**CORRECTION**

# Correction: Why Psychologists Should by Default Use Welch’s t-test Instead of Student’s t-test

Marie Delacre1, Daniël Lakens2, and Christophe Leys1

1. Université Libre de Bruxelles, Service of Analysis of the Data (SAD), Bruxelles, BE

Affiliation

2. Eindhoven University of Technology, Human Technology Interaction Group, Eindhoven, NL

Corresponding author: Marie Delacre (marie.delacre@ulb.ac.be)

This article details a correction to: Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch’s t-test Instead of Student’s t-test. *International Review of Social Psychology*, 30(1), 92–101. DOI: http://doi.org/10.5334/irsp.82

**Correction**

Since the publication of our paper “Why Psychologists Should by Default Use Welch’s *t*-test Instead of Student’s *t*-test”, we have noticed several errors. These mistakes do not affect the main message of our article (Welch’s *t*-test should always be privileged over Student’s *t*-test when we compare groups based on their mean), but some led us to overgeneralize some related findings and others might induce confusion going against the pedagogical aim of the article.

Through this article, we reviewed the differences between Welch’s *t*-test, Student’s *t*-test and Yuen’s *t*-test. We used a simulation plan in order to compare the type I and type II error rate of these three tests when samples are extracted from different distributions that are symmetric or not. In order to assess the type I error rate of the three tests, we created scenarios where two samples were extracted from populations with equal mean. Unfortunately, this is not appropriate in order to assess the Type I error rate of Yuen’s *t*-test when samples are extracted from asymmetric distributions. Indeed, the null hypothesis of Yuen’s *t*-test is that the trimmed means are equal across groups and when distributions are asymmetric, means and trimmed means differ. In conclusion, while we maintain that Welch’s *t*-test has a better control on type I error rate than Yuen’s *t*-test when populations are symmetrically distributed, we are not able to generalize our conclusion to situations where distributions are skewed.

In p.93, we suggest that the *F*-ratio statistic is obtained by computing , where is the sample standard deviation of the *j*th group (*j* = 1,2). However, the *F*-ratio statistic should be obtained by computing the ratio between the largest and the smallest sample standard deviation .

More importantly, in p.93, we confused ***sample*** standard deviation and ***population*** standard deviation in our definition of SDR. The SDR should be defined as the ***population*** standard deviation ratio. We should therefore have written this : “*When SDR > 1, the standard deviation of the second population is bigger than the standard deviation of the first population and when SDR < 1, the standard deviation of the second population is smaller than the standard deviation of the first population.*” The same confusion occurs later when we suggest that in Kester(1969; 1.32 is only an ***estimate*** of SDR). These confusions does not impact our estimations of the power of Levene’s test (see Figure 1) nor other simulations, because SDR was correctly defined in all our simulation scripts.

In Table 1 in p.96, we mention that “*when both variances and sample sizes are the same in each independent group, the t-values, degrees of freedom, and the p-values in Student’s t-test and Welch’s t-test are the same*” (p.96). Looking back, we realize that readers might mistakenly believe that *t*-values, degrees of freedom and *p*-values will be identical when the homoscedasticity assumption is true but actually, *t*-values, degrees of freedom and *p*-values will be identical only if sample *estimates* of standard deviation are identical. This information is not very relevant, as two equal population variances could lead to unequal estimates (and to a lesser extent, two unequal population variances could possibly lead to equal estimates, although it is very unlikely). This example perfectly represents a mistake that we have made several times in this article: we have used the term "group" interchangeably to sometimes describe samples and to sometimes describe populations. From a pedagogical perspective, this can lead to confusion, which is very problematic in our view. On the other hand, this does not alter our conclusions since the confusion was never committed in our simulation scripts.

Finally, although not visible in the article, two errors made in the simulations impacted some conclusions in the Additional File of the article. First, we used different population SD when simulating double exponential distributions than when simulating other distributions, due to a confusion between lambda and sigma when using the "rdoublex" function in R. This mistake had consequences on the assessment of the power of both Welch's *t*-test and Student's *t*-test, and therefore, we erroneously claimed in the Additional File that there is a loss of power with heavy tailed distributions. This point was discussed in a later article : « [Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's *F*-test instead of the Classical *F*-test in One-Way ANOVA](https://www.rips-irsp.com/articles/10.5334/irsp.198/?ref=https://githubhelp.com) » (Delacre et al., 2019). Second, there were an error in the scripts we ran in order to simulate samples extracted from chi-square distributions. As a consequence, we cannot generalize our conclusions to scenarios where sample sizes differ and are extracted from highly skewed distributions. Scripts were corrected and rerun, and tables and conclusions were modified accordingly in the Additional File available on [Github](https://github.com/mdelacre/studentbackup/blob/master/Additional%20file_Updated.docx) (changes from the original version on the IRSP website are indicated in blue).

**Competing interests**

[Competing Interests statement] **I guess there is no Competing Interests ?**

**References**

* Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch’s t-test Instead of Student’s t-test. International Review of Social Psychology, 30(1), 92–101. DOI: <http://doi.org/10.5334/irsp.82>
* Please add any additional references here