Recall a $(1-\alpha)100\%$ confidence interval (CI) for p based on $\hat{p}$ is given by

$$\widehat{p} - z^* \mathrm{SE}(\widehat{p}) \leq p \leq \widehat{p} + z^* \mathrm{SE}(\widehat{p})$$

where z^* solve $P(-z^* \leq Z \leq z^*) = 1-\alpha$.

A CI isn't guaranteed to contain the true value p, rather it has a probability $1-\alpha$ of doing so. Visualizing CIs can help reinforce this.

This project uses a function `plotCI()` to graphically display simulated confidence intervals. When simulating, a known value of p must be specified, although this isn't known in reality.

This function must be downloaded:

```
> source("http://wiener.math.csi.cuny.edu/st/R/plotCI.R")
```

The function can be used in several ways by giving different values to the arguments:

```
> plotCI()                      # shows m=50 95% CIs for n=10, p=0.5
> plotCI(n=100)                 # shows m=50 95% CIs for n=100, p=0.5
> plotCI(p=0.25)                # shows m=50 95% CIs for n=10, p=0.25
> plotCI(conf.level = 0.80)     # Makes 80% CIS, not 95%
```

One can also find CIs for the population mean for some named distributions.

Download the function and use it to answer these questions:

1. Make 50 CIs for $n = 10$ and $p = 0.5$. How many CIs contain p? What percent is this?

2. Make 50 CIs for $n = 200$ and $p = 0.5$. How many CIs contain p? What percent is this?

3. Suppose you generate 100 95% CIs for p. How many of these do you expect to contain p? Why?

4. Does you answer to the last question depend on the sample size n? (the 100 is the value of m)

5. A 95% CI has a width given by $2 * 1.96 * \mathrm{SE}(\widehat{p})$, or 2 margin of errors. Create 50 CIs with $n = 10$ and 50 with $n = 40$. By quadrupling n, how much do you change the margin of error?

6. Estimate the margin of error for a 95% CI when $p = 0.5$ and $n = 500$, $n = 1000$, $n = 2000$. Based on your investigation, what sample size produces a MOE of 3 percentage points.

7. Play around to find what sample size produces a margin of error of 2 percentage points when $p = 0.5$.

8. Do several simulations with the default values $m = 50, n = 10, p = 0.5$. For each record the number of times the CI misses.

   (a) Is the value recorded the same for each simulation, or is it random?

   (b) If the value is random, can you describe the distribution?

   (c) What is the expected number to record each time

   (d) What is the standard deviation of this number?

9. There are built in functions to compute confidence intervals in R. The function `prop.test()` is used with the number of successes and the number of trials. For instance a 95% CI based on a sample of size 1000 of which $\hat{p} = .34$ (or $X = 340$) is

```
> prop.test(340, 1000)

        1-sample proportions test with continuity correction
data:  340 out of 1000, null probability 0.5
X-squared = 101.761, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3108142 0.3704312
sample estimates:
   p
0.34
```

   Look for the line

```
95 percent confidence interval:
 0.3108142 0.3704312
```

   The function `binom.test()` uses the exact binomial distribution for the calculation, not an approximation. It is used in an identical manner.

   Find 95% CIs for p when $n = 800$ and $\widehat{p} = .01$. Use both `prop.test()` and `binom.test()` and compare any differences.

   Repeat with $\widehat{p} = .10$. Are there differences?