Homework 5

Tauqeer Kasam Rumaney

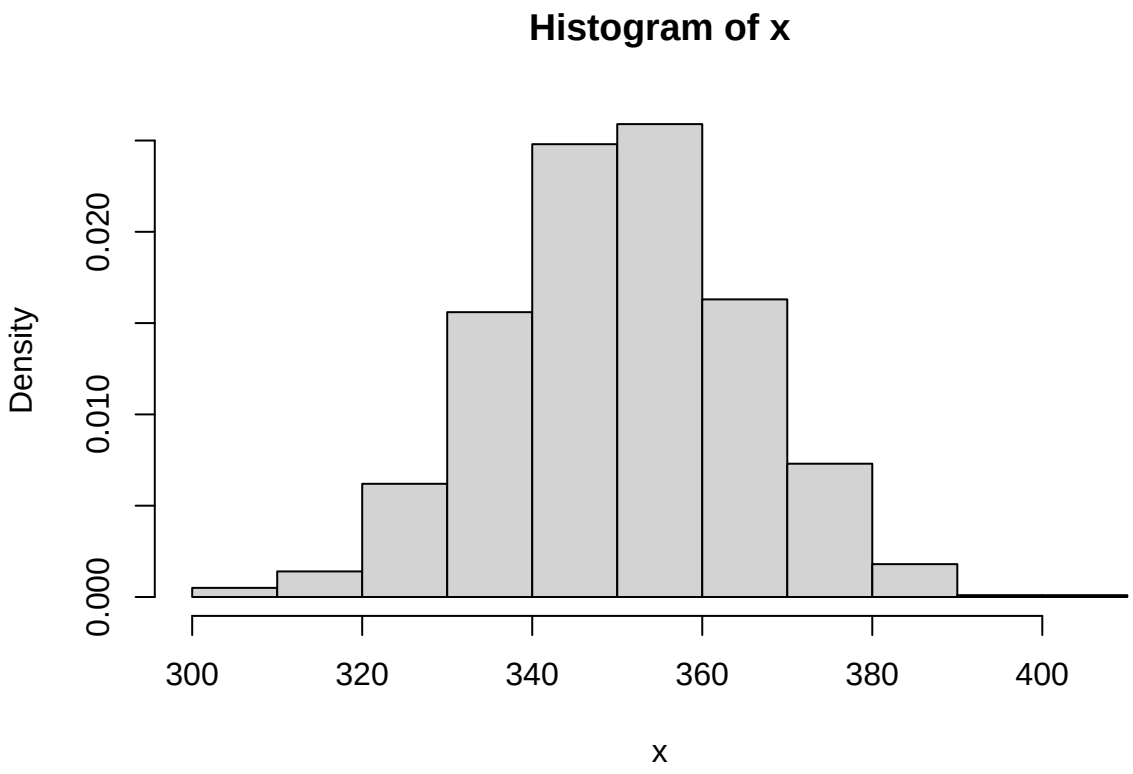
November 20, 2022

Download this R Markdown file, save it on your computer, and perform all the below tasks by inserting your answer in text or by inserting R chunks below. After you are done, upload this file with your solutions on Moodle.

