Permutation Tests and Multiple Comparisons in the Linear Models and Mixed Linear Models, with Extension to Experiments using Electroencephalography

by

Jaromil Frossard

A thesis submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland,
in fulfilment of the requirements for the degree of
PhD in Statistics

Members of the thesis committee:
Prof. Olivier Renaud, Advisor, University of Geneva
Prof. Elvezio Ronchetti, Chair, University of Geneva
Dr. Guillaume Rousselet, University of Glasgow

Thesis No. 69
August 2019

La Faculté d'économie et de management, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par-là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Geneva, August 14th, 2019

Dean
Marcelo OLARREAGA

Impression d'après le manuscrit de l'auteur

Abstract

We present how permutation tests can be applied in experiments using electroencephalography (EEG). First, we present the `permuco` `R` package which allows permutation tests on linear model and repeated measures ANOVA with nuisance variables. It uses several permutation methods and, for comparison of signals, it applies multiple comparisons procedures like the cluster-mass test or the threshold-free cluster-enhancement. Second, we show that most of the permutation methods have a geometrical interpretation. Moreover, we present a real data analysis where the cluster-mass test is used for a full-scalp analysis of EEG data. We also show that using the slopes of the EEG signals in combination to the cluster-mass test produces more powerful tests. Third, asymptotic properties of the F statistic of several permutation methods are derived using the moments of the conditional distribution by permutations. Fourth, we explain why experiments in psychology should often be modelised by a cross-random effects mixed effects model (CRE-MEM) and we show that the assumed correlation structure of the data influences tests of fixed effect parameters. Finally, we propose a general re-sampling framework to analyse EEG data when using CRE-MEM.

Résumé

Cette thèse introduit l'application des tests de permutation aux expériences utilisant l'électroencéphalographie (EEG). Premièrement, nous présentons la librairie `R permuco` qui permet des tests de permutation appliqués aux modèles linéaires et à l'ANOVA à mesures répétées contenant des variables de nuisances. De plus, cette librairie permet l'utilisation de plusieurs méthodes de permutation et utilise des procédures de comparaisons multiples tel que le test de masse de groupes (cluster-mass) ou le renforcement de groupes sans seuil (threshold-free cluster-enhancement) pour la comparaison de signaux. Deuxièmement, nous montrons que les méthodes de permutation ont une interprétation géométrique. De plus, nous présentons une analyse de données où le test de masse de groupes est appliqué sur des données EEG du scalp complet. Ensuite, nous montrons que la pente des signaux combinée au test de masse de groupes améliore la puissance statistique du test. Troisièmement, nous déduisons les propriétés asymptotiques de la statistique F pour plusieurs méthodes de permutation grâce aux moments de la distribution conditionnelle par permutation. Quatrièmement, nous présentons un argumentaire en faveur de la modélisation des expériences en psychologie par les modèles mixtes croisés à effets aléatoires croisés (CRE-MEM) et nous montrons l'importance de la structure de corrélation lors de leur utilisations pour des tests statistiques. Finalement, nous proposons un cadre général de tests de ré-échantillonage pour analyser les expériences en psychologie à l'aide de CRE-MEM.

Contents

Introduction

The following manuscript consists of the work and reflections I produced during 5 years as a PhD candidate. The general aim is to use permutation tests in experiments using electroencephalography (EEG). It is motivated by actual experiments in neuroscience and the complexity of the experimental designs lead me to investigate the topic of cross-random effects mixed-effects model (CRE-MEM) as well. In the following introduction, I first introduce an example of an experiment in neuroscience which motivates the thesis, then some general concepts in permutation tests and how they are applied in neuroscience. Finally, I explain why CRE-MEM is related to experiment in psychology and neuroscience.

Generally, an EEG experiment records the electrical brain activity of participants while they see some stimuli. Each participant spends tens of minutes doing the experiment and a cap of electrodes records their brain activity. During the experiment, they usually perceive stimuli, and usually need to accomplish a task that depends on the type of stimuli. For instance, in the EEG dataset in the `permuco` (Frossard and Renaud, 2018) `R` (Chambers, 2009) package, the participants had to see stimuli (faces) with 2 types of emotions, either angry or neutral, with 2 types of visibility, supraliminal (166 ms) or subliminal (16 ms) and with 2 types of laterality, either displayed at the right or at the left of the screen (Tipura et al., 2017). Neuroscientists want to know if the experimental setting (the design of the experiment) influences the EEG signals and produces signals that are different. Moreover, they are also interested by when (after seeing the stimuli) and where (on which electrode) these differences occur. The ”raw” dataset is usually pre-processed before being analysed. In the pre-processing, we delete trials that are too noisy (e.g. the eye blinks disturb the recording), or filter frequencies that may come from the electric grid (50 Hz). We also adjust the time such that each trial is synchronized; for instance at the event, the precise time when the stimuli is shown. Once the data are ”cleaned”, neuroscientists compute event-related potentials (ERP), which are averages of the signals at each time-point. These averages are grouped by electrode and experimental setting. Figure 1 shows ERPs of the electrode O1 of the experiment described above. We see that before the event some small differences between the experimental settings occur; which are obviously caused only by statistical noise. However, we also see a larger difference between ERPs of pictures that are subliminal (16 ms, represented by the thin lines) and supraliminal (166 ms, represented by the thick lines). The statistical problem is to determine which differences are likely to be generated by random noise or can be attributed to the experimental design.

In that multivariate setting, Karniski et al. (1994) recognize flaws using parametric approaches like MANOVA or repeated measures ANOVA to test the difference between experimental conditions. They argue that testing the differences of EEG recording between two groups using the parametric methods implies assumptions on the EEG data generation process which are not satisfied. For repeated measures ANOVA, the assumptions of sphericity is almost always violated for EEG data. Moreover, using MANOVA

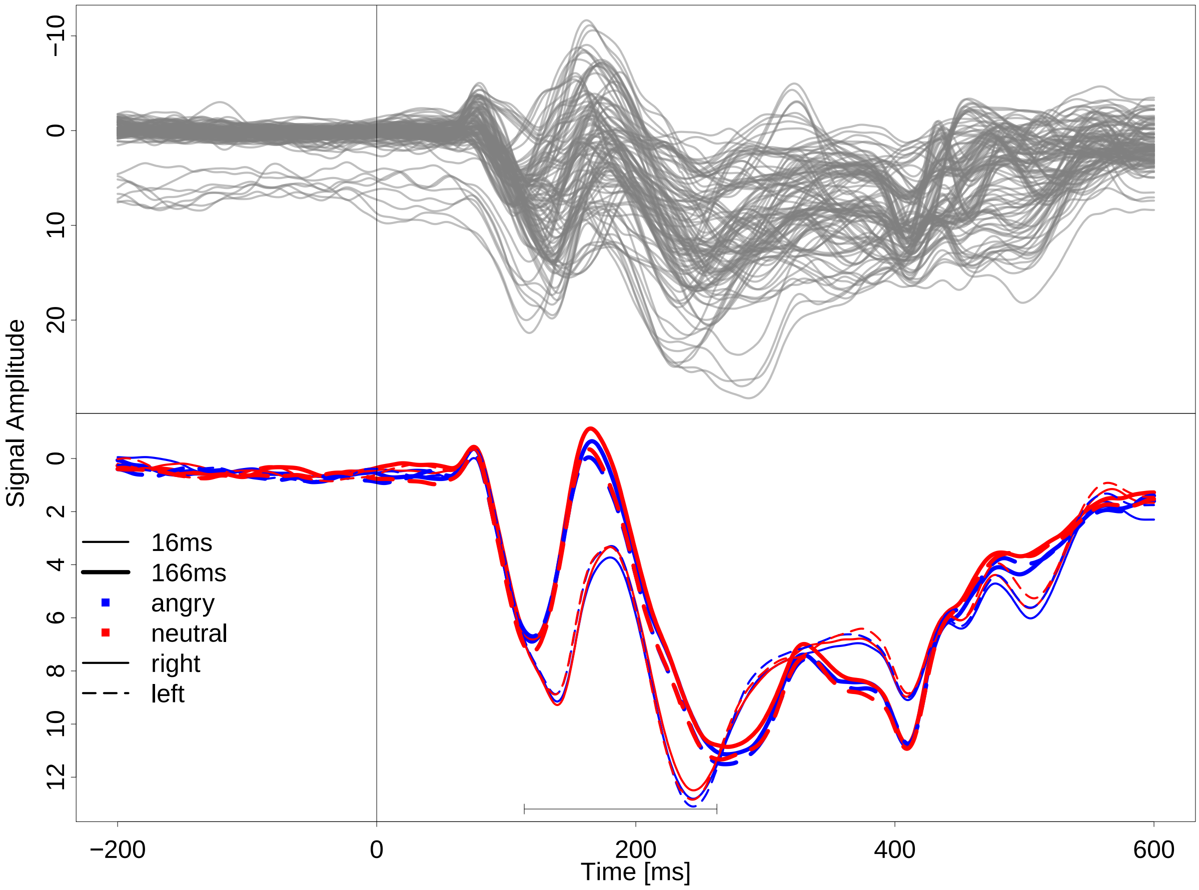

Figure 1: ERP of the electrode O1. In the top panel, one line corresponds to the signal of one participant (out of 15) in one experimental setting (out of 8). In the bottom panel, each line corresponds to the average ERP over all participants in one experimental setting. The vertical line corresponds to the time of the event.

needs the assumption of normality which is also not fulfil in EEG. Using these methods despite the violation of assumption usually increase the type I error rate of the tests. To solve this problem, Karniski et al. (1994) introduce re-sampling methods, more precisely permutation tests, to EEG data analysis to test hypotheses on the difference between signals given the experimental setting.

Arguably, the state-of-the-art statistical methods to detect these effects are based on permutation tests. They are also called exact test and are an old topic in statistics that was presented by Fisher (1935) for simple design like in the two-sample t-test. To understand how permutation tests are applied when analysing complex EEG data, we first recall the basics for observations measured on two groups (or conditions). The rationale behind permutation tests is quite straightforward. In general, a statistical test uses an observed sample to decide if a null hypothesis on the population should be assumed false (e.g. the null hypothesis for a t-test is: two populations have the same mean). The test determines if observing a similar (or more extreme) sample is a likely event under the null hypothesis; it is expressed for permutation tests as an equality of distribution of the populations (e.g. $H_0 : \ F_A = F_B$). If the null hypothesis is true, for each observation, there is no link between the response variable and the membership to one of the populations. In addition, observing our sample is as likely as observing a sample where each group membership and response variable are mixed together; mixing together group membership and response variable creates what we call a permutation of the observed sample. Hence, when we assume that the null hypothesis is true, each permutation of the sample is equally likely. It follows that computing a statistic on each permuted sample produces a set of statistics that are equally likely under the null hypothesis. Using this set of statistics and the statistic computed on the observed sample, we compute a p-value as the proportion of permuted statistics that are more extreme than the observed one.

The explanation above does not need a specific statistic to be valid. Therefore, one advantage of this procedure is that permutation tests do not impose restrictions on the choice of the statistic. Furthermore, they produce an exact test even for unusual statistics, unless there is ties in the values of the statistics. This property is particularly useful when generalized to detect difference between signals as it allows us to combine statistics of time-points.

Formally, permutation tests are also attractive because they only need the assumption of exchangeability of the data under the null hypothesis. This means that, under the null hypothesis, the response variable must have the same multivariate distribution after any permutations. It follows that permutation tests handles skewed, heavy-tailed or unknown distributions. They are said to be robust (Lehmann and Romano, 2008) as they stay exact under non-gaussian distributions (this properties is also refereed as robustness of validity). However, the power of the test is also of interest and, in that case, the choice of the test statistic matters. Indeed, Welch (1987) and Welch and Gutierrez (1988) show that, if data are contaminated by extreme values, using the trimmed mean or the median improves the efficiency of permutation tests compared to a t statistic. Moreover, in presence of outliers, using robust statistic improves the inference as the distribution of interest may not be the full distribution but only the unperturbed distribution (Heritier et al., 2009), i.e. the distribution freed from unrelated and outlying effects. In that case, the robustness properties of permutation tests do not naturally solve this problem. Indeed, a permutation test is robust only when the test statistic has robust properties, i.e. is not be influenced by outlying points. The usual F or t statistics are based on means and sums of squares which are even influenced by a small proportion of outliers. In fact, the robust properties of the statistic are independent of the approach: parametric or by re-sampling. To solve this problem, parametric approaches using M-estimators haven been developed by Huber (1964). These estimators have the property to reduce the influence of outlying data by down-weigthing them and, consequently, are robust to contamination by extreme values. Moreover, their asymptotic distributions are known which allows tests of hypothesis. The inference is done on the uncontaminated distribution while permutation tests (using non-robust statistics) infer on the contaminated distribution. Finally, re-sampling approaches using the bootstrap (Efron, 1979) have been proposed for computing confidence interval (Salibian-Barrera and Zamar, 2002) or testing hypothesis (Salibian-Barrera, 2005) in a robust way in regression models.

For permutation tests, the assumption of exchangeability of the data is not fulfilled even in simple models like factorial ANOVA with only 2 factors and their interaction. Indeed, when we are interested to obtain the p-value for the effect of one factor (or the interaction), all other effects should be consider as nuisance effects that violate the exchangeability assumption as they produce unequal first moment under the null hypothesis of interest. In the example of Figure 1, to test the difference between the 2 levels of visibility, one must take into account that there might be a true effect of the emotion, of the laterality and of all of the higher order interactions. Even under the null hypothesis (in our example: the main effect of visibility is absent), all these effects induce means (first moments) of the observations that are possibly different between the groups or conditions. Without adjusting for these effects, the test of visibility is potentially wrong as it would include other effects. Using a naive permutation approach violates the exchangeability assumption which usually results in an increase of the type I error rate. Several authors have tried to circumvent to this problem. First, authors derived asymptotic properties of the permutation tests under data that are not exchangeable (Pauly et al., 2015). Usually these asymptotic distributions are only derived for specific statistics and the permutation tests lose the flexibility of the choice of the statistic. Secondly, authors proposed to restrict the number of permutations (Pesarin, 2001; Edgington and Onghena, 2007; Anderson and Braak, 2003). In some designs, only a subset of the permutations conserves the exchangeability of the data under the null hypothesis and an exact test is derived using only this subset. The simplest example may be for paired sample design where permutations within each participant are exchangeable under the null hypothesis. These procedures widen the range of models with exact test. However, they are not generalizable to the general linear model framework. The third strategy is changing or decomposing the model before performing unrestricted permutations. Those procedures try to reduce the effect of the nuisance variables but may not create exact tests. In this case, the flexibility of the procedures bares the cost of imprecision of the type I error rate. These permutation methods have been developed by Draper and Stoneman (1966), Dekker et al. (2007), Kennedy (1995), Huh and Jhun (2001), Freedman and Lane (1983) or ter Braak (1992), and Winkler et al. (2014) summarized them in a common notation. In this thesis, we propose new theoretical works on these issues (Chapter 2 and 3) as well as software implementing these methods (Chapter 1).

Moreover, in the top panel of Figure 1, each ERP (or line) corresponds to a given type of stimuli shown to a participant during the experiments. These stimuli should be assumed to be sampled among a population of stimuli as, for instance, the particular images of ”angry” faces used by Tipura et al. (2017) are only a subset of all possible ”angry” faces. Indeed, a new experimenter would certainly select others images of faces for the same ”angry” level for a replication of this experiment. In order, to be able to interpret the inference for the population of faces and not only on the subset of the particular stimuli selected by the experimenter, the statistical analysis must assume a sampling of both participants and stimuli. These 2 different samplings imply some covariances between observations as, for instance, each observation recorded using the same stimulus is probably correlated across participants. Not taking into account these correlations in the statistical model implies an increase of the type I error rate (Bürki et al., 2018) up to 80%. Once again, when using a naive permutation test these correlations violate the exchangeability assumption. However, in that case, it is the covariance of the observations which is not exchangeable any more. In this thesis, Chapter 4 covers this topics in a parametric setting and Chapter 5 proposes permutation methods for EEG data that take into account the variability induced by the 2 samples.

The bootstrap is also a re-sampling methods that allows to produce p-values for hypothesis testing (Efron and Tibshirani, 1994). Depending on the way bootstrap is used, we can derive decision of statistical test from the bootstrap confidence interval or, following Efron (1982) proposal, we can use the bootstrap t, initially proposed in the regression framework. The first solution is not adapted to our setting as we may be interested in the test of multiple parameters simultaneously. Moreover, in the context of EEG data analysis, re-sampling methods need to be used in combination with multiple comparisons procedures. However, the second approach is similar to some permutation methods. Indeed, the proposition of ter Braak (1992), which is one of the permutation method presented in Chapter 1, is directly inspired by the bootstrap t. Moreover, in the two samples problem, Pernet et al. (2015) use the bootstrap t to test the difference between EEG signals and show that it gives similar results than permutation tests in a simulation study. ter Braak (1992) summarizes well the relationship between bootstrap and permutation when it comes to hypothesis testing: ”In both approaches, there is the choice what to permute or to bootstrap: the data values themselves or the residuals?”. Indeed, in principle, all permutation methods presented in Winkler et al. (2014) may be used by replacing permutation (sampling without replacement) by bootstrap (sampling with replacement) without probably inducing a large deviation of the re-sampled distribution and therefore the p-value. However, to our knowledge no specific simulation study have been proposed to evaluate differences between these two re-sampling strategies.

Recent advances make permutation tests especially useful to analyse neuroscience data like EEG signals. First, note that they need a lot of computing power which is the reason that at the time of Fisher (1935) the parametric procedure was more attractive. With the increase in computing power of the late 90's and 2000's, permutation tests showed a renewed interest (David, 2008). Moreover, they are particularly interesting for multiple comparisons procedures (Troendle, 1995; Maris and Oostenveld, 2007; Smith and Nichols, 2009). The field of neuroscience clearly requires multiple comparison procedures as in both EEG and functional magnetic resonance imaging (fMRI) data analysis, a large number of tests is performed. In EEG, each response variable is a signal (or a map when using fMRI) measured at a high frequency (up to 1024Hz) during almost one second. This corresponds to almost 1000 measured for only one electrode and the actual caps record 32 electrodes or more. Usually, one test is performed at each time point for each electrode. In total, a full-scalp EEG experiment often involves more than 10000 tests [1]. For the rest of the discussion and without loss of generality, we focus, for clarity, on only

[1] For fMRI, the number of tests is even higher. The human brain is around $1'200cm^3$, and represented by a 3D grid of small size voxels (around $1mm^3$) which store the BOLD signal. It implies up to 1 million tests for a full brain analysis.

Table 1: Classification of multiple test between the true hypothesis and statistical decision

True hypothesis/Decision	Significant tests	Non-significant tests	Total
True H_0	V	U	m_0
False H_0	S	T	$m - m_0 = m_1$
Total	R	$m - R$	m

one electrode. As shown by Benjamini and Hochberg (1995), it is useful to address the multiple comparisons problem by classifying each test in a two-entries table (see Table 1). When performing m tests, an unknown number $m_0 \leq m$ of tests are under H_0 and an unknown number of tests $m_1 = m - m_0$ are under H_1. For any procedure, the total number of true null hypothesis (m_0) is then split into V type I errors and U correctly non-rejected tests. In the "frequentist" approach, we try to control the number of type I errors. By assuming independent and exact tests using an individual type I error rate α, we have $E(V/m_0) = \alpha$. Hence, the probability of type I error increases with the number m_0 of tests under the null hypothesis. In that setting, we define the family-wise error rate (FWER) as $q = P(V \geq 1)$. In addition, controlling the FWER allows us to be more confident in the total number of rejected tests R as true findings; q is usually chosen with the same value than α. Note that the FWER is not the only measure that is useful to control and Benjamini and Hochberg (1995) shows that controlling the false discovery rate (FDR) defined as $q_{FDR} = E(V/R)$ where $q_{FDR} = 0$ when $R = 0$ leads to powerful methods that reduce the number of type I errors.

In order to control the FWER, well-known and general procedures, like Bonferroni (Dunn, 1958) or Holm (Holm, 1979), have been proposed. However, they are not adapted to the number of tests in neuroscience. Indeed the correction to control the FWER would be so restrictive that almost no test would be declared significant (assuming the usual effect size and constraints on the sample size of EEG experiments). Moreover, the EEG signals have both errors highly correlated and the true effects are also spatially and temporally distributed which is information that is not used by these methods. Permutation tests consider the correlation between the tests naturally. By applying the same permutation for each time-point, we conserve the correlation between tests and, as explained in Figure 2, it is used to improve the multiple comparisons procedure.

Moreover, another advantage to permutation test in neuroscience is that it does not need neither a modelization of the spatio-temporal correlation, nor of the error terms. Eklund et al. (2016) have shown that parametric methods that rely of these modelizations fail to control the FWER for fMRI while permutation tests which do not rely on these assumptions keep a FWER close to the nominal levels.

However, the use of permutation tests in neuroscience is limited when the experimental designs become too complex. Even if permutation methods have been proposed for repeated measures ANOVA (rANOVA) (Kherad-Pajouh and Renaud, 2015), the design of experiment does not always fall into this framework. The recording of the brain activity is made after participants react to stimuli. In addition, neuroscientists want to generalize their findings to both the population of participants and the population of stimuli (e.g. images of faces). In that case, the models need to consider the variability induced by the sampling of participants and also by the sampling of stimuli. The cross-random effects mixed-effect model (CRE-MEM) is the appropriate framework to analyse this type of data.

