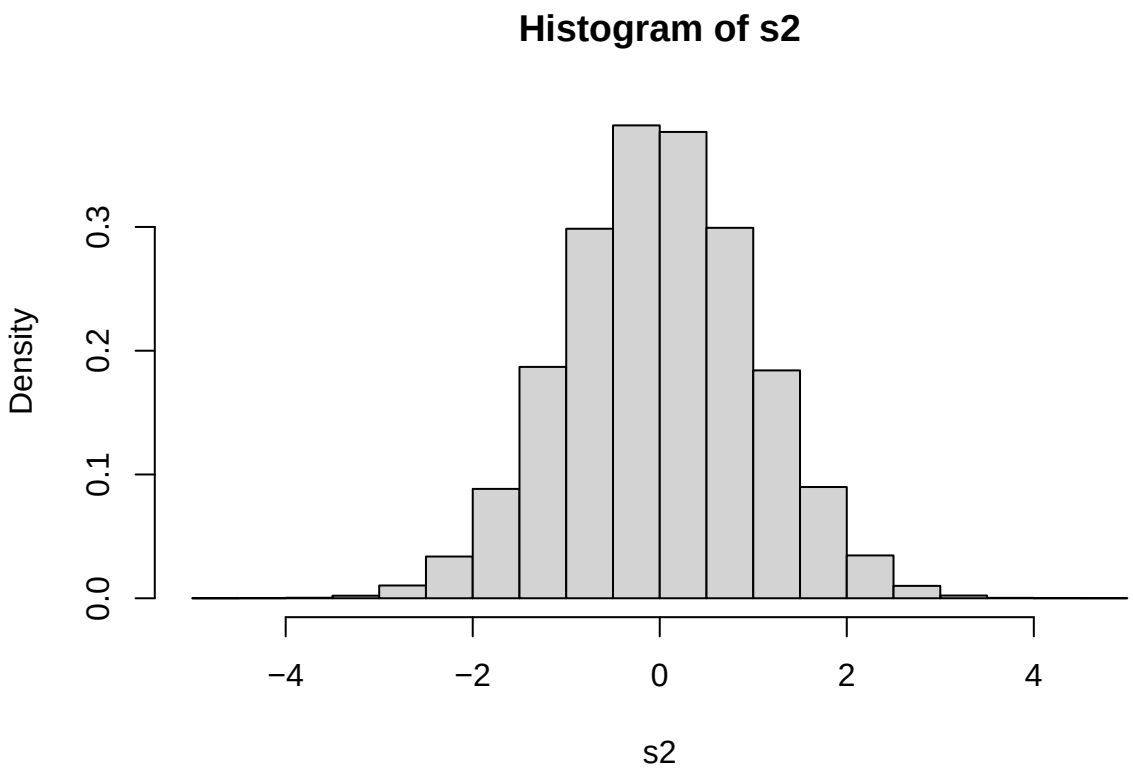
Part (a):

Generate two samples of sizes 100 and 100000 from a standard normal distribution. Make histograms
and QQ-plots, compute the means and standard deviations of the both samples. Explain your findings