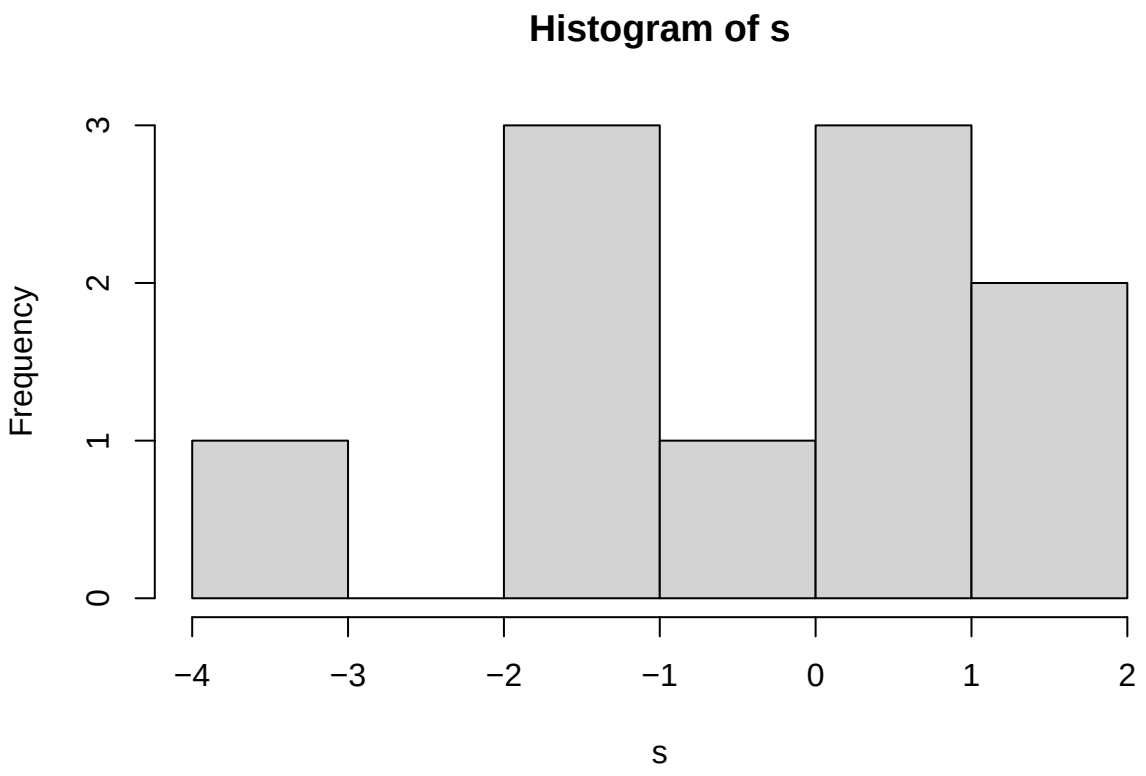
Exercise 1: Probability distributions

Explore the functions *rnorm*, *rt*, *runif*, *rbinom* in R that allow you to generate random numbers from the normal, t-, uniform, and binomial distribution. Compute them with different values, and inspect histograms to visualize their distribution.

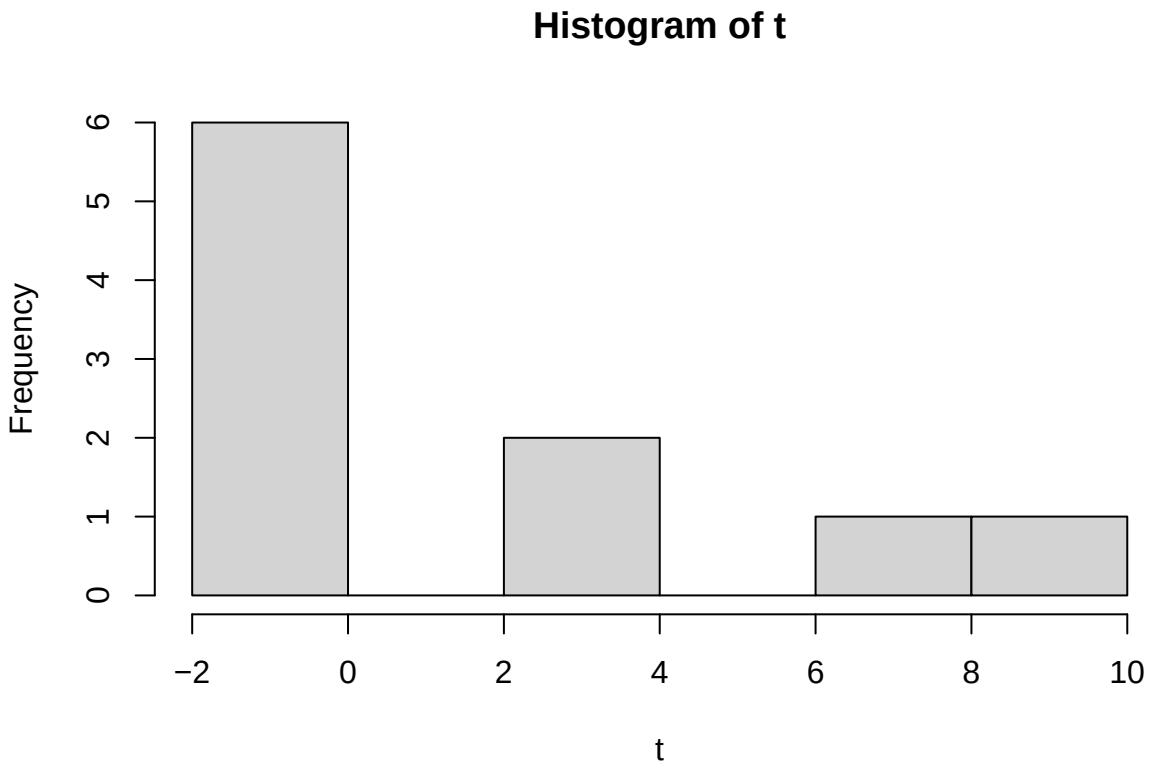
```
x <- rnorm(1000, mean=350, sd=15)
hist(x, probability=TRUE)
```



