



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

MARIE DELACRE 

DANIËL LAKENS 

CHRISTOPHE LEYS 

\*Author affiliations can be found in the back matter of this article

CORRECTION



## ABSTRACT

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: <https://doi.org/10.5334/irsp.82>.

This article has been corrected here:  
<https://doi.org/10.5334/irsp.82>

### CORRESPONDING AUTHOR: Marie Delacre

Université Libre de Bruxelles,  
Service of Analysis of the Data  
(SAD), Bruxelles, BE  
[marie.delacre@ulb.be](mailto:marie.delacre@ulb.be)

### KEYWORDS:

Welch's  $t$ -test; Student's  $t$ -test; homogeneity of variance; Levene's test; Homoscedasticity; statistical power; type I error and type II error

### TO CITE THIS ARTICLE:

Delacre, M., Lakens, D., & Leys, C. (2022). Correction: Why Psychologists Should by Default Use Welch's  $t$ -test Instead of Student's  $t$ -test. *International Review of Social Psychology*, 35(1): 21, 1–3. DOI: <https://doi.org/10.5334/irsp.661>

## 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 review 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  $S_2/S_1$ , where  $S_j$  is the sample standard deviation of the  $j^{\text{th}}$  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  $\frac{\max(S_1, S_2)}{\min(S_1, S_2)}$ .

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  $SDR \approx 1.32$  in Kester (1969; 1.32 is only an **estimate** of SDR). These confusions do 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 this 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 » (Delacre et al., 2019). Second, there was 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 (changes from the original version on the IRSP website are indicated in blue).


## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR AFFILIATIONS

**Marie Delacre**  [orcid.org/0000-0001-5875-1606](https://orcid.org/0000-0001-5875-1606)  
Université Libre de Bruxelles, Service of Analysis of the Data (SAD), Bruxelles, BE

**Daniël Lakens**  [orcid.org/0000-0002-0247-239X](https://orcid.org/0000-0002-0247-239X)  
Eindhoven University of Technology, Human Technology Interaction Group, Eindhoven, NL

**Christophe Leys**  [orcid.org/0000-0002-3832-3097](https://orcid.org/0000-0002-3832-3097)  
Université Libre de Bruxelles, Service of Analysis of the Data (SAD), Bruxelles, BE

## 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: <https://doi.org/10.5334/irsp.82>
- Delacre, M., Leys, C., Mora, Y. L., & Lakens, D.** (2019). Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's *F*-test instead of the Classical *F*-test in One-Way ANOVA. *International Review of Social Psychology*, 32(1), 13. DOI: <https://doi.org/10.5334/irsp.198>
- Kester, S. W.** (1969). The communication of teacher expectations and their effects on the achievement and attitudes of secondary school pupils In: University of Oklahoma. Retrieved from <https://shareok.org/handle/11244/2570>

---

### TO CITE THIS ARTICLE:

Delacre, M., Lakens, D., & Leys, C. (2022). Correction: Why Psychologists Should by Default Use Welch's *t*-test Instead of Student's *t*-test. *International Review of Social Psychology*, 35(1): 21, 1–3. DOI: <https://doi.org/10.5334/irsp.661>

**Submitted:** 24 November 2021    **Accepted:** 24 November 2021    **Published:** 25 November 2022

### COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*International Review of Social Psychology* is a peer-reviewed open access journal published by Ubiquity Press.