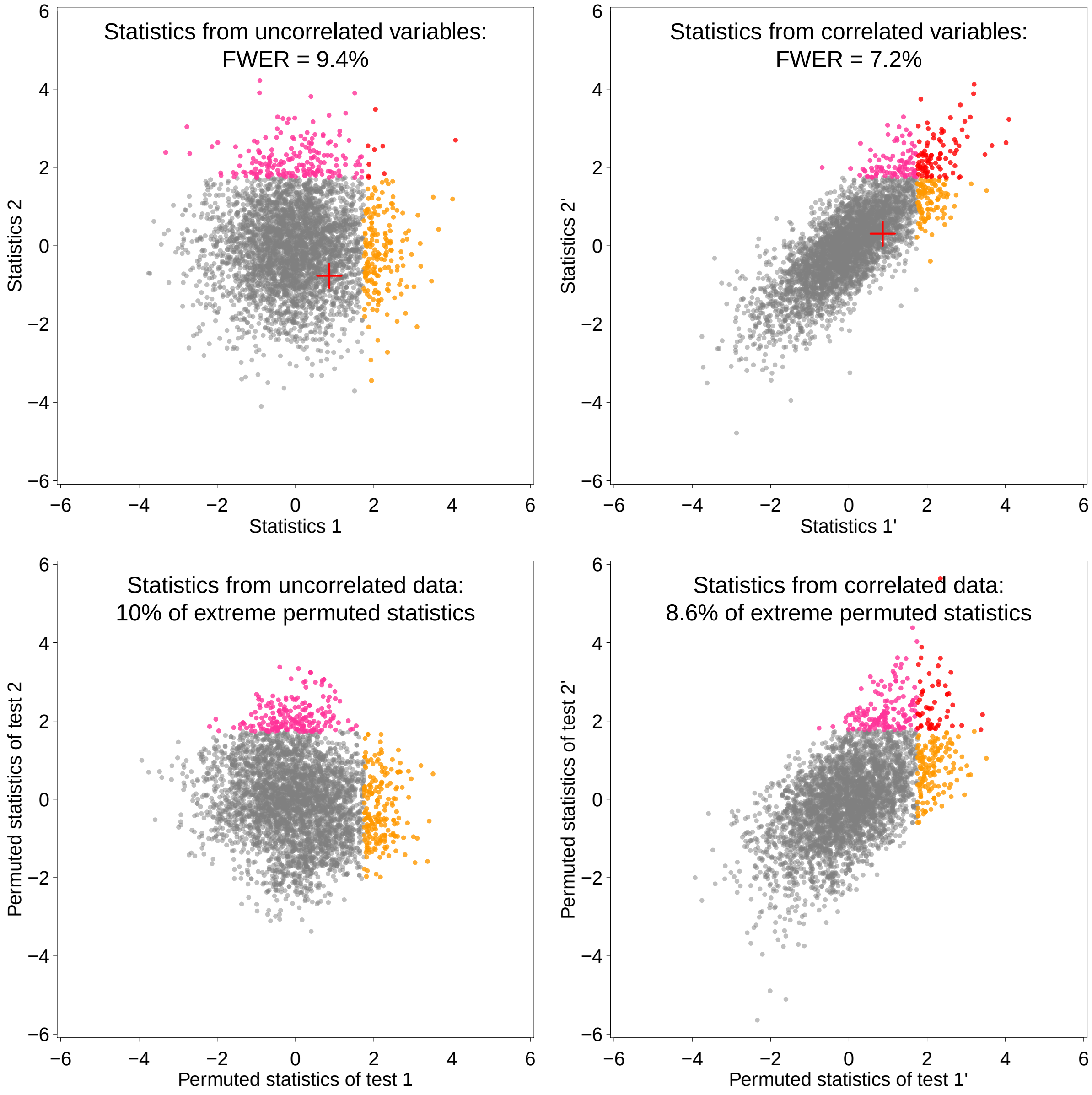

Figure 2: The top-left panel shows pairs of t-statistics under the null hypothesis computed on 4000 samples from uncorrelated variables. Using the 2 uncorrected tests results in a FWER of 9.4%. The correction required to control the FWER should be more stringent than the one applied to the 2 correlated variables (top-right panel, $\rho = .8$) which results to a lower FWER of 7.2%. Without any modelization of the correlation between the 2 tests, the permutation tests recreate multivariate distributions with an appropriate correlation. The bottom-left panel shows the distribution by permutation of the uncorrelated dataset which statistic is marked by the red cross in the top-left panel and the bottom-right panel shows the distribution by permutation of the correlated dataset marked by the cross in the top-right panel. The multivariate distributions by permutation maintain ”naturally” the correlation between the tests and allow us to easily adjust the multiple comparison procedures.

However, even in the univariate (and fully parametric) case, the choice of the model, especially its correlation structure, is still in debate. Several propositions have been made for experiments crossing 2 samples in psychology. The arguments usually confront a design-based approach (the experimental design defines the correlation structure (Barr et al., 2013)) to a data-based approach (the goodness of fit defines the correlation structure (Bates et al., 2015)). Note that both approaches have different influence on the type I error rate, or on the power of the tests.

Nonetheless, Judd et al. (2012) shows that psychologists may not take into account the random effects associated to the stimuli, and instead, were using the well-known repeated measures ANOVA (rANOVA) by averaging the data over the stimuli. It usually implies an increase of the type I error rate. In neuroscience, no method produces both a powerful control of the type I error rate and a modelization of the variability of both participants and stimuli. Bürki et al. (2018) show also an increase of the type I error rate when ignoring the variability of the stimuli in EEG experiment. However, in order to apply permutation tests of a fixed parameter of a CRE-MEM, the assumption of exchangeability is violated by both other fixed effects and by its complex correlation structure. Hence, both the first moment and the covariance of the response variable violate the exchangeability assumption.

Finally, the optimization of a CRE-MEM is a challenging task that requires a lot of computing power. In neuroscience experiments, this optimization must be performed for each test (typically on more than 1000 tests) and it usually creates a practical frontier that also needs to be addressed.

Chapter 1 contains the article ”Permutation Tests for Regression, ANOVA and Comparison of Signals : the `permuco` Package” submitted to the Journal of Statistical Software which presents the `permuco` package. The package proposes several permutation methods for ANOVA, regression and repeated measures ANOVA as well as multiple comparisons procedure like the cluster-mass test or the threshold-free cluster-enhancement (TFCE) for comparison of signals. The `permuco` package is the first software to implement several permutation methods within the same package. The other available packages are usually focused on only one method or only one particular statistic. For research purpose, it has been a necessary task to implement these methods. The project of this package began as a set of functions to test the existing permutation methods to understand their advantages and flaws. It was the natural following step to combine these functions into a single `R` package for the scientific community. The need for such software was present as `permuco` has already been used in several peer review publications through several fields, including ecology (Kern and Langerhans, 2019; Musariri et al., 2018), genetic (Soler et al., 2019; Allen et al., 2018) or psychology (Hartmann et al., 2019; Godfrey et al., 2019) and has been downloaded more than 4000 times (https://ipub.com/dev-corner/apps/r-package-downloads/) since it was made available on CRAN in January 2018.

Chapter 2 presents some complements on the permutation methods and multiple comparisons procedure implemented in the `permuco` package. Section 2.1 describes the geometrical interpretation of some permutation methods. Then, Section 2.2 describes my involvement in the application of cluster-mass test in Cheval et al. (2018) and presents functions for a full scalp cluster-mass test that will be added in a next release of `permuco`. Finally, Section 2.3 presents an extension of the cluster-mass test using the slope of the signals to increase the power of the tests.

Chapter 3 presents some theoretical results on permutation methods. Using linear algebra, we compute analytically, for several permutation methods, the moments of the numerator and denominator of the F statistic over all permutation (conditionally on the sample). In finite sample size, these results suggest some corrections of the permutation methods. In addition, we show the asymptotic properties of the conditional distribution by permutation of the F statistic for several permutation methods.

Chapter 4 presents the article ”The Correlation Structure of Mixed Effects Models with Crossed Random Effects in Controlled Experiments”. This article first presents a classification of variables present in a CRE-MEM and their consequence on the possible correlation structures. We also compare both theoretically and by simulations several correlation structures used in the literature. Moreover, we propose a new one that is the natural extension of the rANOVA to CRE-MEM. Finally, the purpose of this article is also to give experimenters both tools to understand the influence of the correlation structure on their analysis and `R` code to implement them using the `lme4` (Bates et al., 2015) or the `gANOVA` packages (https://github.com/jaromilfrossard/gANOVA).

The work presented in Chapter 4 is necessary and preliminary to applied randomization test in CRE-MEM. Finally, Chapter 5 presents a general framework to apply randomized tests in CRE-MEM with an extension to the design based signals like in EEG experiments.

Chapter 1

Permutation Tests for Regression, ANOVA and Comparison of Signals: the `permuco` Package

The following chapter is the main part of an article submitted to the Journal of Statistical Software.

Abstract. Recent methodological researches produced permutation methods to test parameters in presence of nuisance variables in linear models or repeated measures ANOVA. Permutation tests are also particularly useful to overcome the multiple comparisons problem as they are used to test the effect of factors or variables on signals while controlling the family-wise error rate (FWER). This article introduces the `permuco` package which implements several permutation methods. They can all be used jointly with multiple comparisons procedures like the cluster-mass tests or threshold-free cluster enhancement (TFCE). The `permuco` package is designed, first, for univariate permutation tests with nuisance variables, like regression and ANOVA, including repeated measures ANOVA; and secondly, for comparing signals as required, for example, for the analysis of event-related potential (ERP) of experiments using electroencephalography (EEG). This article describes the permutation methods and the multiple comparisons procedures implemented. A tutorial for each of theses cases is provided.

1.1 Introduction

Permutation tests are exact for simple models like one-way ANOVA and t test (Lehmann and Romano, 2008, pp. 176-177). Moreover it has been shown that they have some robust properties under non normality (Lehmann and Romano, 2008). However they require the assumption of exchangeability under the null hypothesis to be fulfilled which is not the case in a multifactorial setting. For these more complex designs, Janssen and Pauls (2003), Janssen (2005), Pauly et al. (2015) and Konietschke et al. (2015) show that permutation tests based on non exchangeable data can be exact asymptotically if used with studentized statistics. Another approach to handle multifactorial designs is to transform the data before permuting. Several authors (Draper and Stoneman, 1966; Freedman and Lane, 1983; Kennedy, 1995; Huh and Jhun, 2001; Dekker et al., 2007; Kherad Pajouh and Renaud, 2010; ter Braak, 1992) have proposed different types of transformations and Winkler et al. (2014) give a simple and unique notation to compare those different methods.

Repeated measures ANOVA including one or more within subject effects are the most widely used models in the field of psychology. In the simplest case of one single random factor, an exact permutation procedure consists in restricting the permutations within the subjects. In more general cases, free permutations in repeated measures ANOVA designs would violate the exchangeability assumption. This is because the random effects associated with subjects and their interactions with fixed effects imply a complex structure for the (full) covariance matrix of observations. It follows that the second moments are not preserved after permutation. Friedrich et al. (2017) have derived exact asymptotic properties in those designs for a Wald-type statistic and Kherad-Pajouh and Renaud (2015) proposed several methods to transform the data following procedures developed by Kennedy (1995) or Kherad Pajouh and Renaud (2010).

For linear models, permutation tests are useful when the assumption of normality is violated or when the sample size is too small to apply asymptotic theory. In addition they can be used to control the family wise error rate (FWER) in some multiple comparisons settings (Troendle, 1995; Maris and Oostenveld, 2007; Smith and Nichols, 2009). These methods have been successfully applied for the comparison of experimental conditions in both functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) as they take advantage of the spatial and/or temporal correlation of the data.

The aim of the present article is to provide an overview of the use of permutation methods and multiple comparisons procedures using permutation tests and to explain how it can be used in `R` (Chambers, 2009) with the package `permuco`. Note that the presentation and discussion of the available packages that handle permutation tests in related settings is deferred to section 1.5.1, where all the notions are introduced. Appendix A.1 shows a comparison of the relevant code and outputs. But first, Section 1.2 focuses on fixed effect models. It explains the model used for ANOVA and regression and the various permutation methods proposed in the literature. Section 1.3 introduces the methods for repeated measures ANOVA. Section 1.4 explains the multiple comparisons procedures used for comparing signals between experimental conditions and how permutation tests are applied in this setting. Section 1.5 describes additional programming details and some of the choices for the default settings in the `permuco` package. Section 1.6 treats two real data analyses, one from a control trial in psychology and the second from an experiment in neurosciences using EEG.

1.2 The Fixed Effects Model

1.2.1 Model and Notation

For each hypothesis of interest, the fixed effects model (used for regression or ANOVA) can always be written as

$$y = D\eta + X\beta + \epsilon, \tag{1.1}$$

where $\underset{n\times 1}{y}$ is the response variable, $\left[\underset{n\times(p-q)}{D} \quad \underset{n\times q}{X} \right]$ is a design matrix split into the nuisance variable(s) D (usually including the intercept) and the variable(s) of interest X associated with the tested hypothesis. D and X may be correlated and we assume without loss of generality that $\left[D \quad X \right]$ is a full rank matrix. The parameters of the full model $\left[\begin{array}{cc}\underset{1\times(p-q)}{\eta^\top} & \underset{1\times q}{\beta^\top}\end{array}\right]^\top$ are also split into the parameters associated with the nuisance variable(s) η and the one(s) associated with the variable(s) of interest β. ϵ is an error term that follows a distribution $(0, \sigma^2 I_n)$. The hypothesis tested writes

$$H_0 : \beta = 0 \text{ vs. } H_1 : \beta \neq 0. \tag{1.2}$$

The permutation test is exact under the null hypothesis for finite samples if the data are exchangeable under the null hypothesis. This assumption is not fulfilled in model (1.1) as we cannot control the influence of the nuisance term $D\eta$ when permuting. In fact, under the null hypothesis (1.2), the responses follow a distribution $(D\eta, \sigma^2 I_n)$ which are not exchangeable due to the presence of unequal first moments. Pauly et al. (2015) show however that permuting the responses and using a Wald-type statistic is an asymptotically exact procedure in factorial designs. Another approach, which is the focus of this paper, is to transform the data prior to the permutation. Those transformation procedures are what will be called permutation methods. They are described in Chapter 1.2.2 and are implemented in `permuco`.

The permutation of a vector v is defined as Pv and the permutation of the rows of a matrix M as PM where P is a permutation matrix (Gentle, 2007, pp. 66-67). For any design matrix M, its corresponding "hat" matrix is $H_M = M(M^\top M)^{-1}M^\top$ and its corresponding "residuals" matrix is $R_M = I - M(M^\top M)^{-1}M^\top$ (Greene, 2011, pp. 24-25). The full QR-decomposition is:

$$\left[\begin{array}{cc}\underset{n\times n}{M \quad 0}\end{array}\right] = \left[\begin{array}{cc} Q_M & V_M \end{array}\right]\left[\begin{array}{cc} U_M & 0 \\ 0 & 0 \end{array}\right], \tag{1.3}$$

where $\underset{n\times p}{Q_M}$ and $\underset{n\times(n-p)}{V_M}$ define together an orthonormal basis of $\mathbb{R}^n$ and where $\underset{p\times p}{U_M}$ is interpreted as M in the subspace of Q_M. An important property of the QR-decomposition is that $H_M = Q_M Q_M^\top$ and $R_M = V_M V_M^\top$ (Seber and Lee, 2012, pp. 340-341).

1.2.2 Permutation Methods for Linear Models and Factorial ANOVAs

The discussed permutation methods are functions that transform the data in order to reduce the effect of the nuisance variables. They can be computed for all permutations $P \in \mathscr{P}$ where $\mathscr{P}$ is the set of all n_P distinct permutation matrices of the same size. For any permutation matrix P, a given permutation method will transform the observed data $\{y, D, X\}$ into the permuted data $\{y^*, D^*, X^*\}$. The `permuco` package provides several permutation methods that are summarized in table 1.1 using a notation inspired by Winkler et al. (2014).

The default method of `permuco` is the `freedman_lane` method that works as follows: we first fit the "small" model which only uses the nuisance variables D as predictors. Then, we permute its residuals and add them to the fitted values. Theses steps produce the permuted response variable y^* which constitutes the "new sample". It is fitted using the unchanged design D and X. In this procedure, only the residuals are permuted and they are supposed to share the same expectation (of zero) under the null hypothesis. For each permutation, the effect of nuisance variables is hence reduced. Using the above notation, the fitted values of the "small" model can be written as $H_D y$ and its residuals $R_D y$. Its permuted version is pre-multiplied by a permutation matrix, e.g., $PR_D y$. The permuted

Table 1.1: Permutation methods in the presence of nuisance variables. See text for explanations of the symbols.

`method`/Authors	y^*	D^*	X^*
`manly` (Manly, 1991)	Py	D	X
`draper_stoneman` (Draper and Stoneman, 1966)	y	D	PX
`dekker`(Dekker et al., 2007)	y	D	PR_DX
`kennedy` (Kennedy, 1995)	$(PR_D)y$		R_DX
`huh_jhun` (Huh and Jhun, 2001)	$(PV_D^\top R_D)y$		$V_D^\top R_DX$
`freedman_lane` (Freedman and Lane, 1983)	$(H_D + PR_D)y$	D	X
`terBraak` (ter Braak, 1992)	$(H_{X,D} + PR_{X,D})y$	D	X

response variable is therefore simply written as $y^* = H_Dy + PR_Dy = (H_D + PR_D)y$, as displayed in table 1.1. The permuted statistics (e.g. t or F statistics) are then computed using y^* and the unchanged design matrices $D^* = D$ and $X^* = X$.

All the remaining permutation methods are also summarized by the transformation of y, D and X into y^*, X^* and D^* and are explained next. The `manly` method simply permutes the response (this method is sometimes called raw permutations). Even if this method does not take into account the nuisance variables, it still has good asymptotic properties when using studentized statistics. `draper_stoneman` permutes the design of interest (note that without nuisance variables permuting the design is equivalent to permuting the response variable). However, this method ignores the correlation between D and X that is typically present in regressions or unbalanced designs. For the `dekker` method, we first orthogonalize X with respect to D, then we permute the design of interest. This transformation reduces the influence of the correlation between D and X and is more appropriate for unbalanced design. The `kennedy` method orthogonalizes all of the elements (y, D and X) with respect to the nuisance variables, removing the nuisance variables in the equation, and then permutes the obtained response. Doing so, all the design matrices lie in the span of X, a sub-space of observed design X and D. However this projection modifies the distribution of the residuals that lose exchangeability ($R_Dy \sim (0, R_D\sigma^2)$ for original IID data). The `huh_jhun` method is similar to `kennedy` but it applies a second transformation ($V_D^\top$) to the data to ensure exchangeability (up to the second moment, $V_D^\top R_Dy \sim (0, I_{n-(p-q)}\sigma^2)$). The V_D matrix comes from the Equation 1.3 and has a dimension of $n \times (n - (p - q))$. It implies that the P's matrices for the `huh_jhun` method have smaller dimensions. The `terBraak` method is similar to `freedman_lane` but uses the residuals of the full model. This permutation method creates a new response variable y^* which assumes that the observed value of the estimate $\hat{\beta}|y$ is the true value of β. Computing the statistic using y^*, X, D would not produce a permutation distribution under the null hypothesis. To circumvent this issue, the method changes the null hypothesis when computing the statistics at each permutation to $H_0 : \beta = \hat{\beta}|y = (X^\top R_DX)^{-1}X^\top R_Dy|y$. The right part of this new hypothesis corresponds to the observed estimate of the parameters of interest under the full model, and implicitly uses a pivotal assumption. Note that `terBraak` is the only method where the statistic computed with the identity permutation is different from the observed statistic. The notation $R_{D,X}$ means that the residuals matrix is based on the concatenation of the matrices D and X. See section 1.5.2 for advises on the choice of the method.

For each of the methods presented in Table 1.1, permutation tests can be computed using different statistics. For univariate or multivariate β parameters, the `permuco` package implemented a F statistic that constitutes a marginal test (or “type III” sum of square) (Searle, 2006, pp. 53-54). For a univariate $\underset{1\times 1}{\beta}$, one- and two-sided tests (based on a t-statistic) are also implemented. We write the F statistic as:

$$F = \frac{y^\top H_{R_D X} y}{y^\top R_{D,X} y}\frac{n-p}{p-q}. \tag{1.4}$$

When $q = 1$, the t statistic is:

$$t_{St} = \frac{(X^\top R_D X)^{-1} X R_D y}{\sqrt{y^\top R_{D,X} y (X^\top R_D X)^{-1}}}\sqrt{n-p}, \tag{1.5}$$

where the numerator is the estimate of β under the full model. Note that the statistic can be simplified by a factor of $(X^\top R_D X)^{-1/2}$. The two statistics are function of data. They lead to the general notation $t = t(y, D, X)$ when applied to the observed data and to $t^* = t(y^*, D^*, X^*)$ when applied to the permuted data. The permuted statistics constitute the set $\mathscr{T}$ which contains the t^* for all $P \in \mathscr{P}$. We define the permuted p value as $p = \frac{1}{n_P}\sum_{t^*\in\mathscr{T}} I\left(|t^*| \geq |t|\right)$, for a two-tailed t test, $p = \frac{1}{n_P}\sum_{t^*\in\mathscr{T}} I\left(t^* \geq t\right)$, for an upper-tailed t test or an F test and finally $p = \frac{1}{n_P}\sum_{t^*\in\mathscr{T}} I\left(t^* \leq t\right)$, for a lower-tailed t test, where $I(\cdot)$ is the indicator function.

1.3 Repeated Measures ANOVA

1.3.1 Model and Notation

We write the repeated measures ANOVA model in a linear mixed effects form:

$$y = D\eta + X\beta + E^0\kappa + Z^0\gamma + \epsilon, \tag{1.6}$$

where $\underset{n\times 1}{y}$ is the response, the fixed part of the design is split into the nuisance variable(s) $\underset{n\times(p_1-q_1)}{D}$, and the variable(s) of interest $\underset{n\times(p_1)}{X}$. The specificity of the repeated measures ANOVA model allows us to split the random part into $\underset{n\times(p_2^0-q_2^0)}{E^0}$ and $\underset{n\times q_2^0}{Z^0}$ which are the random effects associated with D and X respectively (Kherad-Pajouh and Renaud, 2015). The fixed parameters are $\left[\underset{1\times(p_1-q_1)}{\eta^\top} \quad \underset{1\times q_1}{\beta^\top}\right]^\top$. The random part is $\left[\underset{1\times(p_2^0-q_2^0)}{\kappa^\top} \quad \underset{1\times q_2^0}{\gamma^\top}\right]^\top \sim (0,\Omega)$ and $\epsilon \sim (0, \sigma^2 I)$. The matrices associated with the random effects E^0 and Z^0 can be computed using:

$$E^0 = (D^{0\prime}_{within} * Z^{0\prime}_{\Delta})^\top \text{ and } Z^0 = (X^{0\prime}_{within} * Z^{0\prime}_{\Delta})^\top, \tag{1.7}$$

where D^0_{within} and X^0_{within} are overparametrized matrices and are associated with the within effects in the design matrices D and X. Z^0_Δ is the overparametrized design matrix associated to the subjects and $*$ is the column-wise Khatri-Rao product (Khatri and Rao, 1968). Since the matrices E^0 and Z^0 are overparametrized, it is not convenient to compute their corresponding sum of squares. We need versions that are constrained into their respective appropriate sub-spaces:

$$E = R_{D,X}E^0 \text{ and } Z = R_{D,X}Z^0. \tag{1.8}$$

Table 1.2: Permutation methods in the presence of nuisance variables for repeated measures ANOVA.

method	y^*	D^*	X^*	E^*	Z^*
`Rd_keradPajouh_renaud` (R_D)	PR_Dy		R_DX		R_DZ
`Rde_keradPajouh_renaud` ($R_{D,E}$)	$PR_{D,E}y$		$R_{D,E}X$		$R_{D,E}Z$

The matrices E and Z are respectively of rank $p_2 - q_2$ and q_2 and are the ones used to compute F statistics. Formally, the hypothesis of interest associated with Equation 1.6 writes:

$$H_0 : \beta = 0 \text{ vs. } H_1 : \beta \neq 0. \tag{1.9}$$

1.3.2 Permutation Methods for Repeated Measures ANOVA

Similarly to the fixed effects model, we can test hypotheses using permutation methods (Kherad-Pajouh and Renaud, 2015). The ones that are implemented in the `permuco` package are given in Table 1.2. The two methods are based on a similar idea. By premultiplying the design and response variables by R_D or $R_{D,E}$, we orthogonalize the model to the nuisance variables. This procedure can be viewed as an extension of the `kennedy` procedure (see table 1.1) to repeated measures ANOVA.