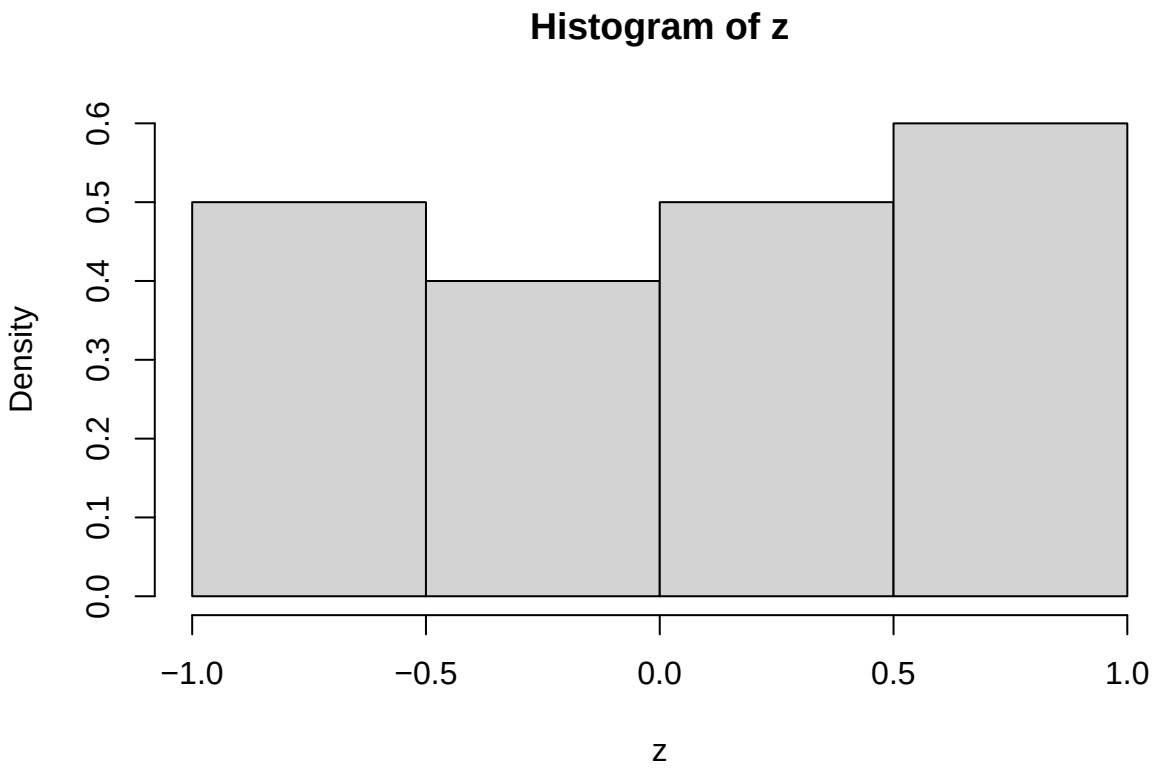

```
s <- rt(10,2)
t <- rt(10,1)
hist(s)
```



