How do my distributions differ? Significance testing for the Overlapping Index using Permutation Test

Ambra Perugini [1], Giulia Calignano [1], Massimo Nucci [2], Livio Finos [3], Massimiliano Pastore [1]

November 29, 2024

Abstract

The present contribution aims to compare both commonly and less commonly used statistical methods in psychological sciences to evaluate their utility in tailored cases. Specifically, the paper proposes applying the Permutation test alongside the Overlapping index to estimate effects of interest in psychological science. Starting from real and openly available data, we simulated different scenarios focusing on residual distribution characteristics. The present contribution provides practical tools for considering, and deciding which statistical methods are useful and sufficient considering the features of data distribution. Subsequently, we present a Simulation study to illustrate the practical implications and reliability of each approach, particularly valuable in scenarios commonly encountered in quantitative psychology, where navigating data characteristics and adhering to or deviating from test assumptions is crucial. The findings underscore the necessity of choosing statistical methods that are resilient to the complexities inherent in psychological data, where assumption violations are often inevitable.

1 Statistical testing choices in Psychology

Methodological choices in cognitive and behavioral sciences aim to combine data richness with data collection feasibility, and at the same time they aim to land on valid interpretation based on reliable and robust statistical methods. Classic examples, like reaction times, demonstrate how specific measures have achieved such an acceptable trade-off, and for example, this is true even by comparing the framework of in lab *vs* online data collection (Semmelmann & Weigelt, 2017). Nevertheless, even in the fortunate case of reaction times which have widespread and solid epistemic rationale of use (Grosjean, Rosenbaum, & Elsinger, 2001; Proctor & Schneider, 2018; Silverman, 2010) significance testing often relies on the rigid application of a few statistical methods that have gained popularity among the scientific community and are perpetrated *perinde ac cadaver* by formal guidelines (Cumming, Fidler, Kalinowski, & Lai, 2012), even if their limits and risks have always been noted in the field of psychology and beyond (Boneau, 1960).

In fact, there is a growing caution against blindly using statistical tools and analytical methods without a deep understanding of their assumptions and implications (Scheel, Tiokhin, Isager, & Lakens, 2021). In other words, it is increasingly apparent that relying solely on significance testing as a trustworthy measure is improbable without considering the assumptions inherent to specific statistical methods, such as the t-test, across various scenarios in psychology. In fact, considering the particular circumstances of application has consistently been crucial advice when deciding on significance testing methods (Fisher, 1925).

The present contribution aims to present a novel approach for statistical testing, in particular for comparing two groups. Specifically, the present work proposes applying the Permutation test (Pesarin & Salmaso, 2010) alongside the Overlapping index (η, Pastore & Calcagnì, 2019) to compare empirical distributions. Importantly, by simulating different scenarios, we compare widely used statistical tests in Psychological Sciences with this novel approach to evaluate the performance in controlling type I error and power. Understanding and applying significance testing properly is crucial to deriving meaningful conclusions from psychological research. Accordingly, we offer an alternative tool to manage the assumptions underlying these analytical approaches, and increase awareness in significance testing in psychology. But above all, and even most importantly, the advantage of the Overlapping Index and its test is to take into account at the same time mean, variance and shape with a single test.

Considering that parametric tests, such as linear models, used for statistical inference require strong assumptions that are unlikely to be respected in the aforesaid field, such as normality and homoscedasticity, alternative methods become optimal. The use of non-parametric methods is particularly beneficial when these assumptions are violated (?, ?). Specifically, when using a t-test to compare two groups or two experimental conditions using a given variable, it functions as a straightforward version of linear regression. This statistical process necessitates assumptions about the residuals, such as their independence and normal distribution, to be met. In cases such as reaction times, these assumptions might be violated if they are not properly addressed. Most importantly, two populations might present the same mean, yet their distributions largely differ in other characteristics, such as variability, leading to genuinely distinct groups (see figure 1).

The remainder of this article is structured as follows. First, we introduce the concept of the Overlapping Index, providing foundational definitions and highlighting its importance. Next, we define the Permutation approach and explore its application to the Overlapping Index, showcasing its relevance in statistical analysis. Subsequently, we present a Simulation study to illustrate the practical implications and trustworthiness of the overlapping index utilizing permutations.

