CrossMark

A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs

Sara Kherad-Pajouh · Olivier Renaud

Abstract Repeated measures ANOVA and mixed-model designs are the main classes of experimental designs used in psychology. The usual analysis relies on some parametric assumptions (typically Gaussianity). In this article, we propose methods to analyze the data when the parametric conditions do not hold. The permutation test, which is a non-parametric test, is suitable for hypothesis testing and can be applied to experimental designs. The application of permutation tests in simpler experimental designs such as factorial ANOVA or ANOVA with only between-subject factors has already been considered. The main purpose of this paper is to focus on more complex designs that include only within-subject factors (repeated measures) or designs that include both within-subject and between-subject factors (mixed-model designs). First, a general approximate permutation test (permutation of the residuals under the reduced model or reduced residuals) is proposed for any repeated measures and mixed-model designs, for any number of repetitions per cell, any number of subjects and factors and for both balanced and unbalanced designs (all-cell-filled). Next, a permutation test that uses residuals that are exchangeable up to the second moment is introduced for balanced cases in the same class of experimental designs. This permutation test is therefore exact for spherical data. Finally, we provide simulations results for the comparison of the level and the power of the proposed methods.

S. Kherad-Pajouh (✉)
Division of Biostatistics, School of Public Health, University of California, Berkeley, CA, USA
e-mail: Sara.kherad@gmail.com

O. Renaud
Methodology & Data Analysis Group, Dept. of Psychology, Boulevard du Pont d'Arve 40,
1205 Geneva, Switzerland
e-mail: Olivier.Renaud@unige.ch

1 Introduction

ANOVA is a technique used to test the equality of means from different treatment groups. Usually the designs with within-subject factors only are called repeated measures ANOVA. Within-subject factors are factors for which measurements are obtained on the same subject over time or under all the conditions. These designs are extremely common in psychological and biomedical studies. Mixed-model designs are another family of useful designs, especially in clinical research, when there are both within-subject and between-subject factors. Between-subject factors occur when each subject is measured on only one level of the factor. Mixed-model designs are also known as split-plot ANOVA (Keppel 1991), or univariate mixed-model designs with the subject and its interaction as a random effects. Sometimes these designs are simply called repeated measures; for example the IBM SPSS menu called *repeated measures* treats split-plot designs as well (IBM Corp 2013).

These kinds of designs have some advantages over other (between-subject) ANOVA methods. For instance in clinical research, fewer subjects are required than for the between-subject ANOVA. Repeated measures designs make efficient use of subjects, both in the practical sense of using fewer subjects than between-subject designs, and in the statistical sense of having less error variance, i.e. more statistical power. The major disadvantage of this kind of design is that usually the models are more complex than for the non-repeated measures designs, since they need to take into account the associations between observations obtained from the same individuals.

It is likely that for this kind of design, the parametric assumptions are not satisfied, like the assumption that all the random parts are normally distributed. In that case, the result of the methods based on parametric assumptions might not be reliable.

Permutation tests or randomization tests date back to Fisher (1935). These tests belongs to the family of distribution-free techniques and are suitable for hypothesis testing. The only requirement for permutation tests is the exchangeability of the observations under the null hypothesis. For a set of continuous random variables $(z_1, z_2, \ldots, z_n)$, exchangeability holds if the joint density f satisfies $f(z_1, z_2, \ldots, z_n) = f\left(z_{\Pi(1)}, z_{\Pi(2)}, \ldots, z_{\Pi(n)}\right)$ for any permutation Π of the indices (Good 1994). Permutation tests can be categorized as either exact or approximate. A test is exact if under the null hypothesis the type I error rate is equal to the nominal level of the test (Good 1994; Anderson and Ter Braak 2003). The exchangeability of the permuted units is the requirement for having an exact permutation test. Note that the permuted units can be the original observations, some form of residuals, or a restricted set thereof (citealtbassospssynchronizedsps2007,bassospsdiscussionsps2006).

For a review of all the alternatives of permutation methods in the ANOVA designs, see Anderson and Ter Braak (2003). For a complete account on the permutation of the residuals, we refer to Kherad-Pajouh and Renaud (2010). In the following, we only mention the literature that deals with repeated measures ANOVA based on permutation tests. Since in the repeated measures ANOVA (or mixed-model designs), there is a correlation between observations within a given subject, the application of permutation tests is more involved. White and Harris (1978) proposed a permutation test for experimental designs, and a computer program which can run up to a four-way analysis of variance with any combination of between- and within-subject factors. The pro-

posed permutation test is an approximate test. Still and White (1981) also developed an alternative F test for ANOVA, using various forms of approximate permutation tests. Good (1994) introduced a form of restricted permutations for ANOVA with only between-subject factors, specifically to test the main effects. Later, Anderson and Ter Braak (2003) applied permutation techniques (permutation of the observations and permutation of the residuals) for fixed effect ANOVA and nested designs. Finally, Mazzaro et al. (2001), Pesarin and Salmaso (2010) and Basso (2009) introduced some techniques for repeated measures designs and unbalanced repeated measures designs. However both the model they considered, and what they called "repeated measures", are different from what we will present here, (they considered only repetition over time). Pesarin (2001) also considered repetition over time in the MANOVA layout for response profiles, where the analysis can be extended to more complex designs and where repetitions are not necessarily time points. In their proposed techniques, repeated measurements, panel data and longitudinal data are synonymous.

The structure of the paper is the following. In Sect. 2, a general model and its matrix formulation are given for balanced and unbalanced repeated measures and mixed-model designs. Section 3 presents the parametric approach for balanced repeated measures and mixed-model designs, and corresponding F-statistics. In Sect. 4, we introduce an approximate permutation test for balanced repeated measures and mixed-model designs, which is based on permutation of the reduced residuals. We also introduce a technique for computing the residuals which is applicable to general repeated measures designs with any number of factors and subjects. The tests are carried out separately, much like in traditional ANOVA F-tests. We present an extension of permutation of the reduced residuals for unbalanced designs with unequal number of repetitions in each cell in Sect. 5. A permutation test for balanced repeated measures and mixed-model designs is introduced in Sect. 6. This method is based on permutation of the residuals under the modified model, which removes the correlation between the reduced residuals. In Sect. 7, we show an application of the introduced methods for repeated measure designs with one and two within-subject factors. In Sect. 8, the application of the introduced method for mixed-model designs (one between-subject factor and one within-subject factor) is considered, and Sect. 9 is dedicated to the simulations that have been performed in order to compare the level and the power of different proposed and existing methods.

2 General model for mixed-model designs and repeated measures ANOVA

The mixed-model designs, as their name suggests, are designs with combination of between-subject and within-subject factors. Consider A as (one of) the between-subject factor(s), B as (one of) the within-subject factor(s), and S the random factor corresponding to the subjects. The structural model for a mixed-model design can be written as

$$
\begin{aligned}
y_{l\cdots i\cdots j\cdots k} = {} & \mu + \alpha_l + \cdots + \pi_{j/l} + \eta_i + \cdots + \alpha\eta_{li} + \cdots + \pi\eta_{ji/l} \\
& + \cdots + \epsilon_{l\cdots i\cdots j\cdots k}, \\
& l = 1, \ldots, a, \ i = 1, \ldots, b, \ j = 1, \ldots, s, \ \text{and} \ k = 1, \ldots, n_{lijk}. \quad (1)
\end{aligned}
$$

