



Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

Exact Multivariate Permutation Tests for Fixed Effects in Mixed-Models

Dario Basso, Livio Finos
Department of Statistical Sciences
University of Padua
Italy

Abstract: A test for the fixed effect in mixed-models is proposed. It is based on permutation strategy and is exact. The testing approach presented is very general and the class of model covered is very broad.

Multivariate responses with different type of variables (e.g. continuous, categorical and ranks) are usually tested with separated models and the overall test are usually reached through Bonferroni-like combinations, i.e. without taking in account the joint distribution of the tests statistics. On the contrary in this approach the joint distribution is immediately obtained and the dependence among tests is taken in account in the overall test.

Keywords: Mixed Model, Permutation Tests

Contents

1	Introduction	1
2	The model	3
3	The test	5
4	Example	8
5	Conclusions	10

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
<http://www.stat.unipd.it>

Corresponding author:
Dario Basso
tel: +39 049 827 4151
dario@stat.unipd.it
<http://www.stat.unipd.it/~livio>

Exact Multivariate Permutation Tests for Fixed Effects in Mixed-Models

Dario Basso, Livio Finos

Department of Statistical Sciences

University of Padua

Italy

Abstract: A test for the fixed effect in mixed-models is proposed. It is based on permutation strategy and is exact. The testing approach presented is very general and the class of model covered is very broad.

Multivariate responses with different type of variables (e.g. continuous, categorical and ranks) are usually tested with separated models and the overall test are usually reached through Bonferroni-like combinations, i.e. without taking in account the joint distribution of the tests statistics. On the contrary in this approach the joint distribution is immediately obtained and the dependence among tests is taken in account in the overall test.

Keywords: Mixed Model, Permutation Tests

1 Introduction

Mixed models are an extension of regression models that allow for the incorporation of random effects. The random effects are usually adopted when the available observations are taken from a random sample of a population, and allow for peculiarities within subjects.

To describe the kind of test considered in this work we use (part of the data) collected by [4] in an experiment of Visual Search for Faces among Objects in Complex Visual Displays. Here twelve adults were recorded under two possible viewing conditions (one face among four/six objects). Each subject was subjected to thirty-two randomly selected images, half for each viewing condition. Two response variable were considered among many others: *first fixation* (Primax) a dichotomous variable that records whether the face is fixed at first sight or not, and the *proportion of face fixation among total fixation time* (Prop), i.e. the proportion of the total trial-time spent looking at the face. The aim of the study was to compare the performances of the subjects under the two experimental conditions.

This paper investigates the problem of testing the within subject fixed effects in a multivariate mixed model. To illustrate the method we have chosen the example above because the within subject two-levels factor is a very simple model, but in general any number and kind of predictors can be used to model the multivariate response. By ‘within’ we mean the effect of an explicative variable inside one sample, whereas we refer to the multi-sample comparison as the ‘between’ effect.

We focus on the within effects although a preliminar proposal to test for the between group effects has been recently introduced in [2] and [3], but this testing strategy is not a goal of this paper.

This is an example of stratified bivariate problem with a continuous (Prop) and a dichotomous (Primafix) variable. Despite the multivariate problem, in the parametric mixed-model framework it is usually assumed that the total variance of the response is given by the sum of the variance of the random effect and that of the error term, which is usually assumed to be common, i.e. the error components are assumed to be i.i.d. random variables. Even with this strong assumption, if one considers a location parameter estimator such as the subject mean under each experimental condition, its variability depends on the number of trials, so in our example heteroscedasticity would arise again a continuous variable, such as Prop. Moreover, if we focus on the dichotomous variable, then the variance of the subject's means is itself a function of an unknown parameter (the true proportion of time the subject looks at the face), even if the experiment is balanced, i.e. when all the subject have the same amount of trials.

