Estimation_Exercice_R.R

Snow

2023-07-11

```
# Exercise on Statistical inference: Estimation
# By Mesli Redhouane
# Date: 11/07/2023

# This report proposes a solution to an exercise related
# to the chapter dealing with statistical estimation,
# and more particularly with the sampling distributions of the sample means.

# Statement of the exercise:
# We have a variable (x), which takes the following values:
# 1, 3, 5, 7, 9.

# 1- Calculate the mean and the standard deviation of the population;
# 2- What is the number of possible combinations of two elements
# among the five elements mentioned above;
# 3- Expose these combinations;
# 4- Extract the sampling distribution of the mean of these combinations;
# 5- Calculate its mean and its standard deviation;
# 6- Calculate the mean and the standard deviation from the formulas
# proposed by the estimation theory.

# Solution:

# Create a sequence of the first 5 odd digits

x <- seq(1,9, by = 2)
x
```

```
## [1] 1 3 5 7 9
```

```
# Calculate mean and standard deviation with R functions

m <- mean(x)                                    # Mean of population
m
```

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

```
sd_sam <- sd(x)                                 # for sample
sd_sam
```

```
## [1] 3.162278
```

```r
sd_pop <- sd(x)*sqrt((length(x)-1)/length(x))  # for population
sd_pop
```

```
## [1] 2.828427
```

```r
# sd_pop <- sqrt(sum((x-mean(x))^2)/length(x)) # second formulas

# Install and attach combinat package

# install.packages("combinat")                 # Install combinat package
# library("combinat")                          # Load combinat package

# Calculate combinations of 2 elements taken from 5

n <- 2
pairs <- combinat::combn(x,n)
pairs
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    1    1    1    1    3    3    3    5    5     7
## [2,]    3    5    7    9    5    7    9    7    9     9
```

```r
pairs2 <- t(pairs)                           # Transpose pairs
pairs2
```

```
##       [,1] [,2]
##  [1,]    1    3
##  [2,]    1    5
##  [3,]    1    7
##  [4,]    1    9
##  [5,]    3    5
##  [6,]    3    7
##  [7,]    3    9
##  [8,]    5    7
##  [9,]    5    9
## [10,]    7    9
```

```r
pairs3 <- cbind(pairs2, rowMeans(pairs2))    # Add meanrows to pairs3
pairs3
```

```
##       [,1] [,2] [,3]
##  [1,]    1    3    2
##  [2,]    1    5    3
##  [3,]    1    7    4
##  [4,]    1    9    5
##  [5,]    3    5    4
##  [6,]    3    7    5
##  [7,]    3    9    6
##  [8,]    5    7    6
##  [9,]    5    9    7
## [10,]    7    9    8
```

```r
pairs4 <- data.frame(pairs3)        # Change pairs3 to dataframe so we can do further manipulations

# install.packages("plyr")
library(plyr)

pairs5 <- count(pairs4, 'X3')       # contract data to have frequency table
pairs5
```

```
##   X3 freq
## 1  2    1
## 2  3    1
## 3  4    2
## 4  5    2
## 5  6    2
## 6  7    1
## 7  8    1
```

```r
# install.packages("plotrix")
library("plotrix")

weighted.hist(pairs5$X3,pairs5$freq) # Draw a histogram

N_smd <- sum(pairs5$freq)            # Calculate sum of frequencies
N_smd
```

```
## [1] 10
```

```r
mx <- weighted.mean(pairs5$X3, pairs5$freq)  # Calculate the mean of sample distribution of mean
mx
```

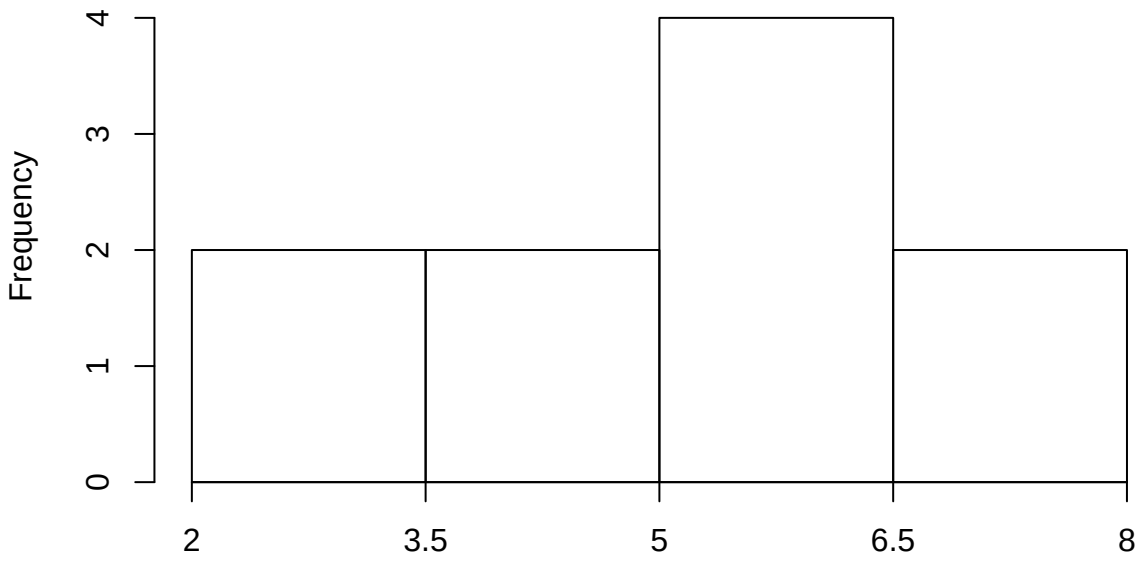
```
## [1] 5
```

```r
# install.packages("Hmisc")
library("Hmisc")
```

```
##
## Attachement du package : 'Hmisc'

## Les objets suivants sont masqués depuis 'package:plyr':
##
##     is.discrete, summarize

## Les objets suivants sont masqués depuis 'package:base':
##
##     format.pval, units
```




```
var_smd <- wtd.var(pairs5$X3, pairs5$freq)   # Calculate variance of Sample distribution of mean
var_smd
```

```
## [1] 3.333333
```

```
sd_smd <- sqrt(var_smd*(N_smd-1)/N_smd)      # Calculate sd of sample distribution of mean
sd_smd
```

```
## [1] 1.732051
```

```
sd_smd_for <- (sd_pop/sqrt(n))*sqrt((length(x)-n)/(length(x) - 1)) # Calculate sd of sample distributio
sd_smd_for
```

```
## [1] 1.732051
```