In this model α_l are the effects of the between-subject factor, η_i are the effects of the within-subject factor and $\pi_{j/l}$ are the effects of the subjects. The interaction terms are given by the concatenation of the symbols ($\alpha\eta_{li}$ and $\pi\eta_{ji/l}$). $N = \Sigma_i \Sigma_j \cdots \Sigma_k n_{ij\ldots k}$ is the total number of observations and it is clear that in the balanced case, $N = abs \cdots n$. It is important to note that we included the interactions between the subject and the within-subject factor(s). It is quite common and highly desirable in psychology, see e.g. Rouanet and Lépine (1970), and this is the model underlying SPSS repeated measures procedure. For various reasons, these terms (interactions between subjects and between-subject factors) are often omitted in the mixed-effect/hierarchical/multilevel model literature as well as in the longitudinal analysis literature. As A is a between-subject factor, there is no interaction term such as $\alpha\pi$ in the model. In our model, α and η are fixed factors. Without loss of generality, we will assume that they satisfy the sigma-restriction constraints, i.e. $\sum_{l=1}^{a} \alpha_l = 0$ and $\sum_{i=1}^{b} \eta_i = 0$. Concerning the random parts in the model, we assume that $\pi \sim (0, \sigma_\pi^2)$ and $\eta\pi \sim (0, \sigma_{\eta\pi}^2)$. Following the pigeon hole model that characterize the distinction between fixed and random effects (Cornfield and Tukey 1956), the interaction terms ($\eta\pi$) of the subject (random effect, π) and the within-subject factor (fixed effect, η) satisfy the constraint $\sum_{i=1}^{b} \pi\eta_{ji/l} = 0$ for $j = 1, \cdots, s$, which introduces a dependence between certain interaction terms for each subject between the levels of the within-subject factor, for more details we refer to Sahai and Ageel (2000). Note that the major dependence or correlation that will be brought out in the next sections is no due to this term but to the subject effects. The extension to more than one within-subject factor and/or more than one between-subject factor of our notation is obvious. Finally, let's mention here that repeated measures designs are special cases of our model (1), where there are only within-subject factors in the model. The conditions mentioned above, hold in this kinds of designs as well. It is not difficult to see that model (1) can be written in matrix form as

$$y = X\beta + Z\gamma + \epsilon,$$

where y contains all observations for all subjects, $X_{N \times p}$ is the design matrix for the fixed part of the model and $Z_{N \times q}$ is the design matrix due to the random part of the model. $\beta_{p \times 1}$ is the unknown vector of parameters for the fixed effects, $\gamma_{q \times 1}$ contains the random part of the model and $\epsilon_{N \times 1}$ is random vector of the error terms. It should be mentioned here that X and Z matrix and corresponding parameters are based on *sigma-restricted* (SR) parametrization, see e.g. Cardinal and Aitken (2006).

Often in practice, the condition of normality of the random parts and of the error term are not satisfied, and the traditional approaches are not suitable. Expressed in the matrix form, we will only suppose that γ, the vector of random effects is distributed as $\gamma \sim (0, G)$, where G is the covariance matrix of random part of the design matrix. The random error term ϵ is distributed as $\epsilon \sim (0, \sigma_\epsilon^2 I_N)$, and is uncorrelated from the other random part γ. Therefore we can write $y \sim (X\beta, \Sigma)$, where $\Sigma = Z'GZ + \sigma_\epsilon^2 I_N$.

Suppose we want to test one part of the parameter, say β_2, where the vector of fixed parameter β is divided in two parts: $\beta' = (\beta_1', \beta_2')$. The corresponding hypotheses are

$$\mathsf{H}_0 : \beta_2 = 0 \quad \text{vs.} \quad \mathsf{H}_1 : \beta_2 \neq 0. \tag{2}$$

The matrix formula for the repeated measures or mixed-model design can be written as the following

$$y = X_1\beta_1 + X_2\beta_2 + Z_1\gamma_1 + Z_2\gamma_2 + \epsilon. \tag{3}$$

In this model $(X_1)_{N \times p_1}$ is the part of design matrix which corresponds to β_1 and $(X_2)_{N \times p_2}$ to β_2. The matrix Z_2 and its corresponding vector of parameters γ_2 are the random effects counterparts of X_2 and β_2. More precisely they contain the interaction between the subject effect and the effect(s) to be tested for within-subject effects and merely the subject effect for between-subject effects. The matrix Z_1 and its corresponding vector of parameters γ_1 contain all the other random effects.

3 Parametric approach for balanced designs

In the case of a balanced design, a genuine test for β_2 is defined as

$$F_{ss} = \frac{SS(\hat{\beta}_2)/p_2}{SS(\hat{\gamma}_2)/q_2}. \tag{4}$$

The F-statistic for test of β_2 (between-subject or within-subject factor) is based on two sum of squares: the one attributed to the effect of β_2 and another one, taken from the random part, that play the role of the denominator, see e.g. Myers and Well (2003) . So $(Z_1)_{N \times q_1}$ can be viewed as the part of design matrix which is not in the denominator of F-statistic to test for (β_2) and $(Z_2)_{N \times q_2}$ is the part which is used.

This test is not the likelihood ratio test, but it can be shown to be an exact test with Gaussian and sphericity assumption. It is the test used in SPSS with the *repeated measures* menu, and S-Plus/R *Error* statement in function aov. Due to the orthogonality implied by the balanced designs, the F-statistic can also be written based on *residual sum of squares*(RSS) as

$$F_{ss} = \frac{[RSS(\hat{\beta}_1) - RSS(\hat{\beta}_1, \hat{\beta}_2)]/p_2}{[RSS(\hat{\beta}_1) - RSS(\hat{\beta}_1, \hat{\gamma}_2)]/q_2}, \tag{5}$$

where the estimates in parenthesis indicate which model is fitted to obtain the corresponding RSS.

Using the same computation as for the numerator in Theorem 2 of Kherad-Pajouh and Renaud (2010), we can rewrite the numerator and the denominator of (5) in a matrix form as

$$F_{ss} = \frac{y'H_{[X_2]}y/p_2}{y'H_{[Z_2]}y/q_2}, \tag{6}$$

where $H_{[X_2]} = X_2(X_2'X_2)^{-1}X_2'$ is the projection matrix of β_2, the fixed part to be tested in Eq. (3). Similarly for the denominator, the projection matrix of random part in the denominator of the F-statistic is $H_{[Z_2]} = Z_2(Z_2'Z_2)^{-1}Z_2'$. Recall that the design

matrices of X_2 and Z_2, are based on *sigma-restricted* (SR) parametrization, so the inverse is always defined.

4 Permutation of the reduced residuals for balanced designs

We now propose an altenative testing procedure, extending the idea of permutation of the residuals under the reduced model or *reduced residuals*, introduced by Still and White (1981) and Freedman and Lane (1983). In some designs like ANOVA with a single error term and nested ANOVA designs, Anderson and Ter Braak (2003) showed that this method is relatively more powerful in comparison with other permutation methods.

In Eq. (3), the random parts η has an expectation equal zero, so the only part which has non-zero expectation under the null hypothesis (2) is the fixed part not tested. So, the y_{ijkl}'s do not share the same expectation under the null hypothesis and so the observations are far from exchangeable.

The idea behind the reduced residuals is to obtain so-called residuals that share the same expectation under the null hypothesis. This is achieved by removing the effect of fixed parameter not tested. For mixed-model and repeated measures designs, the model includes also the random effects. Since the latter have mean zero, we can choose whether to keep or to remove these random parts in the reduced residuals. Now the question is: what are the necessary and sufficient conditions to define the reduced residuals? Here is a more precise definition for the reduced residuals.

Definition 1 Consider the hypothesis on β_2 as in (2). The reduced residuals are defined as $y_{rr} = H_p y$, where H_p is a projection matrix which satisfies the following conditions:

(a) all elements in y_{rr} have zero expectation under the null hypothesis;
(b) the F-statistic of the reduced residuals defined as

$$F_{rr} = \frac{y'_{rr} H_{[X_2]} y_{rr} / p_2}{y'_{rr} H_{[Z_2]} y_{rr} / q_2}, \tag{7}$$

is equal to F_{ss} for the original observations defined in (4).

