R Tutorial for STAT 350 Lab 4

Author: Leonore Findsen; Jeremy Troisi

1) Generate sampling distributions

The usefulness of the Central Limit Theorem (CLT) is through the averaging of random samples to simulate what is happening when many samples are taken from the same distribution.
Even though you will just be running the provided code (with minor modifications), I will outline what is happening below:

1) Generate the appropriate number of SRSs of each of the distributions.

2) Average (by row) the appropriate number of SRSs and put into Column G. For example in the following table, we are averaging six different samples (columns) each with 5 SRS (rows).

	A	B	C	D	E	F
1	-0.30	-1.28	0.24	1.28	1.20	1.73
2	-2.18	-0.23	1.10	-1.09	-0.69	-1.69
3	-1.85	-0.98	-0.77	-2.12	-0.57	-0.40
4	0.13	-0.37	-0.33	-0.37	1.34	-0.09
5	-0.19	-0.51	1.97	0.87	2.38	-0.65

3) From column G, generate histogram, QQ plot, mean and standard deviation.

code:

```
SRS <- 1000 #the number of repeats (not to be changed)

# n: the number of columns that are being averaged over
n <- 1
# I strongly suggest that you change any titles to display the number
# of columns that you use are averaging and the type of distribution.

#calculates the average data
data.vec <- rnorm(SRS*n,mean=0,sd=1) #creates the random data
data.mat <- matrix(data.vec, ncol = n) #separates the data into columns
#apply(matrix, c(1, 2) == c("row", "column"), function)
avg <- apply(data.mat, 1, mean) #performs the averaging

. . .
```

The data that you want to analyze is in the variable 'avg'. You will need to include the rest of the code as appropriate to produce what is required in the assignment. Please see Lab 3 for details.

STAT 350: Introduction to Statistics
Department of Statistics, Purdue University, West Lafayette, IN 47907

R Tutorial for STAT 350 Lab 4

Author: Leonore Findsen; Jeremy Troisi

2) Generate random samples
In Lab 3, we have already discussed how to generate a normal distribution, exponential, Cauchy, and a uniform distribution. Be careful about what the rate is when using the exponential distribution. For the exponential, the parameter in R is λ.

Use the following commands to generate the other two distributions in this lab:

```
data.vec <- rgamma(SRS*n,2,rate=1) #creates the random data
```

```
data.vec <- rpois(SRS*n,2) #creates the random data
```

Remember to use the appropriate data set when you are generating the normal quantile plot and the histogram. I would also suggest changing the title so that you know which plot this is.