HW4.R

xboxv

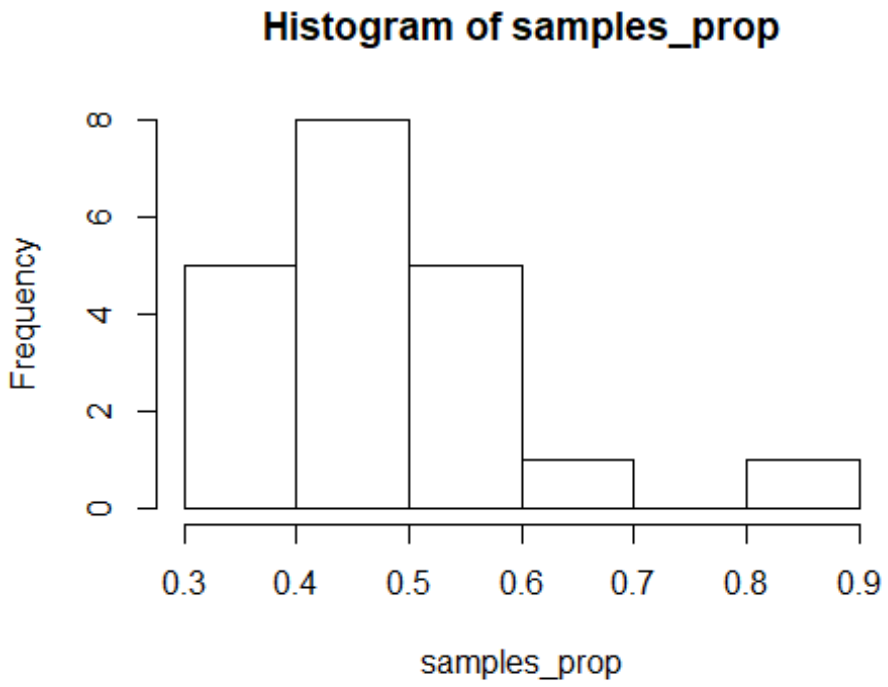
2020-03-01

```
#HW 4
library(moments)

#distribution of a vector
printVectorInfo <-function(vectorInputs){
  mean_value <- mean(vectorInputs)
  median_value <- median(vectorInputs)
  min_value <- min(vectorInputs)
  max_value <- max(vectorInputs)
  std_value <- sd(vectorInputs)
  qt_value <- quantile(vectorInputs, probs = c(0.05, 0.95))
  skw_value <- skewness(vectorInputs)
  cat('Mean : ',mean_value,'\n')
  cat('Median : ',median_value,'\n')
  cat('Min : ',min_value,'  ')
  cat('Max : ',max_value, '\n')
  cat('Std : ',std_value,'\n')
  cat('quatile : ',qt_value,'\n')
  cat('Skewness : ',skw_value,'\n\n')
}

#test the fucntion
x <- c  (1,2,3,4,5,6,7,8,9,10,50)
printVectorInfo(x)

## Mean :  9.545455
## Median :  6
## Min :  1   Max :  50
## Std :  13.72125
## quatile :  1.5 30
## Skewness :  2.620396

#step 2

new_jar <- c('red','blue')
jar <- rep(new_jar,50)
length(which(jar == 'red'))

## [1] 50

new_sample <- 10
samples <- sample(jar,new_sample, replace = TRUE)
```

```r
red_prop <- length(which(samples == 'red'))/new_sample
red_prop
```

```
## [1] 0.4
```

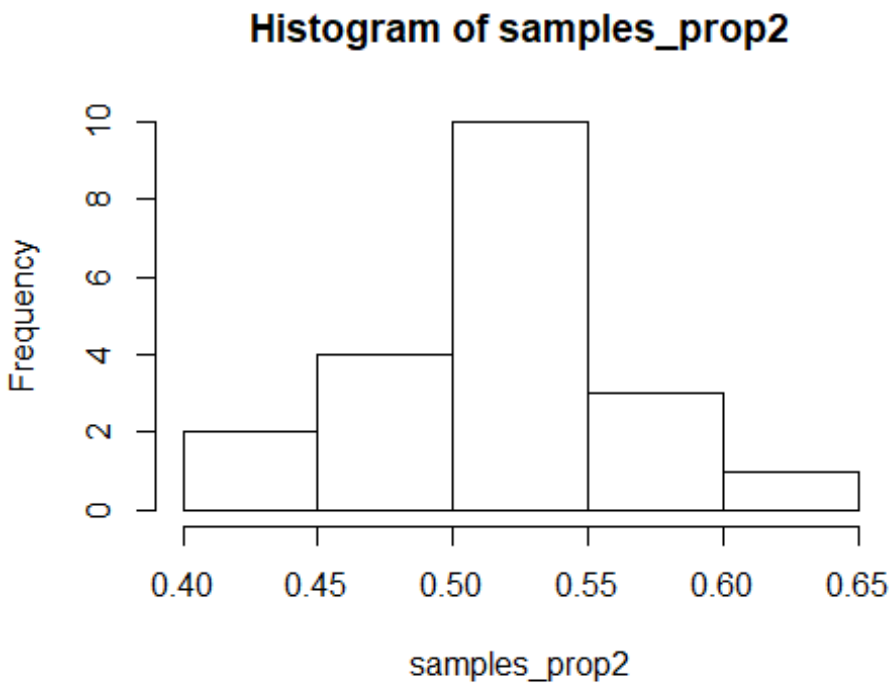
```r
samples_prop <- replicate(20, length(which((sample(jar,new_sample, replace = TRUE)) == 'red'))/new_sample)

hist(samples_prop)
```



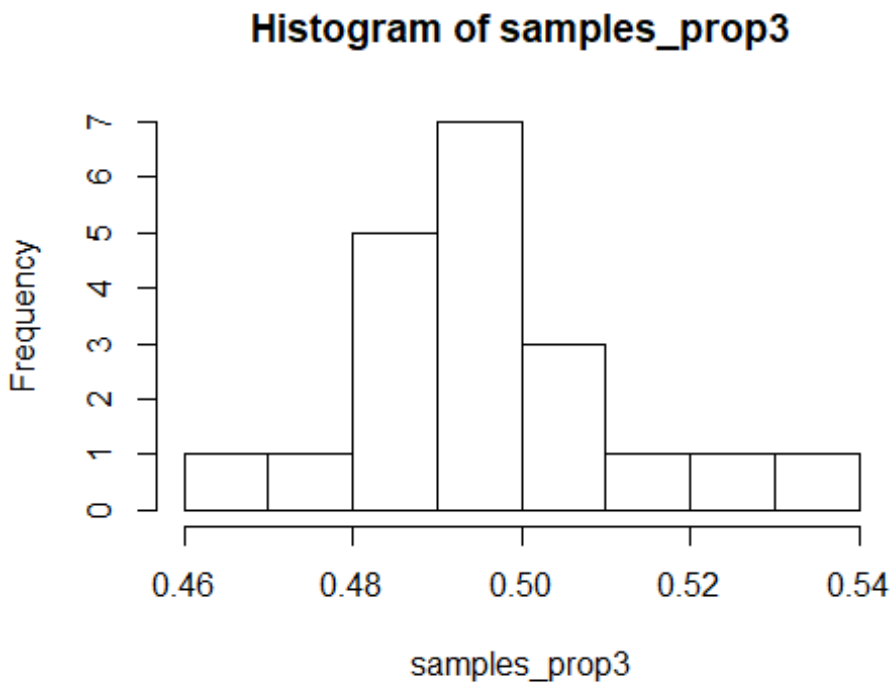

```r
printVectorInfo(samples_prop)
```

```
## Mean :  0.52
## Median :  0.5
## Min :  0.3   Max :  0.9
## Std :  0.1361114
## quatile :  0.3 0.71
## Skewness :  0.7863279
```

```r
#100 times
new_sample <- 100
samples_prop2 <- replicate(20, length(which((sample(jar,new_sample, replace = TRUE)) == 'red'))/new_sample)
hist(samples_prop2)
