Plotting Exercise 1

Regina Zweng

9/26/2017

Imagine that you wish to conduct a study of height of people living in Los Angeles. One strategy is to take a sample of 100 people and compute the average height. Another study design would involve taking a sample of 1000 people and computing the average height. You wish to test the extent to which the different sample sizes affect your estimates of the average height.

To test this, conduct a simple simulation study in R. Assume that the heights of people from LA are normally distributed with a mean of 69 inches and a standard deviation of 10 inches.

In order to conduct the simulation study, you should do the following: 1. Write a function (called "get_heights") in R to draw a sample of individuals (either 100 or 1000) from the population. Hint: You will want to use "rnorm" within your function. Store the random heights that you've generated in a variable called "heights".

```
get_heights<-function(samplesize=100){rnorm(samplesize,mean=69,sd=10)}

heights<-get_heights(samplesize=100)
print(heights)
```

```
##   [1] 63.44214 77.11027 64.17015 66.51804 68.83649 68.17535 84.04827
##   [8] 70.02097 51.07758 57.07811 60.99811 67.94941 74.20481 75.45009
##  [15] 76.54549 55.28628 76.46087 75.79867 62.43491 49.14114 32.02210
##  [22] 75.25314 65.21683 78.90472 55.51858 98.50963 78.88683 63.85907
##  [29] 60.86482 63.04426 80.27819 68.51414 54.95205 84.82524 55.07346
##  [36] 68.00882 54.53066 48.02246 65.63239 69.65435 73.50277 77.99713
##  [43] 64.48134 67.65559 70.91646 57.92881 59.00694 63.76804 86.71800
##  [50] 86.14940 74.52105 77.36754 72.09026 70.96766 56.33517 79.29249
##  [57] 54.83835 58.68126 61.59430 46.58933 60.11355 76.87423 91.30380
##  [64] 58.07376 94.22668 58.44836 71.87496 57.46462 70.93379 69.99631
##  [71] 59.09547 77.30897 83.63886 76.90233 65.88101 79.39729 80.38800
##  [78] 78.98565 75.80698 67.72107 61.71850 71.23078 52.52205 57.14673
##  [85] 66.46570 80.99497 65.05930 83.52971 75.80908 70.29521 61.32555
##  [92] 76.81959 61.14113 74.34009 74.07754 74.80937 84.19930 62.42452
##  [99] 65.42452 81.53581
```

2. Within your function, compute the average height from your "heights" vector.

```
get_heights<-function(sample_heights=100){
  heights<-(rnorm(sample_heights,mean=69,sd=10))
  average_heights<-mean(heights)
  }
get_heights(100)
```

3. Make your function return the average height.

```
get_heights<-function(sample_heights=100){
  heights<-(rnorm(sample_heights,mean=69,sd=10))
  average_heights<-mean(heights)
  return(average_heights)}
get_heights(100)
```

```
## [1] 70.35151
```

4. Use a "for" loop to call your "get_heights" function 1000 times, with taking a sample of size 100 from the population. Save the mean height from each replicate in a vector called "mean_heights_100".

```
mean_heights_100<-rep(NA,1000)
for (i in 1:1000)
  {mean_heights_100[i]<-get_heights(100)}
```