The hypothesis in (1.9) is tested based on the conventional F statistic for repeated measures ANOVA:

$$F = \frac{y^\top H_{R_D X} y}{y^\top H_Z y} \frac{p_2}{p_1}. \tag{1.10}$$

As for the fixed effects model, the statistic is written as a function of the data $t = t(y, D, X, E, Z)$ and the permuted statistic $t^* = t(y^*, D^*, X^*, E^*, Z^*)$ is a function of the permuted data under the chosen method. The p value is defined as in the fixed effect case.

1.4 Signal and Multiple Comparisons

In EEG experiments, researchers are often interested in testing the effect of conditions on the event-related potential (ERP). It is a common practice to test the signals at each time point of the ERP. In that kind of experiments, thousands of tests are typically carried out (e.g., one measure every $2ms$ over 2 seconds) and the basic multiple hypotheses corrections like Bonferroni (Dunn, 1958) are useless as their power is too low.

Troendle (1995) proposed a multiple comparisons method that considers the correlation between the re-sampling data. This method does not specifically use the time-neighbourhood information of a signal but uses wisely the general correlation between the statistics and may be used in more general settings.

Better known, the cluster-mass test (Maris and Oostenveld, 2007) has shown to be powerful while controlling the family-wise error rate (FWER) in EEG data analysis. And recently using a similar idea, the threshold-free cluster-enhancement (TFCE) was developed for fMRI data (Smith and Nichols, 2009) and EEG data (Pernet et al., 2015), but usually presented only with one factor.

All these approaches use permutations and are compatible with the methods displayed in Tables 1.1 and 1.2, as shown next. In addition to multiple comparisons procedures that use permutation, the well-known Bonferroni and Holm (Holm, 1979) corrections and the control of the false positive rate by Benjamini and Hochberg (1995) are also implemented in `permuco`.

1.4.1 Model and Notation

We can construct a model at each time point $s \in \{1, \dots, k\}$ for the fixed effects design as:

$$y_s = D\eta_s + X\beta_s + \epsilon_s, \tag{1.11}$$

where y_s is the response variable for all observations at time s and each of the k models are the same as (1.1). D and X, the design matrices, are then identical over the k time points. The aim is to test simultaneously all k hypotheses $H_0^s : \beta_s = 0$ vs. $H_1^s : \beta_s \neq 0$ for $s \in \{1, \dots, k\}$ while controlling for the FWER through the k tests. Likewise, the random effects model is written:

$$y_s = D\eta_s + X\beta_s + E^0\kappa_s + Z^0\gamma_s + \epsilon_s, \tag{1.12}$$

where each of the k models are defined as in (1.6) and, similarly, we are interested to test the k hypotheses $H_0^s : \beta_s = 0$ vs. $H_1^s : \beta_s \neq 0$ for $s \in \{1, \dots, k\}$.

For both models, we choose one of the permutation methods presented in Tables 1.1 or 1.2 and compute the k observed statistics t_s, the k sets of permutated statistics $\mathscr{T}_s$, which lead to k raw or uncorrected p values.

To correct them, the k sets of permutated statistics $\mathscr{T}_s$ can be analyzed as one set of multivariate statistic. It is done simply by combining the k univariate permutation-based distributions into a single k-variate distribution which maintains the correlation between tests. For each permutation, we simply combine all k univariate permuted statistics $t_1^*, \dots, t_k^*$ into one multivariate permuted statistic $\mathbf{t}^* = [t_1^* \ \dots \ t_k^*]^\top$. The three multiple comparisons procedures described below are all based on this multivariate distribution and take advantage of the correlation structure between the tests.

1.4.2 Troendle's Step-Wise Re-Sampling Method

The method developed by Troendle (1995) takes advantage of the form of the multivariate resampling distribution of the t_s^*. If we assume that t_s is distributed according to T_s then by ordering the observed statistics t_s we obtain $t_{(1)} \leq \dots \leq t_{(s)} \leq \dots \leq t_{(k)}$ with their corresponding k null hypotheses $H_{(1)} \leq \dots \leq H_{(s)} \leq \dots \leq H_{(k)}$. Then Troendle (1995) use the following arguments. First, for all s, controlling the FWER with $P_{H_{(1)},\dots,H_{(k)}}\left(\max_{i\in\{1,\dots,k\}} T_{(i)} \leq t_{(s)}\right) < \alpha_{FWER}$ is a conservative approach. Secondly, if we reject $H_{(k)}$ and want to test $H_{(k-1)}$, we can safely assume that $H_{(k)}$ is false while controlling the FWER. Either $H_{(k)}$ is true and we already made a type I error or was wrong and we can go as if $H_{(k)}$ was absent. We can then update our decision rule for testing $H_{(k-1)}$ by $P_{H_{(1)},\dots,H_{(k-1)}}\left(\max_{i\in\{1,\dots,k-1\}} T_{(i)} \leq t_{(k-1)}\right) < \alpha_{FWER}$. We continue until the first non-significant result and declare all s with a smaller t statistic as non-significant.

This procedure is valid in a general setting and is easly implemented for permutation tests. The permuted sets $\mathscr{T}_s$ is interpreted as a nonparametric distribution of the T_s and based on Troendle (1995), we use the following algorithm to compute the corrected p value:

Algorithm 1 Troendle corrected p value

```
Order the $k$ observed statistics $t_s$ into $t_{(1)} \leq \cdots \leq t_{(s)} \leq \cdots \leq t_{(k)}$
for $i \in \{1, \dots k\}$ do
    Define the null distribution $\mathscr{S}_{(k-i+1)}$ for $t_{(k-i+1)}$ by:
    for each $P \in \mathscr{P}$ do
        Return the maximum over the $k-i+1$ first values $t^*_{(s)}$ for $s \in \{1, \dots, k-i+1\}$
    Define the corrected $p$ value $p_{(k-i+1)} = \frac{1}{n_P} \sum_{t^* \in \mathscr{S}_{(k-i+1)}} I\left(t^* \geq t_{(k-i+1)}\right)$
    Control for a stepwise procedure by:
    if $p_{(k-i+1)} < p_{(k-i+2)}$ and $i > 1$ then $p_{(k-i+1)} \coloneqq p_{(k-i+2)}$
```

1.4.3 Cluster-Mass Statistic

This method has been proposed by Maris and Oostenveld (2007) and is commonly implemented in specialised software of EEG data analysis like LIMO (Pernet et al., 2011). It relies on a continuity argument that implies that an effect will appear into clusters of adjacent timeframes. Based on all time-specific statistics, we form these clusters using a threshold τ as follows (see Figure 1.1). All the adjacent time points for which the statistics are above this threshold define one cluster C_i for $i \in [1, \dots, n_c]$, where n_c is the number of clusters found in the k statistics. We assign to each time point in the same cluster C_i, the same cluster-mass statistic $m_i = f(C_i)$ where f is a function that aggregates the statistics of the whole cluster into a scalar; typically the sum of the F statistics or the sum of squared of the t statistics. The cluster-mass null distribution $\mathscr{M}$ is computed by repeating the process described above for each permutation. The contribution of a permutation to the cluster-mass null distribution is the maximum over all cluster-masses for this permutation. This process is described in Algorithm 2.

Algorithm 2 Cluster-mass null distribution $\mathscr{M}$

```
for each $P \in \mathscr{P}$ do
    Compute the $k$ permuted statistics $t^*_s$ for $s \in \{1, \dots, k\}$.
    Find the $n^*_c$ clusters $C^*_i$ as the sets of adjacent time points which statistic is above $\tau$.
    Compute the cluster-mass for each cluster $m^*_i = f(C^*_i)$
    Return the maximum value over the $n^*_c$ values $m^*_i$.
```

To test the significance of an observed cluster C_i, we compare its cluster-mass $m_i = f(C_i)$ with the cluster-mass null distribution $\mathscr{M}$. The p value of the effect at each time within a cluster C_i is the p value associated with this cluster, i.e. $p_i = \frac{1}{n_P} \sum_{m^* \in \mathscr{M}} I(m^* \geq m_i)$.

In addition to the theoretical properties of this procedure (Maris and Oostenveld, 2007), this method makes sense for EEG data analysis because if a difference of cerebral activity is believed to happen at a time s for a given factor, it is very likely that the time $s+1$ (or $s-1$) will show this difference too.

1.4.4 Threshold-Free Cluster-Enhancement

Although it controls (weakly) the FWER for any a priori choice of threshold, the result of the cluster-mass procedure is sensitive to this choice. The TFCE (Smith and Nichols,

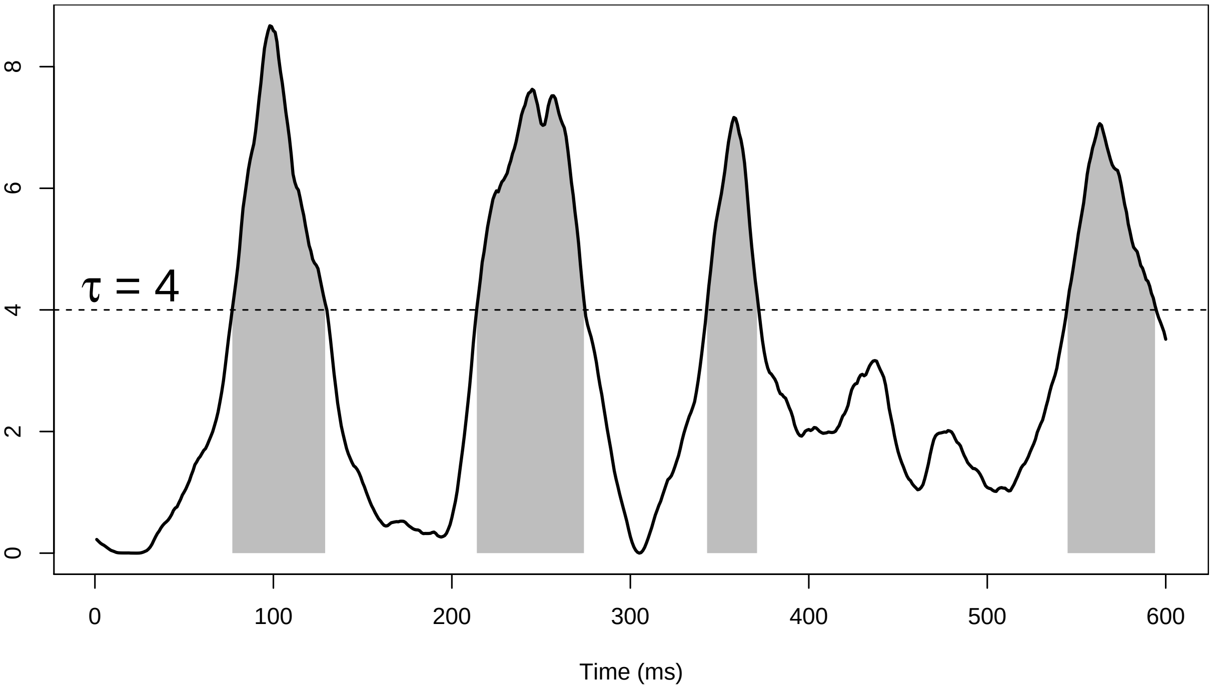

Figure 1.1: Display of the 600 statistics corresponding to the tests on 600 time points. Here 4 clusters are found using a threshold $\tau = 4$. Using the sum to aggregate the statistics, for each cluster i, the shaded area underneath the curve represents its cluster-mass m_i.

2009) is closely related to the cluster-mass but gets rid of this seemingly arbitrary choice. It is defined at each time $s \in [1, \dots, k]$ for the statistics t_s as:

$$u_s = \int_{h=t_0}^{h=t_s} e(h)^E h^H dh, \tag{1.13}$$

where $e(h)$ is the extend at the height h and it is interpreted as the length of a cluster for a threshold of h. E and H are free parameters named the extend power, and the height power respectively. t_0 is set close to zero. Figure 1.2 illustrates how the TFCE statistic is computed for a given time point s.

We construct the TFCE null distribution $\mathscr{U}$ by applying the formula in (1.13) at each time-point of the permuted statistics t_s^* for $s \in \{1, \dots, k\}$ to produce for each permutation, k values u_s^*. Then the contribution of a permutation to $\mathscr{U}$ is the maximum of all k values u_s^*. In practice, the integral in (1.13) is approximated numerically using small $dh \leq 0.1$, (Smith and Nichols, 2009, Pernet et al. (2015)).

At time s, the statistic t_s will be modified using the formula in (1.13). The formula can be viewed as a function of characteristics in the grey area (its area in the special case where both E and H are set to 1).

Algorithm 3 Threshold-free cluster-enhancement null distribution $\mathscr{U}$

```
for each $P \in \mathscr{P}$ do
    Compute the $k$ permuted statistics $t_s^*$ for $s \in \{1, \dots, k\}$
    Compute the $k$ enhanced statistics $u_s^*$ using a numerical approximation of (1.13)
    Return the maximum over the k value $u_s^*$
```

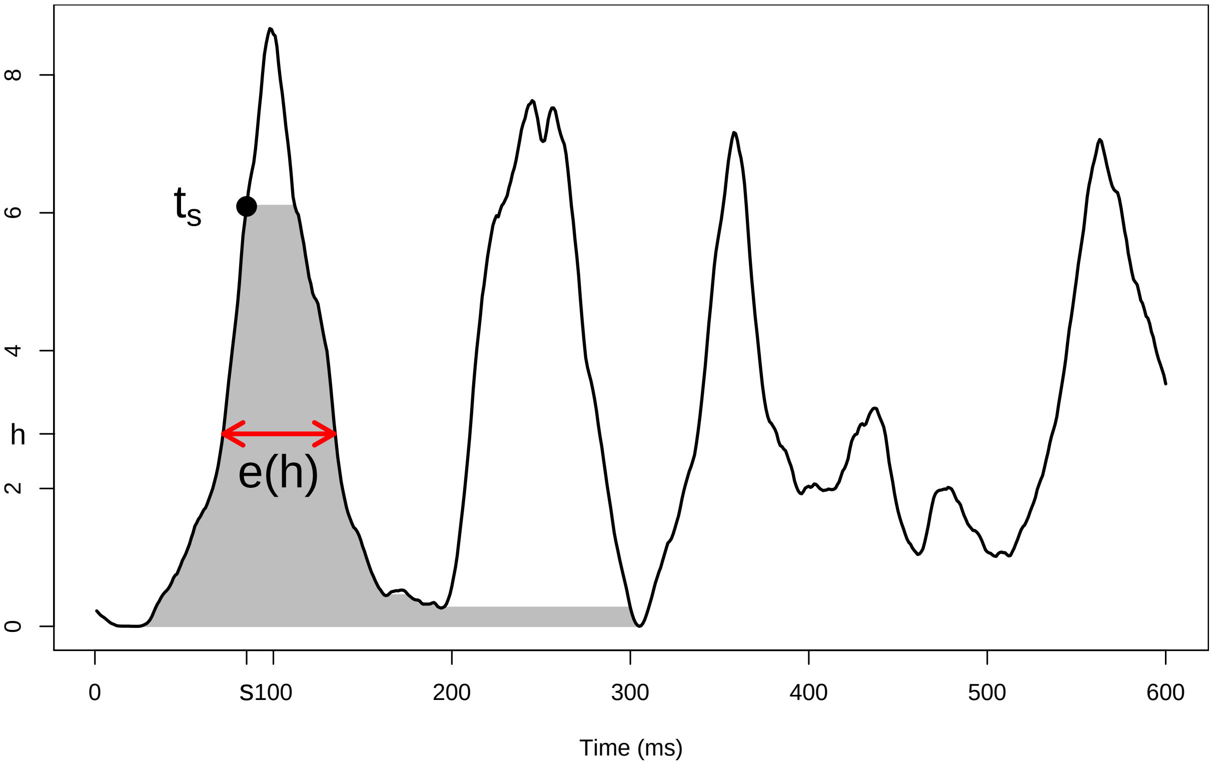

Figure 1.2: The TFCE transforms the statistic t_s using formula in (1.13). The extend $e(h)$, in red, is shown for a given height h. The TFCE statistics u_s at s can be viewed as a function of characteristics in the grey area.

To test the significance of a time point s we compare its enhanced statistics u_s with the threshold-free cluster-enhancement null distribution $\mathscr{U}$. For an F test we define the p value as $p_s = \frac{1}{n_P} \sum_{u^* \in \mathscr{U}} I(u^* \geq u_s)$.

1.4.5 Interpreting Cluster Based Inference

The cluster-mass test and the TFCE are methods based on clustering time-points and the interpretation of significant findings is then not intuitive. First, note that Bonferroni, Holm, the control of the false positive rate and Troendle's method are not based on clustering and do not have these issues. Their interpretation is straight-forwards as we can interpret individually each discovery. For the cluster-mass test the interpretation should be done at a cluster level: a significant cluster is a cluster which contains at least one significant time-point. It follows that the cluster-mass test does not allow the interpretation of the precise location of clusters (Sassenhagen and Draschkow, 2019). Intuitively, the cluster-mass test is a two steps procedure: first, it aggregates time-points into clusters, and then summarizes them using the cluster-mass. The inference is only performed at the second step which looses any information on the location of the clusters. It implies that the interpretation of individual time-point is proscribed. Finally, the TFCE statistic is an integration over all thresholds of cluster statistics (Smith and Nichols, 2009). Therefore, the TFCE does not allow an interpretation of each time-point individually either as it also summarizes statistics using the concept of clusters. It implies that a significant time-point must be interpreted as a time-point being part of at least one significant cluster (among all clusters formed using all thresholds), where a significant cluster contains at least one significant time-point.

1.5 Comparison of Packages, Parameters Choices and Implementation Details

1.5.1 Comparison of Packages

Several packages for permutation tests are available for `R` in CRAN. Since permutation tests have such a variety of applications, we only review packages (or the part of packages) that handle regression, ANOVA or comparison of signals.

For testing one factor, the `perm` (Fay and Shaw, 2010), `wPerm` (Weiss, 2015) and `coin` (Hothorn et al., 2008) packages produce permutation tests of differences of locations between two or several groups. The latter can also test the difference within groups or block, corresponding to a one within factor ANOVA.

The package `lmPerm` (Wheeler and Torchiano, 2016) produces tests for multifactorial ANOVA and repeated measures ANOVA. It computes sequential (or Type I) and marginal (or Type III) tests for factorial ANOVA and ANCOVA but only the sequential is implemented for repeated measures, even when setting the parameter `seqs = FALSE`. The order of the factors will therefore matter in this case. The permutation method consists in permuting the raw data even in the presence of nuisance variables, which correspond to the `manly` method, see Table 1.1. For repeated measures designs, data are first projected into the `"Error()"` strata and then permuted, a method that has not been validated (to our knowledge) in any peer-reviewed journal. Additionally, `lmPerm` by default uses a stopping rule based on current p value to define the number of permutations. By default, the permutations are not randomly sampled but modified sequentially merely on a single pair of observations. This speeds up the code but the quality of the obtained p value is not well documented.

The `flip` package (Finos et al., 2014) produces permutation and rotation tests (Langsrud, 2005) for fixed effects and handles nuisance variables based on methods similar to the `huh_juhn` method of table 1.1. It performs tests in designs with random effects only for singular models (e.g. repetition of measures by subjects in each condition) with method based on Basso and Finos (2012) and Finos and Basso (2014) to handle nuisance variables.

The `GFD` package (Friedrich, Sarah et al., 2017) produces marginal permutation tests for pure factorial design (without covariates) with a Wald-type statistic. The permutation method is `manly`. This method has been shown to be asymptotically exact even under heteroscedastic conditions (Pauly et al., 2015).

To our knowledge, only the `permuco` package provides tests for comparison of signals.

The codes and outputs for packages that perform ANOVA/ANCOVA are given in Appendix A.1.1 and in Appendix A.1.2 for repeated measures. For fixed effects, this illustrates that `permuco`, `flip` and `lmPerm` handle covariates and are based on the same statistic (F) whereas `GFD` uses the Wald-type statistic. It also shows that `flip` is testing one factor at a time (main effect of `sex` in this case) whereas the other packages produce directly tests for all the effects. Also, the nuisance variables in `flip` must be carefully implemented using the appropriate coding variables in case of factors. Note that `lmPerm` centers the covariates using the default setting and that it provides both marginal (Type III) or sequential (Type I) tests.

Concerning permutation methods, only the `manly` method is used for both `lmPerm` and `GFD`, the `flip` package uses the `huh_jhun` method, whereas multiple methods can be set by users using the `permuco` package. Note also that different default choices for the V matrix as implemented in `flip` (based on eigendecomposition) and `permuco` (based on QR decomposition) packages lead to slightly different results (see Table 1.1 for more information on the permutation methods).

Finally, concerning repeated measures designs, `flip` cannot handle cases where measures are not repeated in each condition for each subject, and therefore cannot be compared in Appendix A.1.2. As already said, `lmPerm` produces sequential tests in repeated measures designs and `permuco` produces marginal tests. This explains why, with unbalanced data, only the last interaction term in each strata produces the same statistic.

1.5.2 Permutation Methods

For the fixed effects model, simulations (Kherad Pajouh and Renaud, 2010; Winkler et al., 2014) show that the method `freedman_lane`, `dekker`, `huh_jhun` and `terBraak` perform well, whereas `manly`, `draper_stoneman` and `kennedy` can be either liberal or conservative. Moreover Kherad Pajouh and Renaud (2010) provide a proof for an exact test of the `huh_jhun` method under sphericity. Note that `huh_jhun` will reduce the dimensionality of the data and if $n-(p-q)\leq 7$ the number of permutations may be too low. Based on all the above literature the default method for the `permuco` package is set to `freedman_lane`.

For the random effects model, Kherad-Pajouh and Renaud (2015) show that a more secure approach is to choose the `Rde_keradPajouh_renaud` method.

All $n!$ permutations are not feasible already for moderate sized datasets. A large subset of permutation is used instead, and it can be tuned with the `np` argument. The default value is `np = 5000`. Winkler et al. (2016) recall that with `np = 5000` the 0.95% confidence interval around $p=0.05$ is relatively small: $[0.0443; 0.0564]$. For replicability purpose, the `P` argument can be used instead of the `np` argument. The `P` argument needs a `Pmat` object which stores all permutations. For small datasets, if the `np` argument is greater than the number of possible permutations ($n!$), the tests will be done on all permutations. This can be also be selected manually by setting `type = "unique"` in the `Pmat` functions.

Given the inequality sign in the formulas for the p value described at the end of section 1.2.2, the minimal p value is 1/ `np`, which is a good practice for permutation tests. Moreover this implies that the sum of the two one-sided p values is slightly greater than 1.

The `huh_jhun` method is based on a random rotation that can be set by a random $n\times n$ matrix in the `rnd_rotation` argument. This random matrix will be orthogonalized by a QR decomposition to produce the proper rotation. Note that the random rotation in the `huh_jhun` method allows us to test the intercept, which is not available for the other methods.

1.5.3 Multiple Comparisons

The `multcomp` argument can be set to `"bonferroni"` for the Bonferroni correction (Dunn, 1958), to `"holm"` for the Holm correction (Holm, 1979), to `"troendle"`, see chapter 1.4.2, to `"clustermass"`, see chapter 1.4.3 and to `"tfce"`, see chapter 1.4.4.