The following lemma specifies a set of sufficient conditions on the projection matrix in order to have reduced residuals for repeated measures and mixed-model designs.

Lemma 1 *Any projection matrix H_p satisfying (a) $H_p X_1 = 0$, (b) $H_p X_2 = X_2$, and (c) $H_p Z_2 = Z_2$ generates a suitable reduced residual vector $y_{rr} = H_p y$.*

Due to space constraints, all the proofs are provided as supplementary material. In the following we demonstrate two specific examples of reduced residuals.

Example 1 Consider the projection matrix that only removes the fixed part among all the non-interesting parts to be tested: $H_p = (I - H_{[X_1]})$, where $H_{[X_1]} = X_1(X_1'X_1)^{-1}X_1'$. It is easy to see that $H_p X_1 = 0$. Moreover, since, in the balanced

model, X_1, X_2, Z_1, and Z_2 have orthogonal columns, they span orthogonal sub-spaces, and therefore they remain unchanged by multiplying by H_p. Hence,

$$y_{rr}^{(1)} = (I - H_{[X_1]})y = X_2\beta_2 + Z_1\gamma_1 + Z_2\gamma_2 + \epsilon_{rr}^{(1)},$$

where $\epsilon_{rr}^{(1)} = H_p\epsilon$.

Example 2 A more conservative choice removes all the parts except the ones on the numerator and denominator of the F-ratio. We called it H_{nd} and it is defined as $H_{nd} = H_{[X_2]} + H_{[Z_2]}$. Again note that mutual orthogonality of the columns of X_1, X_2, Z_1, and Z_2 implies that pre-multiplying H_{nd} removes every part, except X_2 and Z_2. Therefore, for this example we get

$$y_{rr}^{(2)} = H_{nd}y = X_2\beta_2 + Z_2\gamma_2 + \epsilon_{rr}^{(2)}. \tag{8}$$

In the following, we use y_{rr} to denote the reduced residuals in its general sense. It is worth mentioning that different projection matrices H_p may yield different values for residuals and different covariance matrices Σ_{rr}. However, they all have zero expectation under the null hypothesis. Additionally, as stated in definition 1, the value of F_{rr} is unique regardless of which projection matrix is being used. Under the null hypothesis the distribution of y_{rr} is

$$y_{rr} \sim (0, \Sigma_{rr}), \tag{9}$$

where $\Sigma_{rr} = H_p\Sigma H_p$ and Σ is the covariance matrix of y. Note that although the elements of y_{rr} have the same expectation, their (variance-) covariance matrix may not be compound-symetric and therefore they would not be exchangeable.

After selecting the suitable reduced residuals, the next step is to perform a permutation test on these residuals in order to make inference about the parameter β_2. The following algorithm shows the required steps.

Algorithm 1 – Choose a suitable statistic based on the reduced residuals y_{rr}. We propose to use the statistic of F_{rr} introduced in (7).
– Freely permute the reduced residuals to obtain "new" residuals y_{rr}^*.
– The permuted statistics is calculated for the "new" residuals y_{rr}^* as

$$F_{rr}^* = \frac{y_{rr}^{*\prime} H_{[X_2]} y_{rr}^* / p_2}{y_{rr}^{*\prime} H_{[Z_2]} y_{rr}^* / q_2}.$$

– The two previous steps should be repeated for a large number of times (say M times). Define the *p*-value for the test on β_2 as

$$p - value = \frac{(\#F_{rr}^* \geq F_{rr})}{M}, \tag{10}$$

which is simply the proportion of F_{rr}^* larger or equal to the original value of F_{rr}.

Note that this method does not follow the traditional approach in permutation tests. Indeed, (the residuals of) the observations are permuted, instead of permuting within the subjects (or blocks). A more traditional approach would be to split the design in strata (corresponding in traditional ANOVA to one between- and (depending on the number of within-subject factors) several within-subject sums of squares (Kirk 1994). We believe that it is worth going beyond the block approach for the following reasons. First, if desired, one could additionally restrict the above permutations to occur only within the subjects, leading to a restricted permutation method. Second, note that this splitting corresponds geometrically to apply several projections that are very similar the above projection. So applying the additional restriction leads to a method that somehow gets back to the splitting. Third, in the mixed model literature, this approach by observation is well accepted and several bootstrap or permutation methods are often done this way (Field and Welsh 2007). Finally, and most importantly, the splitting approach relies heavily on the orthogonality of the subspaces, i.e. on the requirement that the design is fully balanced. Even with a single missing value, the factors do not belong any more to a single stratum. It is therefore not suited to provide a method for unbalanced data, which is our aim in the next section.

5 Permutation of the reduced residuals for unbalanced design

Since the design is not balanced, the matrices X_1, X_2, Z_1 and Z_2 are not orthogonal any more. Therefore the matrix form is slightly more complex. For example, if we use $y_{rr}^{(1)}$ as the reduced residuals, then

$$y_{rr} = X_{2rr}\beta_2 + Z_{1rr}\gamma_1 + Z_{2rr}\gamma_2 + \epsilon_{rr}, \tag{11}$$

where $X_{2rr} = (I - H_{[X_1]})X_2$, $Z_{1rr} = (I - H_{[X_1]})Z_1$ and $Z_{2rr} = (I - H_{[X_1]})Z_2$.

In the case of unbalanced designs with random effects, there is no general and exact F-statistic. In many statistical softwares like SPSS, when there are missing observations, they automatically delete all the observations from the subject(s) with at least one missing value. However this might remove an important proportion of subjects and lead to a comparative loss of power. Alternatively, some F-statistics have been proposed in the unbalanced designs with more than one random effects, e.g. a quasi-F statistic Myers and Well (2003). Another technique is to use so-called weighted or unweighted means for unbalanced designs with random effects with only between-subject effects. For example Hirotsu (1979) proposed an approximate F-test to test the main effects and the interactions in unbalanced ANOVA with random effects, using the test statistics corresponding to those in the balanced case, but where the mean squares are obtained with the unweighted mean approach (Sahai and Ageel (2000)). However, there is no consensus on the choice of the statistic for unbalanced mixed-model designs. We propose to use a statistic that is very close to the balanced case [(compare with Eqs. (5) and (6)].

$$F_{ss} = \frac{\left[\mathsf{RSS}(\hat{\beta}_1) - \mathsf{RSS}(\hat{\beta}_1, \hat{\beta}_2)\right]/p_2}{\left[\mathsf{RSS}(\hat{\beta}_1) - \mathsf{RSS}(\hat{\beta}_1, \hat{\gamma}_2)\right]/q_2} = \frac{y'H_{[X_{2rr}]}y/p_2}{y'H_{[Z_{2rr}]}y/q_2}, \tag{12}$$

where $H_{[X_{2rr}]} = X_{2rr}(X'_{2rr}X_{2rr})^{-1}X'_{2rr}$ and $H_{[Z_{2rr}]}$ is defined as $H_{[Z_{2rr}]} = Z_{2rr}(Z'_{2rr}Z_{2rr})^{-1}Z'_{2rr}$. As in the balanced case, we define the reduced residuals and their corresponding statistics as

$$y_{rr} = \left(I - H_{[X_1]}\right)y \quad \text{and} \quad F_{rr} = \frac{y'_{rr}H_{[X_{2rr}]}y_{rr}/p_2}{y'_{rr}H_{[Z_{2rr}]}y_{rr}/q_2}. \tag{13}$$

Then by freely permuting the elements of y_{rr} and replacing in Eq. (13), like in Algorithm 1, we obtain the permuted statistics and the corresponding p-value for testing the parameter β_2. Note that for the unbalanced case, it is also correct that under the null hypothesis, $y_{rr} \sim (0, \Sigma_{rr})$.