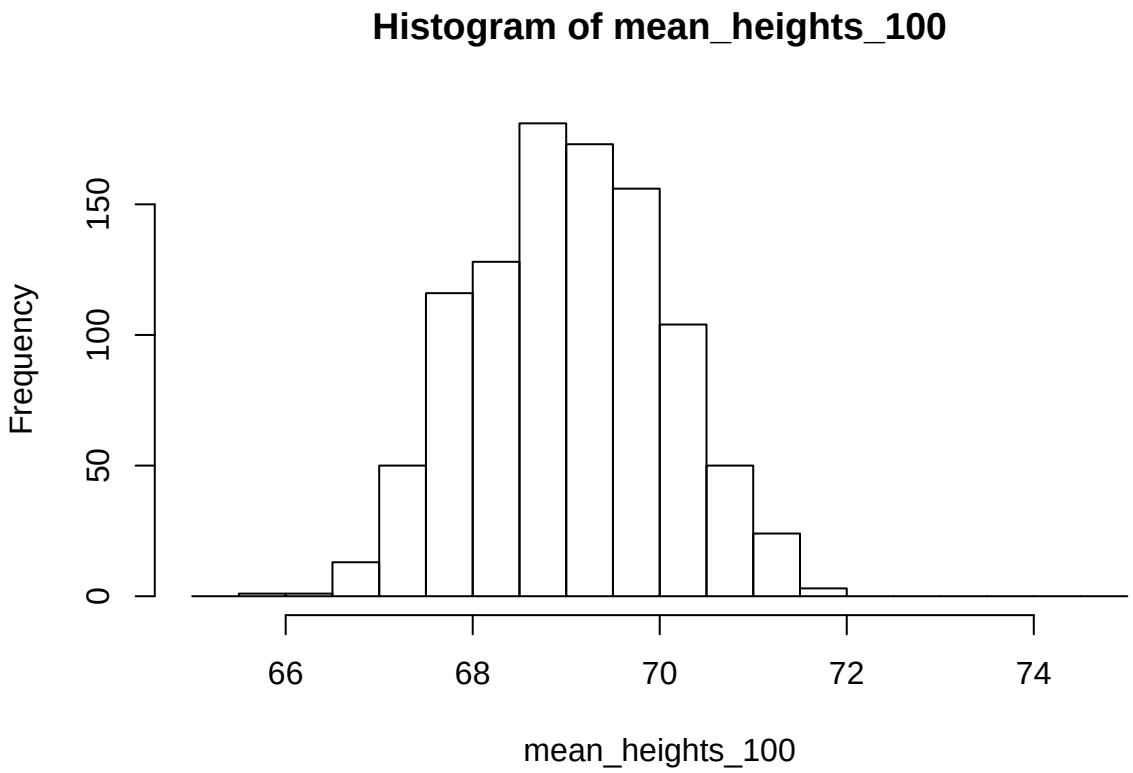
5. Use a "for" loop to call your "get_heights" function 1000 times, with taking a sample of size 1000 from the population. Save the mean height from each replicate in a vector called "mean_heights_1000".

```
mean_heights_1000<-rep(NA,1000)
for (i in 1:1000)
  {mean_heights_1000[i]<-get_heights(1000)}
```

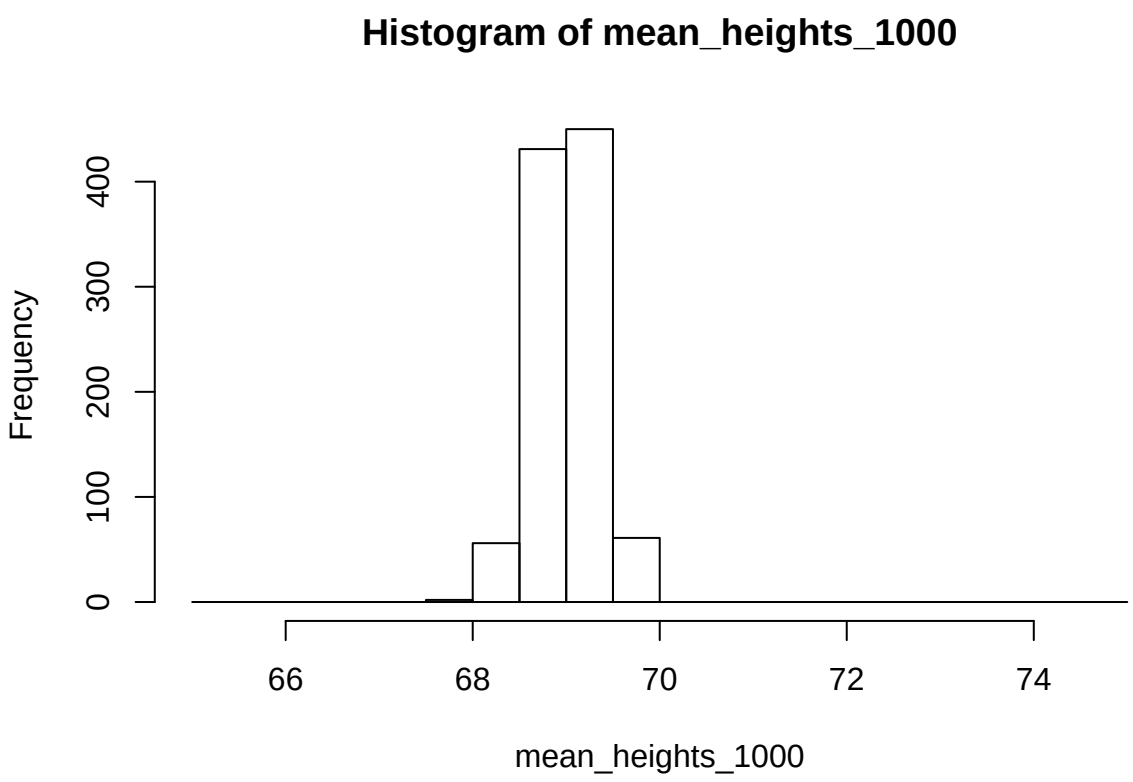
6. Plot a histogram of the distribution of the average heights for your sample size of 100 and 1000 individuals. The two sets of data should be plotted on the same axes. Add a legend. Label the axes. Plot the data from the 100 samples in red and the data from the 1000 samples in blue.

