

Домашна работа 2 по R

Николай Желязков, 82022

Задача 1

```
#Проверяваме дали отхвърля H0
denyZeroHyp = function(n){
  t = table(sample(1:6,size = n,replace = T))
  result = chisq.test(t)$p.value<0.05
  result
}
```

```
#Гледаме какъв процент от всички отхвърлят H0
f1 = function(n){
  s = sum(replicate(10000,denyZeroHyp(n)))/10000*100
  result = paste(s, "%")
  result
}
f1(100)
```

```
## [1] "5.1 %"
```

```
f1(200)
```

```
## [1] "4.73 %"
```

```
f1(400)
```

```
## [1] "5.02 %"
```

Задача 2

```
#Проверяваме дали данните са от нормално разпределение
checkNormality = function(n) {
  x = runif(n, 5, 12)
  hypResults = shapiro.test(x)$p.value > 0.05
  hypResults
}
```

```
#Намираме Колко често заключението на теста е вярно
f2 = function(nBig,n){
  results = replicate(n=nBig,checkNormality(n))
  prop.table(table(results))
}
f2(10000,15)
```

```
## results
## FALSE TRUE
## 0.1339 0.8661
```

```
f2(10000,25)
```

```
## results
## FALSE TRUE
## 0.2872 0.7128
```

```
f2(10000,50)
```

```
## results
## FALSE TRUE
## 0.7476 0.2524
```

```
f2(10000,90)
```

```
## results
## FALSE TRUE
## 0.9886 0.0114
```

Задача 3

```
#Функция четаеща графиката на оценката на бета1
graphics = function(beta1_est){
  hist(beta1_est)
  qqnorm(beta1_est)
}
```

```
f3 = function(n){
  beta1 = 5
  # 4 реда - по 1 за всяка подточка
  beta1_est = matrix(data = NA, nrow = 4, ncol = 10000)
  beta1_sd_err = matrix(data = NA, nrow = 4, ncol = 10000)
  beta1_sd_err = matrix(data = NA, nrow = 4, ncol = 10000)
  leftConfInt = matrix(data = NA, nrow = 4, ncol = 10000)
  rightConfInt = matrix(data = NA, nrow = 4, ncol = 10000)

  #генерираме данните 10000 пъти
  for (k in 1:10000) {
    x = sample(1:10, size = n, replace = TRUE)

    # а)