The widely most used approach in most applied filed is to separately model and test the two responses in an univariate way, later summarizing the two inferences in a global p-value. The univariate models used are usually parametric and MLE approach is used to make inference on the parameter estimates. For linear mixed models the Wald-type test on a single effect rely on a t -student distribution with unknown degrees of freedom. Therefore, even in relatively simple models, inference is available only asymptotically. Also in the more general context of generalized mixed models the inference are limited to an approximate level. When parameter estimation is performed through Full (i.e. not Restricted) Maximum Likelihood approach, LRT approach is available. Even in this case the test controls the type I error only asymptotically and the inference is limited to the two-sided alternative. As an alternative, a p-value can be obtained trough MCMC methods [Mixed-effects modeling with crossed random effects for subjects and items [1]. In this case the prior (multivariate) distribution for all error terms are assumed to be well known and the computational (i.e. time) costs are elevate.

The multivariate inference, that usually combines the p-values of the univariate models in a overall test, shows some weakness as well. The univariate tests cannot be assumed to be independent and the strategy usually available is the Bonferroni correction which is known to be very consevative when the correlation among p-values is strong. An adequate multivariate extention seems to be not finalixed yet for neither the MLE methods nor for the MCMC approach.

A relevant results aimed to overcome part of these weakness is due to [6]. Here we restate the approach in a fully nonparametric way providing an exact multivariate test that is suitable for more general models, that allows for directional alternatives and efficiently combines univariate p-values for a multivariate inference. This is the aim of this paper, which is organized as follows: in Section 2 we introduce the model assumptions and discuss some examples; Sections 3 deals with the proposal, its theoretical justification and some possible extensions. Section 4 is devoted to a practical application and Section 5 to the discussion.

2 The model

In order to illustrate the model, recall that in the example above we have a multivariate response measured repeatedly on a set of N subjects (clusters) under two different conditions (treatments/predictors). The aim of the study is to assess whether the treatment has a significant effect on the response. For the moment, let's also reduce the description to an univariate response, we will discuss its extension to a multivariate framework at the end of the section.

We assume \mathbf{y}_i to be a $n_i \times 1$ vector of measurements such that:

$$\mathbf{y}_i | (\boldsymbol{\mu}_i, \mathbf{R}_i) \sim \phi(\boldsymbol{\mu}_i, \mathbf{R}_i) \quad i = 1, \dots, N; \quad (1)$$

where n_i is the number of observations (trials) on subject i , $\phi(\cdot)$ is a density/probability function,

$$\boldsymbol{\mu}_i = h(\mathbf{X}_i, \boldsymbol{\beta}_i) \quad (2)$$

and \mathbf{R}_i is the variance/covariance matrix of each subject's observations, with $\boldsymbol{\beta}_i$ being a $p \times 1$ vector of coefficients and \mathbf{X}_i a $n_i \times p$ matrix of (fixed) predictors. For instance, in the univariate usual mixed linear model, if normality assumptions hold (hence $\phi(\cdot)$ is the multivariate normal density), $\mathbf{y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$, with $\mathbf{R}_i = \sigma_i^2 \mathbf{I}_{n_i}$, and $\boldsymbol{\mu}_i = \mathbf{X}_i(\boldsymbol{\beta} + \mathbf{u}_i)$, where $\boldsymbol{\beta}$ is a vector of fixed terms, $\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u^2)$ and $COV(\boldsymbol{\varepsilon}_i, \mathbf{u}_i) = \mathbf{0}$. Here \mathbf{u}_i is a $p \times 1$ vector of random effects with variance/covariance matrix $\boldsymbol{\Sigma}_u^2$ that is common for all the subjects and it is assumed to be independent of $\boldsymbol{\varepsilon}_i$. We can write the whole model as:

$$\mathbf{y}_i = \mathbf{X}_i(\boldsymbol{\beta} + \mathbf{u}_i) + \boldsymbol{\varepsilon}_i,$$

so $E[\mathbf{y}_i] = \mathbf{X}_i\boldsymbol{\beta}$ and $V[\mathbf{y}_i] = \mathbf{X}_i V[\boldsymbol{\beta} + \mathbf{u}_i] \mathbf{X}_i' + V[\boldsymbol{\varepsilon}_i] = \mathbf{X}_i \boldsymbol{\Sigma}_u \mathbf{X}_i' + \mathbf{R}_i$.