Your plot should look something like the one shown on the next page.

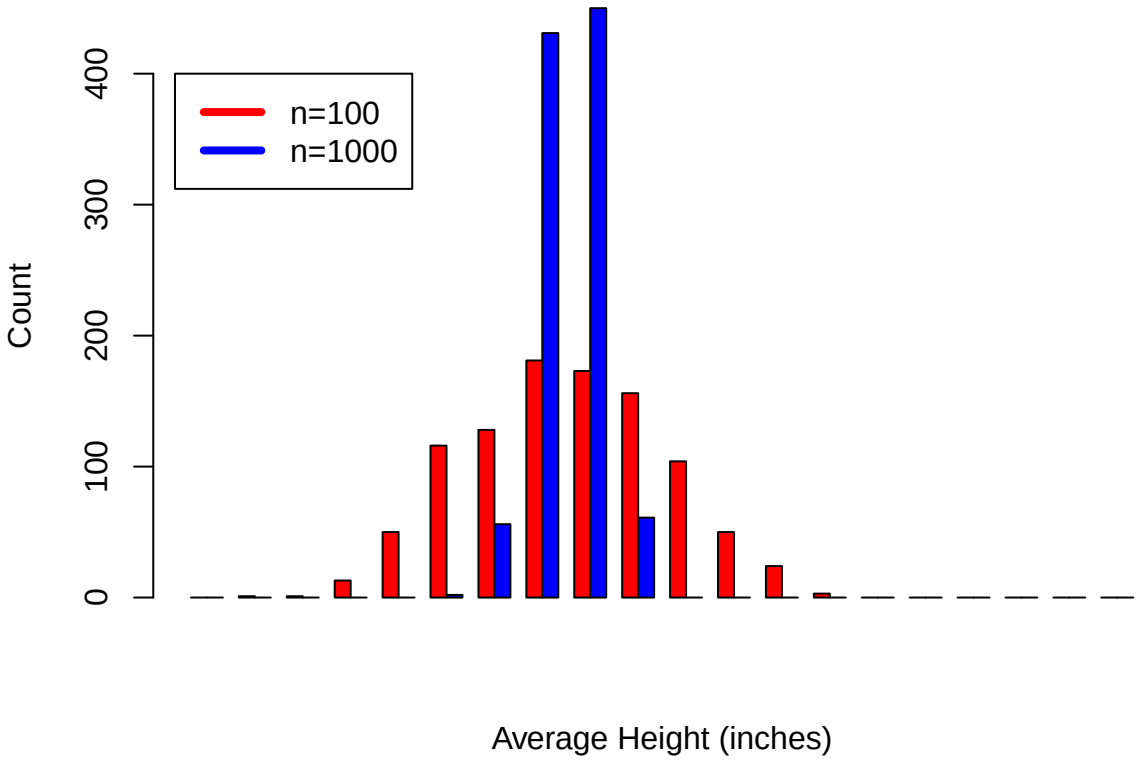
```
bins<-seq(65,75,by=0.5)
counts_hundred<-hist(mean_heights_100,breaks=bins)$counts
```



```
counts_thousand<-hist(mean_heights_1000,breaks=bins)$counts
```




```
par(mfrow=c(1,1),mar=c(4,4,3,2))
barplot(rbind(counts_hundred,counts_thousand),col=c(2,4),beside=T,xlab="Average Height (inches)",ylab="
legend(x=0,y=400,legend=c("n=100","n=1000"),col=c(2,4),lwd=4)
```


n=100
n=1000
Count
0
100
200
300
400
Average Height (inches)