    e = rnorm(n, mean = 0, sd = 2)
    y = beta1 * x + e

    df = data.frame(x, y)
    model = lm(y ~ x, data = df)
    model_summary = summary(model)

    beta1_est[1, k] = model_summary$coefficients[2, 1] # оценка на beta1
    beta1_sd_err[1, k] = model_summary$coefficients[2, 2]
    leftConfInt[1,k] = confint(model)[2,1] #левия край на дов. интервал за beta1
    rightConfInt[1,k] = confint(model)[2,2] #десния край на дов. интервал за beta1

    # б)
    e = runif(n, -3.5, 3.5)
    y = beta1 * x + e

    df = data.frame(x, y)
    model = lm(y ~ x, data = df)
    model_summary = summary(model)
    p = paste("b)")
    beta1_est[2, k] = model_summary$coefficients[2, 1] # оценка на beta1
    beta1_sd_err[2, k] = model_summary$coefficients[2, 2]
    leftConfInt[2,k] = confint(model)[2,1] #левия край на дов. интервал за beta1
    rightConfInt[2,k] = confint(model)[2,2] #десния край на дов. интервал за beta1

    # в)
    v = rexp(n, 0.7)
    w = rexp(n, 0.7)
    e = v - w
    y = beta1 * x + e

    df = data.frame(x, y)
    model = lm(y ~ x, data = df)
    model_summary = summary(model)
    beta1_est[3, k] = model_summary$coefficients[2, 1] # оценка на beta1
    beta1_sd_err[3, k] = model_summary$coefficients[2, 2]
    leftConfInt[3,k] = confint(model)[2,1] #левия край на дов. интервал за beta1
```

```

rightConfInt[3,k] = confint(model)[2,2] #десния край на дов. интервал за бета1

# z)
u = rexp(n, 0.5)
e = u - 2
y = beta1 * x + e

df = data.frame(x, y)
model = lm(y ~ x, data = df)
model_summary = summary(model)

beta1_est[4, k] = model_summary$coefficients[2, 1] # оценка на бета1
beta1_sd_err[4, k] = model_summary$coefficients[2, 2]
leftConfInt[4,k] = confint(model)[2,1] #левия край на дов. интервал за бета1
rightConfInt[4,k] = confint(model)[2,2] #десния край на дов. интервал за бета1