Moreover, to control the false discovery rate using the method proposed by Benjamini and Hochberg (1995), the `multcomp` argument can be set to `"benjamini_hochberg"`. Note that in the `permuco` package, these 6 methods are available in conjunction with permutation, although the first 3 methods are general procedures that could also be used in a parametric setting.

For the `"clustermass"` method, the `threshold` parameter of the cluster-mass statistic is usually chosen by default at the 0.95 quantile of the corresponding univariate parametric distribution; but the FWER is preserved for any a priori value of the `threshold` that the user may set. The mass function is specified by the `aggr_FUN` argument. It is set by default to the sum of squares for a t statistic and the sum for an F. It should be a function that returns a positive scalar which will be large for an uncommon event under the null hypothesis (e.g., use the sum of absolute value of t statistics instead of the sum). It can be tuned depending on the expected signal. For the t statistic, typically, the sum of squares will detect more efficiently high peaks and the sum of absolute values will detect more efficiently wider clusters.

For the `"tfce"` method, the default value for the extend parameter is `E` $= 0.5$ and for the height `H` $= 2$ for t tests and, for F test, it is `E` $= 0.5$ and `H` $= 1$ following the recommendations of Smith and Nichols (2009) and Pernet et al. (2015). The `ndh` parameter controls the number of steps used in the approximation of the integral in (1.13) and is set to 500 by default.

The argument `return_distribution` is set by default to `FALSE` but can be set to `TRUE` to return the large matrices ($n_P \times k$) with the value of the permuted statistics.

The algorithm and formula presented in the previous sections may not be efficient for very large size of data. When available, they are implemented in a more efficient way in `permuco`. For example, to reduce the computing time, the permuted statistics are computed through a QR decomposition using the `qr`, `qr.fitted`, `qr.resid` or `qr.coef` functions.

1.6 Tutorial

To load the `permuco` package:

```
R> install.packages("permuco")
R> library("permuco")
```

1.6.1 Fixed Effects Model

The `emergencycost` dataset contains information from 176 patients from an emergency service (Heritier et al., 2009). The variables are the sex, the age (in years), the type of insurance (private/semiprivate or public), the length of the stay (`LOS`) and the cost. These observational data allow us to test which variables influence the cost of the stay of the patients. In this example, we will investigate the effect of the sex and of the type of insurance on the cost and we will adjust those effects by the length of the stay. To this end, we perform an ANCOVA and need to center the covariate.

```
R> emergencycost$LOSc <- scale(emergencycost$LOS, scale = F)
```

The permutation tests are obtained with the `aovperm` function. The `np` argument sets the number of permutations. We choose to set a high number of permutations (`np`

= 100000) to reduce the variablity of the permutation p values so that they can safely be compared to the parametric ones. The aovperm function automatically converts the coding of factors with the contr.sum which allows us to test the main effects of factors and their interactions.

```
R> mod_cost_0 <- aovperm(cost ~ LOSc * sex * insurance, data = emergencycost,
R>                       np = 100000)
R> mod_cost_0

Anova Table
Permutation test using freedman_lane to handle nuisance variables and
 1e+05 permutations.

                          SS  df        F parametric P(>F)
LOSc               2.162e+09   1 483.4422           0.0000
sex                1.463e+07   1   3.2714           0.0723
insurance          6.184e+05   1   0.1383           0.7105
LOSc:sex           8.241e+06   1   1.8427           0.1765
LOSc:insurance     2.911e+07   1   6.5084           0.0116
sex:insurance      1.239e+05   1   0.0277           0.8680
LOSc:sex:insurance 1.346e+07   1   3.0091           0.0846
Residuals          7.514e+08 168
                   permutation P(>F)
LOSc                          0.0000
sex                           0.0763
insurance                     0.6794
LOSc:sex                      0.1576
LOSc:insurance                0.0233
sex:insurance                 0.8537
LOSc:sex:insurance            0.0847
Residuals
```

The interaction LOSc:insurance is significant both using the parametric p value 0.0116 and the permutation one 0.0233 using a 5% level. However, the difference between these 2 p values is 0.0117 which is high enough to lead to different conclusions e.g., in case of correction for multiple tests or a smaller α level.

If we are interested in the difference between the groups for a high value of the covariate, we center the covariate to the third quantile (14 days) and re-run the analysis.

```
R> emergencycost$LOS14 <- emergencycost$LOS - 14
R> mod_cost_14 <- aovperm(cost ~ LOS14 * sex * insurance, data = emergencycost,
R>                        np = 100000)
R>
R> mod_cost_14

Anova Table
Permutation test using freedman_lane to handle nuisance variables and
 1e+05 permutations.
```

```
                            SS  df        F parametric P(>F)
LOS14              2.162e+09   1 483.4422            0.0000
sex                2.760e+07   1   6.1703            0.0140
insurance          9.864e+05   1   0.2206            0.6392
LOS14:sex          8.241e+06   1   1.8427            0.1765
LOS14:insurance    2.911e+07   1   6.5084            0.0116
sex:insurance      7.722e+05   1   0.1727            0.6783
LOS14:sex:insurance 1.346e+07   1   3.0091            0.0846
Residuals          7.514e+08 168
                   permutation P(>F)
LOS14                         0.0000
sex                           0.0224
insurance                     0.6082
LOS14:sex                     0.1576
LOS14:insurance               0.0233
sex:insurance                 0.6540
LOS14:sex:insurance           0.0847
Residuals
```

For a long length of stay, the effect of sex is significant using the parametric p value p = 0.014 and the permutation one p = 0.0224.

If the researcher has an a priori oriented alternative hypothesis $H_A : \beta_{sex=M} > \beta_{sex=F}$, the `lmperm` function produces one-sided t tests. To run the same models as previously, we first need to set the coding of the factors with the `contr.sum` function before running the permutation tests.

```
R> contrasts(emergencycost$insurance) <- contr.sum
R> contrasts(emergencycost$insurance)
```

```
             [,1]
public          1
semi_private   -1
```

```
R> contrasts(emergencycost$sex) <- contr.sum
R> contrasts(emergencycost$sex)
```

```
  [,1]
F    1
M   -1
```

```
R> modlm_cost_14 <- lmperm(cost ~ LOS14 * sex * insurance,
R>                         data = emergencycost, np = 100000)
R>
R> modlm_cost_14
```

```
Table of marginal t-test of the betas
Permutation test using freedman_lane to handle nuisance variables and
 100000 permutations.
```

```
                      Estimate Std. Error t value parametric Pr(>|t|)
(Intercept)            14217.0     360.17 39.4730              0.0000
LOS14                    845.5      38.45 21.9873              0.0000
sex1                    -894.7     360.17 -2.4840              0.0140
insurance1               169.1     360.17  0.4696              0.6392
LOS14:sex1               -52.2      38.45 -1.3575              0.1765
LOS14:insurance1          98.1      38.45  2.5512              0.0116
sex1:insurance1         -149.7     360.17 -0.4155              0.6783
LOS14:sex1:insurance1    -66.7      38.45 -1.7347              0.0846
                      permutation Pr(<t) permutation Pr(>t)
(Intercept)
LOS14                             1.0000             0.0000
sex1                              0.0152             0.9848
insurance1                        0.6823             0.3177
LOS14:sex1                        0.0796             0.9204
LOS14:insurance1                  0.9868             0.0132
sex1:insurance1                   0.3337             0.6663
LOS14:sex1:insurance1             0.0395             0.9605
                      permutation Pr(>|t|)
(Intercept)
LOS14                               0.0000
sex1                                0.0224
insurance1                          0.6082
LOS14:sex1                          0.1576
LOS14:insurance1                    0.0233
sex1:insurance1                     0.6540
LOS14:sex1:insurance1               0.0847
```

The effect sex1 is significant for both the parametric one-sided *p* value p = 0.007 and the permutation one-sided *p* value p = 0.0152. It indicates that when the length of the stay is high, men have a shorter cost than women.

To test the effect of the sex within the public insured persons (called simple effect), we change the coding of the factors inside the `data.frame` using the `contr.treatment` function and disable the automatic recoding using the argument `coding_sum = FALSE`.

```
R> contrasts(emergencycost$insurance) <- contr.treatment
R> emergencycost$insurance <- relevel(emergencycost$insurance, ref = "public")
R> contrasts(emergencycost$insurance)

             semi_private
public                  0
semi_private            1

R> contrasts(emergencycost$sex) <- contr.sum
R> contrasts(emergencycost$sex)

  [,1]
F    1
M   -1
```

```
R> mod_cost_se <- aovperm(cost ~ LOSc * sex * insurance, data = emergencycost,
R>                          np = 100000, coding_sum = FALSE)
R> mod_cost_se
```

```
Anova Table
Permutation test using freedman_lane to handle nuisance variables and
 1e+05 permutations.

                          SS  df         F parametric P(>F)
LOSc               9.512e+09   1 2126.7539           0.0000
sex                6.092e+07   1   13.6210           0.0003
insurance          6.184e+05   1    0.1383           0.7105
LOSc:sex           1.510e+08   1   33.7708           0.0000
LOSc:insurance     2.911e+07   1    6.5084           0.0116
sex:insurance      1.239e+05   1    0.0277           0.8680
LOSc:sex:insurance 1.346e+07   1    3.0091           0.0846
Residuals          7.514e+08 168
                   permutation P(>F)
LOSc                          0.0000
sex                           0.0004
insurance                     0.6794
LOSc:sex                      0.0000
LOSc:insurance                0.0233
sex:insurance                 0.8537
LOSc:sex:insurance            0.0847
Residuals
```

The sex row can be interpreted as the effect of sex for the `public` insured persons for an average length of stay. Both the parametric $p = 0.0003$ and permutation p value $p = 0.0004$ show significant effect of sex within the public insured persons.

Given the skewness of the data for each case where the permutation test differs from the parametric result, we tend to put more faith on the permutation result since it does not rely on assumption of normality.

1.6.2 Repeated Measures ANCOVA

The `jpah2016` dataset contains a subset of a control trial in impulsive approach tendencies toward physical activity or sedentary behaviors. It contains several predictors like the body mass index, the age, the sex, and the experimental conditions. For the latter, the subjects were asked to perform different tasks: to approach physical activity and avoid sedentary behavior (`ApSB_AvPA`), to approach sedentary behavior and avoid physical activity (`ApPA_AvSB`) and a control task. The dependent variables are measures of impulsive approach toward physical activity (`iapa`) or sedentary behavior (`iasb`). See Cheval et al. (2016) for details on the experiment. We will analyze here only a part of the data.

```
R> jpah2016$bmic <- scale(jpah2016$bmi, scale = F)
```

We perform the permutation tests by running the `aovperm` function. The within subject factors should be written using `+ Error(...)` similarly to the `aov` function from the `stats` package:

```
R> mod_jpah2016 <- aovperm(iapa ~ bmic * condition * time + Error(id/(time)),
R>                         data = jpah2016, method = "Rd_kheradPajouh_renaud")
```

The results are shown in an ANOVA table by printing the object:

```
R> mod_jpah2016
```

```
Permutation test using Rd_kheradPajouh_renaud to handle nuisance
variables and 5000 permutations.

                           SSn dfn      SSd dfd       MSEn      MSEd
bmic                   18.6817   1 106883.5  13    18.6817  8221.808
condition           27878.1976   2 106883.5  13 13939.0988  8221.808
bmic:condition      89238.4780   2 106883.5  13 44619.2390  8221.808
time                  268.8368   1 167304.9  13   268.8368 12869.607
bmic:time             366.4888   1 167304.9  13   366.4888 12869.607
condition:time      21159.7735   2 167304.9  13 10579.8867 12869.607
bmic:condition:time 29145.7201   2 167304.9  13 14572.8601 12869.607
                         F parametric P(>F) permutation P(>F)
bmic                0.0023           0.9627            0.9646
condition           1.6954           0.2217            0.2180
bmic:condition      5.4269           0.0193            0.0230
time                0.0209           0.8873            0.8808
bmic:time           0.0285           0.8686            0.8594
condition:time      0.8221           0.4611            0.4520
bmic:condition:time 1.1323           0.3521            0.3412
```

This analysis reveals a significant p value for the effect of the interaction bmic:condition with a statistic $F = 5.4269$, which lead to a permutation p value $p = 0.023$ not far from the parametric one. For this example, the permutation tests backs the parametric analysis. The permutation distributions can be viewed using the plot function like in Figure 1.3.

```
R> plot(mod_jpah2016, effect = c("bmic", "condition", "bmic:condition"))
```

1.6.3 EEG Experiment in Attention Shifting

attentionshifting_signal and attentionshifting_design are data provided in the permuco package. They come from an EEG recording of 15 participants watching images of either neutral or angry faces (Tipura et al., 2017). Those faces were shown at a different visibility: subliminal (16ms) and supraliminal (166ms) and were displayed to the left or to the right of a screen. The recording is at 1024Hz for 800ms. Time 0 is when the image appears (event-related potential or ERP). The attentionshifting_signal dataset contains the ERP of the electrode O1. The design of experiment is given in the attentionshifting_design dataset along with the laterality, sex, age, and 2 measures of anxiety of each subjects, see Table 1.3.

As almost any ERP experiment, the data is designed for a repeated measures ANOVA. Using the permuco package, we test each time points of the ERP for the main effects and the interactions of the variables visibility, emotion and direction while controlling for the FWER. We perform F tests using a threshold at the 95% quantile, the sum as a cluster-mass statistics and 5000 permutations. We handle nuisance variables with the method Rd_kheradPajouh_renaud:

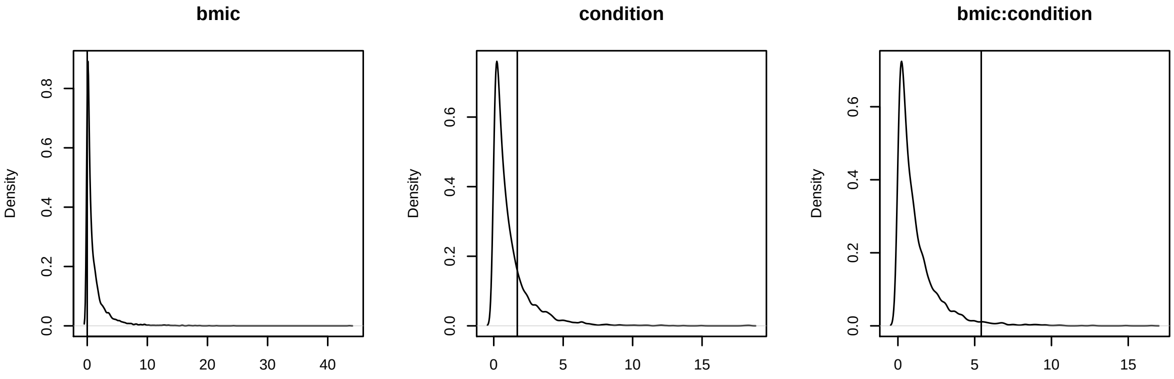

Figure 1.3: The permutation distributions of the F statistics for the effects `bmic`, `condition` and `bmic:condition`. The vertical lines indicate the observed statistics.

Table 1.3: Variables in the `attentionshifting_design` dataset.

Variable name	Description	Levels
`id`	number of identification	15 subjects
`visibility`	time that the image is shown	`16ms 166ms`
`emotion`	emotion of the shown faces	`angry`, `neutral`
`direction`	position of the faces on the screen	`left`, `right`
`laterality_id`	measure of the laterality of the subjects	scale from 25 to 100
`age`	age of the subjects	from 18 to 25
`sex`	sex of the subjects	`male`, `female`
`STAIS_state`	state anxiety score of the subjects	
`STAIS_trait`	trait anxiety score of the subjects	

```
R> electrod_O1 <-
R>   clusterlm(attentionshifting_signal ~ visibility * emotion * direction
R>              + Error(id/(visibility * emotion * direction)),
R>              data = attentionshifting_design)
```

The `plot` method produced a graphical representation of the tests that allows us to see quickly the significant time frames corrected by `clustermass`. The results are shown in Figure 1.4.

```
R> plot(electrod_O1)
```

Only one significant result appears for the main effect of visibility. This effect is corrected using the `clustermass` method. Printing the `clusterlm` object gives more information about all clusters for the main effect of visibility, whether they are driving the significant effect or not:

```
R> print(electrod_O1, effect = "visibility")
```

```
Cluster fisher test using Rd_kheradPajouh_renaud to handle nuisance variables
 with 5000 permutations and the sum as mass function.
```

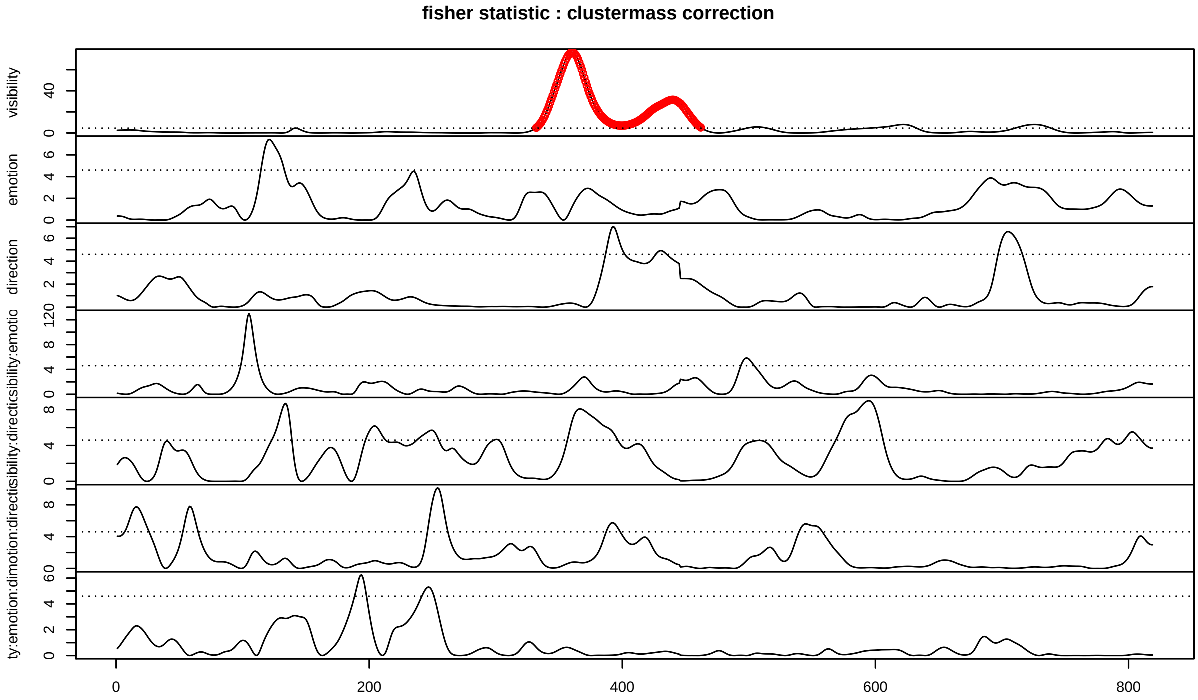

Figure 1.4: The `plot` method on a `clusterlm` object displays the observed statistics of the three main effects and their interactions. The dotted horizontal line represents the threshold which is set by default to the 95% percentile of the statistic. For this dataset, using the `clustermass`, one cluster drives the significant difference for the main effect of visibility as displayed in red. The `print` method gives more details.

```
Alternative Hypothesis : two.sided.

visibility, threshold = 4.60011.
  start end cluster mass P(>mass)
1   142 142     4.634852   0.5048
2   332 462  3559.149739   0.0018
3   499 514    85.019645   0.4060
4   596 632   234.877913   0.2290
5   711 738   191.576178   0.2680
```

There is a significant difference between the two levels of visibility. This difference is driven by one cluster that appears between the measures 332 and 462 which correspond to the 123.7ms and 250.9ms after the event. Its cluster-mass statistic is 3559.1 with an associated p value of 0.0018. The threshold is set to 4.60011 which is the 95% percentile of the F statistic. If we want to use other multiple comparisons procedures, we use the `multcomp` argument:

```
R> full_electrod_O1 <-
R>   clusterlm(attentionshifting_signal ~ visibility * emotion * direction
R>                 + Error(id/(visibility * emotion * direction)),
```

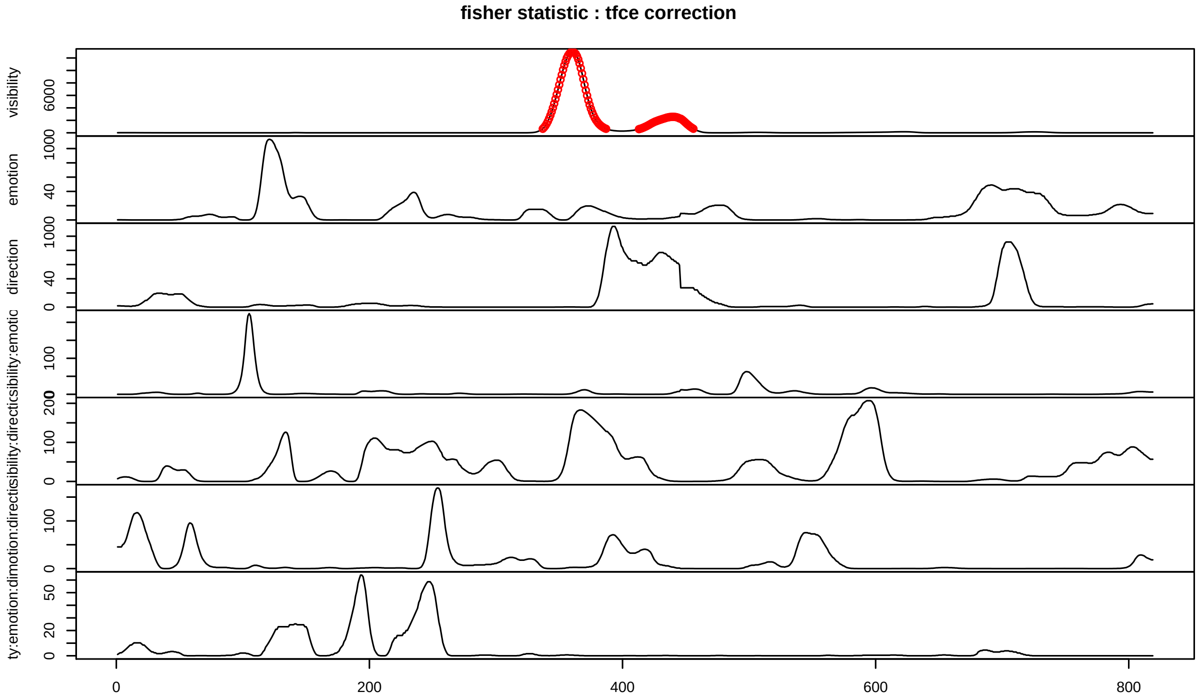

Figure 1.5: Setting the `multcomp` argument to `"tfce"` in the `plot` function will display the TFCE p values. The argument `enhanced_stat = TRUE` shows the TFCE statistics u_s of Equation 1.13.

```
R>              data = attentionshifting_design, P = electrod_O1[["P"]]
R>              method = "Rde_kheradPajouh_renaud",
R>              multcomp = c("troendle", "tfce", "clustermass",
R>                           "bonferroni", "holm", "benjaminin_hochberg"))
```

Note that we retrieve the very same permutations as previous model by using the `P` argument. The computation time for those tests is reasonably low: it takes less than 12 minutes on a desktop computer (i7 3770CPU 3.4GHz, 8Go RAM) to compute the 7 permutation tests with all the multiple comparisons procedures available. To see quickly the results of the threshold-free cluster-enhancement procedure, we set the `multcomp` argument of `plot` to `"tfce"` as shown in Figure 1.5.

```
R> plot(full_electrod_O1, multcomp = "tfce", enhanced_stat = TRUE)
```

The TFCE procedure gets approximately a similar effect. However the time-points around 400 (190 ms) are not part of the significant effect. If the curves in the TFCE plot happen to show some small steps (which is not the case in Figure 1.5) it may be because of a too small number of terms in the approximation of the integral of the `tfce` statistics of Equation 1.13. In that case it would be reasonable to increase the value of the parameter `ndh`.