Theorem 1 *For the original observations, the statistics defined in* (13) *and* (12) *are equivalent, i.e.* $F_{rr} = F_{ss}$.

The proof of the theorem is in the supplementary material. Note that this theorem does *not* state that the null distribution of these statistics are equal. It can be mentioned that other forms of residuals might be defined. For instance, $y_{rr} = (I - H_{[X_1,Z_1,\epsilon]})y$ and corresponding $X_{rr} = (I - H_{[X_1,Z_1,\epsilon]})X_2$ or $y_{rr} = (I - H_{[X_1,Z_1]})y$ and corresponding $X_{rr} = (I - H_{[X_1,Z_1]})X_2$ would still satisfy the above Theorem.

6 Permutation of the modified residuals for balanced designs

This method was introduced by Jung et al. (2006) in balanced ANOVA and then by Kherad-Pajouh and Renaud (2010) in general balanced and unbalanced ANOVA with a single error term and in regression settings. Here we extend this method for mixed-model and repeated measures ANOVA with more than one error term in the balanced case.

The aim of introducing the residuals under the modified model or in brief *modified residuals* is to remove the correlation between the reduced residuals, in order to achieve residuals that are exchangeable up to the second moment. It would follow that if the original data have a spherical pdf, the test based on the modified residuals is exact, see Theorems 3 and 4 in Kherad-Pajouh and Renaud (2010). In other words we want to transform Σ_{rr} to an identity matrix.

Consider the reduced residuals of $y_{rr}^{(2)} = H_{nd}y$ based on model (8). The covariance matrix of y_{rr} is $\Sigma_{rr} = H_{nd}\Sigma H_{nd}$. Since H_{nd} is a projection matrix, the eigen-decomposition of Σ_{rr} can be written as $U^0\Lambda_0 U^{0'} + U\Lambda_{rr}U' = \Sigma_{rr}$, where U^0 contains orthonormal eigen-vectors corresponding to the zero eigen-value, and Λ_0 is a diagonal matrix of the zero eigen-value, or zero matrix. Thus

$$U\Lambda_{rr}U' = \Sigma_{rr}, \tag{14}$$

where U contains orthonormal eigen-vectors corresponding to the non zero eigen-values and Λ_{rr} is a diagonal matrix of the non zero eigen-values. Also by multiplying U' and U in both sides of Eq. (14), using the orthonormality of the columns of U, we get

$$U'\Sigma_{rr}U = \Lambda_{rr}. \tag{15}$$

Let us construct the matrix $V = U\Lambda_{rr}^{-\frac{1}{2}}$.

The matrix V has the same dimension $N \times (p_2 + q_2)$ as the orthonormal matrix U which is related to the geometrical multiplicities of non zero eigen-values of Σ_{rr} (recall that p_2 and q_2 are the number of columns of X_2 and Z_2). In the two following sections, for the two special cases of repeated measures (with one or several within-subject factor(s)) and for mixed-model design (with one within-subject factor and one between-subject factor), we will show that $\Sigma_{rr} = \lambda H_{nd}$ where λ is the value of the only non-zero eigen-value of the matrix Σ_{rr}. We also compute the exact value of λ and show that its multiplicity is equal to $(p_2 + q_2)$. The results can be extended to more general designs as well.

In the case that the reduced residuals are based on H_{nd}, the non-zero eigen-values, or the diagonal elements of Λ, are thus all equal to a unique value, say λ. We propose to use $y_{rr}^{(2)}$, because of this property, since for other forms of y_{rr}, the eigen-values may differ. We can thus write $\Lambda_{rr} = \lambda I_{(p_2+q_2)}$. In this case, the V matrix satisfies the following:

$$V'V = \frac{1}{\lambda}I_{(p_2+q_2)} \quad \text{and} \quad VV' = \lambda H_{nd} \tag{16}$$

The value of λ depends on the design and on several parameters, see the two next Sections. To obtain the modified residuals, we pre-multiply both sides of Eq. (8) by V':

$$y_{mr} = X_{mr}\beta_2 + Z_{mr}\gamma_2 + \epsilon_{mr},$$

where $y_{mr} = V'y_{rr}$ are the residuals under the modified model, or modified residuals, with dimension (p_2+q_2), $X_{mr} = V'X_2$, $Z_{mr} = V'Z_2$ and $\epsilon_{mr} = V'\epsilon_{rr}$. Using matrix V and Eq. (15), the covariance matrix of y_{mr} can be obtained:

$$\text{cov}(y_{mr}) = \text{cov}(\Lambda_{rr}^{-\frac{1}{2}}U'y_{rr}) = \Lambda_{rr}^{-\frac{1}{2}}U'\Sigma_{rr}U\Lambda_{rr}^{-\frac{1}{2}} = I_{(p_2+q_2)}, \tag{17}$$

which implies that under the null, the distribution of y_{mr} is $y_{mr} \sim (0, I_{(p_2+q_2)})$.

Based on the modified residuals, the F-statistic to test of β_2 for the original y_{mr}, can be defined as the following:

$$F_{mr} = \frac{y'_{mr}H_{[X_{mr}]}y_{mr}/p_2}{y'_{mr}H_{[Z_{mr}]}y_{mr}/q_2}. \tag{18}$$

Again, by freely permuting the y_{mr}, we get y^*_{mr} and compute its associated statistic

$$F^*_{mr} = \frac{y^{*\prime}_{mr}H_{[X_{mr}]}y^*_{mr}/p_2}{y^{*\prime}_{mr}H_{[Z_{mr}]}y^*_{mr}/q_2}.$$

We can use the same procedure as Algorithm 1 using F_{mr} to make inference about β_2. By the following theorem, we will show the property of modified residuals, which keeps numerator and denominator of F-statistics for the original modified residuals, equals to the original F-statistics, even though the dimension of y_{mr} and X_{mr} is reduced.

Theorem 2 *For the original observations y, the statistics defined in* Eqs. (4), (7) *and* (18) *are equal.*

$$F_{ss} = F_{rr} = F_{mr}. \tag{19}$$

The proof is in the supplementary material.

7 Application for repeated measures ANOVAs with one or more within-subject factor(s)

First, we consider a simple case of Eq. (1), where there is only one within-subject factor for balanced design. In such situation, the observations can be modelled as

$$y_{ijk} = \mu + \eta_i + \pi_j + (\eta\pi)_{ij} + \epsilon_{ijk}, \qquad i = 1, \dots, b, \qquad j = 1, \dots, s,$$
$$k = 1, \dots, n.$$

This model based on (3) can be expressed in matrix form as

$$y = \mathbb{1}_{N\times 1}\mu + X_\eta\eta + X_\pi\pi + X_{\eta\pi}\eta\pi + \epsilon, \tag{20}$$

where $\mathbb{1}_{n\times m}$ is a $n \times m$ matrix with all elements equal to 1. To test the effect of the within-subject factor B, the hypotheses are

$$\mathsf{H}_0^B : \eta_1 = \eta_2 = \cdots = \eta_b = 0 \text{ vs. } \mathsf{H}_1^B : \text{ not } \mathsf{H}_0^B.$$

Therefore, the working matrices are $X_1 = \mathbb{1}$, $X_2 = X_\eta$ (effect of factor B), $Z_1 = X_\pi$ (subject effect S) and $Z_2 = X_{\eta\pi}$ (interaction effect of BS). More details on these matrices are given in the proof of Theorem 3. Note that y is distributed according to $y \sim (\mathbb{1}\mu + X_2\eta, \Sigma)$, where Σ is the covariance matrix of y. The random parts are based on $\pi \sim (0, \sigma_\pi^2 I_s)$, $(\eta\pi) \sim (0, \text{cov}(\eta\pi))$, and $\epsilon \sim (0, \sigma_\epsilon^2 I_{bsn})$.