```




```
printVectorInfo(samples_prop2)

## Mean :  0.5255
## Median :  0.53
## Min :  0.44   Max :  0.62
## Std :  0.04773557
## quatile :  0.44 0.601
## Skewness :  -0.03211657

#1000 times
new_sample <- 1000
samples_prop3 <- replicate(20, length(which((sample(jar,new_sample, replace =
TRUE)) == 'red'))/new_sample)
hist(samples_prop3)
```




```
printVectorInfo(samples_prop3)
```

```
## Mean :  0.49645
## Median :  0.495
## Min :  0.469   Max :  0.531
## Std :  0.01492252
## quatile :  0.4766 0.5215
## Skewness :  0.4882106
```

```
#air Quality Dataset
temp_data <-airquality
summary(temp_data)
```

```
##      Ozone           Solar.R          Wind             Temp
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
##  NA's   :37       NA's   :7
##      Month            Day
##  Min.   :5.000   Min.   : 1.0
##  1st Qu.:6.000   1st Qu.: 8.0
##  Median :7.000   Median :16.0
##  Mean   :6.993   Mean   :15.8
##  3rd Qu.:8.000   3rd Qu.:23.0
```

```
##  Max.   :9.000   Max.   :31.0
##
```

```r
#remove na
remove_na <- function(df, n=0){
  df[rowSums(is.na(df)) <= n,]
}

temp_data <- remove_na(temp_data)
summary(temp_data)
```

```
##      Ozone           Solar.R           Wind             Temp
##  Min.   :  1.0   Min.   :  7.0   Min.   : 2.30   Min.   :57.00
##  1st Qu.: 18.0   1st Qu.:113.5   1st Qu.: 7.40   1st Qu.:71.00
##  Median : 31.0   Median :207.0   Median : 9.70   Median :79.00
##  Mean   : 42.1   Mean   :184.8   Mean   : 9.94   Mean   :77.79
##  3rd Qu.: 62.0   3rd Qu.:255.5   3rd Qu.:11.50   3rd Qu.:84.50
##  Max.   :168.0   Max.   :334.0   Max.   :20.70   Max.   :97.00
##      Month            Day
##  Min.   :5.000   Min.   : 1.00
##  1st Qu.:6.000   1st Qu.: 9.00
##  Median :7.000   Median :16.00
##  Mean   :7.216   Mean   :15.95
##  3rd Qu.:9.000   3rd Qu.:22.50
##  Max.   :9.000   Max.   :31.00
```

