## Homework 4:

You are expected to **use StatKey for all simulations and statistics problems** in the material in Weeks 4 and 5 (and in the later weeks when we do simulation-based statistics.) Work done using different software for simulations or theoretical distribution methods may not give answers as close as is needed to be counted as correct.

There is a separate document here with the Homework 4 assignment with necessary information about this.

**This information is quite relevant to the first midterm exam, which comes up very soon!**

**For each answer, choose the number that is closest to your answer.**
**(Review the discussion in the Orientation Problems concerning these types of answers.)**

For most of 1-6, use this scenario: In a particular city, most students in each high school come from a specific geographical area, but some students are enrolled who are not from that geographical area. It is known that the population of students this year at Crossroads High School has 14% are classified as being from "out of area." Consider the sampling distribution of the sample proportion of 'out-of-area" students for samples of size *n* from this population.

1. Find the standard deviation of this sampling distribution (also called the "standard error of the sample proportion") for samples of size 100 from this population.
   a. 0.005 b. 0.010 c. 0.015 d. 0.020 e. 0.025 f. 0.030 g. 0.035 h. 0.040 i. 0.045
   **0.035, 0.036, 0.036, 0.033, 0.037 Mean: 0.036**
2. FInd the standard deviation of this sampling distribution (also called the "standard error of the sample proportion") for samples of size 200 from this population?
   a. 0.005 b. 0.010 c. 0.015 d. 0.020 e. 0.025 f. 0.030 g. 0.035 h. 0.040 i. 0.045
   **0.024, 0.025, 0.023, 0.024, 0.025 Mean: 0.024**
3. If we compare the standard error of the sample proportion from this population for sample sizes of size *n* and size *m* where both sample sizes are greater than 100 and *m* is considerably larger than *n*, which one of the following is true? **m < n**
   a. The standard error of the sample proportion for samples of size *m* is about the same as the standard error of the sample proportion for samples of size *n*.
   b. The standard error of the sample proportion for samples of size *m* is smaller than the standard error of the sample proportion for samples of size *n*.
   c. The standard error of the sample proportion for samples of size *m* is larger than the standard error of the sample proportion for samples of size *n*.
   d. It is not possible to determine the relationship between these two standard errors without additional information.

4. **(Ungraded. You will be asked for an answer, but this will count zero points.**
**This is NOT a "multiple choice" problem, but it is a "multiple answer" problem.**
**Read about that in this week's lesson under "Types of questions.")**
Using the population proportion of 0.14, experiment to determine which of these sample sizes appear to have a symmetric-enough sampling distribution for the sample proportions that you would be comfortable using this shape for a "normal dist'n method" of finding a 99% confidence interval.
a. 20 b. 35 c. 200 d. 400

5. In a different high school in this city we want to estimate, with a 90% confidence interval, the proportion of students in the band next year who will be from out-of-area. For the purpose of this estimate, we will consider the students in the band this year as a random sample from all students who have and will attend the high school and be in the band. This year, there are 63 students in the band, 11 of whom are "out-of-area."
Find a 90% confidence interval for that proportion. Then subtract the two endpoints of the interval to report the length of the interval to report here.
a. 0.124 b. 0.160 c. 0.176 d. 0.255
0.254-0.095=0.159 0.254-0.095=0.159

6. In CrossRoads High School, consider this year's students varsity football team a random sample from all the school's varsity football players for 20 years. This year, the varsity football team has 38 students. Of those 9 are from "out-of-area." Test the hypothesis that the proportion of football players from "out-of-area" over the 20 years is higher than 0.14. Find the p-value.
a. 0.013 b. 0.023 c. 0.033 d. 0.045 e. 0.052 f. 0.068

For problems 7-9, consider the StatKey dataset **CocaineTreatment** from the 3rd edition of the Lock text.

7. If we assume the subjects were volunteers who were randomly assigned to the treatments.
i. Is the design of this study appropriate to support generalization to the population of cocaine users?
a. Yes b. No

ii. Is the design of this study appropriate to support evidence for causality if a sufficiently strong effect is found?
a. Yes b. No

8. Test the claim that the proportion of people in the population treated with lithium have a higher proportion of non-relapse than those who are not treated (i.e. treated with a placebo in our sample.) Find the p-value for this test.
a. 0.27 b. 0.36 c. 0.54 d. 0.65 e. 0.72

9. Find a 83% confidence interval for the difference in the proportion of people in the population who are expected to have no relapse between those treated with desipramine and those treated with lithium. Then subtract the two endpoints of the interval to report the length of the interval here.
    a. 0.21 b. 0.35 c. 0.44 d. 0.54

10. Centers of the various distributions.
The "population" in any question about these refers to the population from which the data was taken.

    i. The bootstrap distribution of the sample statistic p-hat is centered at
        a. The population proportion, p,
        b. the value the p-hat from the sample
        c. the value of p in the null hypothesis.

    ii. The sampling distribution of the sample statistic p-hat is centered at
        a. The population proportion, p,
        b. the value the p-hat from the sample
        c. the value of p in the null hypothesis

    iii. The randomization distribution used to test hypotheses about the value of the population proportion is centered at
        a. The population proportion, p
        b. the value the p-hat from the sample
        c. the value of p in the null hypothesis.