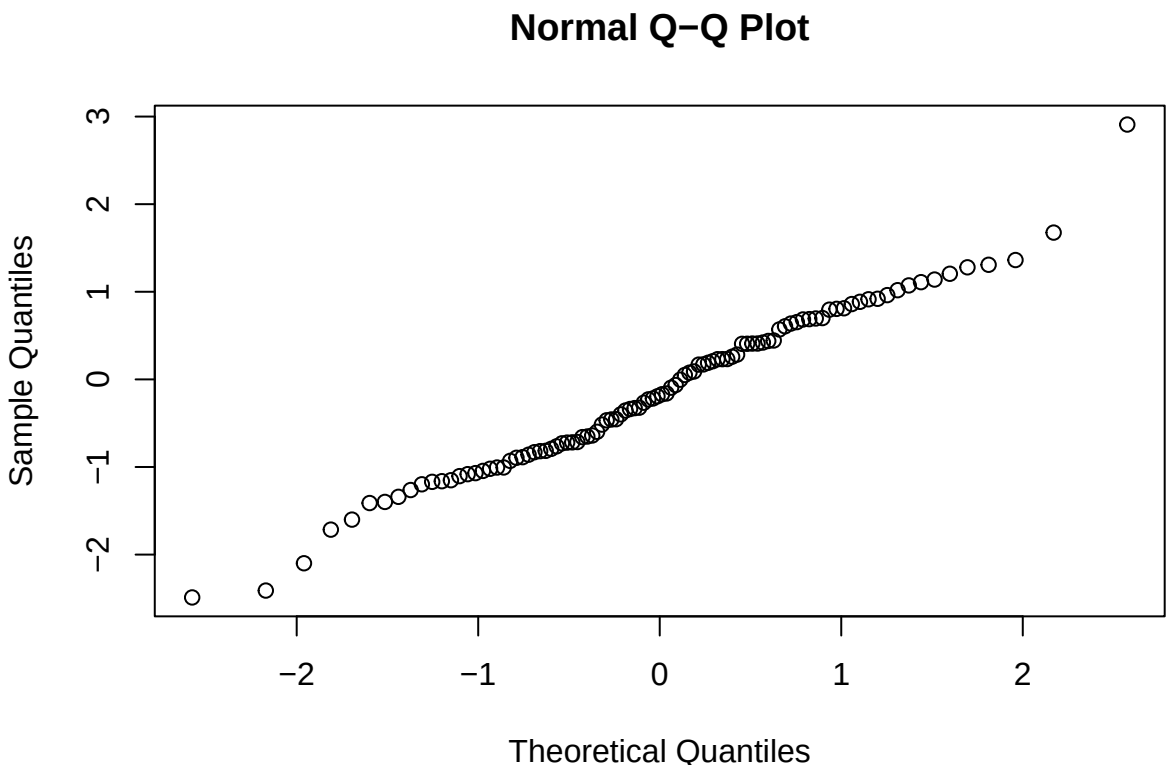
```
# generate two samples from std. norm. dist.
s1 = rnorm(100)
s2 = rnorm(100000)

# make histograms
hist(s1, freq=F)
```



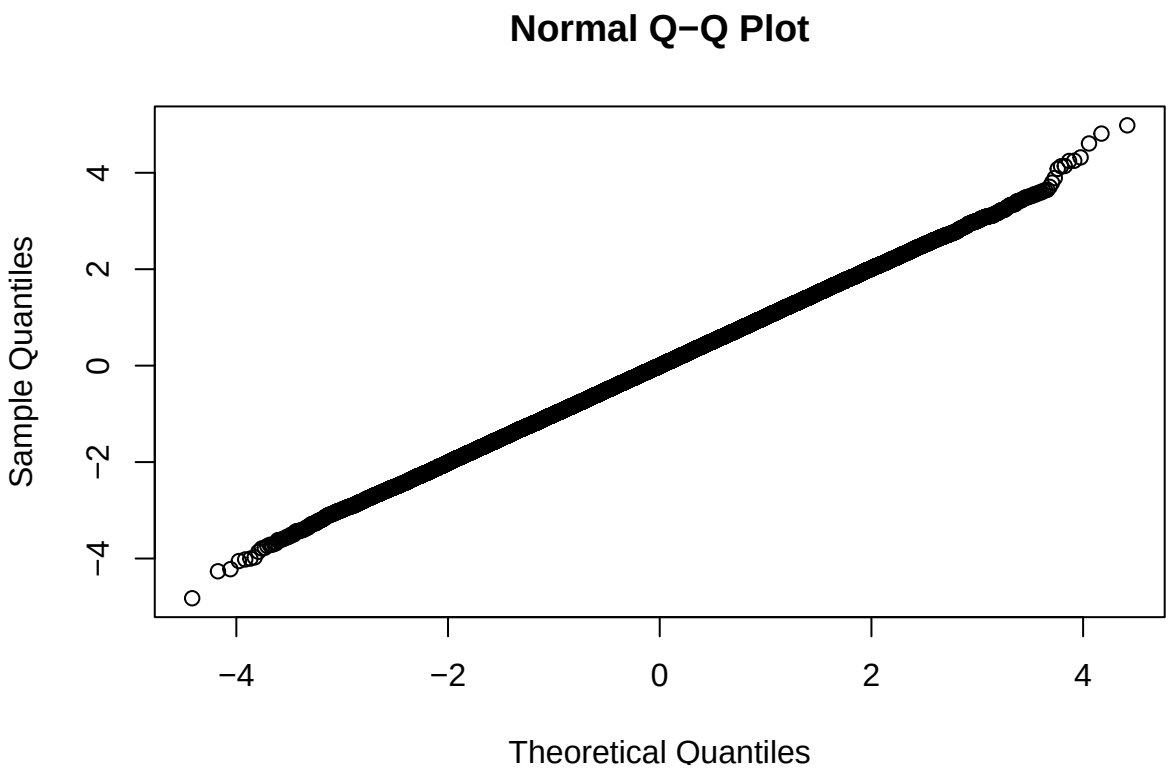

```
hist(s2, freq=F)
```




```
# make qqplots to verify samples appear to come from norm. dist.
qqnorm(s1)
```




```
qqnorm(s2)
```