The rationale behind these steps involves first introducing the concept of the Overlapping Index (η), which is crucial because it provides an intuitive measure of similarity between distributions by quantifying the overlapping area of their empirical density functions, a common question in quantitative psychology. The Permutation approach is then defined and applied to the Overlapping Index, demonstrating how non-parametric methods can offer insights without relying on typical parametric assumptions. Specifically, the Permutation test involves shuffling data points to generate a sampling distribution, allowing the calculation of a p-value and highlighting its utility in assessing the statistical significance when comparing two groups/conditions. Next, the Simulation study uses a set of different combinations of parameters to simulate various scenarios that might meet or violate the assumptions of different statistical tests, modeling a range of conditions reflective of real-world complexities in psychological research. This simulation facilitates the evaluation of the statistical power (probability of correctly rejecting a false null hypothesis) and the type I error rate (likelihood of incorrectly rejecting a true null hypothesis) of each approach. To this end, several statistical tests are compared: the t-test for independent samples, assuming equal variances; the Welch test, which does not assume equal variances; the Wilcoxon test, suitable for ordinal data or when normality is not assumed; the Permutation Test on the Overlapping Index, providing a non-parametric approach to evaluate distributional differences; the F test for examining the homogeneity of variances; and the Kolmogorov-Smirnov test for comparing two distributions regardless of their underlying forms. These results enable researchers to visualize and comprehend the reliability and utility of each approach, particularly valuable in scenarios commonly encountered in quantitative psychology, where navigating data characteristics and adhering to or deviating from test assumptions is crucial.

Finally, we discuss the results, offering insights into the strengths and limitations of the Permutation-based Overlapping index and its potential applications in psychological sciences.

2 Overlapping Index

The overlapping index (η) is an intuitive way to define the area intesected by two or more empirical density functions (Pastore & Calcagnì, 2019). In a simple way, two distributions are similar when their distribution functions overlap, and as η diminishes, the two distributions differ. The η index varies from zero – when the distributions are completely disjoint – and one – when they are completely overlapped (Pastore, 2018). The simple interpretation of the overlapping index (η) makes its use particularly suitable for many applications (?, ?).

More formally, assuming two probability density functions $f_A(x)$ and $f_B(x)$, the overlapping index $\eta : \mathbb{R}^n \times \mathbb{R}^n \to [0, 1]$ is formally defined in the following way:

$$\eta(A, B) = \int_{\mathbb{R}^n} min[f_A(x), f_B(x)]dx \tag{1}$$

where, in the discrete case, the integer can be replaced by summation. As previously mentioned, $\eta(A, B)$ is normalized to one and when the distributions of A and B do not have points in common, meaning that $f_A(x)$ and $f_B(x)$ are disjoint, $\eta(A, B) = 0$. This index provides an intuitive way to quantify the agreement between A and B based on their density functions (Inman & Bradley Jr, 1989).

To quickly illustrate an example of the overlapping area in figure 1 are represented two different empirical densities. In panel [A], are depicted two density distributions, a Normal(10,2) and a Uniform(0,20); note that the two distributions have the same mean (10), but different variance, 4 and 33.3 respectively. In the panel [B] are represented the empirical densities of two random samples of 30 observations drawn from the two populations specified as in panel [A]; the estimated overlapping area being $\eta = 0.46$.

The figure 1 shows how two distributions with almost same mean could still be very different from each other with the overlapping area being $\hat{\eta} = 0.46$. In this case, the t-test focuses on mean differences, therefore correctly does not rejects the null hypothesis, even though the degree of similarity of the two densities is only 46%, in other words, the difference is about 54%. Moreover, we remind that the t-test in this case is far from ideal as the two distributions have different variances.

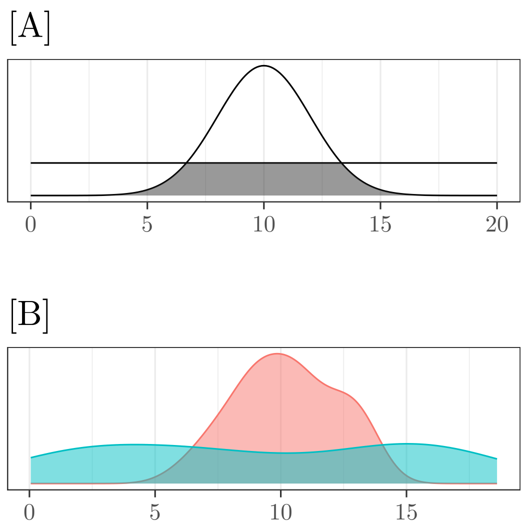

Figure 1: Comparison of a normal distribution and a uniform distribution with same mean.

2.1 Permutation approach

Now we will introduce another approach which does not rely on the assumptions of linear models: the permutation approach. This is a non-parametric statistical method that can be used to determine statistical significance and it is most useful when the assumptions of parametric tests are not met (Pesarin & Salmaso, 2010). What the test does is to rearrange the data in many different ways and recalculates the test statistic each time. If we are thinking about a simple mean comparison (a t-test), the data in the two groups are mixed over and over and the t-value is calculated each time. If the two groups come from the same population, mixing the labels should give similar results to the ones observed. Else, if the two groups come from different populations, mixing tags should lead to very different results. From the empirical density of the permuted values it is possible to calculate the p-value as the probability to obtain an equal or more extreme value compared to the observed one.