The F-statistic to test the within-subject factor B can be written in the well known form using the mean squares as $F_{ss}^B = \frac{\text{MS}_B}{\text{MS}_{BS}}$, or in matrix form as $F_{ss}^B = \frac{y' H_{[X_\eta]} y/(b-1)}{y' H_{[X_{\eta\pi}]} y/(b-1)(s-1)}$, where $H_{[X_\eta]}$ and $H_{[X_{\eta\pi}]}$ are the projection matrices of the sigma-restricted design matrices X_η and $X_{\eta\pi}$.

In order to apply the reduced residuals method, we can use any of two kinds of reduced residuals introduced after Lemma 1 for example $y_{rr}^{(1)} = (I - H_{[\mathbb{1}]})y$ or $y_{rr}^{(2)} = (H_{[X_\eta]} + H_{[X_{\eta\pi}]})y$, and the corresponding F-statistics based on any of these reduced residuals is

$$F_{rr}^{B} = \frac{y_{rr}' H_{[X_\eta]} y_{rr}/(b-1)}{y_{rr}' H_{[X_{\eta\pi}]} y_{rr}/(b-1)(s-1)}. \tag{21}$$

In order to use the modified residuals method, we will only use $y_{rr}^{(2)}$ and in the following section we will show how to construct the covariance matrix and to apply the permutation of the modified residuals method.

7.1 Construction of the covariance matrix of the observations

To use a permutation test based on the modified residuals, we need to find the eigen-values of $\Sigma_{rr} = H_{nd} \Sigma H_{nd}$ to construct the V matrix. It is easy to explicitly construct H_{nd} and a little more complex to find the matrix Σ. To simplify the computation, the construction of the design matrices for the projection matrices are based on ortho-normal contrasts (e.g. Helmert contrasts), which is a special case of sigma-restriction parametrization where all columns are orthonormal. We explain these contrasts in detail in the proofs.

However for construction of covariance matrices, it is much easier to work with design matrices that are over-parametrized, which typically contain only values of 0 and 1. To distinguish between these two parametrizations, we use the notation of X^o for over parametrized design matrices. For more detail about differences between these two kinds of design matrices, we refer to Cardinal and Aitken (2006). The covariance matrix of y which is based on random part of model (20), is constructed based on over-parametrized model and can be written as

$$\Sigma = \sigma_\pi^2 X_\pi^{o'} X_\pi^o + X_{\eta\pi}^{o'} \text{cov}(\eta\pi) X_{\eta\pi}^o + \sigma_\epsilon^2 I_{bsn}, \tag{22}$$

where $X_\pi^{o'} = I_s \otimes \mathbb{1}_{b\times 1} \otimes \mathbb{1}_{n\times 1}$. Note that in repeated measures designs, the constraint $\sum_i (\eta\pi)_{ij} = 0, \forall j$ implies a correlation between the interaction terms $(\eta\pi)_{ij}$. A natural way of keeping such dependency, is to write $(\eta\pi)_{ij} = \omega_{ij} - \frac{1}{b} \sum_{i=1}^{b} \omega_{ij}$. In this case ω_{ij} are independent and identically distributed random variable with $(\omega_{ij}) \sim (0, \sigma_\omega^2)$, where $\sigma_\omega^2 = \frac{b}{b-1} \sigma_{\eta\pi}^2$. This allows us to write $\text{cov}(\eta\pi) = (I_s \otimes \Delta_b) \sigma_\omega^2$ where $\Delta_b = I_b - \frac{1}{b} \mathbb{1}_{b\times b}$. Finally, $X_{\eta\pi}^o = I_s \otimes I_b \otimes \mathbb{1}_{n\times 1}$ and $X_{\eta\pi}^{o'} \text{cov}(\eta\pi) X_{\eta\pi}^o = \sigma_\omega^2 (I_s \otimes \Delta_b \otimes \mathbb{1}_{n\times n})$.

7.2 Eigen-values and Eigen-vectors of Σ_{rr}

Theorem 3 *In the one within-subject factor with possible repetitions in each cell, let $\Sigma_{rr} = H_{nd} \Sigma H_{nd}$, and $H_{nd} = H_{[X_\eta]} + H_{[X_{\eta\pi}]}$, where Σ is defined on Eq. (22). Then $\Sigma_{rr} = \lambda H_{nd}$, and the only non-zero eigen-value of Σ_{rr} is $\lambda = n\sigma_\omega^2 + \sigma_\epsilon^2$ and have multiplicity $(b-1)s$.*

The proof of the theorem is in the supplementary material.

7.3 Estimation of the variance components and the eigen-values in non-oracle designs

To apply the method of permutation of the modified residuals, we need to construct $y_{mr} = V'y_{rr}$, thus V, and therefore we must estimate λ. The value of λ, given on Theorem 3, is based on σ_ω^2 and σ_ϵ^2. Theorem 3 also shows that the columns of H_{nd} are eigen-vectors of Σ_{rr}, and it follows that the columns of X_η and $X_{\eta\pi}$ are the eigen-vectors of Σ_{rr}. The following lemma proposes an estimation for λ.

Lemma 2 *The parameter λ in Theorem 3 can be estimated from*

$$\hat{\lambda} = n v_1 - n v_3,$$

where

$$v_1 = \frac{1}{N} \sum_{ijk} y_{rr}(ijk)^2, \quad v_3 = \frac{1}{n^2 b(b-1)s} \sum_j \sum_{i \neq i'} \sum_{k,k'} y_{rr}(ijk) y_{rr}(i'jk').$$

This estimator is unbiased, that is, $\mathbb{E}[\hat{\lambda}] = \lambda$.

We present the proof of this lemma in the supplementary material.

7.4 Repeated measures with more than one within-subject factor

In the case with two within-subject factors, the model is

$$y_{iljk} = \mu + \eta_i + \gamma_l + \pi_j + (\eta\pi)_{ij} + (\gamma\pi)_{lj} + (\eta\gamma)_{il} + (\eta\gamma\pi)_{ilj} + \epsilon_{iljk},$$
$$i = 1, \ldots, b, \ l = 1, \ldots, a, \ j = 1, \ldots, s, \text{ and } k = 1, \ldots, n.$$

In this model η and γ are fixed effects and both are within-subject factors and π is the random effect due to the subjects. We construct the covariance matrix of Σ_{rr}, based on the constrains which are on the random part of the model, which are: $\sum_{i=1}(\eta\pi)_{ij} = 0, \forall j$ and $\sum_{l=1}(\gamma\pi)_{lj} = 0, \forall j$ and for the three way interactions, $\sum_i (\eta\gamma\pi)_{ilj} = 0 \ \forall l, j$ and $\sum_l (\eta\gamma\pi)_{ilj} = 0 \ \forall i, j$. The covariance matrix for this model can thus be written as

$$\Sigma_{rr} = H_{nd}(\sigma_\pi^2 X_\pi^{o'} X_\pi^o + X_{\eta\pi}^{o'} \text{cov}(\eta\pi) X_{\eta\pi}^o + X_{\gamma\pi}^{o'} \text{cov}(\gamma\pi) X_{\gamma\pi}^o$$
$$+ \ X_{\eta\gamma\pi}^{o'} \text{cov}(\eta\pi\gamma) X_{\eta\gamma\pi}^o + \sigma_\epsilon^2 I_N) H_{nd}. \tag{23}$$

H_{nd} depends on the selected parameter to be tested. We skip the details, as they are similar to the case with one within-subject factor. However, the decomposition of Σ_{rr} can be found in Kherad-Pajouh (2011) and the main result is shown in the following theorem.