The findings suggest that as expected, the both samples appear to come from the standard normal distribution. The bell shaped curves of the histograms, as well as the straight line in the QQ plots provide evidence for this.

Part (b) and (c):

For a standard normal distribution, compute the following 3 probabilities: that an arbitrary outcome
is smaller than 2, that it is bigger then -0.5 and that it is between -1 and 2.

Can you verify the outcomes of b) using only the data from a)?

```
# get prob a single value (from std. norm.) is < -0.5

prob1 = pnorm(-0.5)
sprintf("prob (x < -0.5) = %.4f", prob1)
```

```
## [1] "prob (x < -0.5) = 0.3085"
```

```
prob2 = pnorm(2) - pnorm(-1)
sprintf("prob (-1 < x < 2) = %.4f", prob2)
```

```
## [1] "prob (-1 < x < 2) = 0.8186"
```

```
# note that qnorm does the opposite:
#  qnorm(pnorm(-0.5)) = -0.5

# estimate these probabilities using data from samples:
est1a = sum(s1 < -0.5) / length(s1)
est1b = sum(s2 < -0.5) / length(s2)
sprintf("estimate 1a: %.4f, estimate1b: %.4f", est1a, est1b)
```

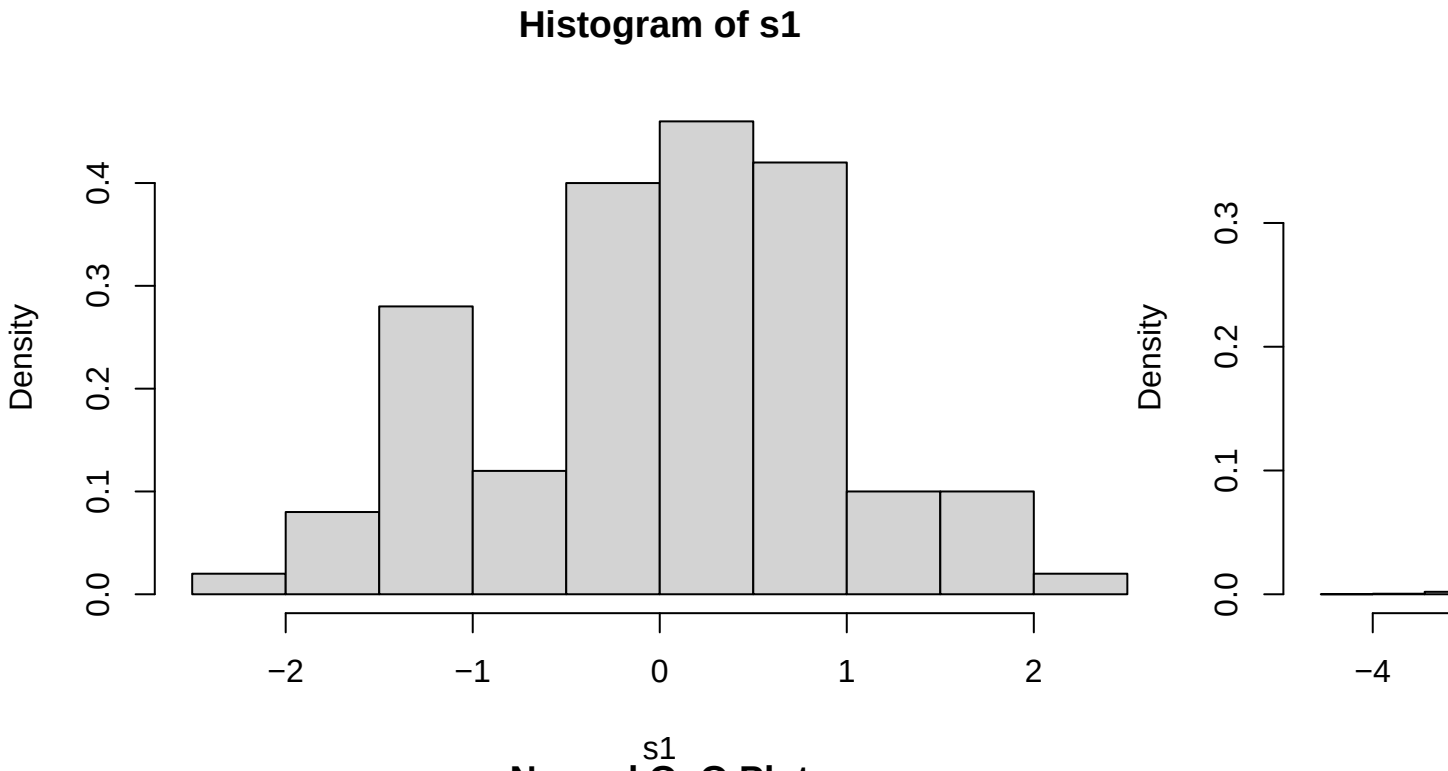
```
## [1] "estimate 1a: 0.3800, estimate1b: 0.3103"
```

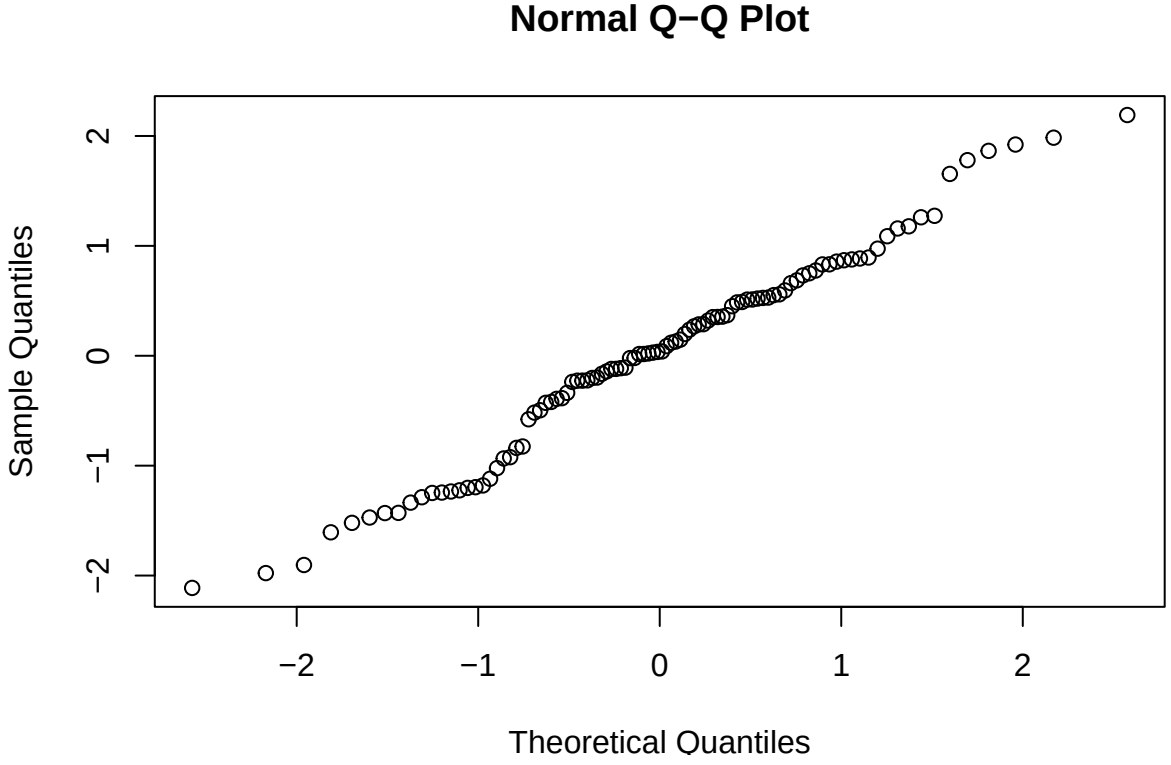
```
est2a = (sum(s1 < 2) - sum(s1 < -1)) / length(s1)
est2b = (sum(s2 < 2) - sum(s2 < -1)) / length(s2)
sprintf("estimate 2a: %.4f, estimate2b: %.4f", est2a, est2b)
```

