Math 247: Relating Confidence Intervals and Hypothesis Tests (Section 8.3)

First, the significance level (for a Two-Tailed Test!) and the confidence level always add up to 1. = 100%

If we want a .05 level of significance, then we should construct a $\boxed{95\%}$ confidence interval.

If we have a 99% confidence interval, we would relate that to a $\boxed{.01}$ level of significance.

A α (*choose a number*) level of significance corresponds to a $1-\alpha$ % confidence level. expressed as a %

Now let's see how **hypothesis tests** and **confidence intervals** relate when actually doing some statistical analysis.

Marijuana grows and salmon. The boom in marijuana grows in Northern California may be having a negative effect on salmon habitat due to water use and polluted run-off. Suppose in 2010, it was found that 8% of small streams that usually support salmon had no juvenile salmon. In 2016, a random sample 100 small streams that support salmon found that 13 of the them had no juvenile salmon. $p = .08$ for 2010 (Why not $\hat{p}$? There is no info about a SAMPLE in 2010!)

Use the data and StatCrunch to find a 95% confidence interval for the proportion of streams ALL small streams that would have no juvenile salmon.

based on the 2016 data $\left\{ \begin{array}{l} n = 100 \text{ streams} = \text{observations} \\ x = 13 \text{ no salmon} = \text{"SUCCESS"} \end{array} \right.$

95% confidence interval results:

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	13	100	0.13	.03363···	.0640···	.1959··· .1959·

Confidence Interval: (.064, .196)

Interpretation: We are 95% confident that between 6.4% and 19.6% of ALL small streams (in this particular region) ~~are~~ were missing the juvenile salmon that normally would be there, in 2016, based on the data.

(Note: We're assuming all conditions are met.)

Does the 2016 data suggest that the proportion of streams where juvenile salmon dying off is different from the proportion in 2010? Explain. NO. it does not.

In 2010, it was reported that 8% of the streams were missing their juvenile salmon.

Based on the 2016 data, we're estimating anywhere from 6.4% to 19.6% of the streams are missing their salmon.

Since the confidence interval includes .08 = 8% we can't rule out the possibility that nothing has changed; i.e. we can't conclude things are different in 2016 relative to 2010.

If we did a hypothesis test on whether the proportion of streams where juvenile salmon were dying off had changed since 2010, based on the 2016 data, what would the hypotheses be?

$H_0: p = .08$ The proportion of all the streams in 2016 that are missing their salmon hasn't changed - it's the same as in 2010.

$H_a: p \neq .08$ There HAS been a change!

Would the confidence interval tell you to reject the null or to not reject the null at the .05 significance level?

CI: All the possibilities

.064 .08 .196

Null

Since the CI "captured" the null, we can't reject it. Conclusion: do NOT reject H_0.

Now use the StatCrunch to perform the hypothesis test to check your answer to the question above. Reflect on what you found.

One sample proportion summary hypothesis test:

Hypothesis test results:

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	13	100	0.13	.0271···	1.843···	.0653

Step 4: Interpret P-value $= .0653 > .05 = \alpha$

Note: We know $\alpha = .05$ since we were using a 95% CI previously.

Fail to reject the null, P-value is too big.

We do <u>not</u> have evidence that there has been a statistically significant change in the percentage (proportion) of streams supporting small salmon

OR

There has not been a significant change in the proportion of streams that support salmon (a bit more concise)