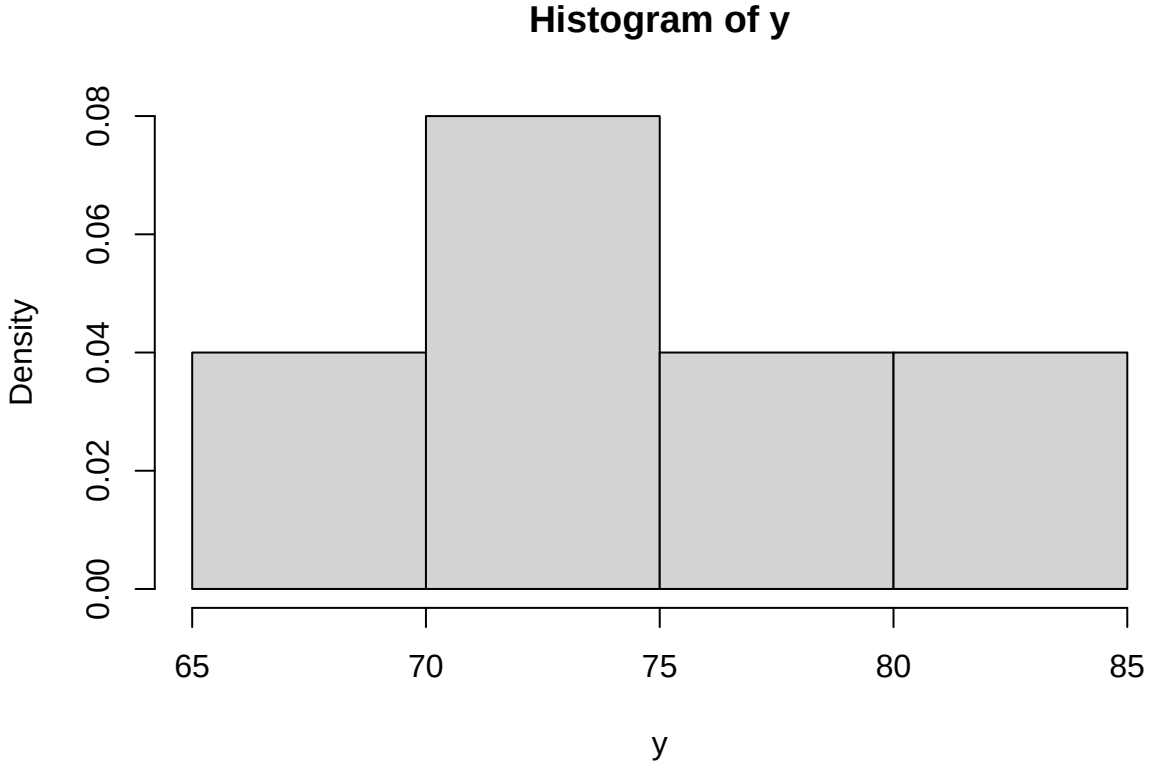

```
hist(t)
```




```
z <- runif(20, min = -1, max = 1)
hist(z, probability=TRUE)
```




```
y <- rbinom(10, size=100, prob=0.75)
hist(y, probability=TRUE)
```



Exercise 2: Odds ratio

In the KiGGS dataset:

```
dat_link <- url("https://www.dropbox.com/s/pd0z829pv2otzqt/KiGGS03_06.RData?dl=1")
load(dat_link)
dat <- KiGGS03_06
```

a) Compute the proportion of mothers that had hypertension during pregnancy. Use the variable 'e0155' which has values "Ja" (yes), "Nein" (No) and "Weiß nicht" (don't know). #202