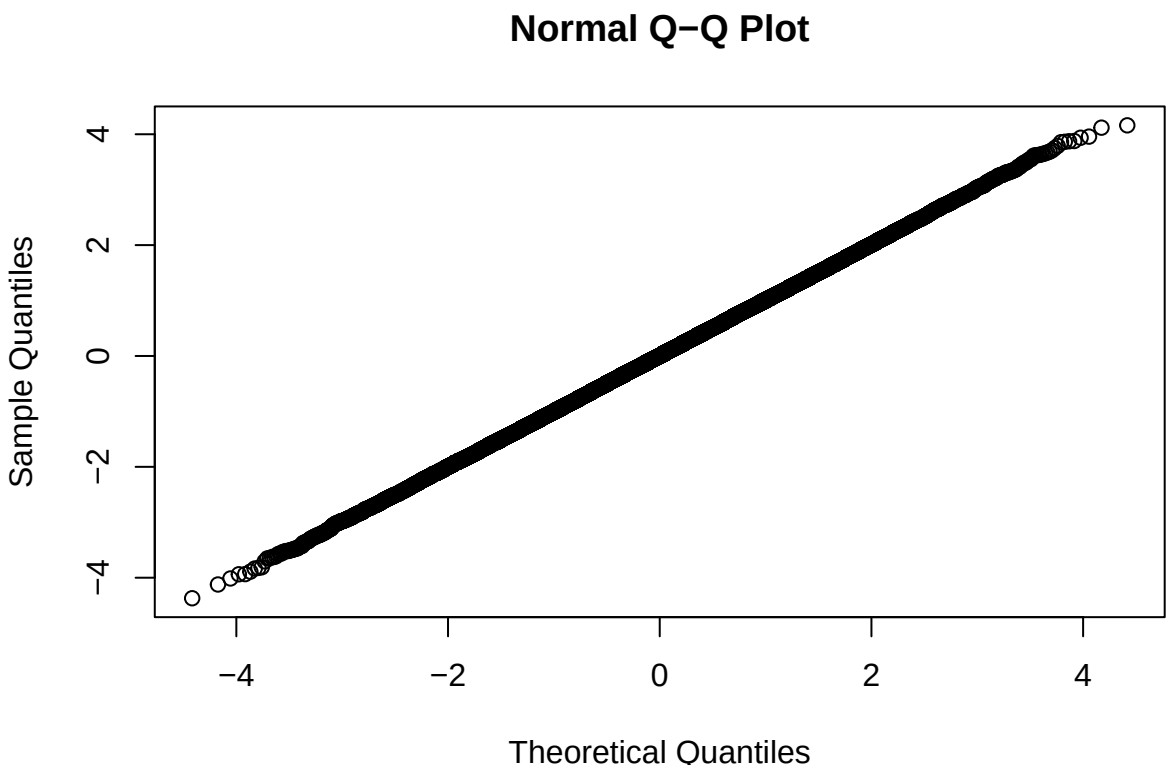
```
## [1] "estimate 2a: 0.7900, estimate2b: 0.8154"
```

