# Stat 67 Project 2

Due: Wednesday, May 25th 1:00pm

Answer questions in order, number them as they are numbered here. Staple multiple sheets together. Put your name in the upper right hand corner and your ID number in the lower right. **Place code in an appendix at the end of the project.** You are welcome to use a language other than R to do your project, but R has several built in functions that might help you.

- Some useful R functions: sample, qqnorm, qqline, hist, matrix, rowMeans, qnorm, pnorm, qt, pt, sqrt
- To count how many times X>Y: sum(X>Y)
- To count how many times X>Y and X<Z: sum(X>Y&X<Z)
- To count how many times X>Y OR X<Z: sum(X>Y|X<Z)

---

Bird Species A is known to have nests that contain a certain number of eggs according to the following distribution:

| Number Eggs in Nest | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Probability | .05 | .05 | .05 | .1 | .15 | .15 | .40 | . 05 |

1. Calculate the population average number of eggs in a nest and the population standard deviation. (Use formulas)
2. Draw a sample of size 100,000 from this distribution and make a histogram.
3. An ecologist studying a local wetland surveys 5 nests of bird species A. She finds that the five nests contain an average of 7.2 eggs per nest with a sample standard deviation of 0.45 eggs per nest. She would like to know if this is unusual.
   (a) Using your own words, explain why you cannot do a t-test in this situation
   (b) Using your own words, explain why you cannot do a z-test in this situation
   (c) Using R, simulate 10,000 samples of 5 nests each. Show a histogram of the averages of the samples.
   (d) Show a qqnorm plot of the averages. Do they look normally distributed?
   (e) Using R and your 10,000 samples of 5 nests each, decide how unusual an average of 7.2 eggs or greater would be.
   (f) Using the 10,000 samples, make 10,000 95% confidence intervals using the z-distribution, $\bar{x}$ calculated from your samples, and $s$ calculuated from your samples. How many contain the true mean (calculated in (1))? Comment on what this shows about using the normal approximation in this situation.
4. How many nests would you need before you would feel comfortable using the normal approximation? Try different numbers of nests until you are happy with the look of the qqplot and the histogram. Show the qqplot and the histogram for the n you have chosen.
   (a) If you are using the n you found in (4), would you use a t-distribution or a z-distribution (normal distribution), when you do inference? Explain why.

(b) Using the n you found in (4), create 10,000 95% confidence intervals.

i. How many do you expect to contain the true mean?

ii. How many contain the true mean (calculated in (1))?

(c) Using the n you found in (4), take 10,000 samples of size n. Carry out 10,000, hypothesis tests at the $\alpha$=0.05 level, for the hypothesis:

$$H_0 : \mu = E[X] \text{ calculated in (1)}$$

$$H_A : \mu \neq E[X] \text{ calculated in (1)}$$

i. How many times do you expect to reject the null?

ii. How many times did you reject the null?

iii. Is a type 1 error possible in this situation? If so, how many type 1 errors are there?

iv. Is a type 2 error possible in this situation? If so, how many type 2 errors are there?

(d) What is the name of the theorem that you used in (4)?