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
table(dat['e0155'])
```

```
## e0155
##         Ja       Nein Weiß nicht
##        202       2306         37
```

b) Create a new variable that is 0 or 1 depending on whether the children are small or tall (think of a good way how to do this) based on the variable 'groeB' (body height). #anyone who's shorter than median value is considerd small

```
#table(dat['groeB'])
summary(dat['groeB'])
```

```
##      groeB
##  Min.   : 56.7
##  1st Qu.:109.0
##  Median :136.0
##  Mean   :132.7
##  3rd Qu.:159.6
##  Max.   :194.4
##  NA's   :87
```

```
dat$abc <- dat$groeB
a <- median(dat$groeB)
a
```

```
## [1] NA
```

```
dat$abc[(dat$groeB <= 136)] <- 0
dat$abc[(dat$groeB >= 136)] <- 1
table(dat$abc)
```

```
##
##    0    1
## 8769 8784
```

c) Then compute the odds ratio that the mother had hypertension during pregnancy (e0155 == “Ja” (yes), versus e0155 == “Nein” (no)) of tall vs. small children.

```
#options(qwraps2_markup = "markdown")
#to find ratio since we need probability of every comibination i.e. (e0155 == "Ja" (yes), versus e0155 =

dat$xyz <- NA
dat$xyz[(dat$groeB <= 100  & dat$e0155 == "Ja")] <- 0
dat$xyz[(dat$groeB >= 100 & dat$e0155 == "Ja")] <- 1
dat$xyz[(dat$groeB <= 100  & dat$e0155 == "Nein")] <- 2
dat$xyz[(dat$groeB >= 100 & dat$e0155 == "Nein")] <- 3

table(dat$xyz)
```

0 1 2 3 198 3 2230 31

```
height <- c('Small', 'Tall')
hypertension <- c('Ja', 'Nein')
data2 <- matrix(c(198, 2230, 3, 31), nrow=2, ncol=2, byrow=TRUE)
dimnames(data2) <- list('Height'=height, 'Hypertension'=hypertension)

data2
```

```
   Hypertension
```

Height Ja Nein Small 198 2230 Tall 3 31

```
library(epitools)