Part (d):

```
part1 = function(mean=0, sd=1) {
  s1 = rnorm(100)
  s2 = rnorm(100000)

  # make histograms
  hist(s1, freq=F)
  hist(s2, freq=F)

  # make qqplots to verify samples appear to come from norm. dist.
  qqnorm(s1)
  qqnorm(s2)

  prob1 = pnorm(-0.5)
  print(sprintf("prob (x < -0.5) = %.4f", prob1))

  prob2 = pnorm(2) - pnorm(-1)
  print(sprintf("prob (-1 < x < 2) = %.4f", prob2))

  # note that qnorm does the opposite:
  #  qnorm(pnorm(-0.5)) = -0.5

  # estimate these probabilities using data from samples:
  est1a = sum(s1 < -0.5) / length(s1)
  est1b = sum(s2 < -0.5) / length(s2)
  print(sprintf("estimate 1a: %.4f, estimate1b: %.4f", est1a, est1b))

  est2a = (sum(s1 < 2) - sum(s1 < -1)) / length(s1)
  est2b = (sum(s2 < 2) - sum(s2 < -1)) / length(s2)
  print(sprintf("estimate 2a: %.4f, estimate2b: %.4f", est2a, est2b))
}

part1()
```

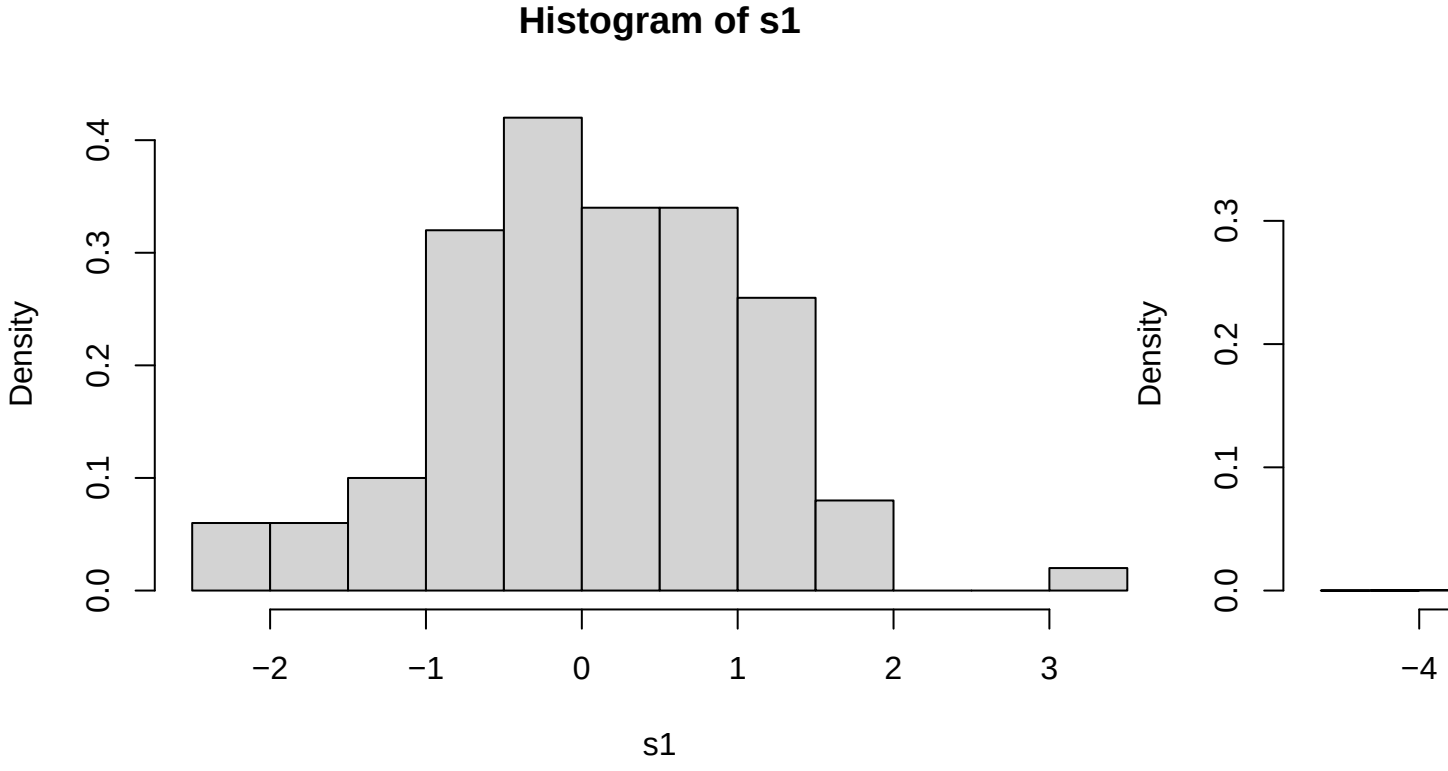

Histogram of s1
Density
0.0
0.1
0.2
0.3
0.4
−2
−1
0
1
2
s1
Density
0.0
0.1
0.2
0.3
−4