2.2 Application of permutation test to the overlapping index

Even though the overlapping index has a simple interpretation, one could argue that it does not provide information on the significance of the parameter η, therefore, we decided to implement permutation testing to offer to the ones interested a value of significance. In particular, we implemented permutations test, to give a tool that tests differences in distributions without assumptions, offering a valid alternative in cases in which traditional assumptions are not met.

If we are reasoning from the perspective of Null Hypothesis Significance Testing (NHST), we should define the null hypothesis as follows: $H_0 : \eta = 1$, meaning that there is complete overlap between the theoretical densities in the two populations from which we sample the data. For this reason, it is more intuitive to work with the complement of η, which is $1 - \eta = \zeta$ which is the area of non-overlap, therefore, defining the null hypothesis as $H_0 : \zeta = 0$, once again meaning that there is no difference between the densities of the two populations. Obviously, this does not change the results, but only the way in which they are interpreted. When testing the difference between the two distributions, we will no longer be working with η, but with the complement ζ.

The algorithm estimates the value of ζ on the observed data ($\hat{\zeta}$). Then, through permutation, the observed values of the two groups are randomly re-assigned to the groups for B times, estimating again the new value of $\hat{\zeta}_b$. The times in which the estimate of $\hat{\zeta}_b$ on permuted data is higher or equal than the one observed on real data is estimated ($\hat{\zeta}_b \geq \hat{\zeta}$) and then the found value is divided by B, returning the p-value. HA SENSO DIRE CHE BISOGNA AGGIUNGERE 1 PERCHE' IL VALORE NON SIA MAI 0? A typical example of data not respecting previously said assumptions is reaction times and for this purpose we present a real case of a dataset available online (**citation of the OSF repository**) on reaction times of word reading of high and low frequency words in English and we implement on the overlapping function the permutation test.

In the figure 2[A] are represented the densities of re-

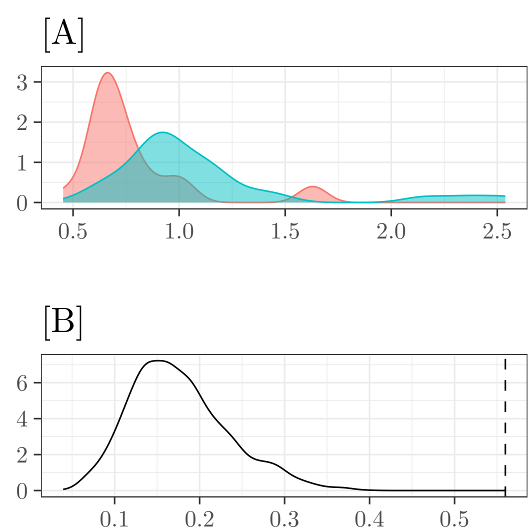

Figure 2: [A] Distribution of reaction times of word reading of high and low frequency words in English, in this case the non-overlapping area is $\hat{\zeta} = 0.56$; [B] Distribution of $\hat{\zeta}$ obtained with 2000 permutations of the data.

action times of word reading of high and low frequency words in English from a sample of two groups of 30 observations each; the two distributions have mean 0.78 and 1.11, standard deviation 0.27 and 0.47, and skewness 2.32 and 2.02 respectively. The obtained value of $\hat{\eta}$ is 0.44, and consequently $\hat{\zeta}$ is 0.56. In figure 2[B] is represented the distribution of the values of $\hat{\zeta}$ obtained with 2000 permutations; we can calculate the *p-value* as follows:

$$p = \frac{(\#\hat{\zeta}_b \geq \hat{\zeta}) + 1}{B + 1} = \frac{1}{2001}$$

Given that $p < .001$, we can conclude that the difference is statistically significant; in this case the t test give the same conclusion $t_{(58)} = 3.34$, $p = .001$, but we remind that the t-test evaluates only the difference between the means and requires assumtions that in this scenario are clearly not met.

3 Simulation study

To evaluate the performance of the permutation test applied to the overlapping index, we performed a simulation study. The aim is to generate data for a set of scenarios distinguishing mean, variance and shape of the populations and compare the ζ perm test to other commonly used tests in terms of type I error control and power.

3.1 Data generation

In the simulation, two density distributions will be compared for many different scenarios. The first distribution will always be a normal standard distribution with $\mu = 0$ and $\sigma = 1$. To simulate data for the second distribution we use the Skew-Normal distribution (Azzalini, 1985), which is defined in the following way: given $\xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}$, then for $y \in \mathbb{R}$ we have

$$\mathcal{SN}(y|\xi,\omega,\alpha) = \frac{1}{\omega\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\xi}{\omega}\right)^2\right]\left[1 + \text{erf}\left(\alpha\left(\frac{y-\xi}{\omega\sqrt{2}}\right)\right)\right] \tag{2}$$

in which

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2} dt$$

is the *error function*. When $\xi = 0$, $\omega = 1$ and $\alpha = 0$ the distribution is a standard normal distribution.