This specification extends the classical mixed linear model allowing the \mathbf{R}_i to be different for each subjects, while usually the assumptions require $\mathbf{R}_i = \mathbf{R} \forall i$. Of course the definition in (1) and (2) comprise much more complex and general models, we will detail some of them later on.

Further we assume that \mathbf{t}_i is an estimator of $\boldsymbol{\beta}_i$ conditionally on each subject is available, so

$$\mathbf{t}_i | \mathbf{u}_i \sim (\boldsymbol{\beta} + \mathbf{u}_i, \boldsymbol{\Sigma}_i),$$

where $\boldsymbol{\Sigma}_i$ is the $p \times p$ variance/covariance matrix of \mathbf{t}_i within each subject. Note that $\boldsymbol{\Sigma}_i$ depends only on \mathbf{R}_i and on \mathbf{X}_i ; for the linear model defined above it holds $\boldsymbol{\Sigma}_i = \sigma_i^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}$. This part of the model describes the within-subject variability. Finally, we assume that

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i \sim (\boldsymbol{\beta}, \boldsymbol{\Sigma}_u),$$

that is the true vector of parameters is drawn from a common distribution for all subjects. This part of the model describes the between-subject variability due to the presence of random effects. Putting all together we have that, unconditionally,

$$\mathbf{t}_i \sim (\boldsymbol{\beta}, \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_i), \quad (3)$$

indeed $E[\mathbf{t}_i] = E[E(\mathbf{t}_i|\mathbf{u}_i)] = E[\boldsymbol{\beta} + \mathbf{u}_i] = \boldsymbol{\beta}$ and $V[\mathbf{t}_i] = V[E(\mathbf{t}_i|\mathbf{u}_i)] + E[V(\mathbf{t}_i|\mathbf{u}_i)] = \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_i$. Note that here the estimators \mathbf{t}_i are obtained conditionally on each subject.

For instance, if the assumptions for a linear model hold on each subjects and $h(\mathbf{X}_i, \boldsymbol{\beta}_i) = \mathbf{X}_i\boldsymbol{\beta}_i$, conditionally on realization of the random effect \mathbf{u}_i the estimators of $\boldsymbol{\beta}_i$ can be obtained through the Ordinary Least Squares, that is $\mathbf{t}_i = (\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{y}_i$. This lead, unconditionally, to:

$$V[\mathbf{t}_i] = (\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i[\mathbf{X}_i\boldsymbol{\Sigma}_u\mathbf{X}'_i + \boldsymbol{\Sigma}_i]\mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1} = \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_i.$$

Let now $\mathbf{t} = [\mathbf{t}'_1, \mathbf{t}'_2, \dots, \mathbf{t}'_N]'$ be the $N \times p$ matrix where each line contains the estimators of the fixed coefficients related to each cluster. The model assumptions can be summarized as:

- (a1) $f(\mathbf{t}_i - \boldsymbol{\beta}_i|\boldsymbol{\beta}_i) = f(-\mathbf{t}_i - \boldsymbol{\beta}_i|\boldsymbol{\beta}_i)$;
- (a2) $g(\boldsymbol{\beta}_i - \boldsymbol{\beta}) = g(-\boldsymbol{\beta}_i - \boldsymbol{\beta})$;
- (a3) $\mathbf{t}_i|\boldsymbol{\beta}_i \perp \boldsymbol{\beta}_i$;
- (a4) \mathbf{t}_i is independent of \mathbf{t}_ℓ for $1 \leq i < \ell \leq N$.

Note that (a1) and (a2) imply the assumption of symmetry of the unconditional joint estimator distributions, $f(\mathbf{t}_i - \boldsymbol{\beta}) = f(-\mathbf{t}_i - \boldsymbol{\beta}) \forall i$ (here f is an unspecified density function). This is ensured if the usual assumptions for mixed models hold, including (a3). This basically corresponds to assuming symmetry and independence for the conditional distribution of the estimators and for the distribution of $\boldsymbol{\beta}_i$; the last assumption deals with the assumption of independence among subjects. Let's also remark that these conditions only focus on the coefficients estimators \mathbf{t} , symmetry is not assumed with respect of the original data \mathbf{y}_i .