Finally, to be able to interpret individually each time-point, we can use the `troendle` multiple comparisons procedure whose results are visualized by plotting the `full_electrod_O1` object. A similar period is significant for the main effect of `visibility`.

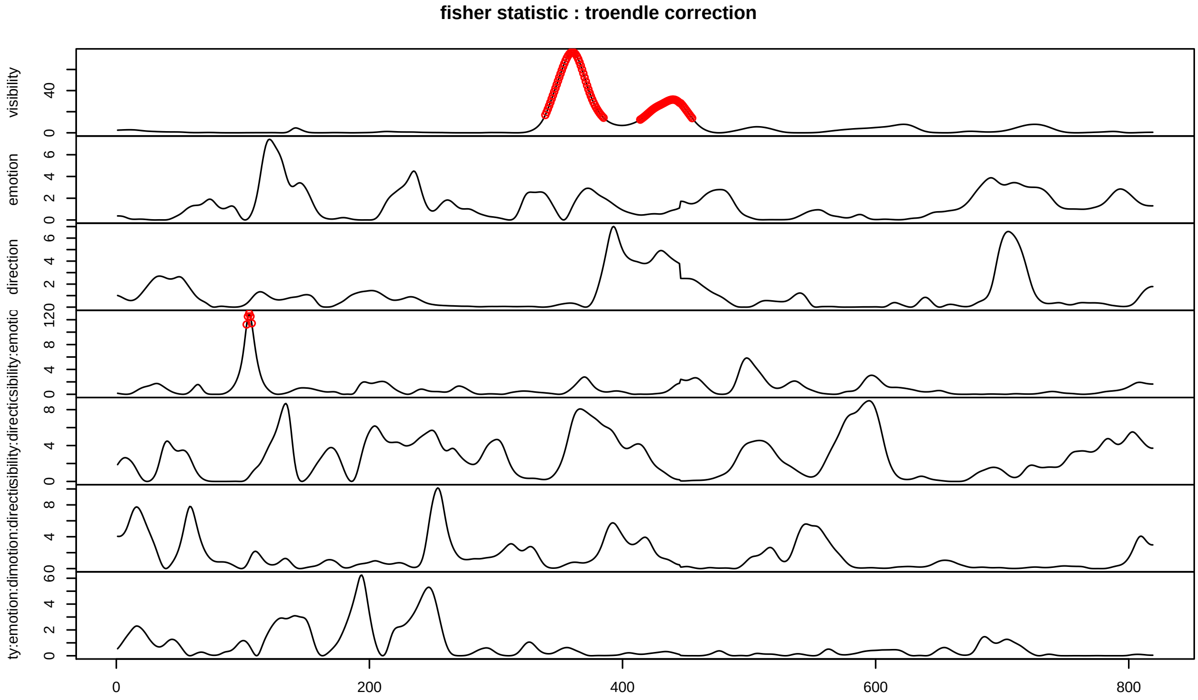

Figure 1.6: Setting the `multcomp` to `"troendle"` will display the `troendle` correction which allows an interpretation of each time-point individually.

```
R> plot(full_electrod_01, multcomp = "troendle")
```

To interpret individually each time-point in Figure 1.6, we extract the significant time-points (with an α level of 5%) using the `summary` method, setting the `multcomp` parameter to `"troendle"`. We find that the main effect of `visibility` begin at 130.6 ms after the event. However, the significant time-points for the interaction `visibility:emotion` are between 100.2 ms and 96.3 ms before the event, which are obviously type I errors.

```
R> tro_sum <- summary(full_electrod_01, multcomp = "troendle")
R>
R> tro_visi_sign <- tro_sum[tro_sum[,"visibility pvalue"] < 0.05,
R>                                         "visibility pvalue"]
R>
R> tro_visem_sign <- tro_sum[tro_sum[,"visibility:emotion pvalue"] < 0.05,
R>                                            "visibility:emotion pvalue"]
```

```
R> tro_visi_sign[1:7]
```

```
 130.6  131.5  132.5  133.5  134.5  135.5  136.4
0.0362 0.0274 0.0118 0.0068 0.0068 0.0068 0.0068
```

```
R> tro_visem_sign
```

```
-100.2  -99.3  -98.3  -97.3  -96.3
0.0448 0.0306 0.0306 0.0306 0.0408
```

1.7 Conclusion

This article presents recent methodological advances in permutations tests and their implementation in the `permuco` package. Hypotheses in linear models framework or repeated measures ANOVA are tested using several methods to handle nuisance variables. Moreover permutations tests can solve the multiple comparisons problem and control the FWER trough cluster-mass tests or TFCE, and the `clusterlm` function implements those procedures for the analysis of signals, like EEG data. Section 1.6 illustrates some real data example of tests that can be performed for regression, repeated measures ANCOVA and ERP signals comparison.

We hope that further developments of `permuco` expand cluster-mass tests to multidimensional adjacency (space and time) to handle full scalp ERP tests that control the FWER over all electrodes. Another evolution will concern permutation procedures for mixed effects models to allows researchers to perform tests in models containing participants and stimuli specific random effects.

Acknowledgement

We are particularly grateful for the assistance given by Eda Tipura that greatly improved this manuscript. She provided original EEG data, as well as many comments coming from her extended reading of the paper; although any errors are our own.

Chapter 2

Complements to the `permuco` Package and Permutation Methods

Chapter 2 presents complements closely related to the content of Chapter 1. Section 2.1 describes the geometry of the permutation methods. Section 2.2 explores a full-scalp EEG data analysis and presents functions that will be added to `permuco` in the next release. Finally, in Section 2.3, we present a new and more powerful approach of the cluster-mass tests using the slopes of the signals.

2.1 Geometrical Interpretation of Permutation Methods

Kherad Pajouh and Renaud (2010) show that permutation methods have a geometrical interpretation. In the following Section, we recall and extend these interpretations to several methods using three-dimensional representations (Soetaert, 2017). Kherad Pajouh and Renaud (2010) produce a graphical representation showing the link between the `kennedy` and the `huh_jhun` methods. Here, we show individual plots for the `manly`, `kennedy`, `freedman_lane` and `huh_jhun` permutation methods. Moreover, this representations inspired us a new transformation of the data that produces the exact same permuted F statistics than the `freedman_lane` permutation method. We show graphically that the similarity between the two transformations are based on the properties of orthogonal projections in the F statistic. Hence, this new transformation may be different than `freedman_lane` using other statistics.

In a regression or ANOVA model as described by Equation 1.1, the space of the geometry is $\mathcal{R}^n$, and all variables, including the response y, the nuisance variables D and the variables of interest X, are set of column-vectors that lie in $\mathcal{R}^n$. Moreover, the design $\begin{bmatrix} D & X \end{bmatrix}$ is a set of column-vectors that spans a subspace of $\mathcal{R}^n$.

Evaluating $\hat{y} = D\hat{\eta}+X\hat{\beta}$ for all possible values $\hat{\beta}$ and $\hat{\eta}$ spans a p-dimensional subspace of $\mathcal{R}^n$. In Figure 2.1, y, is represented by the red vector, X and D are represented by black vectors and all possible candidates for $\hat{y}$ lie on the grey grid. Performing ordinary least squares (OLS) is finding the estimates $\hat{\beta}$ and $\hat{\eta}$ (light blue construction in Figure 2.1) such that they minimize the sum of squares of the residuals $\hat{\epsilon}$ (green vector in Figure 2.1). Moreover, the residuals are computed by the difference between y and $\hat{y}$ and their sum of squares is simply the square of the Euclidean distance between y and $\hat{y}$. In that geometry, the regression, by minimizing the sum of squares of the residuals, finds $\hat{y}$ such as the closest

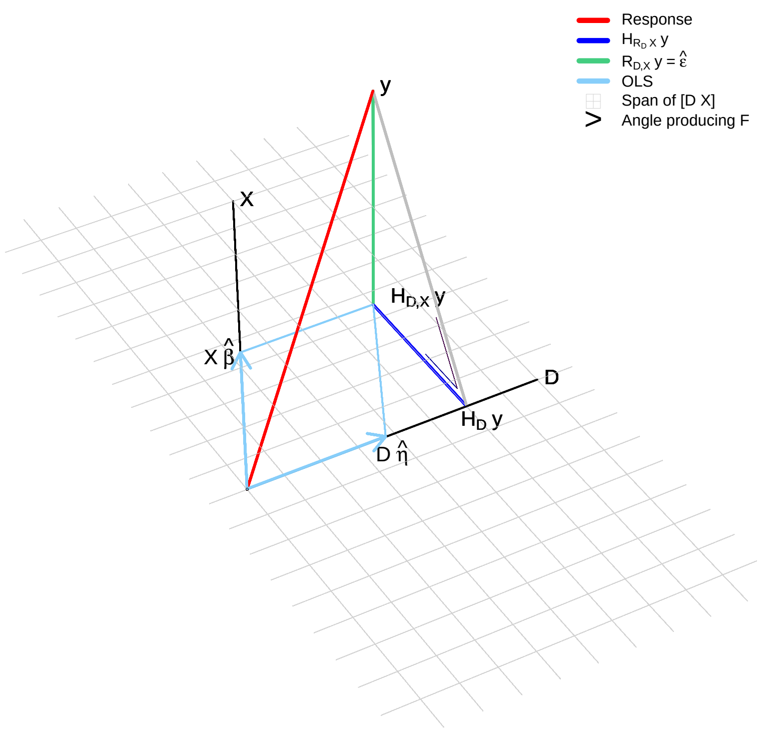

Figure 2.1: Geometry of the regression. The OLS projects y into the span of $[D\ X]$.

point of y in the span of $[D\ X]$. This point is unique and corresponds to the orthogonal projection of y into $[D\ X]$. Geometrically, all points with an equal Euclidean distance to y span a hypersphere centred in y and its square radius corresponds to the residuals sum of squares. In other words, OLS "constructs" the hypersphere of minimal radius that is tangent to the subspace spanned by $[D\ X]$, and this touching point is precisely $\hat{y}$.

Finally, the F statistic also has a geometrical interpretation as both the numerator and denominator correspond to squared lengths of vectors. Using the properties of the projection matrices, we find that the denominator corresponds to the squared norm of the residuals $y^\top R_{D,X} y = (R_{D,X} y)^\top R_{D,X} y = ||R_{D,X} y||^2$ (green line in Figure 2.1). In addition, the numerator corresponds to the squared norm of the difference between the fitted values of the "full" model (using the full design $[D\ X]$) and of the "small" model (using only D in the design) $y^\top H_D y - y^\top H_{D,X} y = y^\top H_{R_D X} y = ||H_{R_D X} y||^2$ (deep blue line in Figure 2.1). These two vectors create a right-angled triangle. The F statistic is a function of an angle (black angle pointing $H_D y$ in Figure 2.1) of this right-angled triangle as $F = \frac{n-p}{p-q} \left(\tan\left(\angle(y - H_D y,\ R_{D,X} y - H_D y)\right)\right)^{-2}$, where tan is the tangent trigonometrical function and $\angle(\cdot)$ denotes the angle between two vectors, here the vectors $y - H_D y$ and $H_{D,X} y - H_D y$.

In this space, the permutation methods transform variables while using the same statistic. Representing the design, the response variable and their transformation as vectors in a 3D space gives some insight of the effect of the permutation methods.

Note that this graphical representation is in a 3D space which implies some simplifications. First, the design D and X are both set of vectors in usual regression models but are graphically represented using only one dimension (one vector). Moreover, the part of the design coding the intercept is the vector $\mathbf{1}$ which is singular when using permutation as

$P\mathbf{1} = \mathbf{1} \; \forall \; P$. Hence, the intercept is literally central in permutation as $H_{\mathbf{1}}y = H_{\mathbf{1}}Py \; \forall \; P$ which implies that permutations are interpreted as rotations "around" $\mathbf{1}$. In a 3D space, both the intercept and other nuisance variables cannot be represented simultaneously as D should be a one-dimensional object and $\mathbf{1} \in D$.

2.1.1 Permutation Methods in the 3D Space

Figure 2.2 represents the `manly` permutation method which is the transformation: $\{y, D, X\} \rightarrow \{Py, D, X\}$. This permutation method does not consider the effect of D and the resampled datasets (red stars) and their corresponding fitted values (red dots) are widespread over the D axis. If the effect of the $D\eta$ is not null, a permutation method that takes fully into account the effect of the nuisance variables should have permuted samples that project into the true value of $D\eta$ (or at least close to it).

Figure 2.3 shows the `kennedy` permutation method which is the transformation: $\{y, D, X\} \rightarrow \{PR_Dy, -, R_DX\}$. It projects everything into the subspace orthogonal to D using the transformation R_D. Since, the vectors that generate the F statistic are merely translated, their norms and therefore the F value are unchanged. However, after the permutations, the vectors PR_Dy do not stay into such a space orthogonal to D and the residuals of the permuted datasets (orange vectors) get longer than expected in a case with nuisance variables. The denominators of the permuted F statistics increase which decrease the values of the permuted F statistics. This results to smaller p-values. We investigate analytically this effect and propose a correction in Section 3.3.

Figure 2.5 shows the `freedman_lane` permutation method which is the transformation: $\{y, D, X\} \rightarrow \{(H_D + PR_D)y, D, X\}$. This method only permutes the residuals of the "small" model (which is the model using only the nuisance variables) and add them to the non-permuted fitted values of the same model. The permuted responses are the one computed by the `kennedy` method and shifted by H_Dy. Doing so, the numerator is similar to the `kennedy` method for each permutation. However, because the permuted data are projected into a larger subspace ($\begin{bmatrix} D & X \end{bmatrix}$), the residuals do not increase as much as in the `kennedy` method. The very same permuted statistics would be produced by adding a new step to the `kennedy` method by pre-multiplying by R_D after permuting or equivalently: $\{y, D, X\} \rightarrow \{R_DPR_Dy, -, R_DX\}$. This transformation does not change the numerator of the F statistic for each permutation in comparison to the `kennedy` or `freedman_lane` methods. However, it reduces the residuals sum of squares of the denominator of the `kennedy` method as R_DPR_Dy lies in a subspace orthogonal to D. Because the distance between PR_Dy and $[D\ X]$ is equal to the distance between R_DPR_Dy and R_DX, the permuted F statistics using the `freedman_lane` method are identical to the one computed using the transformation $\{y, D, X\} \rightarrow \{R_DPR_Dy, -, R_DX\}$.

Figure 2.4 shows the `huh_jhun` method which is the transformation: $\{y, D, X\} \rightarrow \{PV^\top R_Dy, -, V^\top R_DX\}$. As the `kennedy` method, it first projects everything into the subspace orthogonal to D. Then, it defines an orthonormal basis in this subspace before performing the permutations (purple arrows in Figure 2.4). Hence, the permuted responses are limited inside that subspace and the permuted residuals does not increase as much as the `kennedy` method. The new basis is chosen randomly. A random basis implies a random position of the "intercept" vector $\mathbf{1}$. Because the effect of permutation depends on the spatial position of $\mathbf{1}$, it follows that each set of permuted data (and permuted statistic) changes according to the selected basis. As mentioned by Kherad Pajouh and Renaud (2010), choosing a random basis has the effect to almost orthogonalize the

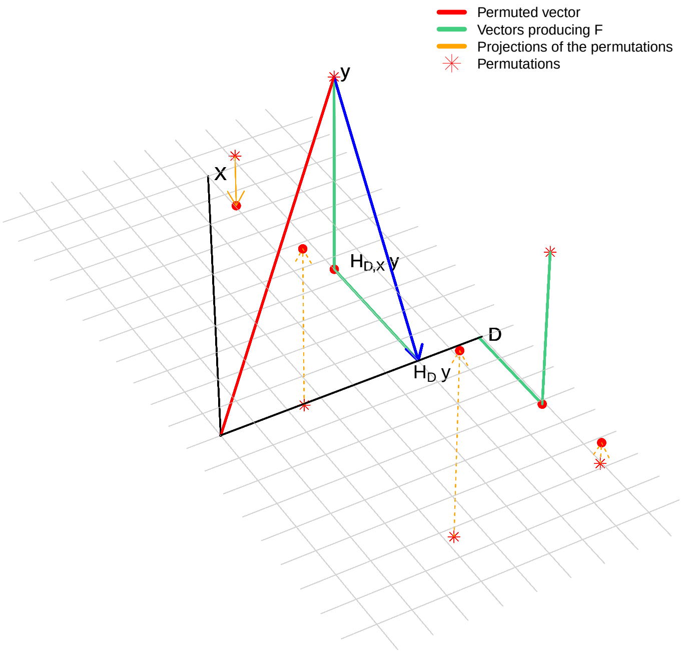

Figure 2.2: Geometry of the `manly` permutation method. It permutes the full vector y represented by the solid red line. This permutation method does not reduce the effect of D before permuting the data.

vectors $V^\top R_D y$ and $V^\top R_D X$ to the intercept with a high probability. Moreover, because the size of the space gets smaller ($n \to n - \text{rank}(D)$), the number of permutations decreases ($n! \to (n - \text{rank}(D))!$).

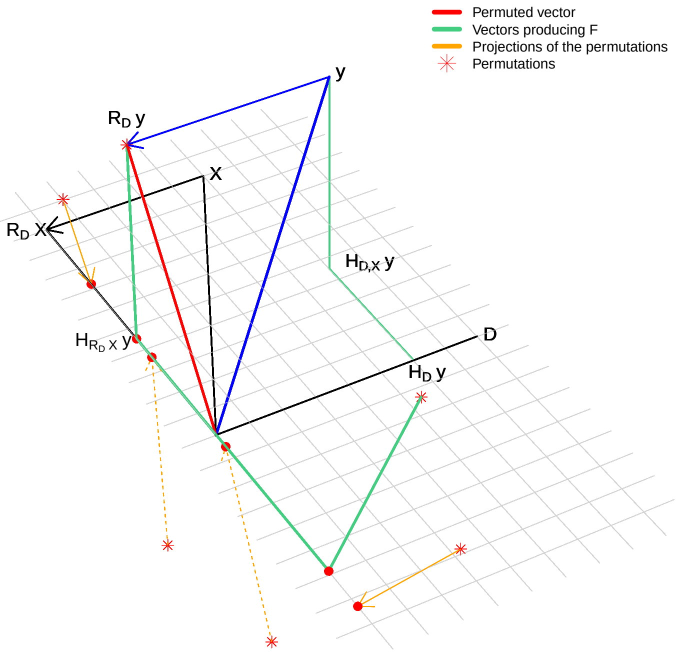

Figure 2.3: Geometry of the `kennedy` permutation method. It orthogonalizes the vectors y, X and D before permuting. The F statistic is produced by the green lines and stays unchanged after the projection orthogonal to D. However, the permutations of R_Dy do not lie in the subspace orthogonal to D.

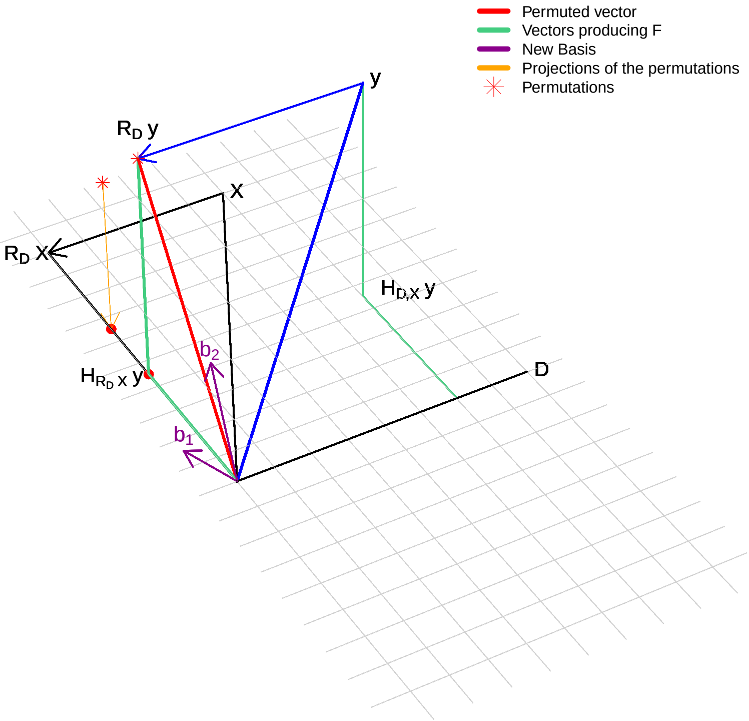

Figure 2.4: Geometry of the `huh_jhun` permutation method. The first step of the `huh_jhun` permutation is similar to the `kennedy` method: a projection of the data into the subspace orthogonal to D. Then, in order to have permuted responses that lie in this subspace, we define a random orthonormal basis of smaller dimension, b_1 and b_2, and perform the permutations inside this subspace. Less permutations are feasible as the subspace is smaller.

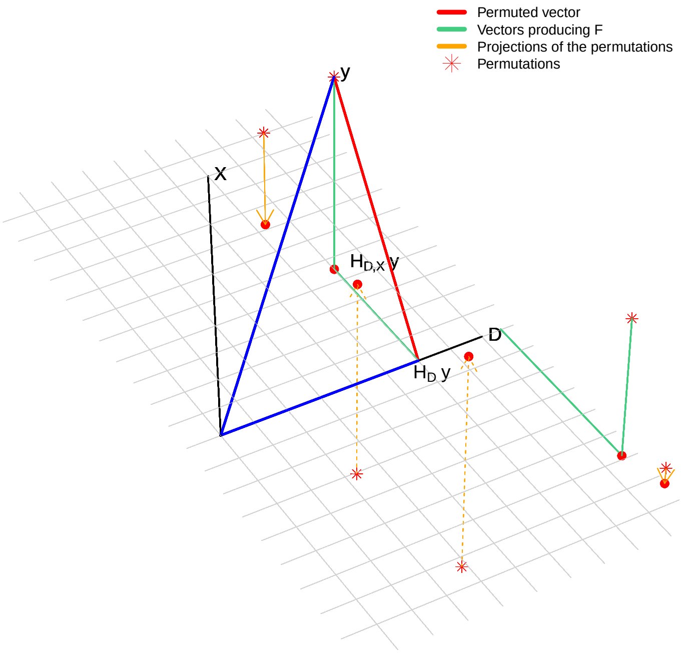

Figure 2.5: Geometry of the `freedman_lane` permutation method. It permutes only the residuals of the "small" model (red line) and add them to the observed fitted values of the same model. The permuted response (red stars) are similar to the `kennedy` method, but shifted by H_Dy. The permuted numerators are then equal in both methods, but the permuted denominators decrease as the permuted responses are projected into a larger subspace ($[D\ X]$ instead of R_DX).

