Worksheet: Inference with proportions

Key formulas:

Suppose that some percentage $p$ of a population has some characteristic. Draw a sample of size $n$, and let $\hat{p}$ denote the percentage of the sample that has the characteristic. The distribution of $\hat{p}$ satisfies the following:

1. The **standard error** of $\hat{p}$ is
   \[
   SE = \sqrt{\frac{p(1-p)}{n}},
   \]
   where $p$ is the true population proportion.

2. The **mean** of $\hat{p}$ is $p$.

3. The **shape** of $\hat{p}$ is roughly normal, assuming the sample size $n$ is large enough. (Specifically: $np \geq 10$ and $n(1-p) \geq 10$.)

4. The **confidence interval** for a single proportion is given by
   \[
   \hat{p} \pm z^* \sqrt{\frac{p(1-p)}{n}}
   \]
   where $z^*$ is the standard normal endpoint to give the desired confidence.

5. The **sample size** needed to achieve a margin of error of size $ME$ with confidence $C$ is
   \[
   n = \left( \frac{z^*}{ME} \right)^2 p(1-p),
   \]
   where $z^*$ is the standard normal endpoint to give the desired confidence. (If $p$ is totally unknown, use $p = .5$.)

6. To test a claim $H_0 : p = p_0$ against an alternative $H_a : p \neq p_0$, use the normalized test statistic
   \[
   z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.
   \]
   Provided $pn \geq 10$ and $n(1-p) \geq 10$, the $P$–value can be read from a standard normal curve.

Exercises:

1. The following represent empirical histograms of $\hat{p}$ for various choices of $n$ and $p$. Which ones seem normal? Do they satisfy the above criterion for a “large enough” sample?
2. Suppose that with a sample of size $n$, your standard error for $\hat{p}$ is 3. How would the standard error change if the sample size were increased to $4n$?

3. Derive the formula for the sample size needed to achieve a specific margin of error with a specific confidence.

4. (6.34) In a survey of 1000 US adults, twenty percent say they never exercise. This is the highest level seen in five years. Find and interpret a 99% confidence interval for the proportion of US adults who say they never exercise. What is the margin of error, with 99% confidence?

5. (6.64) There were 2430 Major League Baseball (MLB) games played in 2009, and the home team won the game in 54.9% of the games. If we consider the games played in 2009 as a sample of all MLB games, test to see if there is evidence, at the 1% level, that the home team wins more than half the games. Show all details of the test.