# STAT511 HW#8

**Reading:** Read Chapter 10 of Ott & Longnecker
**See Canvas Calendar for Due Date.**

40 points total, 2 points per problem part unless otherwise noted.

1. Suppose Y is a binomial random variable with n = 22 and $\pi = 0.72$. Compute the following.
   A. Mean and standard deviation of Y.
   B. $P(Y \leq 16)$
   C. $P(Y < 16)$
   D. $P(16 \leq Y < 18)$
   E. $P(Y = 18)$
   F. $P(Y \geq 18)$
   G. The normal approximation to $P(Y \geq 18)$ without continuity correction.
   H. The normal approximation to $P(Y \geq 18)$ with continuity correction.

2. In a sample of 75 randomly selected students, 51 of them are STEM majors. Express all answers to the following questions as proportions. For this question, do the calculations "by hand".
   A. Give an estimate for the proportion of students who are STEM majors.
   B. Provide a 95% confidence interval for the true proportion of students who are STEM majors. Use the large sample normal approximation (slide 13 of the CH10 notes).
   C. Conduct a hypothesis test (using the large sample normal approximation) with $\alpha$=0.05 to test $H_A$: $\pi$>0.5. Give the Z test statistic, p-value, and conclusion (**4pts**).

3. A factory manager decided to estimate the proportion of defective items. A random sample of 60 items was inspected and it was found that 10 of them are defective.
   A. Give an estimate for the proportion of defective items.
   B. Using R, calculate a **90%** confidence interval for the true proportion of defective items using the normal approximation. NOTES: (1) Use correct=TRUE (default). (2) The R CI will not match a hand calculation for this problem because R uses a different formula.
   C. Using R, calculate a **90%** confidence interval for the true proportion of defective items using the exact binomial method.
   D. Is the sample size large enough for the normal approximation to be valid? Justify your response using the "better" criteria discussed in the notes.

4. A researcher is planning to run a one-way ANOVA to compare 4 different pesticides used on apple trees. The response variable is yield. The researcher wishes to calculate the sample size needed (e.g. number of trees per treatment) to achieve a power level of 0.90 given several different conjectured alternatives. Assume the researcher will have an equal number of trees for each treatment, and believes the standard deviation (within treatment) is $\sigma$ = 10 units. The different conjectured alternatives are:

i. $\mu_1 = 55, \mu_2 = 45, \mu_3 = 45, \mu_4 = 35$
ii. $\mu_1 = 50, \mu_2 = 50, \mu_3 = 40, \mu_4 = 40$
iii. $\mu_1 = 50, \mu_2 = 50, \mu_3 = 45, \mu_4 = 35$

A. What sample size do you recommend to the researcher for each of the above scenarios? (**6 pts**) **Note:** Round your sample size UP to the next integer value!
B. Notice that the conjectured means in the three different scenarios all sum to 180, yet you should have found that quite different sample sizes were needed to achieve a power of 0.90. Which scenario requires the largest sample size? Explain why it requires the largest sample size. It helps to think about the null hypothesis for the ANOVA F-test.