ξ is the location parameter, ω is the scale parameter and α is related to the skewness of the distribution. Therefore, this distribution is suitable to generate data modelling both the distance between means (the effect size), symmetry and variance.

Mean and variance of the Skew-Normal are respectively:

$$\begin{aligned} \mu &= \xi + \omega\gamma\sqrt{2/\pi} \\ \sigma^2 &= \omega^2[1 - (2\gamma^2)/\pi] \end{aligned} \tag{3}$$

in which $\gamma = \alpha/\sqrt{1+\alpha^2}$. Based on the equations (3) we can determine the values to assign to the parameters ξ e ω in function of μ and σ with the equations:

$$\begin{aligned} \xi &= \mu - \omega\gamma\sqrt{2/\pi} \\ \omega &= \sqrt{\sigma^2/[1 - (2\gamma^2)/\pi]} \end{aligned} \tag{4}$$

The Skew-Normal distribution is optimal for our purpose as it allows to have control over parameters of mean, variance, skewness and kurtosis, as shown in figure 3.

3.2 Simulation design

In the simulation we confront two samples extracted from a Skew-Normal, the first one is generated from $\mathcal{SN}(0,1,0)$, which is the Standard-Normal distribution, and the second one from $\mathcal{SN}(\xi,\omega,\alpha)$. Consequently, the first sample derives always from a population with mean 0 and variance 1. To define the various scenarios, we manipulate the parameters of the second population in orther to obtain specific differences in means (δ), standard deviations (σ) and skewness (α). Four factors were sistematically varied ina complete four-factors design as follows:

- $\delta = (0, 0.2, 0.5, 0.8)$; mean of the second population, which corresponds also to the difference between the two groups, the first one has always $\mu = 0$;
- $\sigma = (1, 2, 3)$; standard deviation of the second population;
- $\alpha = (0, 2, 10)$; degree of asymmetry (skewness) of the second population;
- $n = (10, 20, 50, 100, 500)$; sample size, equal in the two samples.

For each of the $4 \times 3 \times 3 \times 5 = 180$ conditions we generated 3000 sets of data on which we performed the analysis.

In figure 4 are graphically represented the 36 scenarios of data generation, the black curves are the first population, always a $\mathcal{SN}(0,1,0)$, and the red curves are relative to the second population $\mathcal{SN}(\xi,\omega,\alpha)$.

For each combination $\delta \times \sigma \times \alpha \times n$, on the generated data were performed the following tests:

- t test for independent samples, assuming equal variance;
- Welch test for independent samples;
- Wilcoxon test for independent samples;
- Permutation test on the complement of the overlapping index, $\zeta = 1 - \eta$, which therefore becomes an index of difference between groups;
- F test of homogeneity of variances;