}
#Създаваме матрица за крайните резултати
tbl = matrix(data = NA, nrow = 4, ncol = 3)
rownames(tbl) = c("a", "б", "в", "г")
colnames(tbl) = c("Колко често доверителният интервал за бета1 съдържа истинската стойнос
т?",
                  "средната дължина на доверителния интервал на базата на 10000 повторени
я",
                  "средното на beta1_est на базата на 10000 повторения")
for (k in 1:4) {
  tbl[k, 1] = paste(sum((beta1 > leftConfInt[k,]) & (beta1< rightConfInt[k,])) / 10000 * 10
0, "%")
  tbl[k, 2] = mean(rightConfInt[k,] - leftConfInt[k,]) # (-beta1_sd_err , +beta1_sd_err)
  tbl[k, 3] = mean(beta1_est[k,])
}
n = paste("N =",n)
print(n)
#Чертаем графиките
graphics(beta1_est)
print(tbl)
}

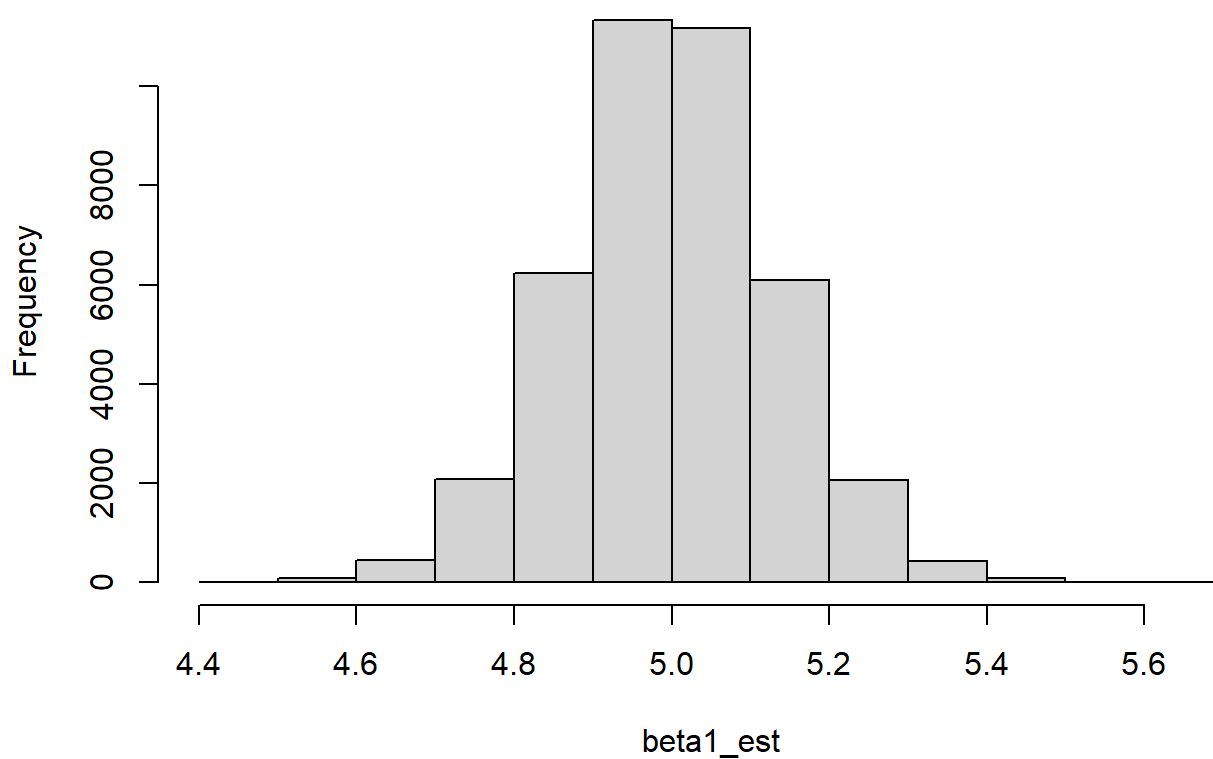
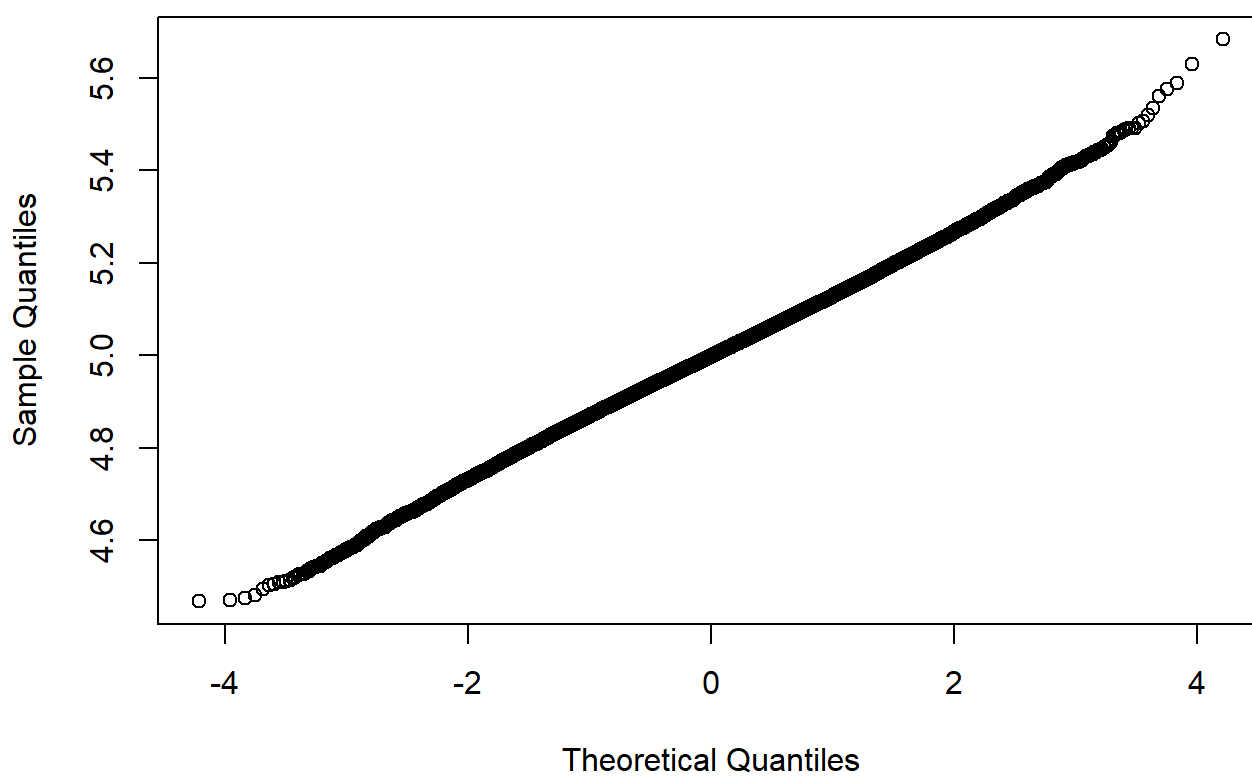
```

```

#Пускаме резултатите за различни стойности на n