The generalized linear model [5] and the generalized linear mixed models are therefore special cases of this model by letting $\boldsymbol{\mu}_i = g^{-1}(\mathbf{X}_i[\boldsymbol{\beta} + \mathbf{u}_i])$ where g is the link function for the distribution of y_{ij} . In these case \mathbf{R}_i shall be more properly written $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\mu}_i)$ since in model specification such as binomial and Poisson regressions the dispersion parameter depends on the true mean effect. This play a direct role on $\boldsymbol{\Sigma}_i$ so that we shall write $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\mathbf{R}_i(\boldsymbol{\mu}_i))$.

This model is also suitable for more general cases, such as rank based statistics. For example if the i th subject is measured under two different conditions, the difference of the ranks means can be considered: $\mathbf{t}_i = \bar{r}_1 - \bar{r}_2$, where $\bar{r}_k = n_k^{-1}r(\mathbf{y}_i)\mathbf{X}_k$, and \mathbf{X}_k is the vector of indicators of the k th experimental condition, $k = 1, 2$, and $r(\cdot)$ is the rank transformation of vector of observations within the subject i . In this example \mathbf{t}_i estimates the true 'difference between average ranks' $\boldsymbol{\mu}_i$, which is a linearized function of predictors in (2), while ϕ in (1) remains a unknown, unspecified function. Outside the estimator \mathbf{t}_i based on linear combinations, further extensions can be considered. As an example in the two-condition-per-subject case we could consider the median difference of all observations among the two conditions. In this case $\boldsymbol{\mu}_i = h(\mathbf{X}_{ij}, \boldsymbol{\theta}_i)$ in (2) cannot be written as linear combination of predictors.

It is also worth to remark that between subjects effects can be taken in account in this framework. The sum of all effects between is included in the overall mean

estimate of each subject i , therefore any generalized linear model with an intercept will accomplish the work.

A further generalization is represented by the multivariate model with q dependent variables: in this case each \mathbf{t}_i becomes a $p \times q$ matrix whose rows contain the estimates of the model for each variable (or as the concatenated vector $\mathbf{t}_i = [t_{11}, \dots, t_{1p}, \dots, t_{1q}, \dots, t_{pq}]$), \mathbf{R}_i becomes an $nq \times nq$ matrix, and $\mathbf{\Sigma}_i$ becomes the $pq \times pq$ variance/covariance matrix of the predictors (Note that dependence among the predictors of each variable is also considered here). The models can be estimated separately for each variable (as it will happen in our example) or jointly (e.g. in multivariate normal models). Assumptions (a1)-(a3) are to be thought with respect to the concatenated vector. The partial tests on each variable's model are done simultaneously and a global p -value can be obtained combining these informations, as we will discuss in the example of section 4.

Within this model we are interested on testing

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

More precisely, we are interested on testing the null hypothesis component-wise

$$H_{0k} : \beta_k = \beta_{0k} \quad \text{against} \quad H_{1k} : \beta_k \neq \beta_{0k}.$$

so that

$$H_0 = \bigcap_{k=1}^p H_{0k}.$$

3 The test

From the assumptions of the previous section, we know that the set of $\{\mathbf{t}_i; i = 1, \dots, N\}$ has independent elements. In spite of this they are not identically distributed even under the null hypothesis, since they have variance matrix $\mathbf{\Sigma}_u + \mathbf{\Sigma}_i$. The special case of $\mathbf{\Sigma}_i = \mathbf{\Sigma}_\ell \forall (i, \ell)$ is reached for example when the subjects have same number of observations ($n_i = n \forall i$), and \mathbf{R}_i is assumed to be the same for all subject (i.e. $\mathbf{R}_i = \mathbf{R}$). In this case a very simple but effective approach uses a parametric one-sample t -test on the k -th columns of \mathbf{t} to test H_{0k} , $k = 1, \dots, p$. When the group estimators \mathbf{t}_i 's are heteroscedastic we need to adopt a more complex test. This is because the estimate of the residual variance becomes biased and this reflects on a wrong reference distribution of the test statistic. The classical maximum likelihood approach for mixed models estimates the residual variance in an unbiased way, but they fail in the pit of defining the correct number of degrees of freedom for the (approximate) distribution of the test statistic.