oddsratio(data2)
```

```
## Warning in chisq.test(xx, correct = correction): Chi-squared approximation may
## be incorrect
```

$data Hypertension Height Ja Nein Total Small 198 2230 2428 Tall 3 31 34 Total 201 2261 2462

$measure odds ratio with 95% C.I. Height estimate lower upper Small 1.0000000 NA NA Tall 0.8764838 0.3081907 3.81214

$p.value two-sided Height midp.exact fisher.exact chi.square Small NA NA NA Tall 0.833785 0.7541289 0.8875484

$correction [1] FALSE

attr(,“method”) [1] “median-unbiased estimate & mid-p exact CI”

Exercise 3 (optional): Confidence intervals

Look at the hypertension variable from exercise 2. Use the binom::binom.confint and the questionr::odds.ratio functions to compute the estimates of the proportion and odds ratio as well as their confidence intervals. (you need to download and load these packages at first).

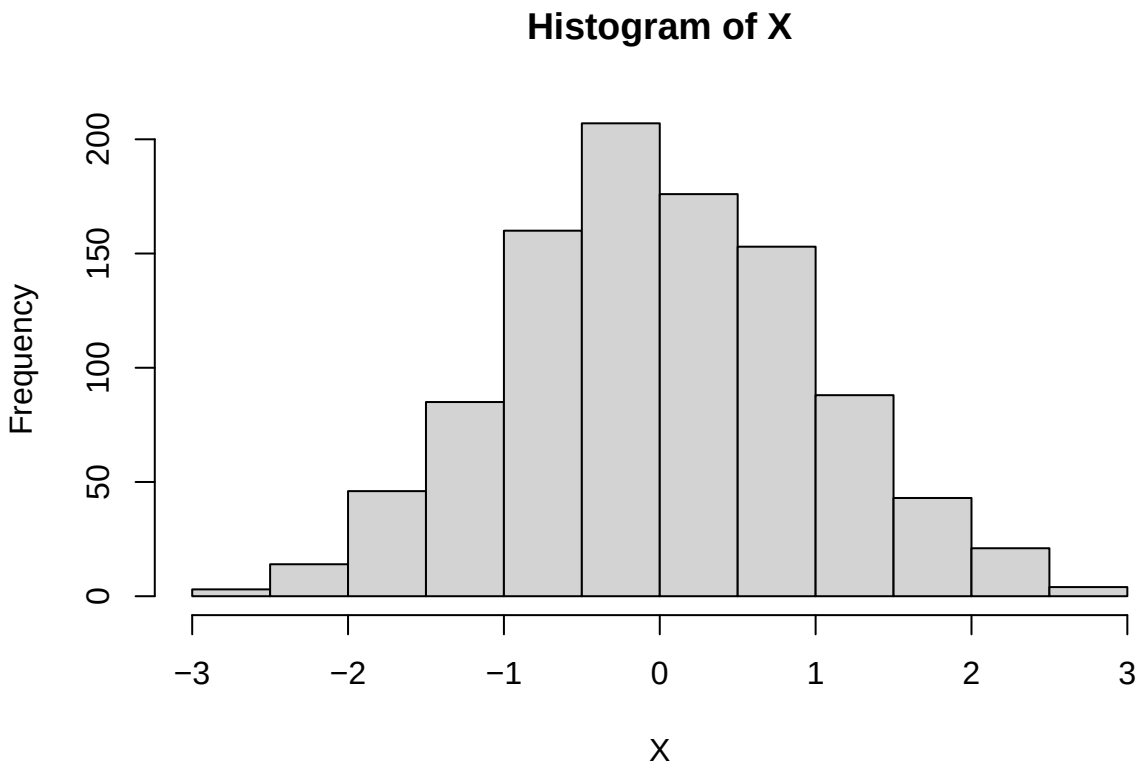
Exercise 4 (optional, advanced): Bootstrap

Adapt the bootstrap implementation in R_5b_estimation_bootstrap.Rmd to compute the bootstrap estimate of the standard error of the variance of a normally-distributed and a t-distributed variable. Are they similar? #NO

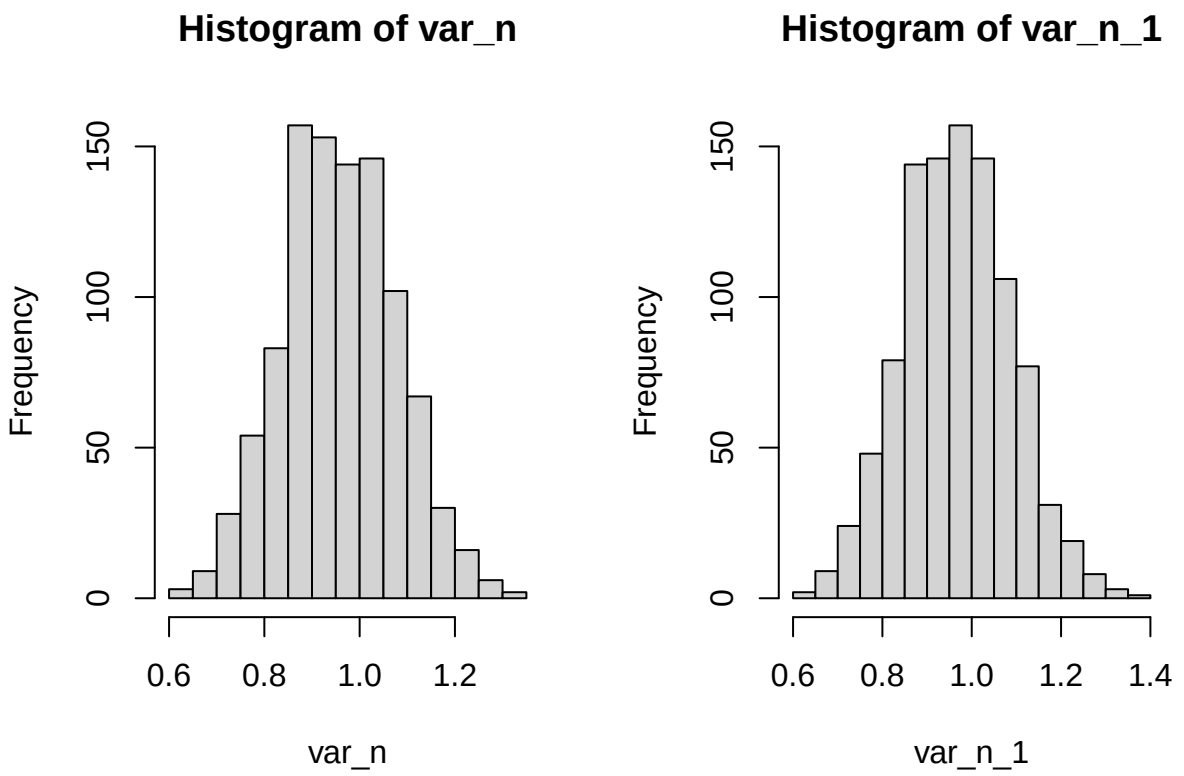
```r
set.seed(7)
X <- rnorm(n = 1000, mean = 0, sd = 1)
var(X)
```