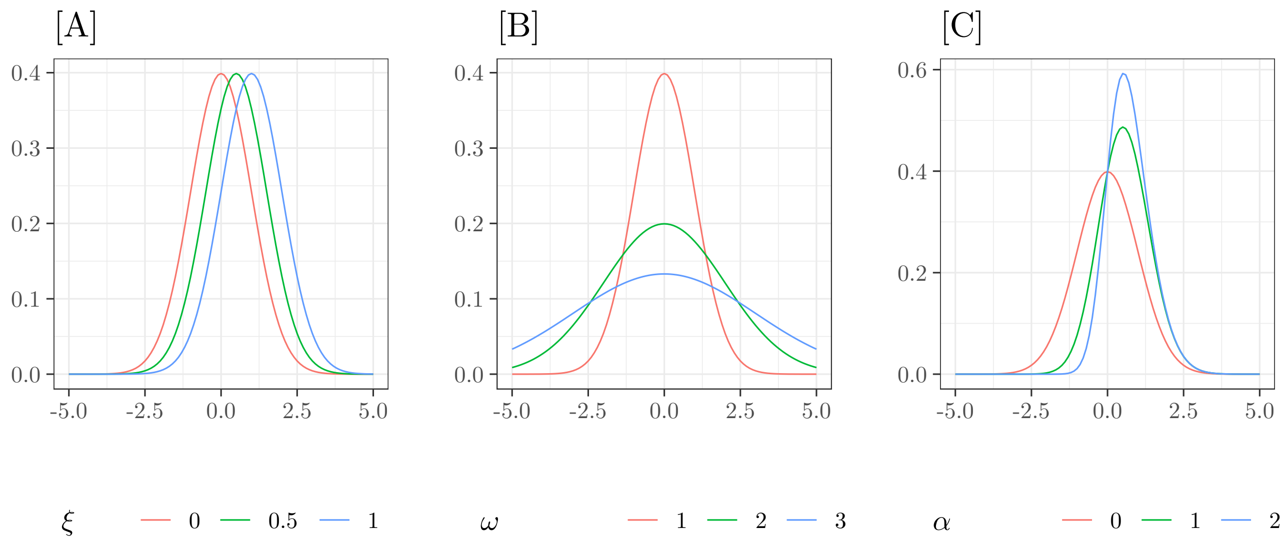

Figure 3: Examples of Skew-Normal distributions (ξ,ω,α); [A] three densities with same variance and shape but different location parameter values ($\xi = 0, 0.5, 1$), [B] three densities with same mean and shape but different scale parameter values ($\omega = 1, 2, 3$) and [C] three densities with same mean and variance but different shape parameter values ($\alpha = 0, 1, 2$).

- Kolmogorov-Smirnov test for comparing two distributions.

The whole procedure generated a total of 540000 datasets as well as 3240000 of statistical tests and corresponding p-values.

3.3 Definition of Statistical tests

We introduce the chosen statistical tests summarizing the specific hypothesis and assumptions for each one.

3.3.1 t test

This is the classic case of a test for independent samples assuming equal variances and the normality of the two distributions:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ with } \sigma_1 = \sigma_2$$

3.3.2 Welch test

This is the t test modified when homogeneity of variances is not respected:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ with } \sigma_1 \neq \sigma_2$$

Also this test assumes the normality.

3.3.3 Wicoxon-Mann-Whitney test

This is the test on ranks which evaluates the following hypothesis without assumptions on distributions:

$$H_0 : P(X_1 > X_2) = P(X_2 > X_1) = 0.5$$

in which X_1 and X_2 are the random variables representing the observations extracted from the two populations.

3.3.4 ζ overlapping test

Since $\zeta = 1 - \eta$, in which η is the area of overlapping of the empirical distributions, the null hypothesis of the test is

$$H_0 : \zeta = 0$$

which implies that the data comes from the same population, or from populations with same shape (mean, variance and skewness) but without specific assumptions.

3.3.5 F test

This is the test of homogeneity of variances

$$H_0 : \sigma_1^2 = \sigma_2^2$$

assuming the normality.

