Assume that squirrel weights are approximately normally distributed with $\mu=1.7$ lbs and $\sigma=0.5$ lbs. Let's simulate a sample of 4 squirrels. Open a new script and start by setting the seed to 338. This ensures that everyone will get the same values (it's "random" but reproducible). Then, we'll simulate randomly selecting 4 squirrels and weighing them using the **rnorm** command.

```
> set.seed(338)
> n_squirrels <- 4
> normal_draws <- rnorm(n = n_squirrels, mean = 1.7, sd = 0.5)
```

Question #1 Using the **print** or **View** function, look at the weights of the four simulated squirrels. Record the four weights below, rounded to three decimal places.

1.529, 2.141, 2.903, 1.725

The next step in our simulation is to compute a statistic that summarizes the sample of four squirrels. In this lab we'll compute both the mean and standard deviation.

```
> mean(normal_draws)
> sd(normal_draws)
```

Question #2 What are the mean and standard deviation of the weights of these four squirrels? Are they exactly equal to the population mean and standard deviation? Are they close?

Mean is 2.074515 and the standard deviation is 0.6085165.

The numbers are fairly close, the mean is off by 0.3

Finally, we will simulate this process many times (1000 to be exact) to obtain 1000 sample means and 1000 sample standard deviations.

```
> n_samples <- 1000
> normal_matrix <- matrix(0, nrow = n_samples, ncol = n_squirrels) #
initialize a matrix containing the actual random draws


> set.seed(338)  # reset seed
> for(i in 1:n_samples){  # a for loop - do this n_samples times
+  normal_matrix[i,] <- rnorm(n = n_squirrels, mean = 1.7, sd = 0.5)
```

```
+ }  # there is actually an easier way to do this in R, but we code it this
way to illustrate conceptually what is actually being done


> mean_vector <- apply(normal_matrix, 1, mean)

> sd_vector <- apply(normal_matrix, 1, sd)
```

This code takes advantage of the **apply** function. In R, the **apply** function indicates to apply a function (the third argument) over a dimension (the second argument: 1 = rows, 2 = columns, etc.) of a matrix or data frame (the first argument). Thus, **mean_vector** is computed by taking the mean of every row of our matrix.

Question #3 Explain the difference between **n_squirrels** and **n_samples**. Which one should we use as our *n* in the formula for computing the standard deviation of the sampling distribution?

n_samples is the amount of times the simulation will run where that number can change to be larger or smaller. n_squirrels is the amount of all squirrels in the given population and cannot change. This is considered the sample.

Question #4 Compute the theoretical mean and standard deviation of the sampling distribution of the mean squirrel weight in samples of size *n* (where *n* is the number you identified in **Question #3**).

Our theoretical would follow the formula for standard deviation is sigma divided by the square root of the number of elements in the population. This will evaluate to 0.5/(4)^1/2 = 0.25. The formula for mean would simply the mean given which is 1.7.

Question #5 What are the actual mean and standard deviation of your sample means? Are they close to what you calculated in **Question #4**? (Hint: use the **mean()** and **sd()** functions on the vector of sample means)

The actual mean is 1.696769 and the actual standard deviation is 0.1924363. The mean value is fine but the standard deviation value is off by around 0.05 but should be fine for all intents and purposes.

Question #6 Using the **pnorm** command and your answers to **Question #4**, find the approximate probability of obtaining a sample mean weight between 0.8 and 1.2 lbs (assuming the sampling distribution is normal). It may help to look at your code from Lab 13.

0.02259102 is the probability for the weight to be in between 1.2 pounds and 0.8 pounds given the sample size of four squirrels.

Question #7 In Lab 13, we computed the probability of obtaining a squirrel weight between 0.8 and 1.2 lbs to be 0.1227. Why do you get a different answer to **Question #6**?

There are significantly more squirrels in the sample for Lab 13.

```
> squirrel_df <- data.frame(mean = mean_vector, sd = sd_vector)

> library(ggplot2)

> squirrel_plot <- ggplot(data = squirrel_df, mapping = aes(x = mean))

> squirrel_density <- squirrel_plot + geom_histogram(aes(y = ..density..),
binwidth = 0.1, center = 1.7) +

+ stat_function(fun = dnorm, n = 101, args = list(mean = mean_sample_means,
sd = sd_sample_means)) +  # replace mean_sample_means and sd_sample_means
with your answers to question 4

+ labs(title = "Simulated Sample Means (n = 4) of Squirrel Weights", x =
"My X-Axis Label", y = "Density")  # fix the x-axis title

> print(squirrel_density)
```

Question #8 The code above creates a density histogram of the sample means of squirrel weights and overlays a normal density curve with estimated mean and sd. Make sure you type the **+** if it appears at the end of a line, but ignore it if it appears at the beginning of a line. Fix the code as indicated in the code's comments, then paste the graph below.

Please see attached graph for solution.

This code finds the number of simulated sample means that are greater than 0.8 but less than 1.2:

```
> sum(mean_vector > 0.8 & mean_vector < 1.2)
```


Question #9 How many of your sample means were between 0.8 and 1.2 lbs? What proportion is that? Is the proportion close to what you calculated in **Question #6**?

The number of squirrels that fall under these criteria is 19. It's proportion is 19/1000 which is 0.019 and is fairly close to the value calculated above.

Question #10 Was the first simulated sample of 4 squirrels unusually heavy? Explain your reasoning. (Hint: Use the **pnorm** command to find the probability of simulating a heavier sample mean of 4 squirrels than the sample mean you got in **Question #2**. Remember that you can add the argument **lower.tail = FALSE** to the command, e.g., **pnorm(0.5, mean = 0, sd = 1, lower.tail = FALSE)**, to find 1 – cumulative proportion instead.)

Since we got a value of 0.06705883, this does not breach the threshold of 1% which would make it unusual. We can conclude that it is highly unlikely that we would have had a heavier first simulated sample of 4 squirrels.