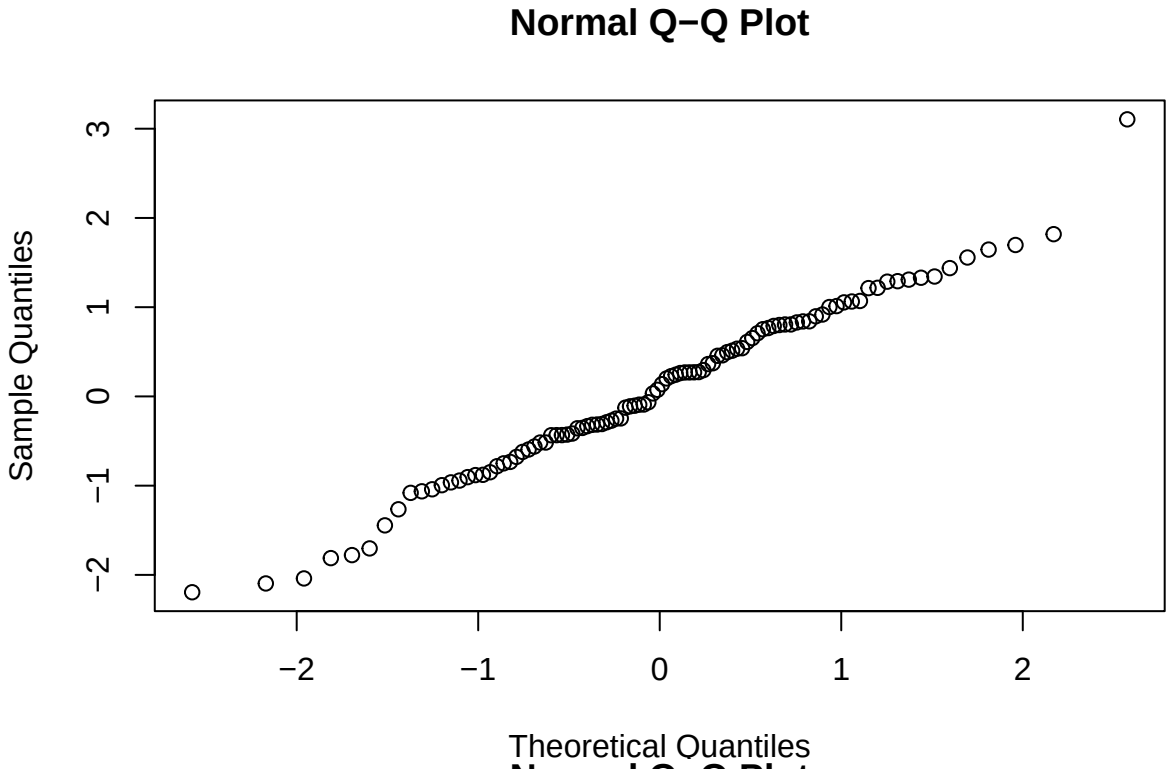

Normal Q−Q Plot
Sample Quantiles
−2
−1
0
1
2
−2
−1
0
1
2
Theoretical Quantiles

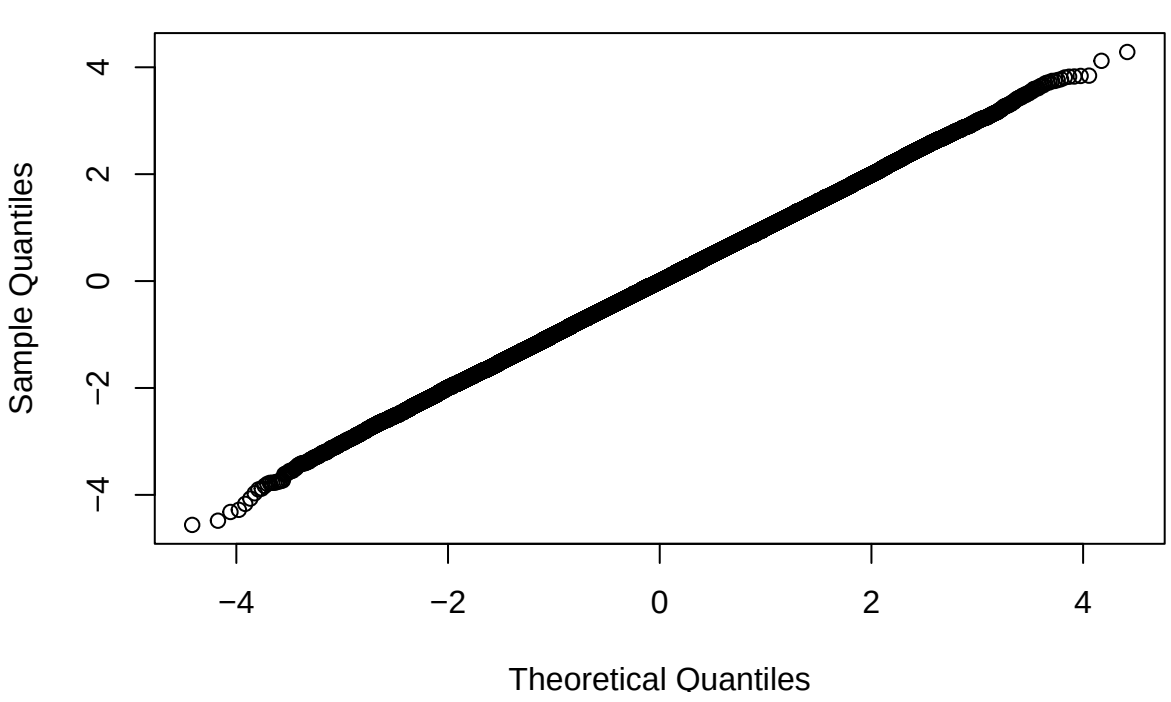



```
## [1] "prob (x < -0.5) = 0.3085"
## [1] "prob (-1 < x < 2) = 0.8186"
## [1] "estimate 1a: 0.2500, estimate1b: 0.3079"
## [1] "estimate 2a: 0.8000, estimate2b: 0.8171"
```

```
part1(mean=3, sd=2)
```








```
## [1] "prob (x < -0.5) = 0.3085"
## [1] "prob (-1 < x < 2) = 0.8186"
## [1] "estimate 1a: 0.2700, estimate1b: 0.3099"
## [1] "estimate 2a: 0.8800, estimate2b: 0.8169"
```

Exercise 2