ns = c(30, 50, 100, 500)
for (n in ns) {
  f3(n)
}

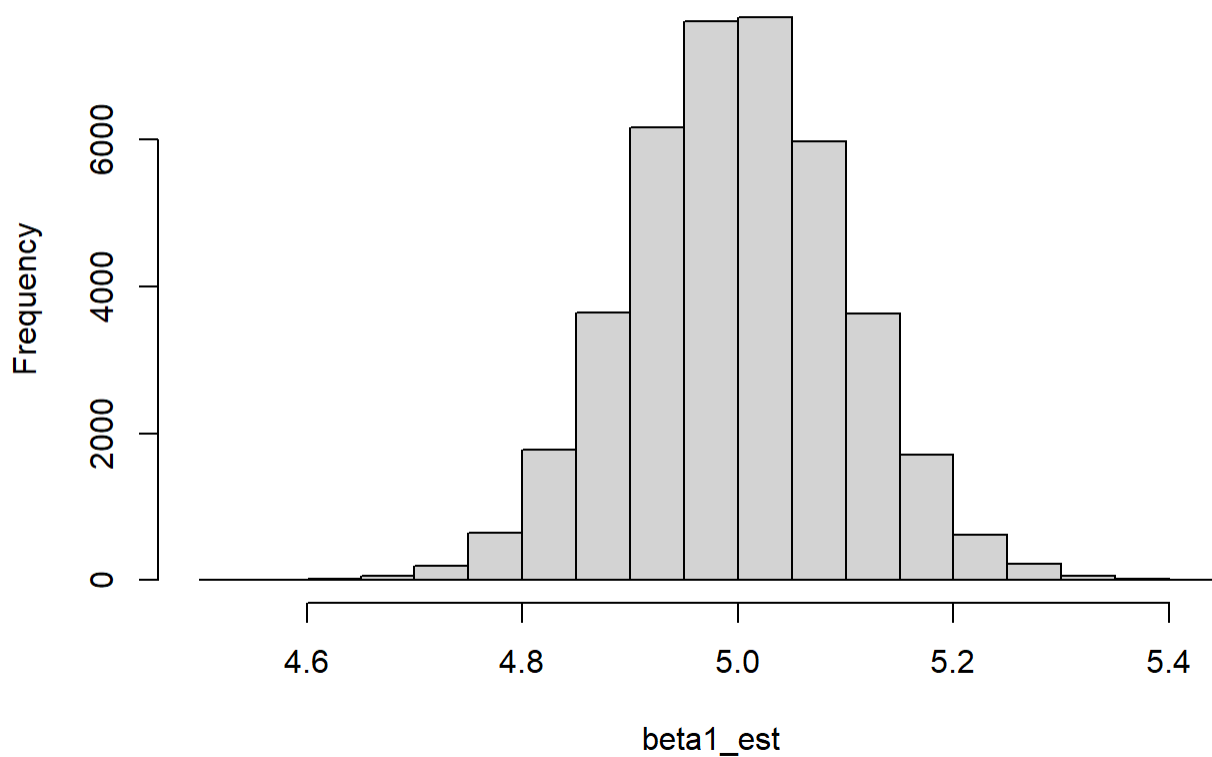
```

```
## [1] "N = 30"
```

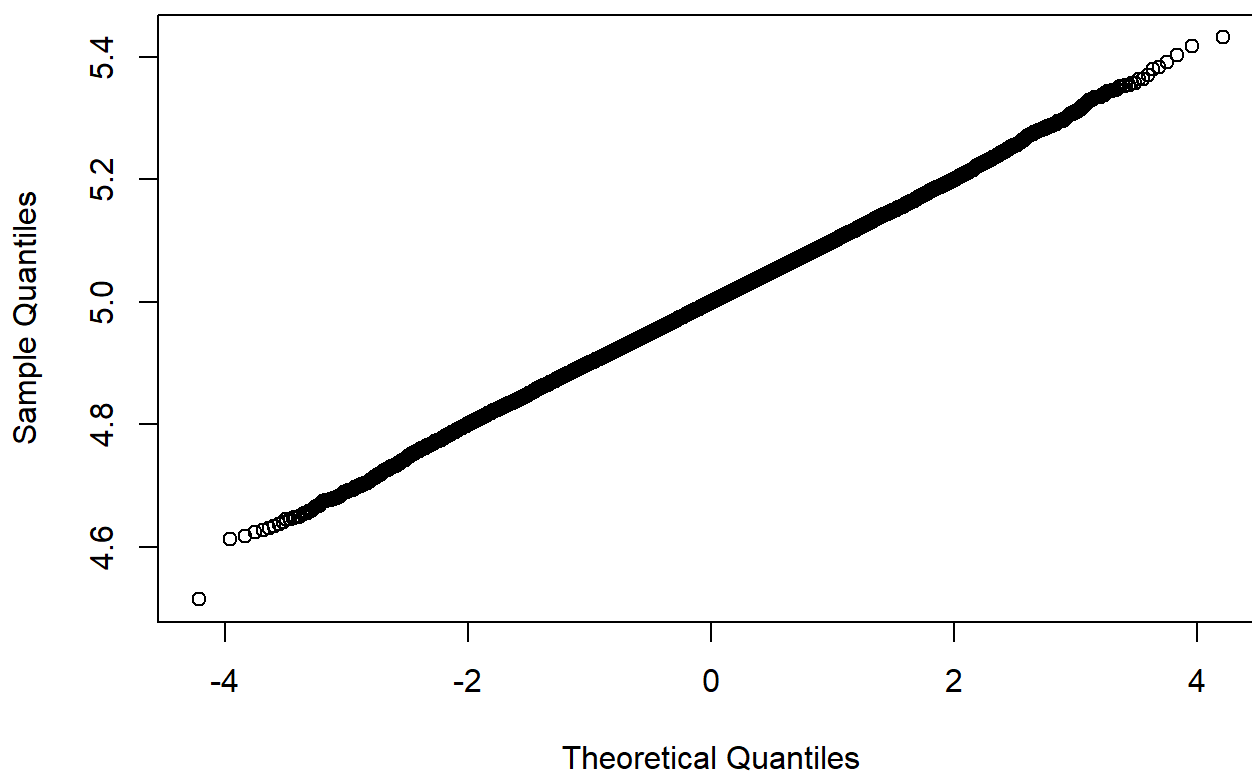
Histogram of beta1_est**Normal Q-Q Plot**

```
##      Колко често доверителният интервал за бета1 съдържа истинската стойност?  
## а) "95.06 %"  
## б) "94.72 %"  
## в) "95.28 %"  
## г) "95.19 %"  
##      средната дължина на доверителния интервал на базата на 10000 повторения  
## а) "0.530465454886926"  
## б) "0.539138801803849"  
## в) "0.530711822334642"  
## г) "0.517781212376251"  
##      средното на beta1_est на базата на 10000 повторения  
## а) "5.00071168066147"  
## б) "4.99735091729263"  
## в) "4.99627122688133"  
## г) "5.00159477064799"  
## [1] "N = 50"
```

Histogram of beta1_est

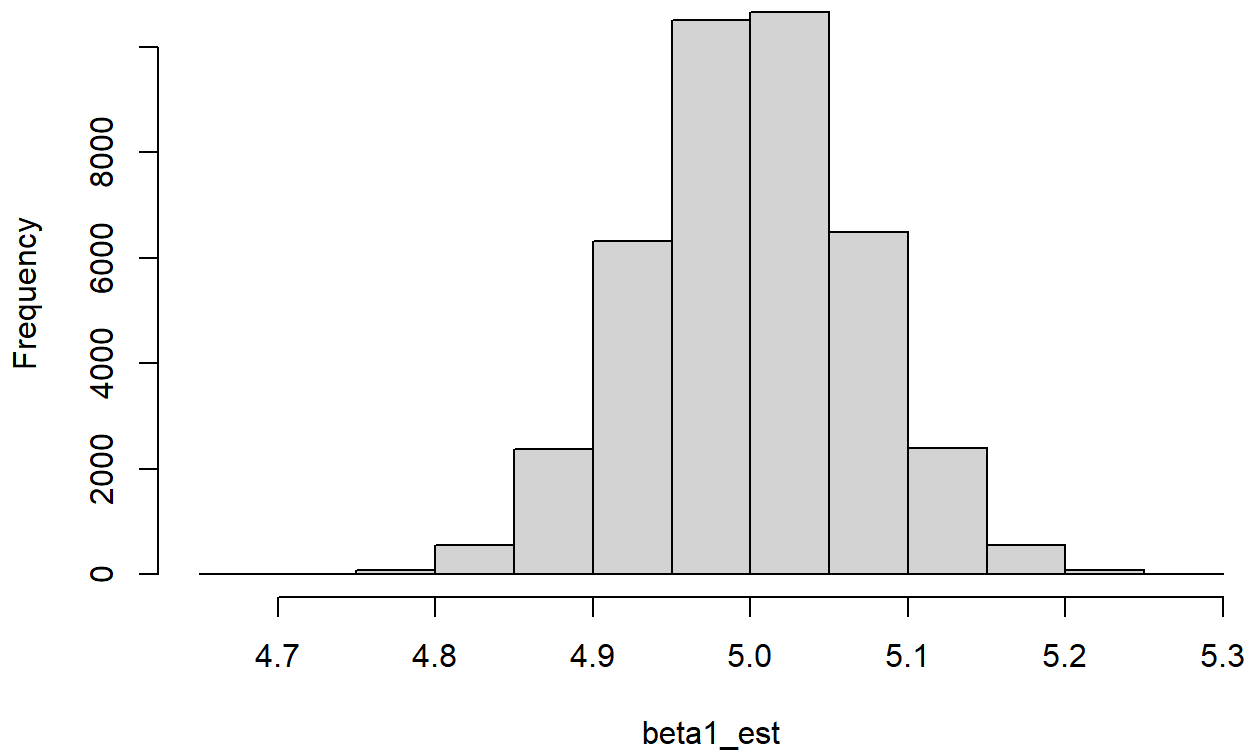


Normal Q-Q Plot

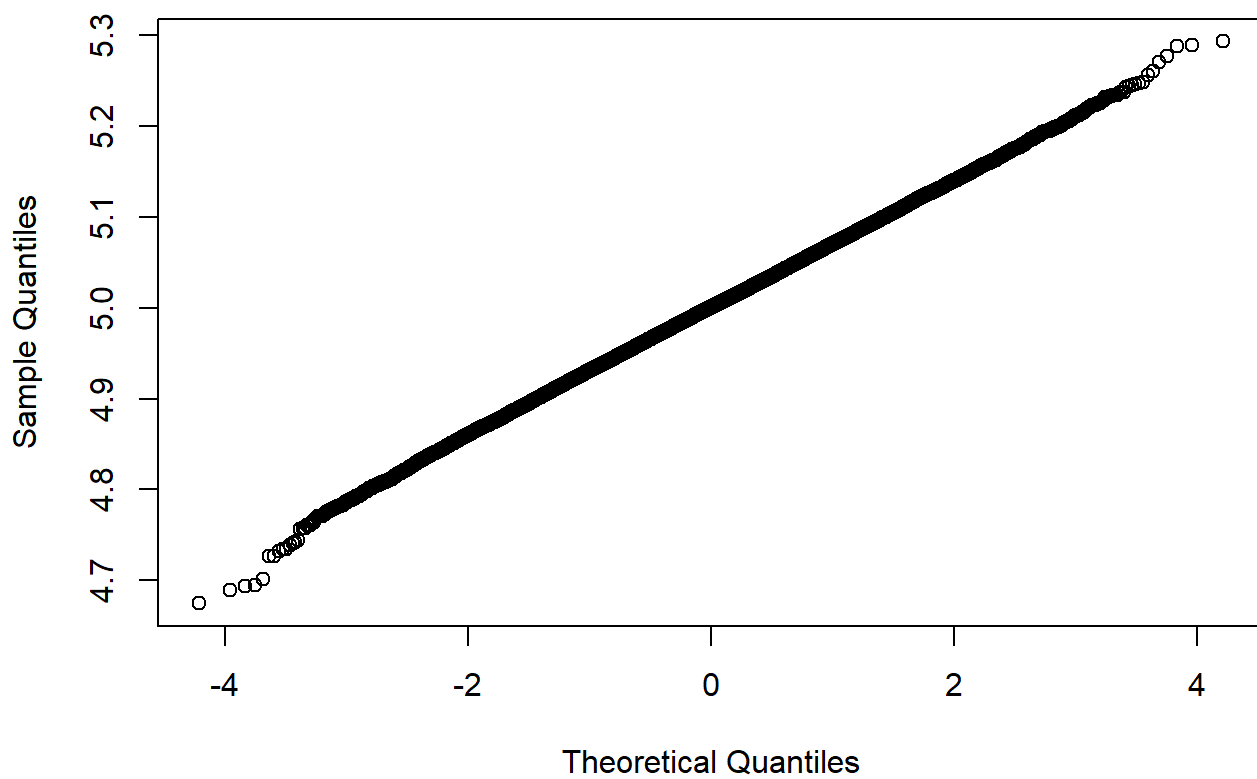


