# R Report 2 (Word Document)

Jeffrey Ugochukwu

6/27/2019

```
set.seed(10)
```

##I

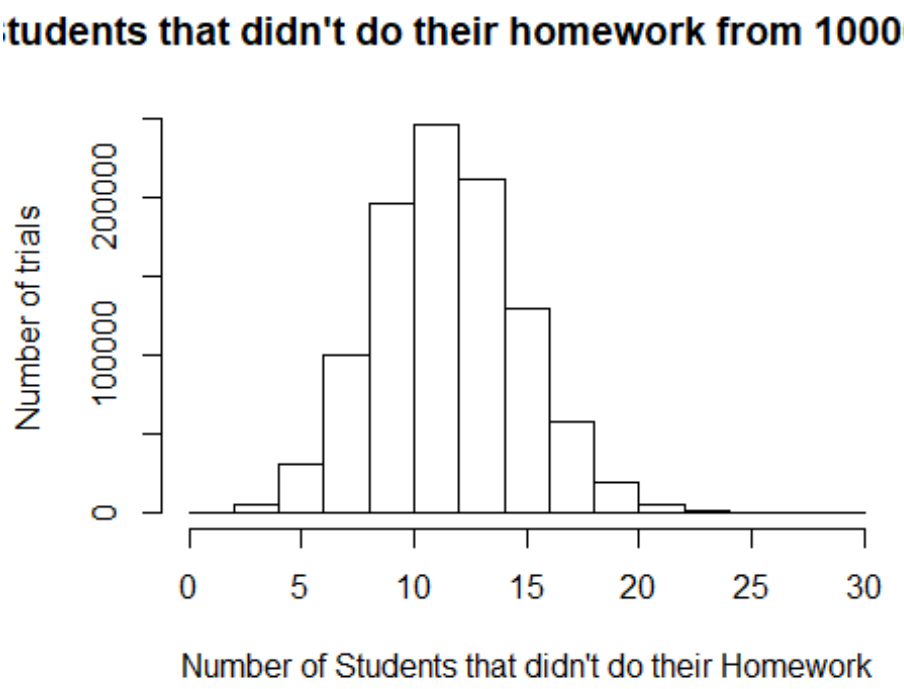
###a

I conducted a binomial distribution in which we have a sample size of 80 students and I'm trying to find the total number of students out of the 80 that didn't do their homework. The probability that a student doesn't do their homework (the probability of success) is 0.15 (15%). The total number of students that didn't do their homework is given below:

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

###b

The histogram that repsresents the number of students that didn't do their homework based on the 1000000 trials is roughly symmetric because even though there's a slight spread of values towards the right of the unimodal peak (10-12 students) which contains the number of studnets that didn't do their homework from a range of 20 to 23 students, majority of the specific ranges for the number of students is fairly symmetric to each other from both sides when you compare it to the unimodal peak.




```
## $breaks
##  [1]  0  2  4  6  8 10 12 14 16 18 20 22 24 26 28 30
##
## $counts
##  [1]    231   4454  29965 100061 195334 245819 211515 129105  57753  19450
## [11]   5089   1020    175     26      3
##
## $density
##  [1] 0.0001155 0.0022270 0.0149825 0.0500305 0.0976670 0.1229095 0.1057575
##  [8] 0.0645525 0.0288765 0.0097250 0.0025445 0.0005100 0.0000875 0.0000130
## [15] 0.0000015
##
## $mids
##  [1]  1  3  5  7  9 11 13 15 17 19 21 23 25 27 29
##
## $xname
## [1] "many_HW"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

###c

The mean and standard deviation of the number of students that didn't do their homework based on the 1000000 trials is given below:

```
## [1] 11.99925
```

```
## [1] 3.195229
```

###d

The probability that all of the students did their homework from the 1000000 trials is given below:

```
## [1] 4e-06
```

###e

The probability that at least 4 students didn't turn in their homework from the 1000000 trials is given below:

```
## [1] 0.998804
```

###f

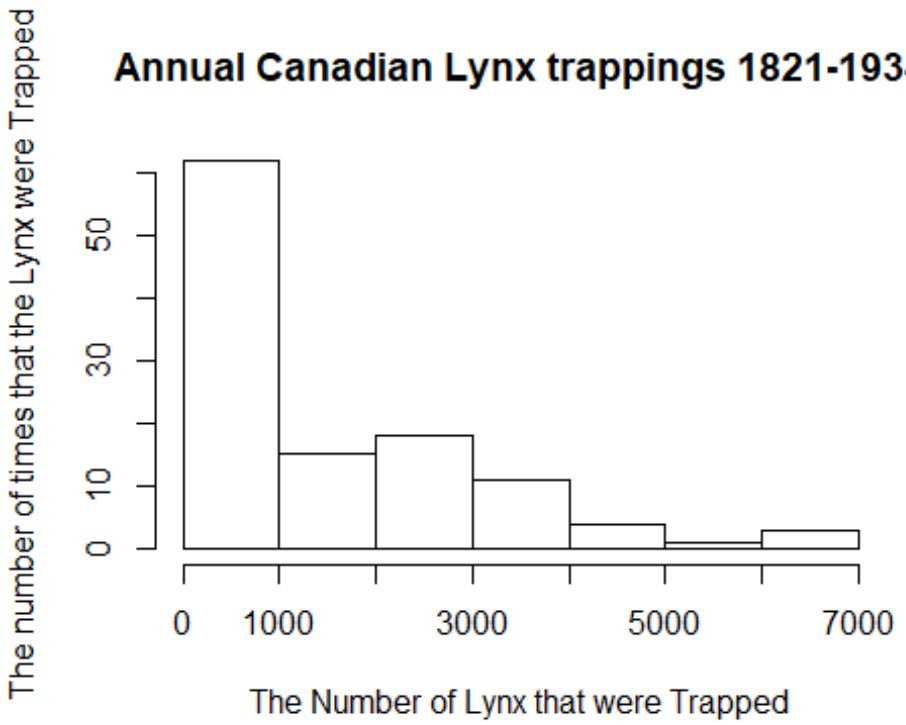
The median for the 1000000 trials on how many students didn't do their homework is given below:

```
## [1] 12
```

##II

###a

The histogram for the annual Canadian Lynx trappings doesn't appear to have a normal distribution since the majority of the ranges are skewed to the right from the unimodal peak (the range of 0 to 1000 lynx that were trapped). The mean and standard deviation of the lynx trappings is given below the histogram itself.




```
## $breaks
## [1]    0 1000 2000 3000 4000 5000 6000 7000
##
## $counts
## [1] 62 15 18 11  4  1  3
##
## $density
## [1] 5.438596e-04 1.315789e-04 1.578947e-04 9.649123e-05 3.508772e-05
## [6] 8.771930e-06 2.631579e-05
##
## $mids
## [1]  500 1500 2500 3500 4500 5500 6500
##
## $xname
## [1] "lynx"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"

## [1] 1538.018

## [1] 2514901
```

###b

With a sample size of 10, the mean of the number of lynx trappings is given below:

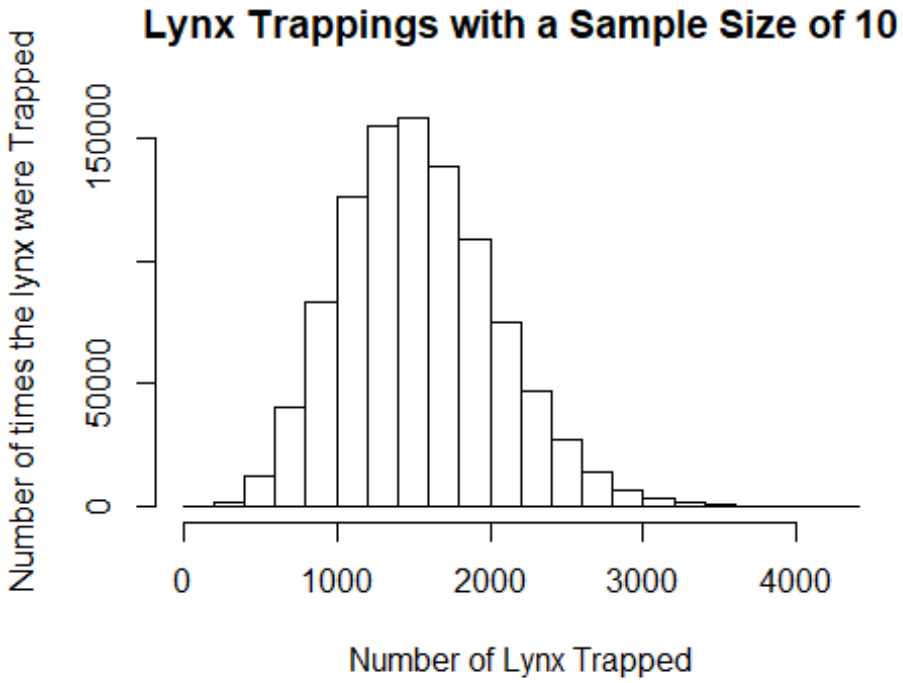
```
## [1] 1349.3
```

###c

The mean and standard deviation of the 1000000 trials of the lynx trappings with a sample size of 10 is given below:

```
## [1] 1537.531
```

```
## [1] 498.6045
```




```
## $breaks
##  [1]    0  200  400  600  800 1000 1200 1400 1600 1800 2000 2200 2400 2600
## [15] 2800 3000 3200 3400 3600 3800 4000 4200 4400
##
## $counts
##  [1]     10   1776  12643  40495  83125 126034 154603 158326 138714 108791
## [11]  75033  47301  27037  14161   6914   3000   1289    477    199     46
## [21]     20      6
##
## $density
##  [1] 0.000000050 0.000008880 0.000063215 0.000202475 0.000415625
##  [6] 0.000630170 0.000773015 0.000791630 0.000693570 0.000543955
## [11] 0.000375165 0.000236505 0.000135185 0.000070805 0.000034570
## [16] 0.000015000 0.000006445 0.000002385 0.000000995 0.000000230
## [21] 0.000000100 0.000000030
```

```
##
## $mids
##  [1]  100  300  500  700  900 1100 1300 1500 1700 1900 2100 2300 2500 2700
## [15] 2900 3100 3300 3500 3700 3900 4100 4300
##
## $xname
## [1] "replicate(1e+06, mean(sample(lynx, 10, replace = TRUE)))"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

###d

With a sample size of 50, the mean of the number of lynx trappings is given below:

```
## [1] 1264.38
```

###e

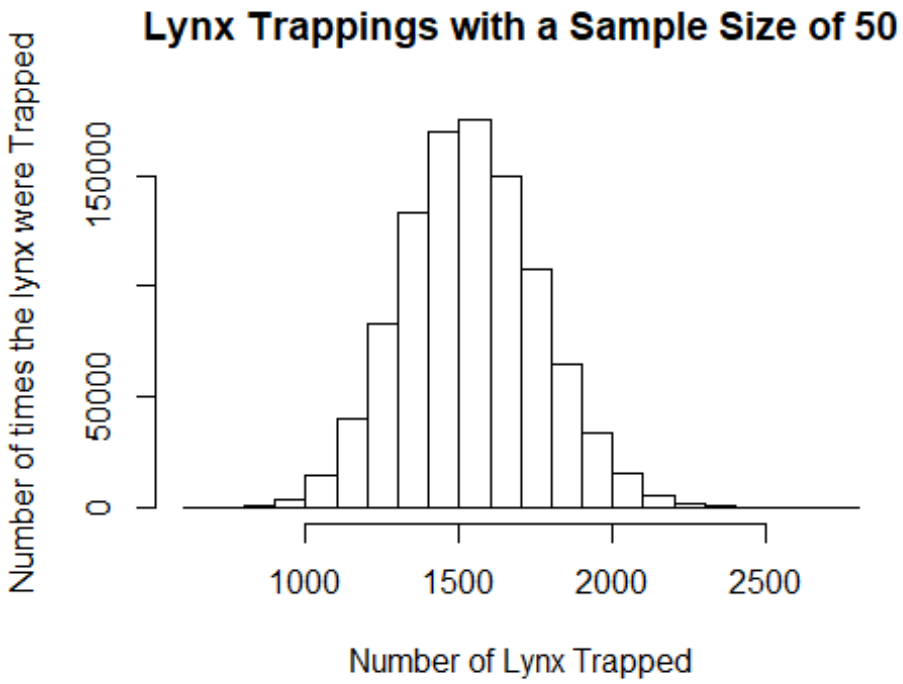
When the sample size of the mean for the number of lynx trappings has increased to 50, the mean roughly stayed the same (with only a slight decrease of 0.172), but the standard deviation drastically decreased in size. This would mean that when you increase the sample size of the sample statistic, your values would be closer to the true mean, which would decrease the variability of the values; therefore, the standard deviation would decrease. The mean and standard deviation of the number of lynx trappings with a sampke size of 50 are given below:

```
## [1] 1537.657

## [1] 223.1118
```