On the contrary, an exact test can be obtained following McNemar's lines: let $T^{ob} = \mathbf{s}'(\mathbf{t} - \mathbf{1}\boldsymbol{\beta}'_0)$ be the observed value of the test statistic, where $\mathbf{1}$ is a $N \times 1$ vector of ones, and \mathbf{s} is a $N \times 1$ vector of values equal to $1/N$.

To define its permutation distribution (or orbit, the set of points with the same likelihood under H_0), define \mathbf{S} to be the collection of 2^N points of the space $\{+1/N, -1/N\}^N$ rearranged in a $2^N \times N$ matrix. Then the permutation space is given by $\mathbf{T} = \mathbf{S}(\mathbf{t} - \mathbf{1}\boldsymbol{\beta}'_0)$. Despite naive, this test is unbiased, consistent and exact. The generic element T^* of the permutation space \mathbf{T} is then given

by the point whose coordinates are the $p \times 1$ vector $\sum_{i=1}^N S_i(\mathbf{t}_i - \boldsymbol{\beta}_0)/N$, where $\Pr\{S_i = -1\} = \Pr\{S_i = 1\} = 1/2$. Note that the inner dependence among the parameter estimators of each subject (typically described by $\boldsymbol{\Sigma}_i$) is maintained since all the components of the vector $(\mathbf{t}_i - \boldsymbol{\beta}_0)/N$ are multiplied by the same sign. This permutation strategy applies also in the multivariate case: here the dependence among the variables is translated into the dependence among each variable's estimators, and we do not require to explicate the remaining elements since the correlations are handled nonparametrically by the permutation strategy. Once the joint (multivariate) permutation space has been obtained, it is straightforward to assess the global null hypothesis on the k th predictor by applying the nonparametric combination of dependent tests (as proposed in).

The observed value of the test statistic for the k th component of $\boldsymbol{\beta}$ is given by $T_k = \sum_{i=1}^N (t_{ik} - \beta_k)/N$, whose expected value and variance are respectively equal to zero and $[\boldsymbol{\Sigma}_u/N + \sum_{i=1}^N \boldsymbol{\Sigma}_i/N^2]_{kk} = [\sum_{i=1}^n V(\mathbf{t}_i)/N^2]_{kk}$, $k = 1, \dots, p$ under H_{0k} . Note that this variance involves only the diagonal elements of the matrix $\sum_{i=1}^n V(\mathbf{t}_i)/N^2$.

The assumptions of symmetry fulfill the exactness of the test despite the correct specification of the model: e.g. if one considers a linear model instead of a linear mixed model. A concern may regard the power of the test since big differences in the $\boldsymbol{\Sigma}_i$ components can affect the sensitivity of the test. Indeed, if we let \mathbf{R}_i vary on each subject, if the design is unbalanced then $\boldsymbol{\Sigma}_i$ will vary for each subject as well. A possible solution is represented by pre-weighting the estimators of each subject by the inverse of their standard errors, similarly to what happens in the General Least Square method, but here we directly weight the estimators rather than the original observations.

If we knew the true value $\boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_i$ for each subject we could standardize the values of \mathbf{T} dividing them by their standard deviations. Since these matrices are generally unknown, we would need to estimate $\text{diag}(\boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_i)$. As it can be easily figured out, there are several alternatives strategies to estimate the residuals variances. They can lead to different inferential conclusions of course. However, the only characteristic of the test from a procedural point of view is if estimate is done assuming H_0 to be true or not. Here we discuss the two choices through the two most simple methods.

