

# HW 4

Due: Wednesday, October 16th by 11:59 pm

## Instructions

- Create your own .Rmd file called last\_first\_hw4.Rmd, replacing “last” and “first” with your first and last name.
- **In the setup chunk, make sure to globally set echo = TRUE, message = FALSE, warning = FALSE.**
- Type up your solutions to the following questions. Do NOT include the questions in your document, only your solutions.
- Upload a pdf/doc AND Rmd.

- 1) The CLT (for means) states that if we have a random sample  $X_1, X_2, \dots, X_n \sim_{iid} N(\mu, \sigma)$ ,

then  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma}{\sqrt{n}})$  then . Prove  $E[\bar{X}] = \mu$  and  $SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$  .

- 2) Generate a random sample of size 500 from an exponential distribution with rate parameter  $\lambda=1$ . Draw a histogram of your sample.
- 3) Comment in detail on the shape, center, and spread of your histogram in complete sentences.
- 4) Generate 5,000 samples of sizes  $n = 5, 30$ , and 100 from an exponential distribution with rate parameter  $\lambda=1$ . For each sample, calculate the mean (so you should have 5,000 sample means for each sample size) and plot the distributions of the sample means. *Make sure to set a seed for reproducibility.*
- 5) As the sample size increases, what happens to the shape of the sampling distribution? Why does this happen according to the CLT?
- 6) Report the mean and standard deviation from each of your sampling distributions in a publication-quality table (no R output). In the same table, include the theoretical mean and standard error for each sample size for comparison.
- 7) In complete sentences, compare the values of the means and standard deviations from your samples to the theoretical values.
- 8) Simulate a small sample of size 10 from an exponential with  $\lambda=1$  and calculate the 95% confidence interval for the mean using both the normal distribution (incorrect method) and the t-distribution (correct method). *Make sure to set a seed for reproducibility.*
- 9) Using the results from above, explain why the t-distribution is necessary when dealing with small samples and unknown population standard deviation in complete sentences.
- 10) Read the ASA Statement on p-values [here](#). In your own words, explain in a paragraph why there can be a problem with using p-values to perform significance tests.
- 11) Was there anything you found difficult with this homework? What topics (if any) do you feel you still need more work on?
- 12) Give yourself a rating for this assignment using the EMRN rubric.

E - Excellent; M - Meeting expectations; R - Revision needed; N - Not accessible