

HW 3

Due: Friday, September 27th by 11:59 pm

Instructions

1. Create your own .Rmd file called last_first_hw3.Rmd, replacing “last” and “first” with your first and last name.
2. In the setup chunk, make sure to globally set echo = TRUE, message = FALSE, warning = FALSE.
3. Type up your solutions to the following questions. Do NOT include the questions in your document, only your solutions.
4. Upload a pdf/doc AND Rmd.

In this assignment, you will explore the theory behind the sampling distribution of a sample proportion, then build upon this in real-world applications.

Theory

- 1) Consider a population proportion, $p = 0.05$. Using the resampling techniques discussed in Lecture 9, generate 10,000 samples of sizes $n = 15, 30$, and 50 . Set a seed before generating these random samples using your birth date (e.g. 01011990).
- 2) Mean and standard error:
 - a) Compute the mean and standard error from each distribution. Display these statistics in a well-organized table.
 - b) Compare the differences (if any) between the means and standard deviations from each distribution.
 - c) Calculate the theoretical mean and standard error from each proposed sampling distribution and compare these values with those from your samples in 1-3 complete sentences.
- 3) Quantiles
 - a) Find the 25th and 75th percentiles from each of your simulated sampling distributions.
 - b) Using the qnorm() function, calculate the true first and third quartiles for each sampling distribution.
 - c) Compare the true quartiles with the empirical quartiles in 1-3 sentences.
- 4) Generate histograms of the three sampling distributions of the sample proportions and display them side-by-side (Optional: add a vertical line in each plot to show the mean and add a legend with the mean and standard error values)
- 5) Generate a QQ plot for each sampling distribution and evaluate the normality of each distribution.
- 6) Using the results above, discuss in 3-5 complete sentences the validity of the Central Limit Theorem for \hat{p} when the population proportion, p , is small.

Application

In August of 2012, news outlets ranging from the [Washington Post](#) to the [Huffington Post](#) ran a story about the rise of atheism in America. The source for the story was a poll that asked people, "Irrespective of whether you attend a place of worship or not, would you say you are a religious person, not a religious person or a convinced atheist?" In this assignment, the parameter of interest is the **true proportion of atheists** of one of the nationalities represented in the survey.

Unset

```
# Load the data
download.file("http://www.openintro.org/stat/data/atheism.RData", destfile =
"atheism.RData")
load("atheism.RData")
```

- 7) Pick one nationality of your choice from the dataset (except Afghanistan and Vietnam because there are too few atheists reported to complete the following questions). Subset the data to only include that nationality and the results from 2012.
- 8) Report the sample size and sample proportion of atheists from your subsetted dataset.
- 9) Conditions:
 - a) Does it seem reasonable to assume that the people who participated in this poll are independent from one another? Comment in a sentence or two.
 - b) Since we do not know the population proportion of atheists of the nationality chosen, use your sample proportion to verify that the success/failure condition of the CLT is met.
- 10) Regardless of whether the conditions are met, calculate a 95% confidence interval for p . Interpret your interval in the context of the problem.