Theorem 4 *In the two within-subject factors design with possible repetitions in each cell, if we are interested to test one of the within-subject factor, say η, let $H_{nd} = H_{[X_\eta]} + H_{[X_{\eta\pi}]}$ and Σ_{rr} defined as in Eq.* (23). *Then $\Sigma_{rr} = \lambda H_{nd}$ and the only non-zero eigen-value of Σ_{rr} is $\lambda = na\sigma_\omega^2 + \sigma_\epsilon^2$ and have multiplicity of $(b-1)s$.*

The proof of the theorem is in the supplementary material. We skip the estimation of λ in this part, as it can be achieved in a similar way as Lemma 2. Similarly, the proof of this Theorem can be generalized to more than two within-subject factors with the only difficulty of having a more complex notation.

8 Application for mixed-model designs

In this section we consider the mixed-model design with one between-subject and one within-subject factor.

$$y_{lijk} = \mu + \alpha_l + \pi_{j/l} + \eta_i + (\alpha\eta)_{li} + \pi\eta_{ji/l} + \epsilon_{lijk}$$
$$l = 1, \ldots, a, \ i = 1, \ldots, b, \ j = 1, \ldots, s, \text{ and}, k = 1, \ldots, n. \tag{24}$$

where α is the between-subject factor, η is the within-subject factor, π is the effect of subjects, which is random, and s is the number of subjects in each levels of between-subject factor. There is obviously no interaction for the subject and the between-subject factor, $\alpha\eta$ is the fixed interaction part, and $\pi\eta$ in the random interaction part. In the parametric approach the F-statistic for a within-subject factor is similar to the one for the repeated measures design and the F-statistic for between-subject factor of A can be written as

$$F_{ss}^A = \frac{\text{MS}_A}{\text{MS}_S} = \frac{y'H_{[X_\alpha]}y/(a-1)}{y'H_{[X_\pi]}y/(s-1)}.$$

The F-statistic for the interaction of within-subject factor and between-subject factor can be written as

$$F_{ss}^{AB} = \frac{\text{MS}_{AB}}{\text{MS}_{ABS}} \quad \text{and} \quad F_{ss}^{AB} = \frac{y'H_{[X_{\alpha\eta}]}y/(a-1)(b-1)}{y'H_{[X_{\alpha\eta\pi}]}y/a(s-1)(b-1)} \tag{25}$$

In order to provide a permutation test based on the reduced residuals, the corresponding statistics F_{rr} should be used. This can be achieved by replacing y_{rr} in the above formulae (and in the unbalanced case, one should also replace $X_{\eta rr}$ and $X_{\alpha\eta rr}$). We skip the details for this method, since it is very similar to the previous cases.

Concerning the definition of the modified residuals in mixed-model designs, we should first construct the covariance matrix of y_{rr} and then find its eigen structure. As mentioned before, the covariance matrix of y is only based on the random part of model (24). Using similar notation as in previous sections, the covariance matrix of the observations can be written as

$$\Sigma = \sigma_\pi^2 X_\pi^{o'} X_\pi^o + X_{\eta\pi}^{o'} \text{cov}(\eta\pi) X_{\eta\pi}^o + \sigma_\epsilon^2 I_{absn}. \tag{26}$$

and then, the covariance matrix of y_{rr} can be written as $\Sigma_{rr} = H_{nd} \Sigma H_{nd}$, where H_{nd} is dependent on the testing parameters.

Theorem 5 *In the mixed-model design with one within- and one between-subject effects and in order to test the between-subject factor, we have $H_{nd} = H_{[X_\alpha]} + H_{[X_\pi]}$ and Σ is based on Eq. (26). Then $\Sigma_{rr} = \lambda H_{nd}$ and the only non-zero eigen-value is $\lambda = nb\sigma_\pi^2 + \sigma_\epsilon^2$ with multiplicity $(a-1) + (s-1)$.*

Theorem 6 *In the mixed-model design, corresponding to (24), in order to test the within-subject factor, we have $H_{nd} = H_{[X_\eta]} + H_{[X_{\eta\pi}]}$ and $\Sigma_{rr} = H_{nd} \Sigma H_{nd}$, where Σ is based on (26). Then $\Sigma_{rr} = \lambda H_{nd}$ and the only non-zero eigen-value is $\lambda = n\sigma_\omega^2 + \sigma_\epsilon^2$, with multiplicity $a(s-1)(b-1) + (b-1)$.*

The proofs of the two above theorems can be found in the supplementary material. We skip the estimation of λ, as it can be achieved in a similar way as in Lemma 2. The proof of these Theorems can be generalized to designs with more factors with the only difficulty of having a more complex notation.

9 Simulation study

In this section, simulation studies are carried out to validate and compare the level and power of the four different methods mentioned in this article (called below Y_{mr}, Y_{rr}, Y and F-test). The permutation of the reduced residuals, based on $y_{rr}^{(2)} = H_{nd}y$ is named Y_{rr}. We compare it to the permutation of raw data, which is based on freely permuting the (raw) data, and using the regular F-statistic given in (6). We also provide the results for the parametric F-test, which is based on Gaussian assumption (and the split-plot approach for mixed-model designs) and uses the F distribution. Of course, this is not a permutation test, but provides the results for the classical approach. Finally, by Y_{mr} we mean permutation of the modified residuals: $y_{mr} = V'y_{rr}$. For this method, it is important to separate out the intrinsic quality of the method with a possible loss of quality (both for the level and the power) due to the estimation of the parameter λ. This leads us to have two modified residuals procedures: an "oracle" case, that we call Y_{mr}^{or}, for which the population value of λ is supposed to be known, and a "non-oracle" case, that we call Y_{mr}^{n-or}, in which we have to estimate λ from the data. Although the oracle case is not a genuine method, it is an important benchmark, both for the non-oracle case, and for the permutation of the reduced residuals (named Y_{rr}). Indeed, if the latter have a similar level and power than the oracle Y_{mr}, this would indicate that the reduced residuals are adequate and that there is no need to try to use the modified residuals or to improve the estimation of the unknown λ.

9.1 Repeated measures ANOVA

To explain how the data are generated, let's first consider the repeated measure simulations. They are based on model (20). In this model, μ and η are the fixed parts. We fixed μ as a vector with all elements equal to 1 (this value has no influence on

the results). The test will concern the fixed parameters η_i. In order to simulate under the null hypothesis and under several alternatives, we consider the variation measure $\theta = \left(\sum_i^b (\eta_i^{\mathrm{orig}})^2 / b \right)^{-1/2}$. In the simulations, the values of η_i that will be used are $(\eta_i^{\mathrm{orig}}) \cdot t / \theta$, where t is a constant that modifies the amplitude of the effects. For instance for estimation of the significance level, t is equal to zero, which makes all the η_i equal to zero, and the power is obtained with values $t > 0$. Concerning the random parts of the model (20), the ones due to π and $\eta\pi$ are generated according to a normal distribution, with the following variances: $\sigma_\pi^2 = 0.2$, $\sigma_{\eta\pi}^2 = 0.5$.

Finally the distribution of the error part is either $\mathcal{N}(0, 1)$, or $U(-\sqrt{3}, \sqrt{3})$ or $\exp(1) - 1$, which all have also a variance equal to 1, i.e. $\sigma_\epsilon^2 = 1$. We also generated data sets from a Generalized Pareto $GP(0.4814, 0.01, 1)$, a more heavy-tailed distribution, with shape parameter 0.4814 and scale parameter 0.01 in order that error terms have variance equal 1. Note that only the errors include skewed or non-normal distributions, not the subject effects.

