Summary of Confidence Intervals and Tests of Significance

- **Sampling Distribution**: The distribution of values that a statistic (e.g., $\bar{x}$ or $\hat{p}$) would take if a sampling procedure were repeated many times. The **Central Limit Theorem** tells us that, for large samples, the sampling distribution will be close to Normal. This means it would be unusual to have a value more than two standard deviations from the mean.

- **Confidence Interval** (CI): A range of plausible values for some unknown parameter, such as a population mean μ or a population proportion or probability p. Our 95% confidence intervals take the form:

$$\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}, \qquad \text{or} \qquad \hat{p} \pm \frac{1}{\sqrt{n}} \qquad \left(\text{or} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

- **95% confidence**: The interval is constructed so that 95% of the possible samples you might have chosen would yield an interval containing the true parameter value (e.g., μ or p).

- **Margin of error** (moe): Our CI's are of the form **point estimate** $\pm$ **moe**. The margin of error for 95% confidence is the largest error you would make for all but the 5% most extreme samples.

- **Confidence Level:** A 95% confidence interval includes values within ± 2 standard deviations of the statistic value (1.96 is the more exact value, but 2 is close enough). A confidence level higher than 95% requires a margin of error larger than 2 standard deviations (e.g., for 99% confidence, the moe is 2.576 standard deviations). With a lower confidence level you can get by with a smaller moe (e.g., for 90% confidence, the moe is 1.645 standard deviations).

- **Statistically Significant**: The difference between the data estimate and the hypothesized value is too large to be due to chance variation. Often you'll see a statement like "the difference was significant ($P = 0.024$)..." The P in parentheses is a "P-value", which tells you how unlikely it was for the difference to have happened by chance.

- **Null and Alternative hypotheses**: The null hypothesis H_o is the hypothesis you might be able to reject, based on your data. The alternative hypothesis H_a is what you will conclude if you reject H_o. These are **hypotheses about a parameter** (e.g., μ or p), not about a statistic (e.g., $\bar{x}$ or $\hat{p}$).

- **P-value**: The probability, computed assuming H_o is true, of observing as much or more evidence for H_a as you did. A small p-value is evidence to reject H_o, because the observed data would have been very unlikely if H_o were true.

- **Significance Level**: You may claim statistical significance at level α if the p-value is less than α. For example, if the p-value is less than 0.05 you may say the result is significant at $\alpha = 0.05$ (the "5% significance level"). If a 95% confidence interval for μ does not include the null hypothesis value, then the test for that value is significant at $\alpha = 0.05$.

Common **incorrect** statements (statements in *italics* are **incorrect**):

- *95% of the population values are in the 95% CI.*

  A confidence interval is an interval estimate for a **parameter** value. Larger samples yield narrower intervals that may include only a small fraction (much less than 95%) of individual values. We are 95% confident the interval contains the true parameter value (μ or p).

- *The p-value is the probability that the null hypothesis is true.*

  The p-value is a probability about a statistic ($\bar{x}$ or $\hat{p}$), computed assuming the null hypothesis is true. A small p-value means the observed statistic would be very unusual if H_o were true, but that's not the same as saying that H_o is probably false.

- *If a difference is statistically significant, then it must be a large difference.*

  Statistically significant means we are quite sure the difference is not 0. Large samples can make a small (trivial) difference statistically significant, so it is important to know the size of the estimated difference along with the p-value. A confidence interval is more useful than the p-value for assessing the possible size of a difference.

  Statistical Significance: We have evidence for some difference.

  Practical Significance: The difference is large enough to care about.

- *Statistically significant means the result is important.*

  Along with the possibility that the difference is very small, it is also possible that bias in the data collection or lurking variables are responsible for the difference. It is important to know how the data were collected in order to judge the importance of the results.

- *Statistically insignificant means H_o is probably true.*

  We either have evidence to reject H_o or else we fail to reject H_o. We never get to conclude H_o is true. A large p-value means the data are not out of line with what we would expect to see if H_o were true. Lack of significance may simply mean that we didn't collect enough data to detect a real difference. The best we can do is to show a very narrow confidence interval that contains the H_o value of the parameter. This means that, if there is a difference, we are confident that it is a very small difference.