```
##      Колко често доверителният интервал за бета1 съдържа истинската стойност?  
## а) "95.12 %"  
## б) "95.04 %"  
## в) "95.27 %"  
## г) "95.19 %"  
##      средната дължина на доверителния интервал на базата на 10000 повторения  
## а) "0.400315768369623"  
## б) "0.405553192105406"  
## в) "0.401548237400706"  
## г) "0.394991082123464"  
##      средното на beta1_est на базата на 10000 повторения  
## а) "4.99926700671118"  
## б) "4.99983400060328"  
## в) "4.9992288099276"  
## г) "4.99926760148843"  
## [1] "N = 100"
```

Histogram of beta1_est

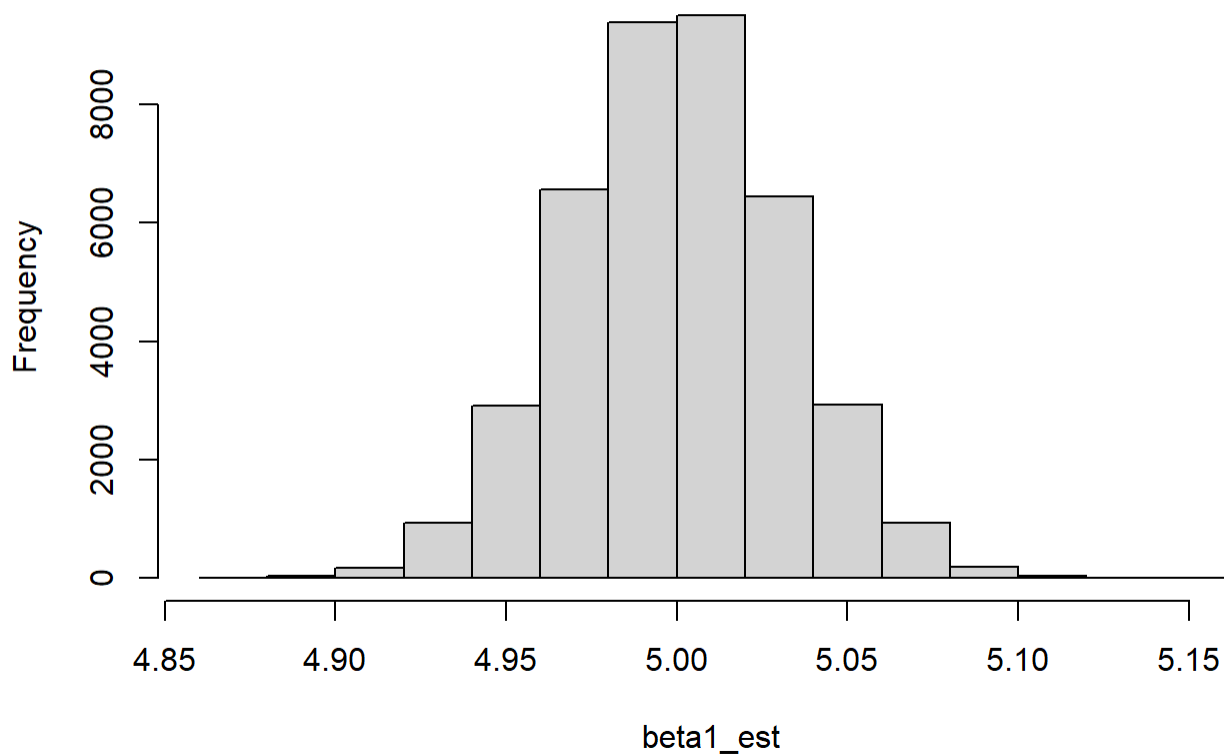


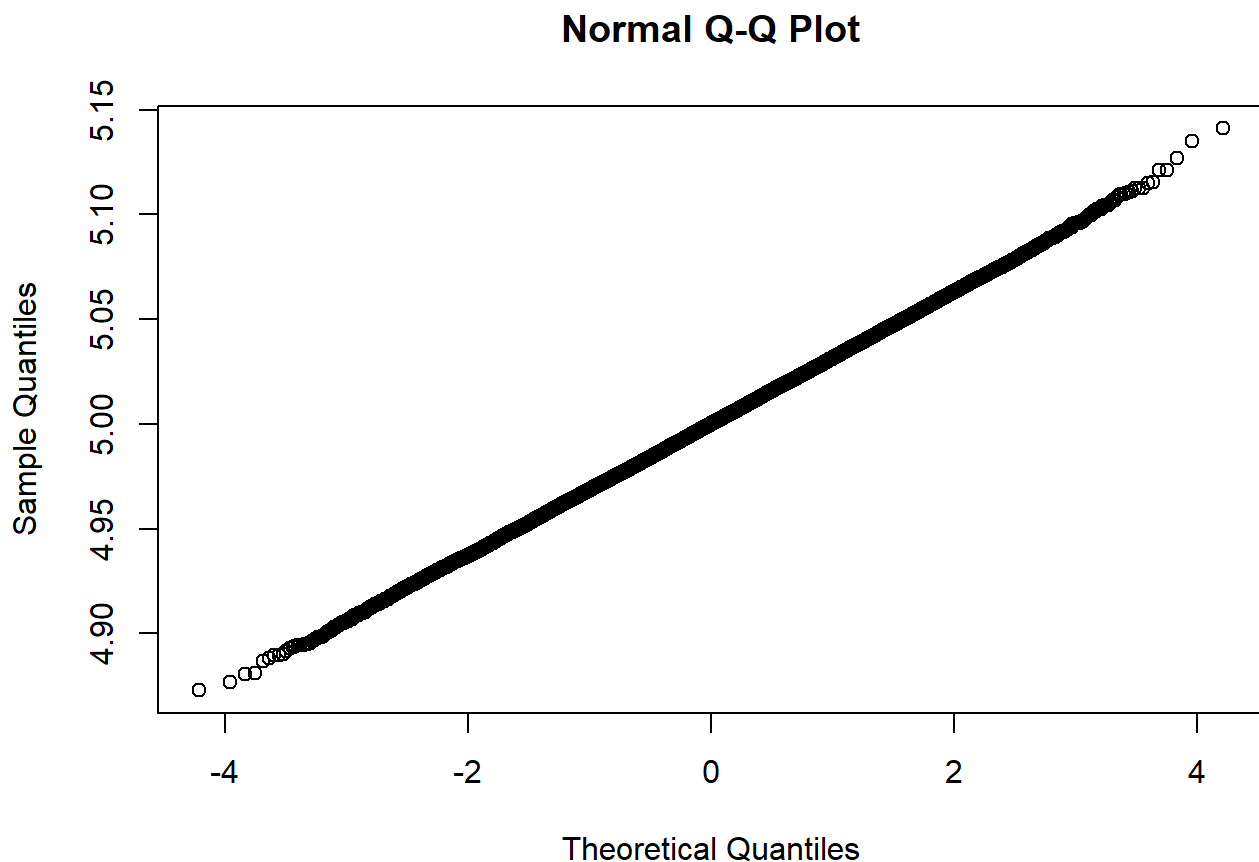
Normal Q-Q Plot




```
##      Колко често доверителният интервал за бета1 съдържа истинската стойност?  
## а) "94.9 %"  
## б) "95.28 %"  
## в) "95.04 %"  
## г) "95.43 %"  
##      средната дължина на доверителния интервал на базата на 10000 повторения  
## а) "0.277638045176584"  
## б) "0.281316725539572"  
## в) "0.279315574780723"  
## г) "0.275879728092943"  
##      средното на beta1_est на базата на 10000 повторения  
## а) "5.00017308331923"  
## б) "5.00053131591181"  
## в) "5.00105631337939"  
## г) "4.99976355376391"  
## [1] "N = 500"
```

Histogram of beta1_est





```
##      Колко често доверителният интервал за бета1 съдържа истинската стойност?  
## а) "94.59 %"  
## б) "94.82 %"  
## в) "95.29 %"  
## г) "95.06 %"  
##      средната дължина на доверителния интервал на базата на 10000 повторения  
## а) "0.122469699059984"  
## б) "0.123797061221086"  
## в) "0.123650377082785"  
## г) "0.122294769335971"  
##      средното на beta1_est на базата на 10000 повторения  
## а) "4.99972543873396"  
## б) "4.99970645699569"  
## в) "5.00015951759973"  
## г) "5.00039039853086"
```