

16.1.1: Assumptions of the Test(s)

All statistical tests make assumptions about your variables, data, and the distributions that they come from. Usually, it's not a huge deal if your sample's data doesn't fit some of the assumptions. Chi-square is not one of those tests. For the chi-square tests discussed so far in this chapter, the assumptions are:

- *Expected frequencies are sufficiently large.* All of the expected frequencies need to be reasonably big for the statistics to use the correction distribution in the background. How big is reasonably big? Opinions differ, but the default assumption seems to be that you generally would like to see all your expected frequencies larger than about 5, though for larger tables you would probably be okay if at least 80% of the the expected frequencies are above 5 and none of them are below 1 (meaning, no categories are empty). However, from what Dr. Navarro has been able to discover, these seem to have been proposed as rough guidelines, not hard and fast rules; and they seem to be somewhat conservative (Larntz, 1973).
- *Data are independent of one another.* One somewhat hidden assumption of the chi-square test is that you have to genuinely believe that the observations are independent. Here's what I mean. Suppose I'm interested in proportion of babies born at a particular hospital that are boys. I walk around the maternity wards, and observe 20 girls and only 10 boys. Seems like a pretty convincing difference, right? But later on, it turns out that I'd actually walked into the same ward 10 times, and in fact I'd only seen 2 girls and 1 boy. Not as convincing, is it? My original 30 *observations* were massively non-independent because I really only had 3 observation. Obviously this is an extreme (and extremely silly) example, but it illustrates the basic issue. Non-independence messes things up. Sometimes it causes you to falsely reject the null, as the silly hospital example illustrates, but it can go the other way too. To give a slightly less stupid example, let's consider what would happen if we asked 50 people to select 4 cards. One possibility would be that *everyone* selects one heart, one club, one diamond and one spade. This is highly non-random behavior from people, but in this case, I would get an observed frequency of 50 for all four suits. For this example, the fact that the observations are non-independent (because the four cards that you pick will be related to each other) actually leads to the opposite effect... falsely retaining the null.

If you happen to find yourself in a situation where independence is violated, it may be possible to use the McNemar test (which we'll discuss) or the Cochran test (which we won't). Similarly, if your expected cell counts are too small, check out the Fisher exact test (which we also won't discuss).

Our first stop in the tour of Chi-Square is the Goodness of Fit test. See you there!

Reference

Larntz, K. Small-sample comparison of exam levels for Chi-Squared Goodness-of-Fit statistics. *Journal of the American Statistical Association*, 73 (362), 253-263.

This page titled 16.1.1: Assumptions of the Test(s) is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michelle Oja.

- 12.5: Assumptions of the Test(s) by Danielle Navarro is licensed CC BY-SA 4.0. Original source: <https://bookdown.org/ekothe/navarro26/>.
- Current page by Michelle Oja is licensed CC BY-SA 4.0.