Checking Your Stats

For others to be able to build on your work, it is important that the statistical information in your article is accurate and complete. People often do not have completely reproducible workflows, so they copy and paste information between software programs and the articles they write. Even in completely reproducible documents programming errors happen that might introduce errors.

Statcheck is software that automatically extracts statistics from articles and recomputes their *p*-values, as long as statistics are reported following guidelines from the American Psychological Association (APA). It checks if the reported statistics are coherent: Given the test statistics and degrees of freedom, is the reported p-value accurate? If it is, that makes it less likely that you have made a mistake (although it does not prevent coherent mistakes!) and if it is not, you should check if all the information in your statistical test is accurate. For a detailed overview of what statcheck can and can not do, check the manual.

Statcheck is not perfect, and it will make Type 1 errors where it flags something as an error when it actually is not. Nevertheless, it is an easy to use tool to check your articles.

Go to http://statcheck.io/ and upload a PDF, word document, or HTML file. You can check the option to try to take the possibility of one-sided tests into account. You will see a Table similar to the one below, or statcheck was able to find statistical tests in the file you uploaded. I used a 2018 article I co-authored for this example.

	Source	Statistical Reference	Computed p Value	Consistency
1	Schietecat et al. - 2018 - Pre (...)	t(41) = 5.59, p < .001	0.00000	Consistent
2	Schietecat et al. - 2018 - Pre (...)	t(41) = 2.49, p = .017	0.01692	Consistent
3	Schietecat et al. - 2018 - Pre (...)	F(1, 39) = 0.345, p = .56	0.56034	Consistent
4	Schietecat et al. - 2018 - Pre (...)	t(19) = 2.82, p = .011	0.01094	Consistent
5	Schietecat et al. - 2018 - Pre (...)	t(20) = 1.02, p = .318	0.31991	Consistent
6	Schietecat et al. - 2018 - Pre (...)	t(29) = 0.4, p = .69	0.69209	Consistent
7	Schietecat et al. - 2018 - Pre (...)	t(29) = 2.06, p = 0.024	0.04847	Consistent
8	Schietecat et al. - 2018 - Pre (...)	F(1, 70) = 8.63, p = .004	0.00447	Consistent
9	Schietecat et al. - 2018 - Pre (...)	t(40) = 0.39, p = .70	0.69861	Consistent
10	Schietecat et al. - 2018 - Pre (...)	t(40) = 2.49, p = 0.008	0.01703	Inconsistency
11	Schietecat et al. - 2018 - Pre (...)	F(1, 81) = 11.63, p = .001	0.00101	Consistent
12	Schietecat et al. - 2018 - Pre (...)	t(34) = 17.44, p < .001	0.00000	Consistent
13	Schietecat et al. - 2018 - Pre (...)	t(38) = -20.49, p < .001	0.00000	Consistent
14	Schietecat et al. - 2018 - Pre (...)	t(38) = 14.82, p < .001	0.00000	Consistent
15	Schietecat et al. - 2018 - Pre (...)	t(32) = 4.12, p < .001	0.00025	Consistent
16	Schietecat et al. - 2018 - Pre (...)	t(34) = 0.57, p = .573	0.57243	Consistent
17	Schietecat et al. - 2018 - Pre (...)	t(34) = 11.70, p < .001	0.00000	Consistent
18	Schietecat et al. - 2018 - Pre (...)	t(37) = 3.9, p < .001	0.00039	Consistent

Showing 1 to 18 of 18 entries

Previous 1 Next

We can see one test is flagged as inconsistent because the *p*-value is not in line with the test statistic. In this case this is a Type 1 error – the specific test that is found by statcheck is an equivalence test, which is one-sided, but the word 'one-sided' is not mentioned in the article. Statcheck is not able to identify this test as a one-sided test, but the *p*-value is correct.

I have found small errors in my own published articles using statcheck. The better tools we get to check our work, and the more open and transparent we work, the more likely it is that people will find errors in our work. This is to be expected. It is almost impossible to not make any errors in your career. In the past, we were less likely to be able to identify errors, because tools to identify errors did not yet exist, or because people did not share their data and code. We will need to get used to dealing with the fact that we all make errors, and that others will identify them. If we compare scientific articles to software, then we see that software is regularly updated because bugs are found. We should similarly expect that humans make errors – scientists are no exception. All we can do is to try our best, and to respond appropriately if errors are identified.

It is good to develop procedures in your workflow that try to minimize errors. Running your article through statcheck before you submit it is recommended.

References:

Nuijten, M. B., Hartgerink, C. H. J., Van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985-2013). Behavior Research Methods, 48(4), 1205-1226. doi:10.3758/s13428-015-0664-2