To construct the modified residuals (Y_{mr}^{or}), we generated the covariance matrix according to the variances for different random parts of the model and based on Eq. (22). For the modified residuals in non-oracle case (Y_{mr}^{n-or}), we additionally estimated λ.

In Table 1, we generate data for the model which is typical for psychological data. It is constituted by ten subjects $s = 10$, one within-subject factor with four levels $b = 4$ and one observation per cell $k = 1$. The original values for η are chosen as $\eta^{\mathrm{orig}} = [2, -2, 4, -4]$ and satisfy the sigma-restriction constraint. In Table 2, an unbalanced repeated measures design is considered with $s = 3$ subjects . The null hypothesis concerns the effect of η with three levels $b = 3$. The original values are $\eta^{\mathrm{orig}} = [4, -2, -2]$. There are two repetitions in each cell except for the last cell with only one observation. In this model, since the parametric F is not defined for this case and the only proposed method is the permutation of the reduced residuals, we only consider this method in this simulation.

9.2 Mixed-model design

In the Tables 3 and 4, the simulation study is related to a mixed-model design as in model (24) with one within-subject factor with three levels $b = 3$, and one between-subject factor with two levels $a = 2$. There are three subjects in each level of between-subject factor and one repetition per cell. The null hypothesis concerning the within-subject factor $H_0 : \eta_1 = \eta_2 = \eta_3$ is tested in Table 3, and the null hypothesis concerning the between-subject factor $H_0 : \alpha_1 = \alpha_2$ is found on Table 4.

The original values for $\eta^{\mathrm{orig}} = [3, -2, -1]$ and for $\alpha^{\mathrm{orig}} = [1, -1]$. As in the previous case, in order to estimate the level and power of each method, we used the variation measure θ. The variance of the random part of the model are $\sigma_\pi^2 = 0.5$, $\sigma_{\eta\pi}^2 = 0.2$ and $\sigma_\epsilon^2 = 1$. We used the same covariance matrix of the random part as in the previous case, since only the within-subject factor has interaction with the random effect of the subjects, so adding a between-subject factor does not change the random structure.

Table 1 Simulation results for a balanced repeated measures ANOVA with one within-subject factor B, with $b = 4$ levels, $s = 10$ subjects and $k = 1$ repetition per cell. Proportion of rejection to test the main effect based on 5000 replications. The column with $t = 0$ corresponds to the level, and the other ones to power with increasing effect sizes

	Stat.	$t = 0$	$t = 0.3$	$t = 0.6$	$t = 0.9$	$t = 1.2$	$t = 1.5$
$\mathcal{N}(0, 1)$	Y_{mr}^{or}	0.050	0.134	0.305	0.642	0.877	1.000
	Y_{mr}^{n-or}	0.049	0.113	0.289	0.603	0.844	1.000
	Y_{rr}	0.048	0.129	0.284	0.605	0.860	1.000
	Y	0.049	0.122	0.273	0.604	0.848	1.000
	F-test	0.048	0.137	0.302	0.632	0.861	1.000
$U(-\sqrt{3}, \sqrt{3})$	Y_{mr}^{or}	0.051	0.144	0.285	0.682	0.877	1.000
	Y_{mr}^{n-or}	0.048	0.113	0.229	0.563	0.794	1.000
	Y_{rr}	0.048	0.133	0.254	0.655	0.850	1.000
	Y	0.046	0.112	0.223	0.554	0.778	0.998
	F-test	0.050	0.137	0.272	0.662	0.856	1.000
$\exp(1) - 1$	Y_{mr}^{or}	0.049	0.122	0.323	0.586	0.871	0.990
	Y_{mr}^{n-or}	0.048	0.108	0.294	0.559	0.816	0.998
	Y_{rr}	0.047	0.1129	0.308	0.578	0.853	0.999
	Y	0.045	0.092	0.258	0.525	0.751	0.999
	F-test	0.043	0.097	0.254	0.514	0.757	0.989
$GP(0.4814, 0.01, 1)$	Y_{mr}^{or}	0.049	0.096	0.259	0.529	0.829	1.000
	Y_{mr}^{n-or}	0.050	0.090	0.235	0.495	0.804	1.000
	Y_{rr}	0.050	0.103	0.288	0.583	0.883	1.000
	Y	0.0481	0.095	0.268	0.535	0.751	1.000
	F-test	0.047	0.089	0.254	0.522	0.737	1.000

Table 2 Simulation results for an unbalanced repeated measures ANOVA with one within-subject factor B with $b = 3$ levels, $s = 3$ subjects and $k = 2$ repetitions in each cell except last cell with one observation. Proportion of rejection to test the main effect based on 5000 replications. The column with $t = 0$ corresponds to the level, and the other ones to power with increasing effect sizes

	Stat.	$t = 0$	$t = 0.3$	$t = 0.6$	$t = 0.9$	$t = 1.2$	$t = 1.5$
$\mathcal{N}(0, 1)$	Y_{rr}	0.050	0.134	0.305	0.642	0.877	1.000
$U(-\sqrt{3}, \sqrt{3})$	Y_{rr}	0.049	0.113	0.289	0.603	0.844	1.000
$\exp(1) - 1$	Y_{rr}	0.048	0.129	0.284	0.605	0.860	1.000
$GP(0.4814, 0.01, 1)$	Y_{rr}	0.049	0.113	0.263	0.588	0.845	1.000

In all simulations we used permutation tests based on 1000 permutations. For all settings, we used 5000 Monte-Carlo replications.

10 Discussion

In all the settings of the present simulations, as expected, the type I error is very close to the nominal level of 0.05, and the power of the method Y_{mr}^{or} is relatively higher

Table 3 Simulation results for a balanced mixed-model ANOVA with one within-subject factor B with $b = 3$ levels and one between-subject factor A with $a = 2$ levels, $s = 3$ subjects and $k = 1$ repetition in each cell. Proportion of rejection to test the main effect of the within-subject factor B based on 5000 replications. The column with $t = 0$ corresponds to the level, and the other ones to power with increasing effect sizes

	Stat.	$t = 0$	$t = 0.3$	$t = 0.6$	$t = 0.9$	$t = 1.2$	$t = 1.5$
$\mathcal{N}(0, 1)$	Y_{mr}	0.051	0.118	0.308	0.598	0.846	0.955
	Y_{rr}	0.052	0.105	0.268	0.557	0.812	0.944
	Y	0.054	0.107	0.275	0.569	0.824	0.945
	F-test	0.051	0.104	0.268	0.562	0.815	0.944
$U(-\sqrt{3}, \sqrt{3})$	Y_{mr}	0.049	0.099	0.291	0.563	0.834	0.971
	Y_{rr}	0.048	0.096	0.275	0.556	0.824	0.969
	Y	0.053	0.097	0.262	0.541	0.826	0.955
	F-test	0.052	0.095	0.261	0.531	0.807	0.949
$\exp(1) - 1$	Y_{mr}	0.049	0.114	0.333	0.611	0.849	0.950
	Y_{rr}	0.048	0.106	0.300	0.584	0.839	0.938
	Y	0.046	0.106	0.307	0.571	0.828	0.944
	F-test	0.046	0.109	0.302	0.576	0.820	0.940
$GP(0.4814, 0.01, 1)$	Y_{mr}	0.049	0.095	0.283	0.555	0.794	0.945
	Y_{rr}	0.050	0.116	0.323	0.589	0.845	0.988
	Y	0.048	0.104	0.304	0.571	0.828	0.964
	F-test	0.049	0.092	0.276	0.544	0.788	0.940

than the other permutation approaches. The power of Y_{mr}^{n-or}, as expected, is a little lower than Y_{mr}^{or}. Note that already with a medium-sized sample, the modified residuals methods and the approximate test Y_{rr} are comparable with the permutation of raw data and the parametric F, especially in the non Gaussian cases. Also based on these results, the permutation of the reduced residuals, which is an approximate test, has a relatively good level close to 0.05 and good power for all designs and factors tested. This result is in contrast with the results of Kherad-Pajouh and Renaud (2010), who found that for very small sample sizes the modified residuals outperform the reduced residuals in ANOVA with a single error term, i.e. with no repeated measures. This is probably due to the two following reasons. First, the number of observations is almost necessarily larger in repeated measures designs, since each subject has to be measured on each level. Second, the correlation remaining in the reduced residuals is probably less serious for repeated measures designs than for designs for independent observations since in the former correlation between observations are anyway present through the effect of the subjects. In the heavy tailed distribution of GP we still have the same results that all methods have almost level close to 0.05. The power of Y_{mr}^{n-or} is still lower than Y_{mr}^{or} and in general Y_{mr} has a little less power than the reduced residual model due to the error terms from heavy tailed distribution.