```
## [1] 0.964956
```

```r
hist(X)
```




```r
var_n <- var_n_1 <- NULL
var_n1 <- var_n1_1 <- NULL
for(i in 1:1000){
  X_sample_i <- sample(X, size = 100, replace = FALSE)
  var_n[i] <- var(X_sample_i)*(100-1)/100
  var_n_1[i] <- var(X_sample_i)
}
par(mfrow = c(1,2))
hist(var_n)
hist(var_n_1)
```

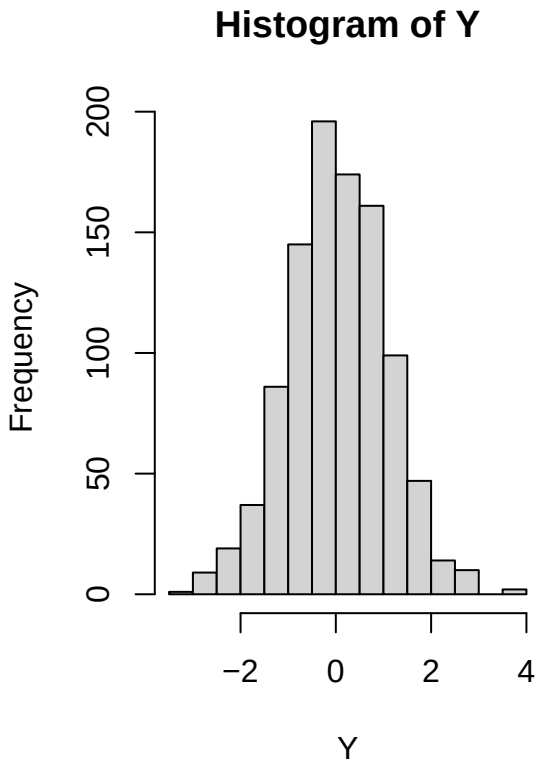



```
set.seed(7)
Y <- rt(1000, 100)
var(Y)
```

```
## [1] 1.082297
```

```
hist(Y)

for(i in 1:1000){
  Y_sample_i <- sample(Y, size = 100, replace = FALSE)
  var_n1[i] <- var(Y_sample_i)*(100-1)/100
  var_n1_1[i] <- var(Y_sample_i)
}
par(mfrow = c(1,2))
```




```
hist(var_n1)
hist(var_n1_1)
```