Inference Guide – Categorical Data

Proportions (z) Success/Failure

4-steps needed for inference problems:

1. Parameters

2. Plan - Think
   - Decide what inference procedure.
   - List the assumptions and check the conditions.
   - Specify the model / name the test
     "*Because the conditions are satisfied, I can model the sampling distribution of the __________ with a _______ model and perform a _________________.*"

3. Mechanics - Show
   - Write down the statistics
   - Draw curve showing sampling model - mark parameters & statistics & shade tail(s).
   - Calculate the value of the test statistic - show the formula, substitute all the proper values, and give the final result.
   - Find the Confidence Interval, P-Value, etc.

4. Conclusion – Tell what you've learned w/ "4Cs")
   - Interpret the confidence interval in context - *"I'm 95% confident, based on this sample, that the proportion of all auto accidents that involve teenage drivers is between 12.7% and 18.6%."*
   - Link the P-value to the decision about the null hypothesis and interpret that decision in the proper – *"The high P-value indicates that these results could be reasonably explained by sampling error, so I fail to reject the null hypothesis. We do not have evidence of a change in the percentage of ________."*

One Sample

XX% Confidence Interval

H_0:

H_A: ( )

H_A: p >or< p_0 ( )

A1 Individuals/data independent.

C1 SRS and n < 10% population.

A2 Sample large enough to approximate SDM w/ Normal model.

C2 Successes ≥ 10 and Failures ≥ 10.

proportion, Normal model

One-proportion

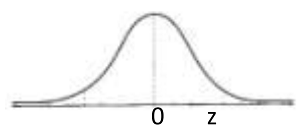
z-interval z-

n = , $\hat{p}$ = n = , $\hat{p}$ = , p_0 =

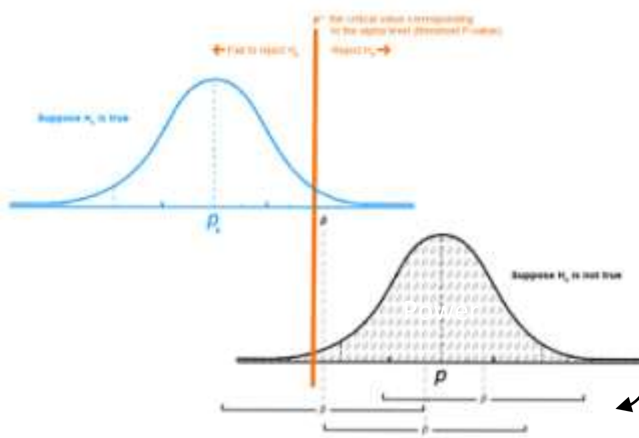
$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$z^* = \left| invNorm\left(\frac{1 - confidence\ level}{2}\right) \right|$$

$$\hat{p} \pm \underbrace{z^* \times SE(\hat{p})}_{ME}$$



= normalcdf(,)



Sample

XX% Confidence Interval

H_0:

H_A: (2 Tailed)

H_A: $p_1 - p_2$ >or< 0 (1 Tail)

A0

C0

A1 Individuals/data in each group Independent.

C1 are SRS and n < 10% populations OR random allocation.

A2 groups large enough.

C2 Successes ≥ 10 and Failures ≥ 10.

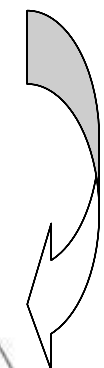
of proportions, Normal model

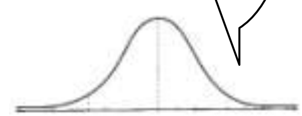
-proportion

z- z-

n_1 = , $\hat{p}_1$ = n_1 = , $\hat{p}_1$ =

n_2 = , $\hat{p}_2$ = n_2 = , $\hat{p}_2$ =





XX% of all random samples will yield confidence intervals that capture the true parameter value.