

13.2: Why tests of assumption are important

Introduction

Note that R (and pretty much all statistics packages) will calculate t -tests or ANOVA or whatever test you ask for, and return a p -value, but it is up to you to recognize that the p -value is accurate only if the assumptions are met. Thus, you can always estimate a parameter, but interpret with caution. The great thing about statistics is that you can directly evaluate whether assumptions hold.

Violation of the assumptions of normality or equal variances can lead to Type I errors occurring more often than the 5% level. That means you will reject the null hypothesis more often than you should! If the goal of conducting statistical tests on results of an experiment is to provide confidence in your conclusions, then failing to verify assumptions of the test are no less important than designing the experiment correctly in the first place. Thus, the simple rule is: know your assumptions, test your assumptions. Evaluating assumptions is a learned skill:

- conduct proper data exploration
- use specialized statistical tests to evaluate assumptions
 - data normally distributed? e.g., histogram, Shapiro-Wilk test
 - groups equal variance? e.g., box-plot, Levene's median test
- evaluate influence of any outlier observations

What to do if the assumptions are violated?

This is where judgement and critical thinking apply. You have several options. First, if the violation is due to only a few observations, then you might proceed anyway, in effect invoking the **Central Limit Theorem** as justification. Second, you could check your conclusions with and without the few observations that seem to depart from the trend in the rest of the data set — if your conclusion holds without the “outliers”, then you might conclude that your results are robust). Third, you might apply a **data transform**, reasoning that the distribution from which the data were sampled was log-normal, for example. Applying a log transform (natural log, base 10, etc.,) will tend to make the variances less different among the groups and may also improve normality of the samples. Fourth, if a nonparametric test is available, you might use it instead of the parametric test. For example, we discuss the [Levene's test of equal variances](#) as a better choice than the parametric F -test. Additionally, there are many nonparametric alternatives to parametric tests. For example, t -tests are parametric and their non-parametric alternatives include tests like [Wilcoxon and Mann-Whitney](#). ANOVA, too is parametric, and its nonparametric alternative version is called **rank ANOVA** (see also [Kruskal-Wallis](#)). See [Chapter 15 for nonparametric tests](#). Finally, a resampling approach could be taken, where the data themselves are used to generate all possible outcomes like the Fisher Exact test; with large sample size, bootstrap or Jackknife procedures are used ([Chapter 19](#)).

For now, let's introduce you to the kinds of nonparametric statistical tests for which the t -test is just one example. For the independent sample t -test, our first method to account for the possible violation of equal variances is a parametric test, Welch's variation of the Student's t -test. Instead of the pooled standard deviation, Welch's test accounts for each group's variance in the denominator.

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The degrees of freedom for the Welch's test are now

$$df \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2 \cdot v_1} + \frac{s_2^4}{n_2^2}}$$

where df for Student's t -test was $n_1 + n_2 - 2$. Note that Welch's test requires normal distributed data. Note also that in R Commander, the default option for conducting the t -test is Welch's version, not the standard t -test (Fig. 13.2.1).

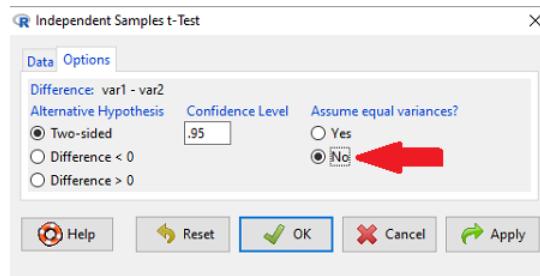


Figure 13.2.1: Copy and Paste Caption here. (Copyright; author via source)

See discussions in Olsen (2003), Choi (2005), and Hoekstra et al (2012).

Questions

Please read article:

Hoekstra, R., Kiers, H., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in psychology*, 3, 137. [Link to article](#)

From the article and your readings in Mike's Biostatistics Book, answer the following questions:

1. What are some consequences if a researcher fails to check statistical assumptions?
2. Explain why use of graphics techniques may be more important than results of statistical tests for checks of statistical assumptions.
3. Briefly describe graphical and statistical tests of assumptions of (1) normality and (2) equal variances
4. Pick a biological research journal for which you have online access to recent articles. Pick ten articles that used statistics (e.g., look for "t-test" or "ANOVA" or "regression"; exclude meta analysis and review articles — stick to primary research articles). Scan the Methods section and note whether or not you found a statement that confirms if the author(s) checked for violation of (1) normality and (2) equal variances. Construct a table.
5. Review your results from question 3. Out of the ten articles, how many reported assumption checking? How does your result compare to those of Hoekstra et al; (2012)?

This page titled [13.2: Why tests of assumption are important](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.