sendwithus A/B test cheat sheet

Chi-Squared Test for testing difference of proportions

These are your Observed Values (O_i's)

successes: number of successful events.
fails: number of unsuccessful events.
totals (n_i): total number of events.

	Variant A	Variant B	Variant C
successes	35	55	75
fails	40	70	25
totals	75	125	100

Compute pooled proportion

This is basically an average.

$$p = \frac{\text{sum of successes}}{\text{sum of totals}} = \frac{45 + 55 + 65}{75 + 125 + 100} = \frac{165}{300} = \boxed{0.55}$$

Compute Expected Values (E_i's)

expected success: $(p)n_i$
expected fail: $(1 - p)n_i$

If a variant performed like the average, here's how many successes and fails it would have had given number of total events.

	Variant A	Variant B	Variant C
successes	0.55(75) = 41.25	0.55(125) = 68.75	0.55(100) = 55
fails	(1 - 0.55)(75) = 33.75	(1 - 0.55)(125) = 56.25	(1 - 0.55)(100) = 45

* Note: All expected values must > 5 to have enough data to run the Chi-Square test.

Compute "Chi-Parts" (X^2_i's)

$$X^2_i = \frac{(O_i - E_i)^2}{E_i}$$

Chi-Parts are an indication of how statistically different the observed value is from the expected value.

	Variant A	Variant B	Variant C
successes	$\frac{(40 - 41.25)^2}{41.25} = 0.038$	$\frac{(50 - 68.75)^2}{68.75} = 5.11$	$\frac{(60 - 55)^2}{55} = 0.45$
fails	$\frac{(35 - 33.75)^2}{37.5} = 0.042$	$\frac{(50 - 56.25)^2}{56.25} = 0.69$	$\frac{(40 - 45)^2}{45} = 0.556$

Sum the Chi-Parts to get a test statistic

$$X^2 = 0.038 + 0.042 + 5.11 + 0.69 + 0.45 + 0.556 = \boxed{6.886}$$

Compare test statistic to magical critical value

A Chi-Square critical value is found using a table or a calculator. Plug in a probability (we use 0.95) and degrees of freedom (number of variants - 1). The critical value for three variants at probability 0.95 is **5.9915.**

If the test statistic is greater than the critical value, there is significant statistical evidence that not all variants are the same. Once we know that, we can move on to Marascuillo's Procedure to determine if there is a winner (if one variant has a statistically significant higher success rate than the others).

6.886 > 5.9915. Let's move on!