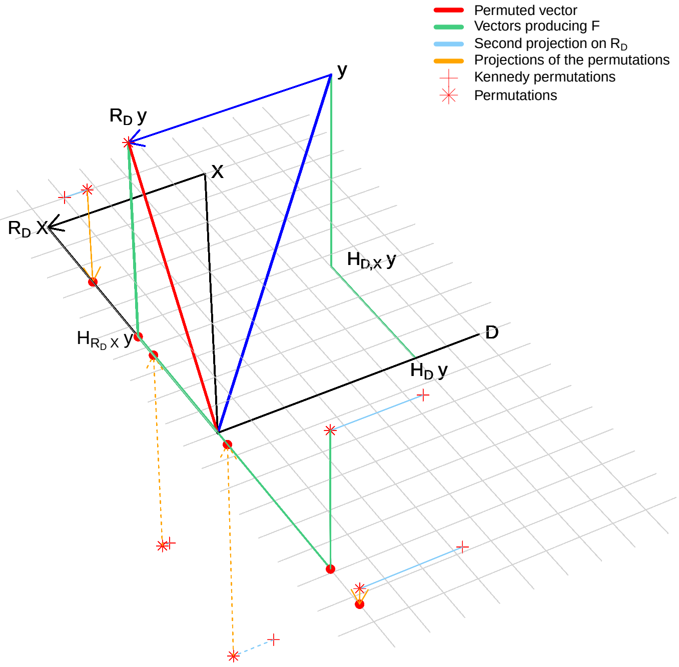

Figure 2.6: Geometry of the `freedman_lane` permutation method using the transformation $\{y, D, X\} \rightarrow \{R_DPR_Dy, -, R_DX\}$. We project the `kennedy` method (red "+" signs) into R_D before computing the statistics. Both approach of the `freedman_lane` method produces the same permuted residuals (orange vectors).

2.1.2 Note on the `dekker` and `terBraak` Permutation Methods

The `dekker` permutation method is the following transformation of the data: $\{y, D, X\} \rightarrow \{y, D, PR_D X\}$. It permutes only the design and a 3D visualization may not be helpful to understand its effects. However, it works by first orthogonalizing the variables of interest X to the nuisance variables D and then permuting the orthogonalized variables of interest. Using the QR decomposition of the full design, the very same permuted F statistics are obtained with the transformation: $\{y, D, X\} \rightarrow \{y, Q_{[D,X]:1\ldots q}, PQ_{[D,X]:q+1\ldots p}\}$, where $Q_{[D,X]}$ is an orthonormal basis of $[D\ X]$ computed by QR decomposition, $Q_{[D,X]:1\ldots q}$ is a set of orthonormal vectors spanning D and $Q_{[D,X]:q+1\ldots p}$ is a set of orthonormal vectors spanning $R_D X$ (see Section 3.4.1 for more details). Permuting $Q_{[D,X]:q+1\ldots p}$ or $R_D X$ is identical when computing the F statistic as it is based on projection into that space, which are independent on the bases of the space. Moreover, permuting different basis that spans the same space creates new set of vectors that span the same space: if two basis M and N span the same space, they both create the same projection matrix $H_M = H_N$ and, for each permutation matrix P, we have $H_{PM} = PH_M P^\top = PH_N P^\top = H_{PN}$. This means that permuting $R_D X$ spans the same space than permuting $Q_{[D,X]:q+1\ldots p}$ for all permutations. The `dekker` permutation method is computationally intensive as its needs to compute a QR decomposition for each permutation and this second approach may reduce its computing time. Hence, a clever implementation that re-uses the set of vectors $Q_{[D,X]:1\ldots q}$ for each permutation when computing the QR-decomposition may save computing time.

The `terBraak` permutation method is the following transformation of the data: $\{y, D, X\} \rightarrow \{(H_{D,X} + PR_{D,X})y, D, X\}$. However, as explained in Section 1.2.2, for each permutation, we compute the statistic under a different null hypothesis, $H_0 : \beta = \mathbf{b}$, where $\mathbf{b} = \left(X^\top R_D X\right)^{-1} X^\top R_D y$ which corresponds to the parameter evaluated on the observed values. To consider this change of null hypothesis in the model, we change the model to test $H_0 : \beta' = 0$ such that:

$$\begin{aligned} y &= X\beta + D\eta + \epsilon \\ y &= X(\beta' + \mathbf{b}) + D\eta + \epsilon \\ y - X\mathbf{b} &= X\beta' + D\eta + \epsilon. \end{aligned} \tag{2.1}$$

Then, the `terBraak` permutation method is identical to the transformation $\{y, D, X\} \rightarrow \{(H_{D,X} - X(X^\top R_D X)^{-1} X^\top R_D + PR_{D,X})y, D, X\}$ without the change of null hypothesis. Note that, this transformation modifies the data such that when $P = I$ the statistic is 0 and the original dataset (e.g. before applying the `terBraak` transformation) should be used to compute the test statistics.

2.2 Full Scalp Data Analysis

Section 2.2 describes my involvement in the analysis of the EEG data of the experiment of Cheval et al. (2018). It reports some of the practical challenges we encountered during the analysis and the solutions we brought.

In Appendix B, we provide `R` code to download the EEG data from the Zenodo repository and to reproduce the analysis and tests of Cheval et al. (2018). The main functions are compiled in the `clustergraph` package (https://github.com/jaromilfrossard/clustergraph) that I have written and will be added to the `permuco` package once a set of user-friendly EEG data importation and manipulation functions are designed.

2.2.1 Experimental design

The goal of the experiment in Cheval et al. (2018) is to understand if humans are naturally attracted to physical activity. The approach-avoidance framework and the manikin task (Mogg et al., 2003; Krieglmeyer and Deutsch, 2010) is implemented to test this hypothesis. In that framework, the participants of the experiment must perform a task when seeing stimuli. The experimenter asks them to move a virtual manikin either in the direction of the stimuli (approach) or in the opposite direction (avoid). If participants are attracted to the stimuli, it is hypothesised that they show, on average, a faster reaction time when approaching rather than avoiding the stimuli. Furthermore, the tendency inverses if they are repulsed by the stimuli.

For the experiment of Cheval et al. (2018), experimenters show images describing physical activity (PA), sedentary behaviour (SED) as well as neutral images (neutral). Moreover, the tasks of the participants are either to approach (Approach) or to avoid (Avoidance) the stimuli. Furthermore, some features of the participants are recorded like a measure of the usual physical activity of the participants which is used to adjust the observed effects. The EEG signal is recorded on the full scalp using a 64 electrodes cap during over a second for each trial. In addition, 800ms after the event, the participants begin to engage in their movements which may disrupt the EEG recording. Hence, we only perform the test during the period from 0 to 800ms after the event.

The goal of the analysis of the ERP is to detect if the design influences the average ERP, where it might occur (which electrodes) and at which time (after showing the stimuli). Without any prior information on the part of the brain and on the time of the potential effect, the solution is to test for each time point at each electrode and to use a powerful multiple comparisons procedure. The main hypothesis of the psychologists lies in the interaction between the type of stimuli and the task. They postulate that, relative to neutral condition, the effect of the type of stimuli (PA and SED) is different depending on the type of task (approaching or avoiding).

It is a typical experimental design in psychology where participants must react to stimuli. This type of design should be analysed using a cross-random effects mixed-effects model (CRE-MEM). However, for this analysis, we face many challenges and choose a repeated measures ANCOVA model and test. For the analysis of ERP, we must perform more than thousands of tests and then use a multiple comparisons procedure. Scaling CRE-MEM to this size creates both statistical and computational problems. First, we must select the appropriate correlation structure of the data for all the tests. Using the same correlation structure would produce many convergence errors and adapting it for each test would cause problems when interpreting the results. Moreover, no multiple comparisons procedure is available and powerful enough for CRE-MEM. Finally, the optimization is computationally intensive and difficult to scale at this number of tests. Hence, we choose to average the signals over the stimuli which amounts to treating them as fixed effects. In order to interpret the results relatively to the neutral stimuli, we transformed the observed signals by taking the difference between the signals in the physical activity (PA) and neutral conditions, and sedentary behaviour (SED) and neutral conditions, signal for each participant. We use these differences of signals as response and perform one test at each time, each electrode.

Each test is a repeated measures ANCOVA, with two factors (task and type of stimuli) and one covariate which is a self-reported measure of moderate-to-vigorous physical activity (MVPA). In order to decompose the interaction effect, we perform two additional simple effects tests. The first one corresponds to the effect of stimuli within the

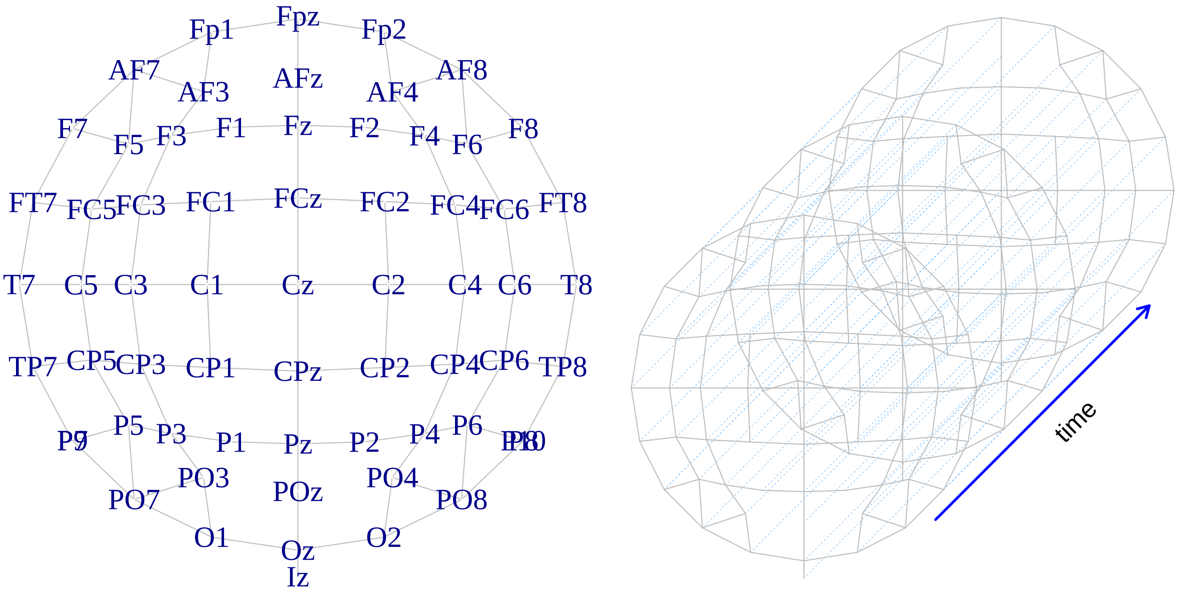

Figure 2.7: Graph of adjacency of the electrodes. On the left panel, the graph defines the spatial adjacency between electrodes. On the right panel, the spatial graph is reproduced for each time point (here only 3) and bind together according to the electrode defining the spatiotemporal adjacency of the cluster-mass test.

”Approach” level of the task only, and the second one the effect of stimuli within the ”Avoidance” level only.

2.2.2 Implementing the cluster-mass test

The p-values are computed using a permutation test and the method proposed for rANOVA by Kherad-Pajouh and Renaud (2015) to handle the nuisance variables. We use the cluster-mass test to control the FWER which is powerful when the effects are adjacent and is relatively fast to compute (in comparison to the TFCE).

In a full-scalp cluster-mass test, hypotheses are distributed on the space (the electrodes on the scalp) and time. The cluster-mass test as implemented in the `permuco` package only handle one electrode measured on multiple time-points; this means that clusters are computed using only time adjacency. To consider spatiotemporal data, we must also define the space adjacency. A connected graph (left panel of Figure 2.7) is the appropriate object to represent the spatial adjacency: the electrodes are represented as the vertices and the adjacency relationship between two electrodes by an edge. To control the FWER in the cluster-mass test, the graph should be defined a priori. It may be defined using prior information of the relationship between electrodes or between brain regions. Without any prior information, the Euclidean distance between electrodes is used for the analysis by Cheval et al. (2018). Two electrodes are declared adjacent if their Euclidean distance is smaller than δ, which is the smallest distance that produces a connected graph[1]. In Cheval et al. (2018), we found the value $\delta = 35mm$ which produces the spatial adjacency defined by the graph in the left panel of Figure 2.7. To define spatiotemporal adjacency,

[1] A connected graph implies no disconnected sub-graph. Having sub-graphs implies that some tests cannot, by design, be in the same cluster, which is not a useful assumption for this analysis.

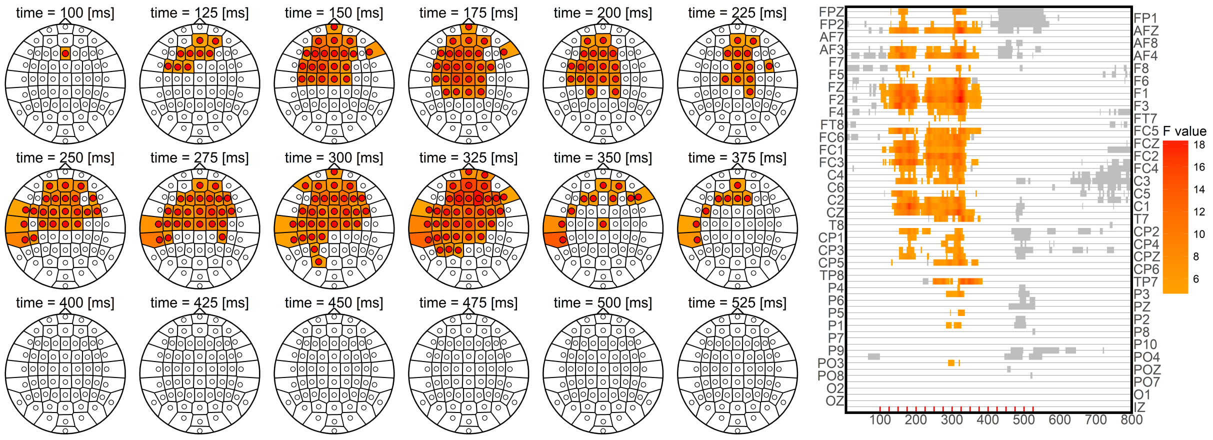

Figure 2.8: Effect of interaction between the task and the stimuli. The 18 figures on the left represent the only significant cluster for selected time-points. The panel on the right shows all clusters: the one in color is significant and the others are in grey. We first see that the effect began around 100ms in the front electrodes and after 250ms the electrodes of the left becomes also part the cluster. The effect ends between 375ms and 400ms. Figure published in Cheval et al. (2018).

the spatial graph is then reproduced for each time-point with edges between all pairs of two vertices (tests) associated to the same electrode when they are temporally adjacent. In the right panel of Figure 2.7, a graph defines spatiotemporal adjacency for 64 electrodes and 3 temporal measures. The full graph that defines the spatiotemporal adjacency has then a total of vertices equal to the number of tests (#electrodes × #time). Note that, in this structure, no difference exists between adjacency in time or space.

From a computational perspective, finding a cluster in this structure becomes simple as we use all the tools developed for the analysis of graphs (Csardi and Nepusz, 2006). As a reminder, a cluster is defined as a set of adjacent statistics that are above the predefined threshold. After performing all tests, we map them on the spatiotemporal graph. We then delete all the vertices which statistics are below the threshold. This produces a new graph composed by multiple connected components. Then, each connected component is interpreted as a spatiotemporal cluster. Finally, for each connected component, we compute the cluster-mass statistic using the sum (or sum of squares) of statistics of that particular connected component.

The cluster-mass null distribution is computed by permutations while maintaining spatiotemporal correlations among tests. Permutations must be performed without changing the position of electrodes nor mixing time-points. Concretely, after transforming the responses using the permutation method in Kherad-Pajouh and Renaud (2015), they are sorted in a three-dimensional array. It has the design (participants × experimental conditions) in the first dimension, time in the second one and electrodes in the third one. Then, only the first dimension is permuted to create a re-sampled response (or 3D array). Doing so, it does not reorder time-points, neither electrodes, therefore, the spatiotemporal correlations are maintained within each permuted sample.

2.2.3 Graphical representation

The representation of the results is multi-dimensional, because tests are spread in space and time. Moreover, it is also useful to interpret different effects in the same time, typically the simple effects and the interaction. In order to help readers to understand this complexity, we produce different graphical representations.

First, we produce graphics where, for each effect, the tests for all electrodes and all time points are represented in a two-dimensional image with the time in the X-axis and electrodes in the Y-axis (right panel of Figure 2.8). Electrodes are sorted from the one in the back (bottom Y) to the one in the front (top Y). The significant clusters are represented in a colour-scale and the non-significant one in grey. The white pixels are tests which statistic are below the threshold. The colour-scale represents the value of the univariate statistics. This representation gives a good general idea of the size and timing of the clusters but does not give a good understanding of its spatial position. Hence, we add graphics where the electrodes are spatially represented using their (2-dimensional) theoretical position, but only for some selected time-points.

Then, to be able to analyse in parallel all the effects, we can focus the graphical representation on single electrodes. Some electrodes were selected as they seem central and typical of the activated region. We represented the observed ERP (average for each condition), the ERP relative to the neutral condition (actual means used for the test) and a third visual representation indicating for the selected electrode all the effects and when they are part of a significant cluster. Figure 2.9 shows this representation for the electrode FCz.

2.3 Extending Cluster-Mass Test using the Slope of Signals

This proposition has been presented as a poster at the congress of the International Society of Non-Parametric Statistics (ISNPS) in 2018 in Salerno. It is mainly developed for tests on a single electrode and multiple time measures.

2.3.1 Motivation Example

In ERP data, physiologically, one process can only happen on multiple adjacent time-points and the cluster-mass test is powerful to detect it. It works by grouping these adjacent time-point into one cluster. Then, it produces one common test statistic and one common inference for all time-points in the cluster. Moreover, the common inference for all time-points in the cluster is relevant because we assume that the full cluster is created by the same underlying process. This procedure is powerful and controls the family-wise error rate when the same process is composed of adjacent time-points but loses power otherwise.

However, a true process in an ERP may happen with a positive difference between conditions followed by a negative difference (or vice versa). Moreover, it may be relevant to assume that these two separate differences are caused by the same underlying mechanism to which the voltage is first higher in one condition, immediately followed by a time when the voltage is lower. This type of processes is not well detected with the usual cluster-mass test because the timing of the process is split into two smaller clusters. In addition, the two clusters come from the same underlying mechanism and it would be relevant to

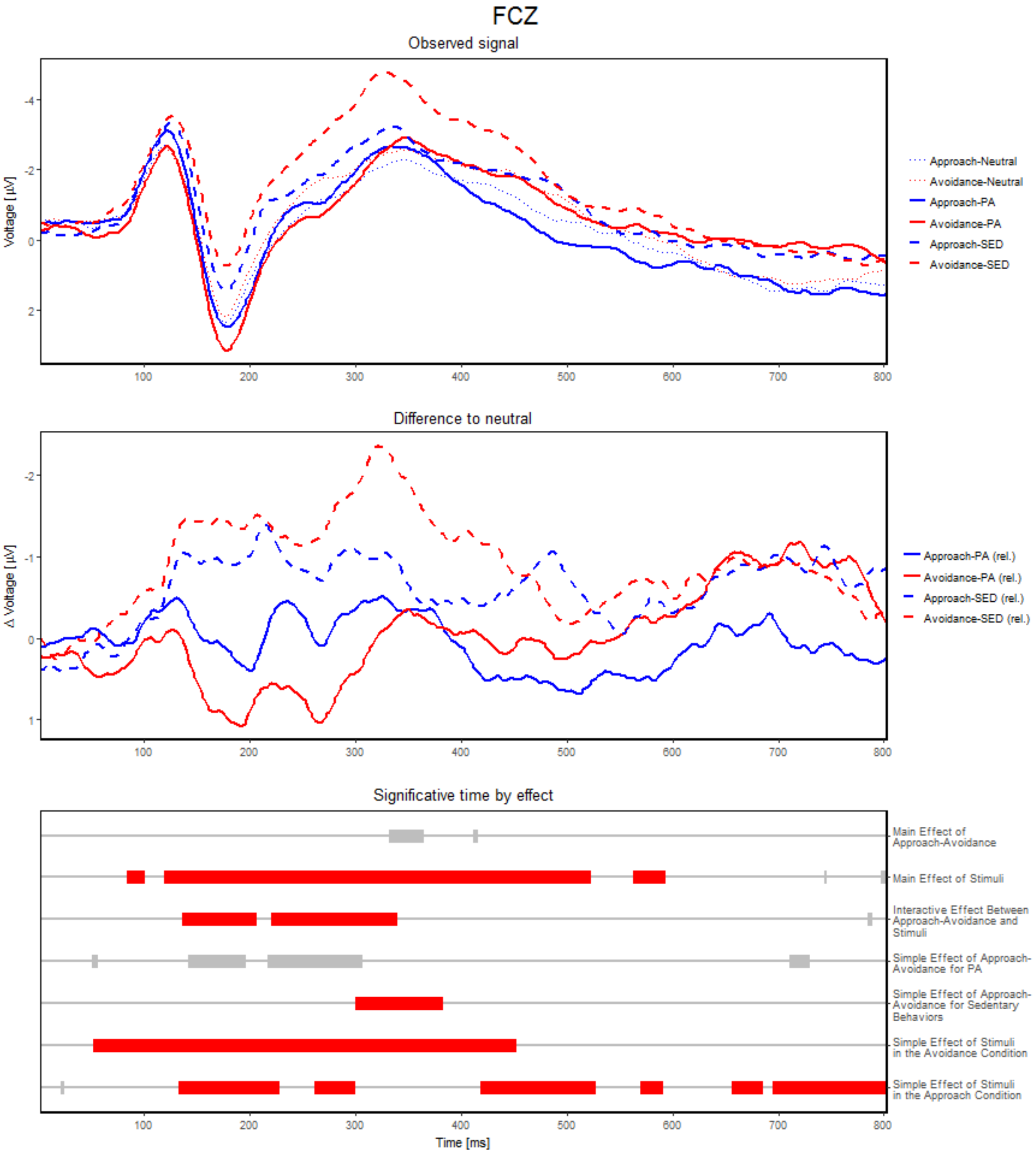

Figure 2.9: All tested effects for the electrode FCz. The top panel shows the ERP in each experimental condition. The middle panel shows the difference to the neutral condition (the signals on which the tests are performed). The bottom panel shows which time-points are in a significant cluster depending on the effect. Figure published in Cheval et al. (2018).

