Iams

Inference Guide – Categorical Data
Proportions (z) Success/Failure

4-steps needed for inference problems:

1. Parameters/Hypotheses
   - define parameters
   - Write the null hypothesis
   - Write the alternative hypothesis and if 1 or 2-tailed test
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/ "4C's")
   - 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 context – *"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

$H_0: p = p_0$

$H_A: p \neq p_0$ (2 Tailed)

$H_A: p >or< p_0$ (1 Tailed)

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

n = , $\hat{p}$ =

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

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

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

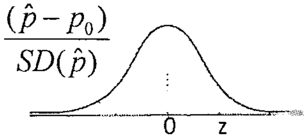
z-test

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

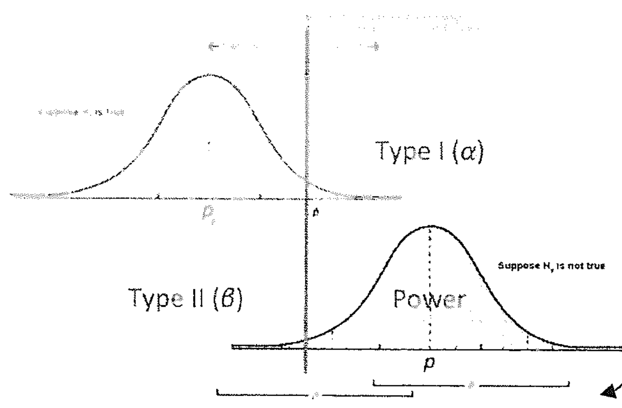
$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}}$$

$$N\left(p_0, \sqrt{\frac{p_0 q_0}{n}}\right)$$

$$z = \frac{(\hat{p} - p_0)}{SD(\hat{p})}$$



P-value = normalcdf(,)



Two Sample

XX% Confidence Interval

$H_0: p_1 - p_2 = 0$

$H_A: p_1 - p_2 \neq 0$ (2 Tailed)

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

A0 Groups are independent.
C0 (Think about the design/how data collected).
A1 Individuals/data in each group Independent.
C1 Both are SRS and n < 10% populations OR random allocation.
A2 Both groups large enough.
C2 Both Successes ≥ 10 and Failures ≥ 10.

difference of 2 proportions, Normal model
Two-proportion

z-interval

n_1 = , $\hat{p}_1$ =
n_2 = , $\hat{p}_2$ =

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \times SE(\hat{p}_1 - \hat{p}_2)$$

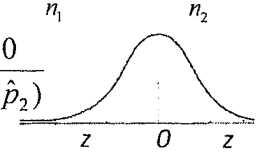
z-test

n_1 = , $\hat{p}_1$ =
n_2 = , $\hat{p}_2$ =

$$\hat{p}_{pooled} = \frac{\#Success_1 + \#Success_2}{n_1 + n_2}$$

$$SE_{pooled}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_{pooled}(\hat{p}_1 - \hat{p}_2)}$$



P-value = 2*normalcdf(,)

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