Given that the two residual methods give similar results, that the reduced residuals are less complex to obtain, that they have slightly higher power for the heavy tailed

Table 4 Simulation results for a balanced mixed-model ANOVA with one within-subject factor B with $b = 3$ levels and one between-subject factor A with $a = 2$ levels, $s = 3$ subjects and $k = 1$ repetition in each cell. Proportion of rejection to test the main effect of the between-subject factor A based on 5000 replications. The column with $t = 0$ corresponds to the level, and the other ones to power with increasing effect sizes

	Stat.	$t = 0$	$t = 0.3$	$t = 0.6$	$t = 0.9$	$t = 1.2$	$t = 1.5$
$\mathcal{N}(0, 1)$	Y_{mr}	0.049	0.119	0.306	0.596	0.827	0.912
	Y_{rr}	0.049	0.117	0.328	0.612	0.846	0.960
	Y	0.042	0.085	0.251	0.492	0.709	0.858
	F-test	0.051	0.118	0.328	0.618	0.848	0.962
$U(-\sqrt{3}, \sqrt{3})$	Y_{mr}	0.052	0.122	0.310	0.596	0.816	0.952
	Y_{rr}	0.053	0.131	0.326	0.606	0.859	0.955
	Y	0.048	0.103	0.253	0.479	0.709	0.841
	F-test	0.054	0.121	0.321	0.588	0.803	0.937
$\exp(1) - 1$	Y_{mr}	0.052	0.132	0.300	0.586	0.816	0.912
	Y_{rr}	0.054	0.131	0.326	0.606	0.849	0.955
	Y	0.048	0.103	0.253	0.479	0.709	0.841
	F-test	0.056	0.131	0.311	0.578	0.803	0.938
$GP(0.4814, 0.01, 1)$	Y_{mr}	0.050	0.090	0.293	0.566	0.808	0.935
	Y_{rr}	0.051	0.095	0.334	0.605	0.854	0.988
	Y	0.052	0.096	0.324	0.566	0.828	0.954
	F-test	0.051	0.096	0.300	0.557	0.798	0.930

distributions, and more importantly that the reduced residuals allow for unbalanced designs as well, we suggest to use the reduced residual methods.

Finally, as expected, the simulation results show that in cases where all random parts have normal distributions, the parametric F-test is correct. It is however important to note that the proposed permutation methods are quite close in terms of level and power even in this parametric settings. However, the permutation methods outperform the parametric F-test quite consistently when the error term is not normal, see e.g. Table 1, $\exp(1) - 1$ at $t = 0.6$, $t = 0.9$ and $t = 1.2$.

11 Conclusion

In this paper at first we introduced an approximate permutation test based on the reduced residuals for repeated measures and mixed-model designs. Based on this method, we can test any within-subject or between-subject factor, including the interactions of within- and between-subject factors. This method is extendable to the unbalanced designs with no empty cell as well. We also introduced a permutation test called modified residuals that is exact up to the second moment. It is based on a modification on the previous approximate test that removes the correlations between the residuals. This method is introduced for balanced repeated measure and mixed-model designs, and can be used to test any within-subject or between-subject factor. We also

provide three examples for which we go trough the details of the two proposed methods. Finally, we did simulations in order to compare the level and power of the proposed methods. The method based on the reduced residuals seems to be a good compromise between precision (close to nominal level), power, and complexity of the method. It is therefore advocated in practice, even as a diagnostic to check if the classical approach using a parametric test gives sensitive results.

References

Anderson M, Ter Braak C (2003) Permutation tests for multi-factorial analysis of variance. J Stat Comput Simul 73(2):85–113

Basso D (2009) Permutation tests for stochastic ordering and ANOVA: theory and applications with R. Springer, New York

Basso D, Salmaso L (2006) A discussion of permutation tests conditional to observed responses in unreplicated 2^m full factorial designs. Commun Stat Theory Methods 35(1):83–97. doi:10.1080/03610920500437277

Basso D, Chiarandini M, Salmaso L (2007) Synchronized permutation tests in replicated IxJ designs. J Stat Plan Inference 137(8):2564–2578. doi:10.1016/j.jspi.2006.04.016

Cardinal R, Aitken M (2006) ANOVA for the behavioural sciences researcher. Lawrence Erlbaum Associates, New Jersey

Cornfield J, Tukey JW (1956) Average values of mean squares in factorials. Ann Math Stat 27(4):907–949. doi:10.1214/aoms/1177728067

Field CA, Welsh A (2007) Bootstrapping clustered data. J Roy Stat Soc B 69:369–390

Fisher R (1935) The design of experiments. Oliver & Boyd, Edinburgh

Freedman D, Lane D (1983) A nonstochastic interpretation of reported significance levels. J Bus Econ Stat 1(4):292–298

Good P (1994) Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer, New York

Hirotsu C (1979) An F approximation and its application. Biometrika 66(3):577–584

IBM Corp (2013) IBM SPSS statistics for Windows, Version 22.0. Armonk, NY

Jung BC, Jhun M, Song SH (2006) A new random permutation test in ANOVA models. Stat Pap 48:47–62

Keppel G (1991) Design and analysis: a researcher's handbook. Prentice-Hall, Englewood Cliffs

Kherad-Pajouh S (2011) Permutation tests for experimental designs, with extension to simultaneous EEG signal analysis. PhD thesis, University of Geneva. url: http://archive-ouverte.unige.ch/unige:15743

Kherad-Pajouh S, Renaud O (2010) An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova. Comput Stat Data Anal 54(7):1881–1893. http://dx.doi.org/10.1016/j.csda.2010.02.015

Kirk R (1994) Experimental design: procedures for the behavioral sciences. Wadsworth Publishing Co Inc., Belmont

Mazzaro D, Pesarin F, Salmaso L (2001) Permutation tests for effects in unbalanced repeated measures factorial designs. In: Advances in model-oriented design and analysis: proceedings of the 6th international workshop on model-oriented data analysis. Springer, Berlin, p 183

Myers J, Well A (2003) Research design and statistical analysis. Lawrence Erlbaum, New Jersey

Pesarin F (2001) Multivariate permutation tests: with applications in biostatistics. Wiley, Chichester

Pesarin F, Salmaso L (2010) Permutation tests for complex data: theory, applications and software. Wiley, New York

Rouanet H, Lépine D (1970) Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. Br J Math Stat Psychol 23:147–163

Sahai H, Ageel M (2000) The analysis of variance: fixed, random, and mixed models. Birkhauser, Boston

Still A, White A (1981) The approximate randomization test as an alternative to the *F* test in analysis of variance. Br J Math Stat Psychol 34(2):243–252

White A, Harris P (1978) RANOVA: a fortran IV program for four-way Monte Carlo analysis of variance. Behav Res Methods 10(5):732