We assume we have an estimate of $\text{diag}(\boldsymbol{\Sigma}_i)$ (e.g. the variances of the parameter estimators in subject i , say $\hat{\sigma}_{ik}^2$, $k = 1, \dots, p$), then we try to estimate $\boldsymbol{\Sigma}_u$. We make use of the Moments Method under the null hypothesis (i.e. the mean of each estimator is equal to 0). In this case $V(\mathbf{T}) = E[\mathbf{T}'\mathbf{T}] = E[\sum_{i=1}^N \mathbf{t}_i \mathbf{t}_i'] / N^2$. Then

$$\text{diag} \left(E \left[\sum_{i=1}^N \mathbf{t}_i \mathbf{t}_i' \right] \right) = \text{diag} \left(N\boldsymbol{\Sigma}_u + \sum_{i=1}^N \boldsymbol{\Sigma}_i \right).$$

Let $\sigma_{u_k}^2$ denote the k element of $\text{diag}(\boldsymbol{\Sigma}_u)$; for each element in the diagonal we get

$$\sum_{i=1}^N t_{ik}^2 = \left(N\sigma_{u_k}^2 + \sum_{i=1}^N \sigma_{ik}^2 \right)$$

where σ_{ik}^2 is the true variance of t_{ik} . By replacing the σ_{ik}^2 's with their estimates we

obtain:

$$\hat{\sigma}_{u_k}^2 = \left(\sum_{i=1}^N t_{ik}^2 - \sum_{i=1}^N \hat{\sigma}_{ik}^2 \right) / N,$$

where the estimate is forced to zero in case it is negative. This implies that no random effect is estimated to be present on the k th component of the model. Note that the estimate of $\sigma_{u_k}^2$ is permutationally invariant (i.e. it is not affected by the permutations of the estimate's signs), indeed $[S_i(t_{ik} - \beta_k)]^2 = (t_{ik} - \beta_k)^2$ wheter S_i is positive or not. We call this the *null* estimate of the variance, since it is obtained by assuming that H_{0k} holds.

Once the estimators of the variance are available, one can consider the test statistic: $\mathbf{T}' = \mathbf{s}'\tilde{\mathbf{t}}_i$, where each element of $\tilde{\mathbf{t}}_i$ is the standarize version of the elements of \mathbf{t}_i :

$$\tilde{t}_{ik} = \frac{t_{ik}}{\sqrt{\hat{\sigma}_{uk}^2 + \hat{\sigma}_{ik}^2}} = \frac{\sqrt{N}t_{ik}}{\sqrt{\sum_{i=1}^N t_{ik}^2 - \sum_{\ell \neq i} \hat{\sigma}_{\ell k}^2 + (N-1)\hat{\sigma}_{ik}^2}}$$

Note that this transformation just modifies the variances of the conditional distributions, but does not affect the means under H_0 (which is always null), so the test remains exact even in case the variances are over/under estimated. This is the case when a mixed effect model is chosen instead of a simple linear model. Moreover, the denominator of \tilde{t}_{ik} is permutationally invariant.

The alternative, possible, strategy is to estimate the estimators variances when the alternative hypothesis is assumed to hold. This is because, when the alternative is true, the variances are over-estimated. In this case the estimates of the random effect variance components are based on the variances of \mathbf{t}_i , rather than on their second moment. Using again the moment estimation, the k th diagonal elements becomes:

$$\sum_{i=1}^N t_{ik}^2 - N\bar{t}_k^2 = \left(N\sigma_{u_k}^2 + \sum_{i=1}^N \hat{\sigma}_{ik}^2 \right), \quad \bar{t}_k = \frac{1}{N} \sum_{i=1}^N t_{ik}.$$

Applying this choice on the estimates, the (observed) modified coefficient estimators become equal to:

$$\tilde{t}_{ik} = \frac{\sqrt{N}t_{ik}}{\sqrt{\sum_{i=1}^N (t_{ik} - \bar{t}_k)^2 - \sum_{\ell \neq i} \hat{\sigma}_{\ell k}^2 + (N-1)\hat{\sigma}_{ik}^2}}.$$

Now, since the value of \bar{t}_k changes at each permutation, the denominator must be re-estimated at each permutation.