```r
printVectorInfo(temp_data$Ozone)
```

```
## Mean :  42.0991
## Median :  31
## Min :  1   Max :  168
## Std :  33.27597
## quatile :  8.5 109
## Skewness :  1.248104
```

```r
printVectorInfo(temp_data$Wind)
```

```
## Mean :  9.93964
## Median :  9.7
## Min :  2.3   Max :  20.7
## Std :  3.557713
## quatile :  4.6 15.5
## Skewness :  0.4556414
```

```r
printVectorInfo(temp_data$Temp)
```

```
## Mean :  77.79279
## Median :  79
## Min :  57   Max :  97
## Std :  9.529969
## quatile :  61 92.5
## Skewness :  -0.2250959
```

```r
#sapply function
sapply(temp_data, printVectorInfo)
```

```
## Mean :  42.0991
## Median :  31
## Min :  1   Max :  168
## Std :  33.27597
## quatile :  8.5 109
## Skewness :  1.248104
##
## Mean :  184.8018
## Median :  207
## Min :  7   Max :  334
## Std :  91.1523
## quatile :  22 310
## Skewness :  -0.4862466
##
## Mean :  9.93964
## Median :  9.7
## Min :  2.3   Max :  20.7
## Std :  3.557713
## quatile :  4.6 15.5
## Skewness :  0.4556414
##
## Mean :  77.79279
## Median :  79
## Min :  57   Max :  97
## Std :  9.529969
## quatile :  61 92.5
## Skewness :  -0.2250959
##
## Mean :  7.216216
## Median :  7
## Min :  5   Max :  9
## Std :  1.473434
## quatile :  5 9
## Skewness :  -0.2912679
##
## Mean :  15.94595
## Median :  16
## Min :  1   Max :  31
## Std :  8.707194
## quatile :  2 30
## Skewness :  -0.01283216

## $Ozone
## NULL
##
## $Solar.R
## NULL
```

```
##
## $Wind
## NULL
##
## $Temp
## NULL
##
## $Month
## NULL
##
## $Day
## NULL
```