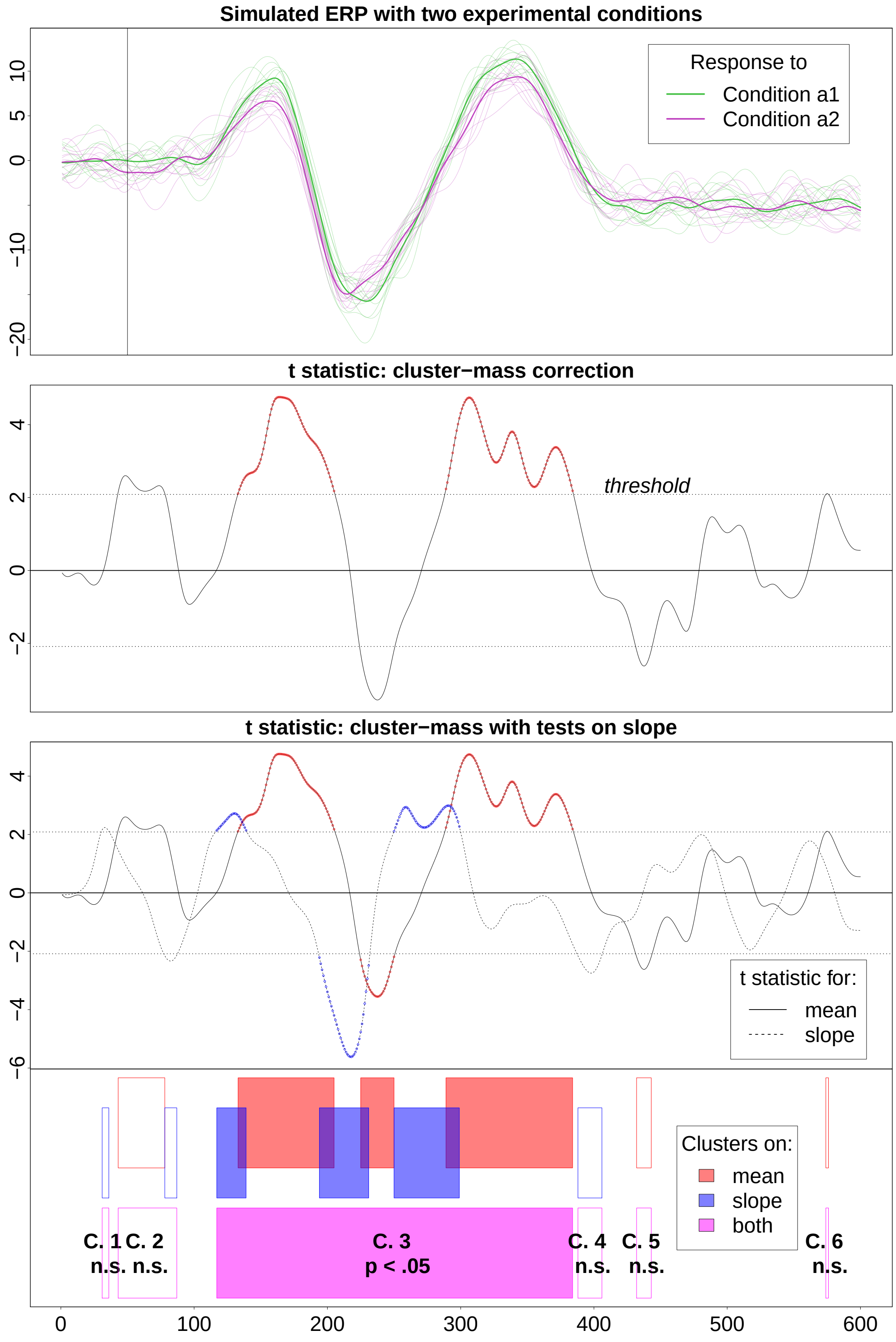

Figure 2.10: Cluster-mass test extended using the slopes. The top panel shows simulated ERP in two experimental conditions. The second panel show the classical cluster-mass test which detects only 2 smaller clusters. The third panel shows statistical signals on the raw signals and on their slopes: when the tests on the raw signals are below the threshold, the tests on their slopes become above the threshold and 3 clusters are bound together.

bind them and to make common inference for these two segments. It would also result in a more powerful test by combining their cluster-masses. For this purpose, we propose to use the smoothed slope of the signals to bind this type of clusters together when relevant.

In brief, to have a positive difference followed by a negative difference, the slope of the signals must become non-null and detecting an effect on the slopes is used to bind two smaller clusters (see Figure 2.10). First, we compute the test statistics of the effect of a factor on the raw signals and on their smoothed slopes which produces two statistical signals. With these two statistical signals, we compute a cluster-mass test using the usual adjacency based on time, but also adjacency based on the ”derivative”: for the same time-point, the test on the raw signals is adjacent to the test on their slopes). Hence, when the average effect is high, the statistics based on the raw signals is high and the statistics computed on the slopes is close to 0 (see Figure 2.10, $t \approx 170$ and $t \approx 320$). In the other hand, between a positive spike and negative spike on the statistical signal of the raw data, the statistics computed on the smoothed slopes is high, and the one computed on the raw signals is close to 0 (see Figure 2.10, $t \approx 220$ and $t \approx 270$). Finally, for the negative spike, the statistic based on the raw signals is high again (in absolute value) (see Figure 2.10, $t \approx 240$). When using the slope, the full effect (from the positive spikes to the negative spikes) is hence detected as part of one large cluster. The clusters computed on the test statistics on the raw signals are bind by the clusters found for the tests on their slopes which is located between the two alternating spikes.

2.3.2 Model and Hypothesis

Formally, we assume a regression model for the signals as in Equation 1.11 in continuous time and we write it together with the corresponding model for their slopes:

$$y_s = D\eta_s + X\beta_s + \epsilon_s, \tag{2.2}$$

$$\dot{y}_s = D\dot{\eta}_s + X\dot{\beta}_s + \dot{\epsilon}_s, \tag{2.3}$$

where $\dot{y}_s$ is the vector of the slopes of the signals at time s, $\dot{\eta}_s = \frac{\partial \eta_s}{\partial s}$ is the derivative effect of the nuisance variables, $\dot{\beta}_s = \frac{\partial \beta_s}{\partial s}$ is the derivative effect of the interest variables and $\dot{\epsilon}_s$ is an error term. Note that the design matrices D and X are similar to the one introduced in Equation 1.11 and correspond to the nuisance variables and the variables of interest, respectively.

Physiologically, one can assume that null or alternative effects happen during intervals. Then, for any open interval I, if $\beta_s = 0 \; \forall s \in I$ it implies that $\dot{\beta}_s = 0 \; \forall s \in I$. Hence, the null hypothesis $H_0^s : \beta_s = 0$ implies the null hypothesis on the effect of the slopes, $H_{0,\partial}^s : \dot{\beta}_s = 0$. As EEG data are always discrete the equivalent is to test simultaneously H_0^s and $H_{0,\partial}^s \; \forall s \in \{1, \dots, k\}$ using permutation tests; we might use the permutation methods described in Table 1.1 in case of nuisance variables. Moreover, this procedure is similar for a repeated measures ANOVA and a model for the slopes containing random effects may be written following Equation 1.12.

2.3.3 Slope Estimation

When recording EEG, we only observe the raw signals and to use this new procedure the slopes must be estimated. The time differences of the signals are too noisy to be used as reliable slopes estimates and they first must be smoothed. Non-parametric estimation like local polynomial (Fan and Gijbels, 1996) implemented in the locpol package

(Cabrera, 2018) or smoothing splines (Green and Silverman, 1993) are used to produce this estimate. However, both methods need to be tuned by one parameter which controls the smoothness of the estimated curves. To produce two similar signals, we select the smoothing parameter such that the raw signals and their smoothed slopes have approximately the same roughness κ. For a standardized signal Y_s with $s \in \{1, \dots, k\}$, we define its roughness as an estimation of the variability of its discrete second derivative. First, we define $\Delta Y_s = Y_{s+1} - Y_s$ and $\Delta^2 Y_s = \Delta Y_{s+1} - \Delta Y_s$, and then we define the roughness of a curve using:

$$\kappa(Y_s) = \frac{1}{k-3} \sum_{s=1}^{k-2} \left(\Delta^2 Y_s - \text{avg}_s \left[\Delta^2 Y_s \right] \right)^2, \tag{2.4}$$

where $\text{avg}_s \left[\Delta^2 Y_s\right] = \frac{1}{k-2} \sum_{s=1}^{k-2} \Delta^2 Y_s$ is the average of the second derivative of the signal. Then, the response variables of the signal of the n observations for the k time-points are written in a $n \times k$ matrix $\begin{bmatrix} y_1 & \dots & y_s & \dots & y_k \end{bmatrix} = \begin{bmatrix} y_{[1]} & \dots & y_{[i]} & \dots & y_{[n]} \end{bmatrix}^\top$ such that $y_{[i]}$ is the signal of the observation i. In other words, when storing the n signals of length k in a large matrix of size $n \times k$, y_s corresponds to its sth column and $y_{[i]}$ to its ith row. In the same way, we define the slope signals for all observations by $\begin{bmatrix} \dot{y}_1 & \dots & \dot{y}_s & \dots & \dot{y}_k \end{bmatrix} = \begin{bmatrix} \dot{y}_{[1]} & \dots & \dot{y}_{[i]} & \dots & \dot{y}_{[n]} \end{bmatrix}^\top$. Then, for each $i \in \{1, \dots, n\}$ observation, the discrete signal is an observed vector $y_{[i]}$ of length k. However, $\dot{y}_{[i]}$ is not directly observed but estimated with local polynomial or smoothing splines using $y_{[i]}$. The smoothing parameter is set for all n signals such that: $\frac{1}{n} \sum_{i=1}^{n} \kappa(y_{[i]}) = \frac{1}{n} \sum_{i=1}^{n} \kappa(\dot{y}_{[i]})$. Hence, the average roughness of the raw signals $y_{[i]}$ is equal to the average roughness of their slopes $y_{[i]}$.

The rationale behind this procedure is based on the fact that, for the classical cluster-mass test, the statistical signal must be smooth enough to create large clusters. If using the cluster-mass test on the original signals ($y_{[i]}$) is an appropriate multiple comparisons procedure, it means that the raw signals are also smooth enough. Hence, we match its smoothness to the slope signals ($\dot{y}_{[i]}$) in order to have slope signals smooth enough for a cluster-mass test.

2.3.4 Simulation Study

The simulation study shows the advantages and limits of the use of the slopes in the cluster-mass test. We keep the design simple in order to highlight the difference between the classical cluster-mass test and its extension described previously. We simulate signals using a design with 2 groups, with $n = 22$ participants and 600 time-points[2]. A t statistic is used to test the hypothesis. The slopes are computed using smoothing splines with a smoothing parameter such that the roughness of the original signals match the one of their slopes. The FWER of both methods is, as expected, close to the nominal level (for the classical cluster-mass test, $\hat{q} = .053$ with 95% CI [.046; .060], and for its extension using the slope, $\hat{q} = .050$ with 95% CI [.044; .058]). Moreover, we simulate data with a true effects similar to a ”wave” shape, with 3 spikes (positive, negative and then positive as shown in the top panel of Figure 2.11). In Figure 2.11, we see the average power of the test for each time-point using 3 different effect sizes. When the effect size on the original signals is small (Figure 2.11, second panel from the top), the average power of

[2] The error are simulated using an exponential autocorrelation function $\rho(\tau) = -3(\tau/30)^2$ (Abrahamsen, 1997) and $\sigma = 1.5$

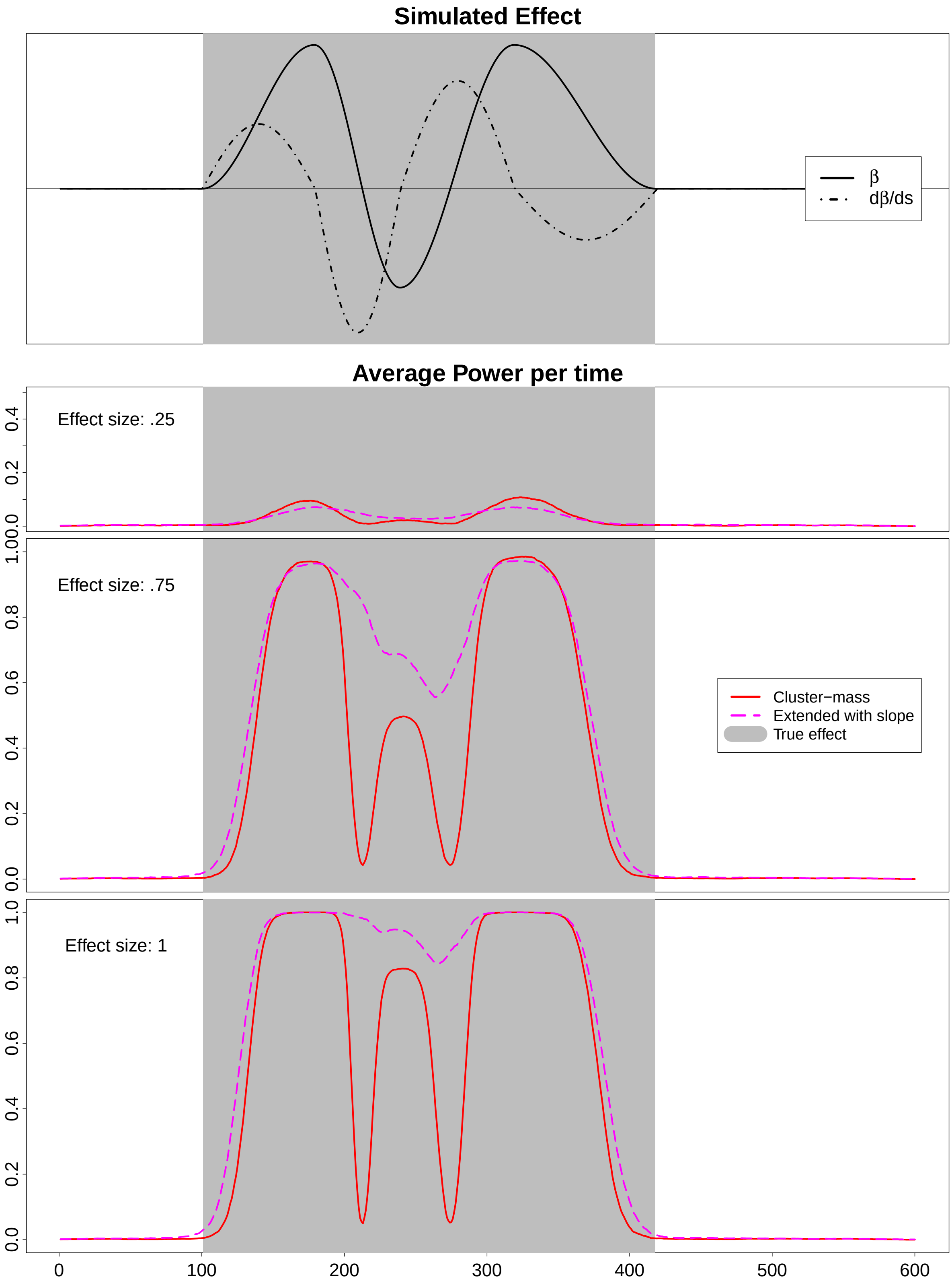

Figure 2.11: Average power of the classical cluster-mass test and its extension using slopes. The top panel represents the shape of the true effect and its slope. The bottom three panels show the average power for effect size (β's) multiplied by .25, .75 and 1. The average power is higher between spikes when using the slopes.

Table 2.1: The False Positive Rate corresponds to the average rate of discovery for the time-points under the null hypothesis (1ms to 100ms and 419ms to 600ms, or white areas in Figure 2.11). The False Positive Rate for the 10 time-points near the true effect (from 96ms to 100ms and 419ms to 423ms) is more influenced when using the slope. Finally, the true discovery rate corresponds to the average rate of discovery for the time-points under the alternative (between 101ms and 418ms or gray area in Figure 2.11). Confidence intervals are computed using Agresti and Coull (1998).

Method	Effect .25	Effect .75	Effect 1.0
False Positive Rate			
Cluster-mass test	.0027 [.0014;.0049]	.0022 [.0011;.0042]	.0020 [.0010;.0040]
Extension using slope	.0039 [.0023;.0064]	.0042 [.0026;.0068]	.0043 [.0026;.0069]
False Positive Rate (10 nearest)			
Cluster-mass test	.0041 [.0025;.0067]	.0038 [.0022;.0063]	.0040 [.0024;.0065]
Extension using slope	.0052 [.0033;.0080]	.0118 [.0088;.0156]	.0162 [.0127;.0206]
True Discovery Rate			
Cluster-mass test	.0390 [.0335;.0455]	.5080 [.4928;.5237]	.6364 [.6217;.6515]
Extension using slope	.0375 [.0321;.0439]	.6499 [.6353;.6649]	.7904 [.7779;.8031]

both methods is small and using the slopes does not results in an increase of the average power. However, when the effect size becomes larger (two bottom panel of Figure 2.11), the transition between positive and negative spike is more often declared significant when using the slopes. However, due to the smoothing of the slopes, more false positive tests are measured at the edge of the true effects (see Table 2.1, section ”False Positive Rate (10 nearest)”). This may be a negative counterpart of the cluster-mass test using the slopes as it increases the number of false positive. It results in clusters that may be larger than the true effect. In our simulation settings, the average false positive rate is still low even for the tests near the true effect (for the 10 nearest tests and high effect size: $FPR = 0.0162$) but it may depend on the shape of the true effect. However, from a practical perspective, neuroscientists may not be interested by the precise position (in time) of the edge of the clusters and this uncertainty may be a reasonable trade-off for an increase in power.

2.4 Conclusion

In Section 2.1, we explain that permutation methods have a geometrical representation. It helps to understand links between the permutation methods. However, the permutation methods which modify the design (like `dekker`) are not well adapted to this graphical representation. Indeed, in this case, the whole plan $[D\ X]$ rotates for each permutation which is more complicated both to represent graphically and also to understand using a 3D graphic. However, a further exploration of this graphical representation may come from the interpretation of the F statistic as a function of an angle. Using this interpretation, permuting the response variables modify only one vector of this angle while permuting the design modifies the other one. A better understanding of the effect of permutations on this angle may leads to a clever graphical representation of the methods permuting the design, especially for the `dekker` method.

In Section 2.2, we describe a real data analysis of a full scalp EEG experiment using the cluster-mass test. We explain some challenges encounter from implementing the cluster-mass test to the graphical representation of the results. We plan to implement the functions used for this data analysis in the next release of the `permuco` package.

In Section 2.3, we present a cluster-mass test using the slopes of the signals to increase the power of the test. This method has still to be investigated in details to be useful for real data applications. Indeed, it has the drawback to increase the false positive rate. Moreover, the choice of the smoothing parameter when using splines or local polynomial is actually not made on an optimality criterion. A better understanding of the effect of the smoothing parameter on the false positive rate could lead to a clever choice of the smoothing parameters.

Chapter 3

Finite Sample and Asymptotic Properties of the Conditional Distribution by Permutations

3.1 Introduction

In this Chapter, we introduce theoretical findings on permutation methods for regression and factorial designs. We propose a formalization of the distribution by permutation as a conditional distribution given the observation of the response variable. We then show that the expectation and variance of the conditional distribution by permutation can be computed analytically. These findings are first applied to investigate the conditional distribution by permutation of the F statistic for finite sample size. This approach is general and is applied using several permutation methods including the one introduced by Manly (1991), Kennedy (1995), Freedman and Lane (1983) or ter Braak (1992). It allows to produce a correction of the permutation distribution of the `kennedy` method. Similarly to Pauly et al. (2015) which found the asymptotic distribution for a Wald type statistic in the Behrens–Fisher problem in a factorial design, we then derive the asymptotic distribution by permutation for the F statistic in a similar setting. Finally, we show the validity of several permutation methods as we give the asymptotic distribution of the F statistic not only for the `manly` permutation method but also for the `kennedy`, `freedman_lane` and `terBraak` permutation methods.

3.2 Regression and ANOVA Model

The general setting is a regression model (which includes ANOVA) which equation is written as:

$$Y = D\eta + X\beta + \epsilon, \tag{3.1}$$

where Y is the response random variable, X are the interest variables associated to the effects of interest β and D are the nuisance variables associated to the nuisance effects η. We assume without loss of generality that the design $[D\ X]$ is of full rank and with variables that may be correlated. Moreover, we assume that the intercept is part of the nuisance variables ($\mathbf{1} \in D$). Finally, ϵ is a random error following a distribution $\epsilon \sim (0, \Omega)$.

In the model of Equation 3.1, we are interested in testing the hypothesis:

$$H_0 : \beta = 0. \tag{3.2}$$

We use a F statistic that we write:

$$F_Y = \frac{Y^\top H_{R_D X} Y \ / \ (p-q)}{Y^\top R_{D,X} Y \ / \ (n-p)}, \tag{3.3}$$

where $H_.$, $R_.$ are respectively the "hat" matrix and "residuals" matrix of the design in subscript. For a general full rank design matrix M, the "hat" matrix is $H_M = M(M^\top M)^{-1}M^\top$ and the "residuals" matrix is $R_M = I - H_M$. In our notation, the subscript D, X means the column-wise binding of the matrices D and X, or $[D \ X]$. If the error term follows a homoscedastic normal distribution ($\epsilon \sim N(0, \sigma_\epsilon^2 I)$), we derive the well known distributions under the null hypothesis:

$$\sigma_\epsilon^2 Y^\top H_{R_D X} Y \sim \chi^2(p-q) \tag{3.4}$$

$$\sigma_\epsilon^2 Y^\top R_{D,X} Y \sim \chi^2(n-p), \tag{3.5}$$

which implies:

$$F_Y \sim F(p-q, n-p). \tag{3.6}$$

Moreover it implies that $(p-q)F_Y$ asymptotically converges to a $\chi^2(p-q)$ (Seber and Lee, 2012). Note that the numerator and denominator of the statistic have both the same expectation of σ_ϵ^2 in the parametric case since the expectation of a χ^2 is the degrees of freedom.

3.3 Finite Sample Conditional Distribution by Permutation

In the following Section, we show that the conditional expectation of the numerator and denominator of the F statistic can be computed analytically for the `manly`, `kennedy`, `freedman_lane` and `terBraak` method of Table 1.1. Cox and Hinkley (1979) already proposed computing moments of the permutation distribution when testing linear dependency between two variables. The results we present in the next Section allow to identify the problem with the `kennedy` permutation method when using the F statistic as it is well known that the `kennedy` method increases the type I error rate in finite sample size (Anderson and Legendre, 1999). Moreover, they imply a natural correction for the `kennedy` method. In addition, computing conditional expectations gives a good insight why the `manly` performs well in simulation studies despite not reducing the effect of nuisance variables. Finally, using a simulation study, we highlight the problem of the `kennedy` method and show the improvement of the type I error rate using this natural correction.

Given y, the equation of the model 3.1 is:

$$Y | y = y = D\eta + X\beta + e, \tag{3.7}$$

where y is the observed realisation of Y and e is the unknown realisation of ϵ.

Lemma 1. For the linear model of Equation 3.1, over all permutations, the conditional expectation of the numerator and denominator of the F statistic given the observation of y are summarized in the following Table.