Summarizing the possibilities we have illustrated, we can conclude that

- i) The original test has been thought on the fixed coefficient estimators as they are. The permutation strategy involves independent changes on the signs of the vectors \mathbf{t}_i ; this test is exact although it might be little sensitive;
- ii) A first modification contemplates to weight the initial coefficient estimators by the inverse of their standard error, when H_0 is assumed to be true. The permutation strategy is obtained by modifying the signs of the vectors $\tilde{\mathbf{t}}_i$; therefore, the weighting process is applied only at the beginning.

- iii) A further modification contemplates to weight the coefficient estimators at each permutation, since the estimates of their standard errors change at each permutation.

We remark that, the whole procedure can be repeated by forcing $\Sigma_u = 0$, i.e. by considering (generalized) linear models instead of (generalized) mixed linear models. This make strategy ii) equal to strategy iii). More remarkably, the choice does not affect the exactness of the test, this clarifies why we assert that this test is not affected by wrong model assumptions, while MLE methods do.

Which kind of test to use depends on the model assumptions which are, in general, subjective, therefore we suggest the original version in case the design is balanced and the within subject variability can be assumed to be equal, whereas the solution (ii) would be more adequate when random effects can be considered negligible, and solution (iii) covers all the remaining cases (some further studies are required to compare solutions ii and iii). Whatever will be the choice, we remark that these test are all exact, although they might differ in performances (i.e. power).

4 Example

Recall the motivating example of section 1: we have 12 adults measured in $n_i = 32$ trials (randomly selected images) under two experimental conditions. The predictor ‘Item’ is coded as 0 = one face among 4 objects and 1 (6 objects) on two variables. The first variable (Primafix) is dichotomous (1 = ‘looking at the face first’) and hence a GLM with two parameters:

$$Y_{ij} \sim Bi(1, \pi_{ij}) \quad \text{logit}(\pi_{ij}) = \beta_{0i} + \beta_{1i}x_{ij} \quad j = 1, \dots, 32,$$

is taken as the model for each subject, where $x_{ij} = 0/1$ depending on the treatment, and $\beta_i = [\beta_{0i}, \beta_{1i}]'$ is the vector of parameters of the i th subject. The second variable (Prop) is continuous and therefore a simple linear regression model $Z_i = \gamma_{0i} + \gamma_{1i}x_{ij} + \varepsilon_{ij}$ has been chosen here. We remark that the estimates of the parameters are done conditionally within each subject (i.e. conditionally to the random effect realization). The estimates of the parameters and their standard deviations (the square roots of the diagonal of $\hat{\Sigma}_i$) have been obtained from the **R** functions `glm` and `lm` separately for ‘Primafix’ and ‘Prop’, respectively. A comparison with the MLE solution has been obtained with the `lmer` function of the `lme4` package (<http://lme4.r-forge.r-project.org/>). We considered two nested models with the same random effects and compare them through the likelihood ratio test (LRT); the estimates of the effect have been done with the full likelihood approach. Table 1 reports the estimates of the slopes for each model and each subject. Note that the twelveth sybject appear to behave as an outlier ($\hat{\beta}_{1,12}$ is much bigger in absolute value than the slopes estimated in the other subjects, and so is its standard deviation). The assumption of within subject’s homoschedasticity $\mathbf{R}_i = \mathbf{R}$ seems to be inadequate for at least this subject.

Table 2 shows the p -values obtained with each testing procedure introduced in the previous section for each model and each variable. We focus the attention on the slopes, as usual. The p -values obtained with the ‘naive’ solution (i) show that the effect of Item is significantly decreasing the proportion of time that the subjects

