

CHI SQUARE EXPECTED FREQUENCIES, DF, AND STANDARDIZED RESIDUALS

To calculate the expected frequencies for a contingency table: for each cell multiply its row's total frequency by its column's total frequency, and divide by the total number of observations: $f_{\text{ROW}} * f_{\text{COL}} / N$.

To calculate the df for a contingency table: $df = (\# \text{rows} - 1) * (\# \text{columns} - 1)$.

To say which cells contribute most to the observed departure from expected values, i.e., contribute most to the chi-square being significant (if it is): Where exactly are your observed frequencies not matching up with the expected? For this you look at the standardized residuals that SPSS provides for each cell. Any absolute value of 2 or larger would indicate that cell is an important contributor to the overall chi-square being significant. The reason is that those values are essentially like z-scores, so a value of 2 means it's 2 standard deviations above the mean, and you only have to be 1.96 standard deviations above the mean to be in the extreme 5% of the z (standard normal) distribution. It's kind of like being "significant". The way the standardized residuals are calculated is that they're each the square root of the cell's contribution to the chi-square calculation where you add up $(f_o - f_E)^2 / f_E$ for each cell. The square root of the cell's contribution, the "standardized residual", is $\sqrt{[(f_o - f_E)^2 / f_E]}$ or $(f_o - f_E) / \sqrt{f_E}$.

TALKING ABOUT ODDS AND ODDS RATIOS IN PLAIN ENGLISH

Kind of an addendum to my odds and probabilities handout, attempting to make the English clear when talking about odds.

Odds is (are?) a ratio of two probabilities.

An odds ratio (OR) is a ratio of two odds.

Check the general discussion at

<http://web9.uits.uconn.edu/lundquis/Odds and probabilities.doc>

A calculated example is at the top of this Excel spreadsheet:

<http://web9.uits.uconn.edu/lundquis/pavlov rosenthal chisq.xls>

The tricky thing with talking about odds is that you can express the same thing in multiple ways, all of which are legitimate. Some may just be clearer for your particular context, is all. (I haven't even settled on whether to make "odds" singular or plural and I don't know what the APA says. The scissors is on the desk. The scissors are on the desk. That kind of thing. Mostly plural, but I'm not consistent.)

There are "odds in favor" and, same thing from a different perspective, "odds against". If you just say the "odds of", it means "odds in favor of". Like, the "odds of" a horse winning a race are pretty clearly the "odds in favor", not against.

If you have two possible outcomes A and B (where B is actually just "not-A", as in "food vs. no-food" or "hat vs. no-hat") then you could either say odds (in favor) of A = probability of A / probability of B

or

odds against A = probability of B / probability of A

So the "odds against" is just the inverse of the "odds in favor".

You could also focus on B (and think of A as "not B", which is necessarily also true). Then the "in favor / against" terminology indicates the inverse ratios of when you focused on A.

odds (in favor) of B = probability of B / probability of A

or

odds against B = probability of A / probability of B

"In favor" being the inverse of "against" still applies, you're just drawing attention to B rather than A. You'll decide rhetorically whether you're trying to draw attention to A or B, and whether you want to express the odds in favor of it or against it.

Sometimes people choose whether to express odds as "odds in favor" or "odds against" based on whether the ratio comes out to be greater or less than 1. You can say "the odds (in favor) are 9-to-4," or reduce it to "the odds (in favor) are 2.25". But if it's "the odds (in favor) are 4-to-9 (or .444)," it may sound more natural to say "the odds AGAINST are 9-to-4 (or 2.25)," just to put the bigger number first. This is not a rule, this is about what sounds better to you. Just don't switch from talking about the odds in favor of A to the odds against A in the same conversation (or paper). You'll lose your audience, like I'm losing you right now.

In conversation it's probably more natural to give odds in the "9-to-4" or "4-to-9" format, but in an article you would reduce it to the "2.25" or ".444" format, where the "to-1" is implied -- e.g., saying "odds are 2.25" means "odds are 2.25-to-1" (equivalent to 9-to-4, obviously).

Odds RATIOS describe the change in odds in different situations or circumstances.

Imagine circumstance C and circumstance D: circumstance C could mean a bell has rung for Pavlov's dog, and D could mean it has not rung; or C could mean it's winter time and D could mean it's summer time, if you're looking at the odds of wearing a hat.

You can calculate the odds in favor of A (let's focus on that and ignore the other terminological options) under circumstance C, and then also calculate it under circumstance D. If it's different under those two conditions, you could make a ratio of the two "odds in favor of A" and that would be an odds ratio (OR).

OR = odds of A given C / odds of A given D

For example, maybe under condition C, the odds (in favor) of A are 9-to-4, or 2.25. And maybe under condition D, the odds (in favor) of A are 2-to-5, or .4. Then the odds ratio would be the C odds divided by the D odds, or $2.25 / .4 = 5.62$.

You could state the the odds of A given C are 5.62 times the odds of A given D. Or you could be more explicit and include the implicit "greater": the odds of A given C are 5.67 times GREATER THAN the odds of A given D.

Or again, you could say it upside-down, because it would be equivalent to say that the odds of A given D are 5.62 times LESS THAN the odds of A given C.

Or instead of going to the "less than" phrasing, you could use take the inverse to find the odds ratio in the other direction, focusing on the D circumstance. Notice that the odds ratio of 5.62 is saying that

odds of A given C / odds of A given D = 5.62

so

odds of A given C = 5.62 * odds of A given D

so

odds of A given C * 1/5.62 = odds of A given D

so

the odds of A given D are 0.178 times the odds of A given C.

Oh, you want to see how all that would look if you had focused on B? Fine.

odds of B given C = 4-to-9 = .444

odds of B given D = 5-to-2 = 2.5

odds ratio = .444 / 2.5 = 0.178

The odds of B given C are 0.178 times the odds of B given D.

The odds of B given D are 5.62 times the odds of B given C.

The funny thing about the language is that different expressions sounds okay depending on whether the odds ratio is greater or less than one. Again, these aren't rules - follow along with your intuition and you'll agree:

You can say "the odds of A given C are 5.62 times GREATER THAN the odds of A given D" and you can equivalently say "the odds of A given D are 5.62 times LESS THAN the odds of A given C." And if you say it that way, the "greater than" is optional but the "less than" is definitely required.

But if you're looking at an odds ratio less than one, it's weird to include either "greater than" or "less than" and you just say "times": "the odds of A given D are 0.178 TIMES the odds of A given C." I mean, maybe you could get away with "0.178 times less than" even though it's strange, but definitely not "0.178 times greater than", which would make no sense since it's not greater. And maybe you could say "the odds of A given D are 0.178 OF the odds of A given C" instead of "times" to indicate the fraction, but you wouldn't say "5.62 OF the odds" when the odds ratio is greater than one.

Your best guide as to how to say all this is, are you conveying what you mean to convey, in a way people will understand and approve of grammatically? Hey, good guide for everything, actually.