MATH 140 **Name** ______________________ **Date:**

Worksheet: Confidence Intervals for a proportion

The Central Limit Theorem for Sample Proportions

When we collect a sufficiently large sample of n independent observations from a population with population proportion of p, the sampling distribution of $\hat{p}$ will be nearly normal:

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

The sample size is typically considered sufficiently large when the success-failure condition has been met: $np \geq 10$ and $n(1-p) \geq 10$, but n does not exceed 10% of the population.

1. According to the 2020-2021 Linfield University Fact Book, In the Fall of 2020, 36% of all Linfield undergraduates were US Students of Color.

   (a) Is 36% a parameter or a statistic?

   (b) According to the central limit theorem, what, approximately, is the sampling distribution for $\hat{p}$, the sample proportion of US students of color in an independent sample of size 100 Linfield students?

   (c) According to the sampling distribution for $\hat{p}$, about how likely would it be to observe a sample of 100 Linfield students with 20 or fewer US students of color? To answer this question, convert to z-scores and use the `pnorm()` function in R, use your calculator, or use a probability table to find the probability.

2. Suppose that in a large city 62% of voters support a particular candidate for mayor. Use the CLT to estimate the probability that the sample proportion $\hat{p}$ supporting the candidate in an independent sample of $n = 200$ voters is within .04 of the population proportion of .62. In other words, what is the approximate probability that $\hat{p}$ is between .58 and .66?

Brief Intermission True or False: If we know a value x is within 2 units of y, it is correct to say that y is within 2 units of x.

This is true, of course. Thinking again about the previous question, we can say that the probability that $\hat{p}$ is within .04 of p is the same as the probability that p is within .04 of $\hat{p}$.

This is a useful shift in perspective because in practice we generally don't know the population proportion p, but we estimate it with a sample proportion $\hat{p}$. *The CLT gives us a way to attach a likelihood that $\hat{p}$ and p are close to each other.*

A Confidence Interval for p

Since the sampling distribution for $\hat{p}$ is approximately normal, 95% of this distribution is within 1.96 standard deviations of the mean, which is p. Knowing this, we define a 95% confidence interval for p to be:

$$\hat{p} \pm 1.96 \cdot \mathrm{SE},$$

where $\mathrm{SE} \approx \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

3. Suppose we want to estimate the proportion of voters in a large city who support a particular candidate for mayor. In an independent sample of $n = 200$ voters it turns out that 108 of the 200 voters support this candidate. Determine a 95% confidence interval for the true proportion of voters in the city who support this candidate.

4. Based on your previous answer, do you think over half the voters in the city support this candidate for mayor?

5. Suppose a second independent sample, taken closer to the election, finds that 94 of the 190 people asked support the candidate. Determine a 95% confidence interval for the true proportion of voters in the city who support this candidate.