Subject	PrimaFix (Dichotomous)		Prop (Continuous)	
	\hat{b}_1	s.e. (\hat{b}_1)	$\hat{\gamma}_1$	s.e. ($\hat{\gamma}_1$)
1	-0.310	0.790	-0.059	0.046
2	0.278	0.747	-0.107	0.055
3	-0.788	0.735	-0.174	0.042
4	-0.511	0.719	-0.115	0.052
5	-1.695	0.909	0.084	0.068
6	-0.260	0.722	-0.108	0.070
7	0.000	0.763	0.001	0.054
8	-0.678	0.837	-0.015	0.077
9	0.260	0.722	-0.105	0.051
10	-0.537	0.738	-0.032	0.060
11	-0.278	0.747	-0.062	0.043
12	17.858	4432.593	-0.099	0.067

Table 1: Slopes coefficients and related s.e. within each subject for both responses.

spend looking at the face, whereas there seems not to be any significant effect on the PrimaFix variable. This might be explained recalling that the test statistic is basically a mean of the slope estimates, and that subject twelve has an opposite, very strong behaviour, compared with the other subject. Also note that the MLE results confirm the extremely significant influence of ‘Item’ on ‘Prop’, but not on ‘PrimaFix’.

	Naive	Null Estimate	Alternative Estimate	LRT
PrimaFix	0.9902	0.0253	0.0254	0.1151
Prop	0.0088	0.0078	0.0059	0.0037
Overall	0.0383	0.0007	0.0007	0.0074

Table 2: P-values for univariate and overall null hypotheses

The second part of the table reports the results of our test when the estimates of the slope had previously been standardized under H_0 (proposal (ii) - ‘Null Estimate’): note that here the test on β_1 becomes highly significant, showing the positive effect of taking into consideration each estimator’s variance, i.e. assuming the observations to be heteroschedastic between subjects. In fact, if the twelfth subject is not included in the study, the MLE solution of ‘PrimaFix’ would become significant as well (and so would the ‘Naive’ solution).

A similar result is obtained with proposal (iii), named ‘Alternative Estimate’ in the table, probably meaning that here the random effect variability is negligible with respect to each subject’s variability.

Finally, our global test’s results rely on Fisher’s combination [7] applied to the joint permutation distribution of the partial p -values, while the combined p -value from MLE has been obtained through Bonferroni’s combination, i.e. without taking into account the dependency among the variables.

5 Conclusions

The tests proposed here depict a very general approach for univariate and multivariate tests. As detailed in section 2), the class of models covered here is very broad and the generalized theoretical linear mixed-models are just a special case of it. Rank-based and more complex models are covered by this approach as well. Also multivariate responses - possibly mixing categorical, ranks and continuous responses- are included in the model without further extensions.

All the tests proposed are exact. In this perspective, Re: confereze mix and permute they are more safe since they mitigates the consequences of miss-modeling, while in parametric approach this misspecification can lead to dramatic consequences on type I error control. Of course the quality of the estimates of $\hat{\Sigma}_u$ and $\hat{\Sigma}_i$ will affect the power and further works will be addressed on definition of more reliable estimators.

Last but not least, the test is immediately extended to multivariate inference since the joint distribution of the test statistics is always available without further assumptions on the model.

References

- [1] R.H. Baayen, D.J. Davidson, and D.M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412, 2008.
- [2] D. Basso, L. Finos, and G. Altoé. A permutation test for linear mixed models. In *Proceedings of the 45 SIS Scientific Meeting*, Padua - Italy, June 14 - 15 2010.
- [3] D. Basso, L. Finos, and Altoé G. A permutation test for random effect models. In *Proceedings of the Applied Statistics 2010 Conference*, Bled - Slovenia.
- [4] Di Giorgio, Turati, and Simion. In *International Conference on Infant Studies*, Baltimore, Maryland (USA), 2010.
- [5] P. McCullagh and John A. Nelder. *Generalized Linear Models, Second Edition (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 2 edition, August 1989.
- [6] S. Mériaux, A. Roche, G. Dehaene-Lambertz, B. Thirion, and J. B. Poline. Combined permutation test and mixed-effect model for group average analysis in fMRI. *Hum Brain Mapp*, 27(5):402–410, May 2006.
- [7] F. Pesarin. *Multivariate permutation test with application to biostatistics*. Wiley, Chichester, 2001.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: <http://wp.stat.unipd.it>