###f

A noticable difference between the histogram of the 1000000 trials for lynx trappings with a sample size of 10 versus the histogram of the 1000000 trials for lynx trappings with a sample size of 50 is that the histogram with the sample size of 50 is more symmetric than the histogram with a sample size of 10. The reason for this occurence is more on the fact that as you increase the sample size of a sample statistic, your values would be closer to the true mean, which would make the distribution more normal since you're decreasing the amount of varaition bewteen your values from the mean. The histogram with a sample size of 10 is slighty skewed to the right due to the fact that the small sample size will increase variation between the range of the values from the true mean of the sample. The histogram for the sample size of 50 is given below.




```
## $breaks
##  [1]  600  700  800  900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900
## [15] 2000 2100 2200 2300 2400 2500 2600 2700 2800
##
## $counts
##  [1]      4     75    706   3787  14588  40232  82783 133010 169499 175427
## [11] 149980 107314  65149  33697  15060   5782   2087    609    161     41
## [21]      8      1
##
## $density
##  [1] 0.00000004 0.00000075 0.00000706 0.00003787 0.00014588 0.00040232
##  [7] 0.00082783 0.00133010 0.00169499 0.00175427 0.00149980 0.00107314
## [13] 0.00065149 0.00033697 0.00015060 0.00005782 0.00002087 0.00000609
## [19] 0.00000161 0.00000041 0.00000008 0.00000001
##
## $mids
##  [1]  650  750  850  950 1050 1150 1250 1350 1450 1550 1650 1750 1850 1950
## [15] 2050 2150 2250 2350 2450 2550 2650 2750
##
## $xname
## [1] "replicate(1e+06, mean(sample(lynx, 50, replace = TRUE)))"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

##Appendex

```
set.seed(10)
SIM_HW = c("Not_HW","HW")
set.seed(10)
a = c(sample(SIM_HW,80,replace = TRUE,prob = c(0.15,0.85)))
NO_HW = length(a[a=="Not_HW"])
NO_HW
simulate_HW = function(x = 0){
  SIM_HW = c("Not_HW","HW")
  a = c(sample(SIM_HW,80,replace = TRUE,prob = c(0.15,0.85)))
  NO_HW = length(a[a=="Not_HW"])
  NO_HW
}
many_HW = sapply(1:1000000,simulate_HW)
many_HW_hist = hist(many_HW, main = "Students that didn't do their homework
from 1000000 trials", xlab = "Number of Students that didn't do their
Homework", ylab = "Number of trials")
many_HW_hist
many_HW_mean = mean(many_HW)
many_HW_mean
many_HW_sd = sd(many_HW)
many_HW_sd
all_HW = length(many_HW[many_HW == 0])
all_prob = all_HW/1000000
all_prob
four_atleast_no_HW = length(many_HW[many_HW >= 4])
four_prob = four_atleast_no_HW/1000000
four_prob
many_HW_median = median(many_HW)
many_HW_median
lynx = as.numeric(lynx)
lynx_hist = hist(lynx, main = "Annual Canadian Lynx trappings 1821-1934",
xlab = "The Number of Lynx that were Trapped", ylab = "The number of times
that the Lynx were Trapped")
lynx_hist
lynx_mean = mean(lynx)
lynx_mean
lynx_var = var(lynx)
lynx_var
lynx_sample = sample(lynx,10,replace = TRUE)
lynx_sample_mean = mean(lynx_sample)
lynx_sample_mean
many_lynx_10_mean = mean(replicate(1000000,mean(sample(lynx,10,replace =
TRUE))))
many_lynx_10_mean
many_lynx_10_sd = sd(replicate(1000000,mean(sample(lynx,10,replace = TRUE))))
many_lynx_10_sd
many_lynx_10_hist = hist(replicate(1000000,mean(sample(lynx,10,replace =
TRUE))),main = "Lynx Trappings with a Sample Size of 10", xlab = "Number of
```

```
Lynx Trapped", ylab = "Number of times the lynx were Trapped")
many_lynx_10_hist
lynx_sample = sample(lynx,50,replace = TRUE)
lynx_sample_mean = mean(lynx_sample)
lynx_sample_mean
many_lynx_50_mean = mean(replicate(1000000,mean(sample(lynx,50,replace =
TRUE))))
many_lynx_50_mean
many_lynx_50_sd = sd(replicate(1000000,mean(sample(lynx,50,replace = TRUE))))
many_lynx_50_sd
many_lynx_50_hist = hist(replicate(1000000,mean(sample(lynx,50,replace =
TRUE))), main = "Lynx Trappings with a Sample Size of 50", xlab = "Number of
Lynx Trapped", ylab = "Number of times the lynx were Trapped")
many_lynx_50_hist
```