3.3.6 Kolmogorov-Smirnov test

This test compares the cumulative distributions

$$H_0 : F(X_1) = F(X_2)$$

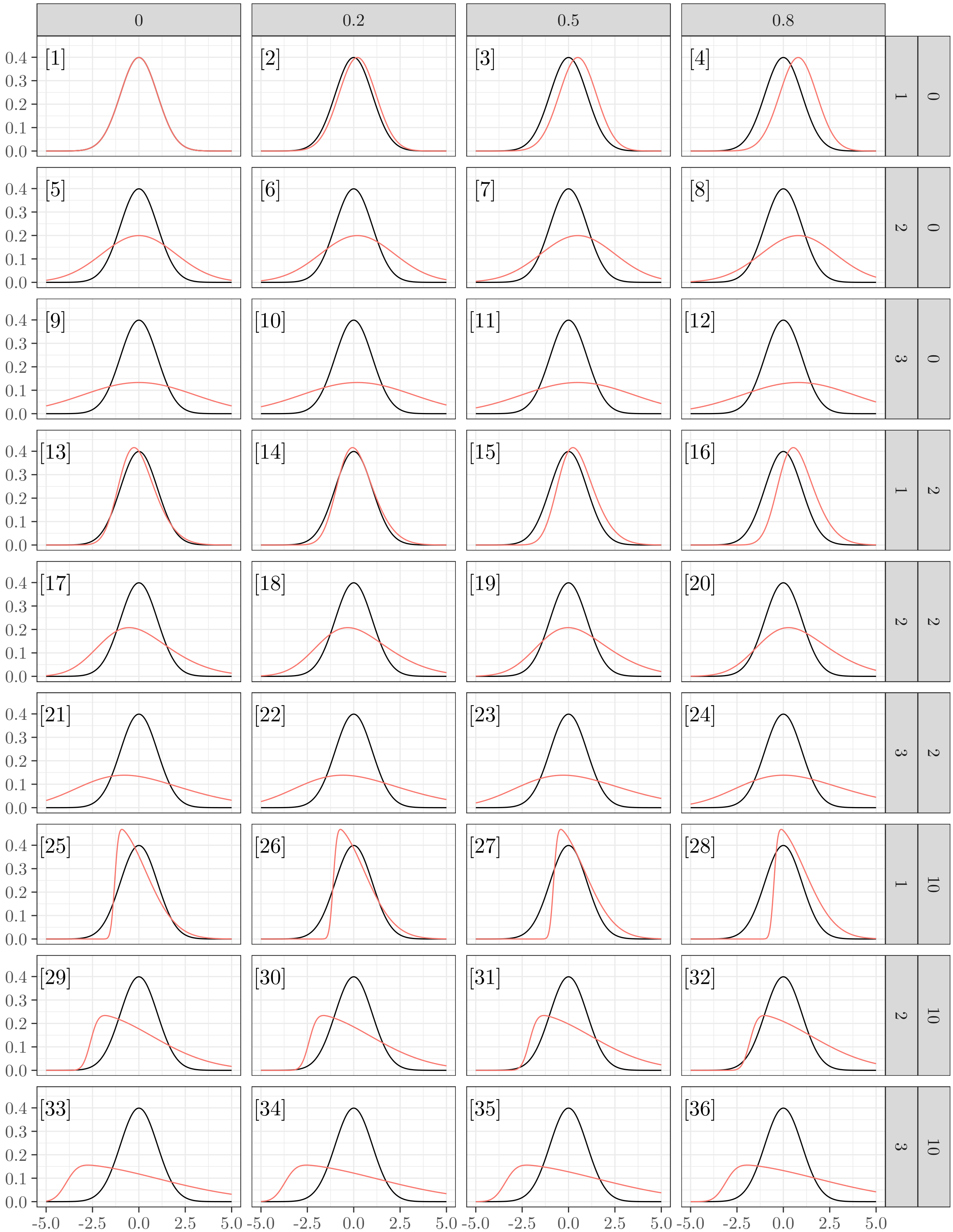

Figure 4: Generative data distributions in function of δ (column panels) and σ (row panels). The black curves are the first sample, $\mathcal{SN}(0,1,0)$, the red ones represent second sample.

without assumptions.

We considered panel [1], figure 4, the scenario in which all null hypothesis are true and assumptions are respected. Consequently, we computed type I error by counting how many times the test is significant in this scenario, and the power by counting how many times it will be significant in all other scenarios.

4 Results

Figure 5 represents the correlation matrix between the p-values for the different tests in all experimental conditions. Two subgroups are clearly visible: the first group with tests on mean and ranks, and the second one on tests about the shape, the F test is not correlated with the others.

The high correlation between the ***t*-test** and **Welch test** reflects their similar objectives, particularly in testing mean differences. However, the lower correlations between parametric and non-parametric tests, such as the **Wilcoxon** and **KS** tests, indicate that these tests capture different aspects of the data (e.g., ranks or distribution shapes rather than means). The **permutation-based tests** exhibit intermediate correlations with both parametric and non-parametric methods, indicating that their results may align with both types depending on the underlying data structure.

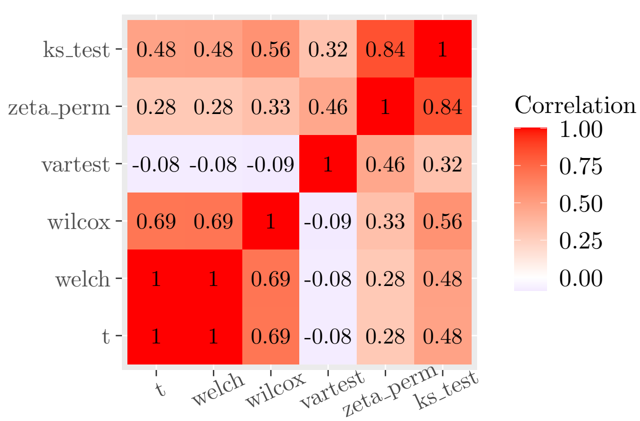

Figure 5: Correlation matrix among p-values ($N = 540000$).

4.1 Type I error and power

In figure 6, is represented type I error in panel [A] and power in panel [B], for all scenarios it is evaluated when H_0 is either true or false respectively. Panel [A] shows how they all control well enough for type I error, except for the F test. The ζ perm test outperforms all other tests in terms of power, already from small sample sizes, once more, the F test is the exception, as it is a test on variance.

5 Discussion

The present analysis evaluated the performance of various statistical significance tests across simulated scenarios that altered whether the null hypothesis (H_0) is true or false and whether the assumptions required by each test are met. The tests include both parametric (such as the T-test, Welch test, and F-test for variance) and non-parametric methods (such as the Wilcoxon Signed-Rank, Kolmogorov-Smirnov, and permutation-based approaches including the ζ permutation test). Through these scenarios, we assess each test's robustness, Type I error control, and power under ideal and non-ideal conditions.

In **Scenario A** (true H_0, assumptions met), the T-test, Welch test, and permutation-based tests maintain Type I error close to the nominal level of 0.05, with the Welch test showing minor variability due to its robustness to variance differences. Non-parametric tests (Wilcoxon, KS, ζ-permutation) are conservative, slightly undershooting the nominal error—a typical trade-off for robustness.

In **Scenario B** (false H_0, assumptions met), power is key. The Welch test excels as sample sizes grow, especially with unequal variances, while permutation-based tests (T-test, ζ) show strong power, making them effective in detecting true effects without strict distributional requirements. Non-parametric tests display moderate power, better suited for cases where shape differences are of interest rather than mean differences.

Scenario C (true H_0, assumptions violated) tests robustness. The Welch and permutation-based tests maintain Type I control under assumption violations, whereas parametric T and F-tests struggle, particularly with heteroscedasticity. Non-parametric tests (Wilcoxon, KS) remain conservative and reliable, showing resilience to non-normality.

Finally, in **Scenario D** (false H_0, assumptions violated), the Welch test and permutation-based tests (T-test, ζ) stand out with high power and robustness, ideal for real-world data with heteroscedasticity or non-normality. Non-parametric tests retain robustness but lack power, while the F-test proves unreliable under these challenging conditions.

Further analysis of p-value correlations among the tests provides additional insight into their relationships. High correlation between the T-test and Welch test, for instance, reflects their similar objectives and shared focus on mean differences. However, lower correlations between parametric and non-parametric methods, such as the Wilcoxon and KS tests, indicate that these tests capture distinct aspects of the data, such as ranks or dis-

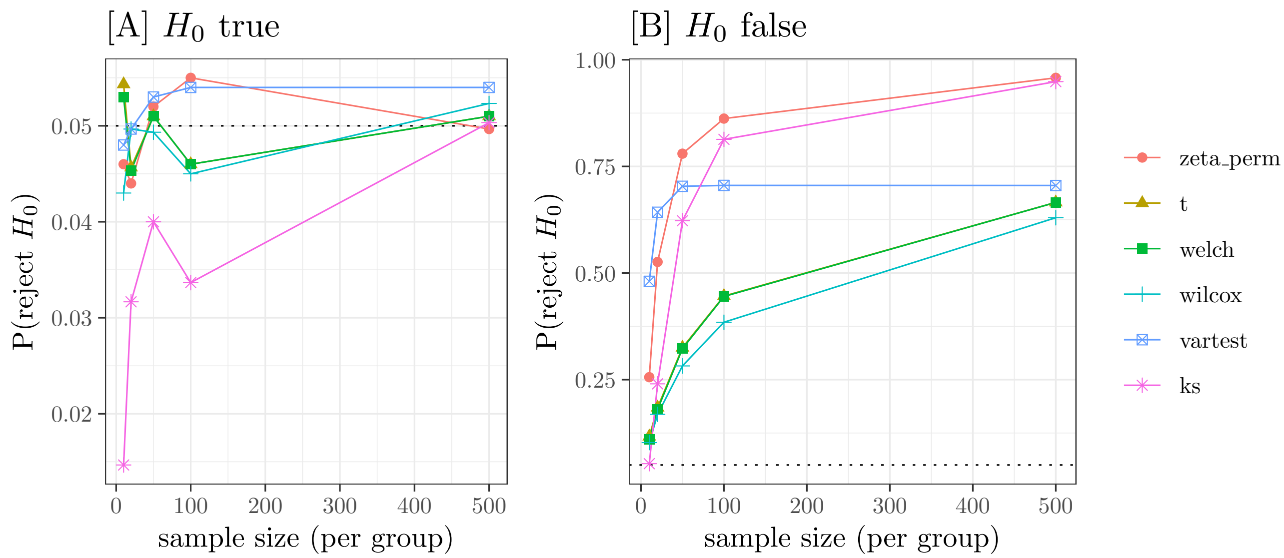

Figure 6: Control of type I error [A] and power [B] in the various tests.

tribution shapes rather than means. The permutation-based tests show intermediate correlations with both parametric and non-parametric methods, which suggests that they may align with either type of test depending on the underlying data structure.

Advantages and Limitations of the ζ Permutation Test

The ζ **permutation test**, designed to measure the degree of overlap between distributions, has specific advantages and limitations. Its main strength lies in its robustness to distributional assumptions, as it calculates p-values through permutations rather than relying on parametric assumptions like normality or equal variance. This makes it particularly useful in scenarios where other tests may fail due to assumption violations, providing a conservative Type I error rate when H_0 is true and robust power when H_0 is false.

However, the ζ permutation test's conservative nature may limit its sensitivity in detecting small mean differences, especially when distributions overlap substantially. Its design focuses on distributional overlap rather than mean differences, which means it may lack power relative to parametric tests that specifically target mean shifts. In scenarios where the primary effect is a shift in central tendency rather than overlap, the ζ permutation test may not be the optimal choice.

The ζ permutation test is indeed a valuable tool for non-parametric inference, particularly when distributional assumptions do not meet those required by common statistical test e.g. t-test. These are particularly relevant points given that in psychological sciences studies often involve small sample sizes, and relying on small changes in location parameters like the mean can be risky. For example, small samples are highly susceptible to the influence of extreme values, which can skew the mean and lead to misleading conclusions about effect sizes. On the contrary, as demonstrated in simulations, the ζ permutation test is less prone to being dramatically impacted by extreme values, as it directly measures the distributional overlap between groups rather than focusing solely on mean differences. This characteristic makes the ζ permutation test particularly valuable in small-sample contexts, like in psychological science where robustness to outliers is critical for obtaining reliable insights into group differences.

6 Conclusion

By exploring alternative scenarios, the study offers practical indication to operate a shift in the philosophical approach to data analysis and significance testing. In fact, the Overlapping index forces the functional interpretation of the results to move beyond significance testing alone (Pastore, 2018; Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016; Gelman, 2018). In psychological research, considering the distribution of data rather than relying solely on significance testing offers a deeper, more nuanced understanding of results. Traditional significance testing doesn't provide information about the nature or magnitude of that effect. By visualizing and considering the entire distribution of data, researchers can observe the spread, central tendency, and shape of the data, which often reveal valuable insights about variability and individual differences within the sample. As presented in Figure 1, reporting a mean difference without an understanding of the data's variability could misrepresent the consistency or generalizability of the observed effect. Therefore, incorporating distributional analyses allows psychologists to present a fuller picture of their findings, improving both interpretability and transparency in their research conclusions.

Moreover, the present study further underscores the necessity of selecting the most suitable statistical tools contingent on the specific characteristics of the data and the assumptions inherent in the analytical techniques employed. Such a switch in the philosophical approach to data analysis in psychological sciences (Vasishth & Gelman, 2021) may improves the robustness and validity of psychological research findings, allowing for more aware interpretations and generalizations. We stress this by making open available data and material so that such an approach might be useful for a wide range of psychologists interested in increasing the understandability of their results.

The findings underscore the necessity of choosing statistical methods that are resilient to the complexities inherent in psychological data, where assumption violations are often inevitable. Tests like the Welch and ζ permutation tests exemplify robust alternatives that accommodate data with unequal variances or non-normal distributions, offering reliable results even when classic parametric conditions are unmet. In this way, these tools extend the flexibility of significance testing, enabling a nuanced understanding of effects in psychology.

Ultimately, statistics in psychology should reflect both theoretical knowledge and an appreciation for the distributional nuances of psychological variables. Rather than a rigid application of conventional methods, statistical analysis should be a deliberate choice that aligns with the nature of the data and the research question. Approaches such as the overlapping index and permutation-based methods embody this principle, capturing the depth and complexity of psychological effects in a way that is both methodologically rigorous and sensitive to the real-world structure of psychological phenomena.

Legenda

η is the area of overlap

ζ is the area of non overlap, therefore $1-\eta$

μ is the parameter of the mean of the normal standard

σ is the standard deviation of the normal standard

α determins the simmetry of the skew-normal

ξ is the mean value of the skew-normal

ω determines the variance of the skew-normal

δ is the difference between the two means

Ethical considerations

Ethical approval was not required

Conflicting interest

The authors declare no conflict of interests.

Funding statement

No Funding supported this project.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 171–178.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological bulletin*, *57*(1), 49.

Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the american psychological association publication manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, *64*(3), 138–146.

Fisher, R. A. (1925). Theory of statistical estimation. In *Mathematical proceedings of the cambridge philosophical society* (Vol. 22, pp. 700–725).

Gelman, A. (2018). The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin*, *44*(1), 16–23.

Grosjean, M., Rosenbaum, D. A., & Elsinger, C. (2001). Timing and reaction time. *Journal of Experimental Psychology: General*, *130*(2), 256.

Inman, H. F., & Bradley Jr, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics-theory and Methods*, *18*(10), 3851–3874.

Pastore, M. (2018). Overlapping: a R package for estimating overlapping in empirical distributions. *Journal of Open Source Software*, *3*(32), 1023.

Pastore, M., & Calcagnì, A. (2019). Measuring distribution similarities between samples: a distribution-free overlapping index. *Frontiers in psychology*, *10*, 1089.

Pesarin, F., & Salmaso, L. (2010). The permutation testing approach: a review. *Statistica*, *70*(4), 481–509.

Proctor, R. W., & Schneider, D. W. (2018). Hick's law for choice reaction time: A review. *Quarterly Journal of Experimental Psychology*, *71*(6), 1281–1299.

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, *16*(4), 744–755.

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*, 1241–1260.

Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *The American journal of psychology*, *123*(1), 39–50.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.

Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*, *59*(5), 1311–1342.