	Manly	Freedman-Lane	terBraak	Kennedy
Numerator	$\frac{1}{n-1}y^\top R_{\mathbf{1}}y$	$\frac{1}{n-1}y^\top R_D y$	$\frac{1}{n-1}y^\top R_{D,X}y$	$\frac{1}{n-1}y^\top R_D y$
Denominator	$\frac{1}{n-1}y^\top R_{\mathbf{1}}y$	$\frac{1}{n-1}y^\top R_D y$	$\frac{1}{n-1}y^\top R_{D,X}y$	$\frac{1}{n-1}\frac{n-p+q}{n-p}y^\top R_D y$

The conditional expectation in Lemma 1 are equal for the numerator and denominator using the `manly`, `freedman_lane` and `terBraak` methods (like the parametric distribution under normality and homoscedasticity assumptions). However, for the `kennedy` method, the conditional expectation of the denominator is larger than the one in the numerator. Note that $y^\top R_{\mathbf{1}}y$ is the sum of squares of the empirical variance of the vector y and $y^\top R_D y$, $y^\top R_{D,X}y$ are the sum of squares of the residuals of the "small" and "full" model, respectively. Loosely speaking, `manly` inflates the numerator compared to the parametric expectation but this is compensated by the same inflation in the denominator. This is formalized in Section 3.4.

3.3.1 Proof of Lemma 1 for the `manly` Permutation Method

Given the observation y of the response variable in Equation 3.1 and the set of all $n!$ permutation matrices of size $n \times n$: $\mathcal{P} = \{P_1, \dots, P_{n!}\}$, the `manly` permutation method is summarized by the transformation of the data $\{y, D, X\} \to \{Py, D, X\}$. Using this transformation, we define the conditional multivariate distribution by permutation given the observation y by assigning the same probability to all permutations of the vector y:

$$\Pr\left((Y^*|y) = Py\right) = \frac{1}{n!} \;\forall P \in \mathcal{P}. \tag{3.8}$$

Then, the conditional expectation over all permutations of $Y^*|y$ is simply:

$$\mathrm{E}_{\mathcal{P}}\left[Y^*|y\right] = \sum_{P\in\mathcal{P}} \Pr\left((Y^*|y) = Py\right) Py = \sum_{P\in\mathcal{P}} \frac{1}{n!} Py = H_{\mathbf{1}}y = \frac{1}{n}\mathbf{1}\mathbf{1}^\top y. \tag{3.9}$$

Moreover, by using the result of Equation C.2, we also compute its conditional variance over all permutations:

$$\begin{aligned}\mathrm{Var}_{\mathcal{P}}\left[Y^*|y\right] &= \mathrm{E}_{\mathcal{P}}\left[Y^*Y^{*\;\top}|y\right] - \mathrm{E}_{\mathcal{P}}\left[Y^*|y\right]\mathrm{E}_{\mathcal{P}}\left[Y^{*\;\top}|y\right] \\ &= \frac{1}{n!}\sum_{P\in\mathcal{P}} Pyy^\top P^\top - H_{\mathbf{1}}yy^\top H_{\mathbf{1}} = \frac{1}{n-1}R_{\mathbf{1}}y^\top R_{\mathbf{1}}y.\end{aligned} \tag{3.10}$$

Finally, using the definition of the conditional distribution by permutation in Equation 3.8 and the property derived in Equation C.3 of Appendix C.1.1, we compute the conditional expectation of the numerator of the F statistic:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{p-q}Y^{*\top}H_{R_DX}Y^*|y\right] = \frac{1}{p-q}\frac{1}{n!}y^\top\left(\sum_{P\in\mathcal{P}} P^\top H_{R_DX}P\right)y = \frac{1}{n-1}y^\top R_{\mathbf{1}}y, \tag{3.11}$$

Table 3.1: Type I error rate of tests of one or two parameters simultaneously (column $p-q$) for a regression models with 2 different sample sizes (column n). If the ratio $\frac{n-p+q}{n-p}$ increases (column ratio), the discrepancy to the nominal level of the type I error of the `kennedy` method also increases. Correcting the `kennedy` method by the inverse of the ratio produces a type I error rate closer to the nominal level (`Ken. Corr.`); like the `manly`, `terBraak` and `freedman_lane` methods (and the parametric test). Confidences interval are computed using Agresti and Coull (1998). Bold font corresponds to nominal level (5%) within the confidence interval and red font corresponds to confidence interval above the nominal level.

n	p-q	ratio	Parametric	Manly	Freedman-Lane	terBraak	Kennedy	Ken. Corr.
10	1	1.8	**.045**	**.043**	**.047**	**.045**	.121	**.053**
			[.039;.052]	**[.037;.050]**	**[.041;.054]**	**[.039;.052]**	[.111;.132]	**[.046;.060]**
10	2	1.6	**.050**	**.048**	**.050**	**.050**	.119	**.055**
			[.043;.057]	**[.042;.055]**	**[.044;.058]**	**[.043;.057]**	[.109;.129]	**[.048;.063]**
20	1	1.27	**.048**	**.047**	**.048**	**.049**	.073	**.046**
			[.042;.056]	**[.041;.054]**	**[.042;.056]**	**[.043;.056]**	[.066;.082]	**[.040;.054]**
20	2	1.2	**.050**	**.050**	**.051**	**.051**	.069	**.048**
			[.044;.058]	**[.044;.058]**	**[.045;.059]**	**[.044;.058]**	[.062;.077]	**[.042;.055]**

and, for its denominator:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{n-p}Y^{*\top}R_{D,X}Y^{*}|y\right]=\frac{1}{n-1}y^{\top}R_{1}y. \tag{3.12}$$

In Appendix C.4 we derive the conditional distribution by permutation of the numerator and denominator for the `kennedy`, `freedman_lane` and `terBraak` which are reported in Lemma 1.

3.3.2 A Modification of the `kennedy` Method

For the `manly`, `freedman_lane` and `terBraak` methods, both numerator and denominator of the F statistic have the same expectation, similarly to the parametric setting. For the `kennedy` method, the numerator and the denominator do not share the same conditional expectation and we expect that it could lead to incorrect type I error for small sample. However, we can correct the conditional distribution in order to have same expectation by dividing the permuted statistics by the factor $\frac{n-p+q}{n-p}$ (but letting unchanged the observed statistic). We use a simulation study to test the effect of the correction of the `kennedy` method.

Extensive simulations for several permutation methods, under several conditions have already been proposed. Anderson and Legendre (1999) propose simulation study showing the type I error rate of the `kennedy`, `freedman_lane` and `terBraak` methods and already highlight the problems using the `kennedy` method. Kherad Pajouh and Renaud (2010) add the `huh_jhun` method to their simulation study and shows the type I error rate under several distributions of the error terms. They show that `huh_jhun` method keeps a type I error rate close to the nominal level even under non-normality. Moreover, Winkler et al. (2014) test, in addition, the `manly` and `dekker` (noted "Smith" and attributed to O'Gorman (2005) and Nichols et al. (2008)) permutation method, and also use the F

statistic and a new statistic which adjusts the variance within each group. In addition to several designs and error terms, they also test shuffling (coin-flip) instead of permuting the data. They conclude that the `dekker` and `freedman_lane` methods perform well overall.

Our simulation study proposes only to highlight the effect of the ratio $\frac{n-p+q}{n-p}$ on the `kennedy` method. Hence, we choose to keep all parameters fixed while varying this ratio. We simulated regression models, with 4 correlated continuous variables, a normal homoscedastic uncorrelated random error ($\sigma = 1$). 1 or 2 parameters are tested simultaneously and 2 different sample sizes are simulated. In Table 3.1, we first see that increasing the ratio $\frac{n-p+q}{n-p}$ increases the type I error rate of the `kennedy` method and has not any obvious effect on the other methods. Moreover, dividing the distribution by permutation of the F statistics by $\frac{n-p+q}{n-p}$ corrects the `kennedy` method and produces a type I error rate close to the nominal level. Finally, all methods with the expectation of the numerator and denominator equal, including the `manly`, `freedman_lane`, `terBraak`, the corrected `kennedy` methods and also the parametric distribution under normality, keep the type I error rate close to the nominal level. All these findings confirm that the computation of the conditional moments as described in Section 3.3 is a valid tool to investigate the distribution by permutation in finite sample size of a statistic given a permutation method.

3.4 Asymptotic of the Conditional Distribution by Permutation

In the two-samples problem, let us first suppose that we are interested in testing the null hypothesis $H_0 : F_1 = F_2$ where F_1 and F_2 are distributions of the two populations. It is well known that for any statistic, a permutation test is exact meaning that it attains exactly its actual level, except for small discrepancies caused by the discreteness, including ties in the values of the statistics (Efron and Tibshirani, 1994). However, if we are interested in testing a null hypothesis that is specific to parameters, (e.g. the equality of expectation of two distributions) results are more complex. Asymptotic properties have already been studied by Romano (1990) for the comparison of 2 means. He shows that, when using a statistic that is basically the difference of two means, the test is not valid for 2 groups of different sizes and different variances. However, Chung and Romano (2013) show that using studentized statistics leads to valid tests. A similar result is proven by Pauly et al. (2015) using a Wald type statistic for testing contrasts in a factorial design which confirms the importance of studentized statistic in the Behrens-Fisher problem. Pauly et al. (2015) found the asymptotic distribution of the Wald statistic using permutations without restriction which we later called the `manly` permutation method. In the following Section, we show that the F statistics is also a valid test for the same designs than Pauly et al. (2015) and the proof holds for the `manly`, `kennedy`, `freedman_lane` and `terBraak` permutation methods.

The findings in Section 3.3 are general in a regression or ANOVA setting. However, in the following Section, we must restrict the design. We assume a factorial design $[D\ X]$ with G groups or cells. This assumption include all (fixed-effect) one-way or factorial ANOVA designs. The proportion of observations coming from these groups are $\hat{\kappa}_1, \ldots, \hat{\kappa}_G$ such that $\hat{\kappa}_1 + \cdots + \hat{\kappa}_G = 1$ and these proportions converge to positive values. Moreover, we assume uncorrelated heteroscedastic error, with the same variance within each cell.

Table 3.2: Asymptotic convergence by permutation method. $\sigma^2_{D\eta}$ and σ^2_ϵ are defined in Equation 3.21 and 3.22, respectively.

	Manly	Freedman-Lane	terBraak	Kennedy
Numerator	$\frac{\sigma^2_{D\eta}+\sigma^2_\epsilon}{p-q}\chi^2(p-q)$	$\frac{\sigma^2_\epsilon}{p-q}\chi^2(p-q)$	$\frac{\sigma^2_\epsilon}{p-q}\chi^2(p-q)$	$\frac{\sigma^2_\epsilon}{p-q}\chi^2(p-q)$
Denominator	$\sigma^2_{D\eta}+\sigma^2_\epsilon$	σ^2_ϵ	σ^2_ϵ	σ^2_ϵ
F	$\frac{1}{p-q}\chi^2(p-q)$	$\frac{1}{p-q}\chi^2(p-q)$	$\frac{1}{p-q}\chi^2(p-q)$	$\frac{1}{p-q}\chi^2(p-q)$

Altogether, it adds the following assumptions:

$$[D\ X] \text{ is a factorial design with } \mathbf{1} \in D, \tag{3.13}$$

$$\begin{gathered}\hat{\kappa}_1, \dots, \hat{\kappa}_G \rightarrow \kappa_1, \dots, \kappa_G \text{ when } n \rightarrow \infty, \\ \text{where } \kappa_g > 0 \ \forall\ g \in\ 1, \dots, G\end{gathered} \tag{3.14}$$

and finally,

$$\begin{gathered}\epsilon = [\epsilon_1 \ \dots \ \epsilon_i \ \dots \ \epsilon_n] \sim \left(0, \operatorname{diag}(\sigma^2_{g(1)}, \dots, \sigma^2_{g(i)}, \dots, \sigma^2_{g(n)})\right), \\ \text{where } \sigma^2_{g(i)} = \sigma^2_g < \infty \ \forall\ g \in 1, \dots, G,\end{gathered} \tag{3.15}$$

with $\sigma^2_{g(i)}$ the variance of the $g(i)$th group (or cell of a factorial design) and $g(i) \in 1, \dots, G$ is a function returning the cell number of the observation i. The function $g(i)$ is added to link the observations with the cell and is simplified by the subscript g to designate the cell of the factorial design when it is possible. In a factorial design without empty cell, the number of parameters is equal or greater than the number of variances ($p \geq G$) and, within each cell, all variances σ^2_g are equal.

Theorem 1. Under the model in Equation 3.1, if the hypothesis in Equation 3.2 is true and if the conditions 3.13, 3.14 and 3.15 are met, the conditional distribution by permutation of $(p-q)F_{Y^*}$ given the observation y converges to a $\chi^2(p-q)$ when $n \rightarrow \infty$, for the `manly`, `kennedy`, `freedman_lane` and `terBraak` permutation methods. These converging distributions are the same as the (unconditional) distribution under normality and homoscedasticity assumptions. The details of the asymptotic distributions of the numerator and denominator are presented in Table 3.2.

The proof of Theorem 1 uses properties of the "hat" matrix and of the QR decomposition which we first recall in Section 3.4.1.

In Section 3.4.2, we proove Theorem 1 for the `manly` permutation method and in Appendix C.4.1, C.4.2 and C.4.3, we prove Theorem 1 for the `kennedy`, `freedman_lane` and `terBraak` methods respectively which asymptotic results are summarized in Table 3.2.

3.4.1 Properties of the "Hat" Matrix and of the QR Decomposition

For a general full rank design matrix M, the diagonal elements $H_{M:[ii]}$ of the "hat" matrix $H_M = M(M^\top M)M^\top$ are mostly used to detect leverage points (Hoaglin and Welsch, 1978) in the regression setting. However, in a factorial design, the "hat" matrix of the full design H_M computes the fitted values of the response variable which correspond to the means of each group. This implies that the individual coding of the factors (for testing main effects or simple effects, etc) does not influence the "hat" matrix of the full design H_M. Without empty cell, the number of groups is equal to the rank of the full design matrix ($\text{rank}(M) = p = G$). The "hat" matrix of the full factorial design is then block-diagonal when the observations are sorted by cells and is simply $\text{diag}(H_{\mathbf{1}_{n\hat{\kappa}_1}}, \dots, H_{\mathbf{1}_{n\hat{\kappa}_g}}, \dots, H_{\mathbf{1}_{n\hat{\kappa}_p}})$, where the $\hat{\kappa}_g$ corresponds to the proportion of data coming from the cells g and $\mathbf{1}_{n\hat{\kappa}_g}$ is a vector of 1s of length $n\hat{\kappa}_g$. In this example, $H_{\mathbf{1}_{n\hat{\kappa}_g}} = (n\hat{\kappa}_g)^{-1}\mathbf{1}_{n\hat{\kappa}_g}\mathbf{1}_{n\hat{\kappa}_g}^\top$ as $\mathbf{1}_{n\hat{\kappa}_g}^\top\mathbf{1}_{n\hat{\kappa}_g} = n\hat{\kappa}_g$. Finally, each diagonal element of H_M is simply $H_{M:[ii]} = (n\hat{\kappa}_{g(i)})^{-1}$.

The QR decomposition of M is especially useful as it defines an orthonormal basis M. The projection used in the F statistic may be computed in this new basis while using properties of an orthonormal basis. Algebraically, the QR decomposition produces $M = Q_M U_M$, such that Q_M is an orthonormal basis of the span of M and U_M is an upper diagonal square matrix. Geometrically, a projection on the span of M is the same as the projection on the span of Q_M as:

$$\begin{aligned} H_M &= Q_M U_M \left(U_M^\top Q_M^\top Q_M U_M\right)^{-1} U_M^\top Q_M^\top \\ &= Q_M U_M \left(U_M\right)^{-1} \left(U_M^\top\right)^{-1} U_M^\top Q_M^\top \\ &= Q_M Q_M^\top = Q_M \left(Q_M^\top Q_M\right) Q_M^\top = H_{Q_M}, \end{aligned} \tag{3.16}$$

using the orthonormal properties of Q_M, $Q_M^\top Q_M = I$.

The Grahm-Schmidt process (Pursell and Trimble, 1991; Seber and Lee, 2012) is used to compute the QR decomposition. It decomposes a matrix M into $Q_M U_M$. However, we are mainly interested by the properties of the Q_M. Using the Grahm-Schmidt process, the Q_M matrix may be written using "residuals" matrices. The jth column of the Q_M matrix is:

$$Q_{M:[j]} = \begin{cases} M_{[j]}/\sqrt{M_{[j]}^\top M_{[j]}} \text{ for } j = 1 \\ R_{M_{[1\dots j-1]}} M_{[j]}/\sqrt{M_{[j]}^\top R_{M_{[1,\dots,j-1]}} M_{[j]}} \text{ for } j > 1, \end{cases} \tag{3.17}$$

where the square brackets in subscript indicate a selection of columns. Moreover, the denominators are the norm of the numerators in order to produce normalized column vectors of Q_M. Each step of the algorithm computes the residuals of a regression and normalizes them. When the design is split into nuisance and interest variables as in Equation 3.1 ($M = [D\ X]$), the Grahm-Schmidt process also implies the following equality $Q_{D,X:[q+j]} = Q_{R_D X:[j]}\ \forall\ j \in 1, \dots, p-q$, provided that D is of full rank.

In addition, using again the orthonormality property of Q_M, the "hat" matrix of the full design $[M]$ is rewritten using the p vectors $Q_{M:[j]}$ such that:

$$H_M = Q_M Q_M^\top = \sum_{j=1}^{p} Q_{M:[j]} Q_{M:[j]}^\top. \tag{3.18}$$

It follows that the QR decomposition helps us to rewrite the sum of squares used in the F statistic. For the denominator, we have the following form:

$$\begin{aligned} \frac{1}{n-p} y^\top R_M y &= \frac{1}{n-p}\left(y^\top y - y^\top H_M y\right) \\ &= \frac{1}{n-p} y^\top y - \sum_{j=1}^{p}\left(\frac{1}{\sqrt{n-p}} Q_{M:[j]}^\top y\right)^2, \end{aligned} \tag{3.19}$$

where $Q_M = \left[\begin{array}{ccccc} Q_{M:[1]} & \dots & Q_{M:[j]} & \dots & Q_{M:[p]} \end{array} \right]$ is an orthonormal basis of M. Moreover, if the first column of M codes the intercept implies that $Q_{M:[1]} = \frac{1}{\sqrt{n}}\mathbf{1}$.

Finally, in a factorial design, the construction of the QR decomposition described in Equation 3.17 implies that each observation (i.e. line) belonging to the same cell has the same element of $Q_{M:[j]}$ $\forall j$ (i.e. $Q_{M:[ij]} = Q_{M:[i'j]}$ whenever $g(i) = g(i')$).

3.4.2 Proof of Theorem 1 for the `manly` Permutation Method

Using, model assumptions 3.13, 3.14 and 3.15, we first deduce some intermediary convergence results under the null hypothesis:

$$\frac{1}{n}\mathbf{1}D\eta \to \mu_{D\eta} < \infty, \tag{3.20}$$

$$\frac{1}{n}\eta D^\top R_{\mathbf{1}} D\eta \to \sigma^2_{D\eta} < \infty, \tag{3.21}$$

$$\frac{1}{n}eR_{\mathbf{1}}e \to \sigma^2_\epsilon < \infty. \tag{3.22}$$

The convergence 3.20 and 3.21 holds for a factorial design under assumptions in Equations 3.14 and 3.15, $\mu_{D\eta}$ is the population overall mean and $\sigma^2_{D\eta}$ is weighted dispersion of the group means. Finally, σ^2_ϵ is the weighted average of the error variances $\sigma^2_\epsilon = \sum_{g=1}^G \kappa_g \sigma^2_g$.

Then, using the QR decomposition, we write the numerator of the F statistic:

$$\frac{1}{p-q} Y^{*\ \top} H_{R_D X} Y^* = \sum_{j=1}^{p-q} \left(\frac{1}{\sqrt{p-q}} Q^\top_{R_D X:[j]} Y^* \right)^2, \tag{3.23}$$

and its denominator:

$$\frac{1}{n-p} Y^{*\ \top} R_{D,X} Y^* = \frac{1}{n-p} Y^{*\ \top} Y^* - \sum_{j=1}^{p} \left(\frac{1}{\sqrt{n-p}} Q^\top_{D,X:[j]} Y^* \right)^2. \tag{3.24}$$

Theorem 1 is proven by computing the asymptotic of each term of the denominator (Equation 3.24) and each term of the numerator (Equation 3.23) of the F statistic.

We first show that conditional distribution of the denominator converges in probability when assuming the null hypothesis. The first part of the denominator ($\frac{1}{n-p} Y^{*\ \top} Y^*$) converges in probability as:

$$\mathrm{E}_\mathcal{P}\left[\frac{1}{n-p} Y^{*\top} Y^* | y \right] = \frac{1}{n-p} y^\top y \to \mu^2_{D\eta} + \sigma^2_{D\eta} + \sigma^2_\epsilon, \tag{3.25}$$

where details are given in Appendix C.3 and

$$\mathrm{Var}_\mathcal{P}\left[\frac{1}{n-p} Y^{*\top} Y^* | y \right] = 0, \tag{3.26}$$

as $Y^{*\top} Y^* | y = y^\top y$ for all permutations.

For the elements of the second part of the denominator ($\frac{1}{\sqrt{n-p}} Q^\top_{D,X:[j]} Y^*$ in Equation 3.24), we assume that the intercept is coded in the first column of the matrix D. Using the Grahm-Schmidt process, the QR decomposition of the design is such that $Q_{D,X:[1]} = \frac{1}{\sqrt{n}}\mathbf{1}$. Moreover, all vectors $Q_{D,X:[j]}$ for $j \in 2, \dots, p$ are orthogonal to $\mathbf{1}$ such that $Q^\top_{D,X:[j]}\mathbf{1} = 0$ for $j \in 2, \dots, p$. Using these properties and Equation 3.9, we find convergence in probability of the second part of the denominator by computing:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}}Q^\top_{D,X:[j]}Y^*|y\right] = \frac{1}{n\sqrt{n-p}}Q^\top_{D,X:[j]}\mathbf{1}\mathbf{1}^\top y = \begin{cases} \frac{1}{\sqrt{n(n-p)}}\mathbf{1}^\top y \to \mu_{D\eta} \text{ for } j = 1 \\ 0 \text{ for } j \in 2, \dots, p \end{cases} \tag{3.27}$$

Then, the convergence of the conditional variance of $\frac{1}{\sqrt{n-p}}Q^\top_{D,X:[j]}Y^*$ is derived:

$$\begin{aligned}\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}}Q^\top_{D,X:[j]}Y^*|y\right] &= \frac{1}{n-p}Q^\top_{D,X:[j]}\mathrm{Var}_{\mathcal{P}}\left[Y^*|y\right]Q_{D,X:[j]} \\ &= \frac{1}{(n-p)(n-1)}y^\top R_{\mathbf{1}}yQ^\top_{D,X:[j]}R_{\mathbf{1}}Q_{D,X:[j]} \\ &= \frac{1}{(n-p)(n-1)}y^\top R_{\mathbf{1}}y\left(1-\frac{1}{n}\left(\mathbf{1}^\top Q_{D,X:[j]}\right)^2\right) \to 0, \end{aligned} \tag{3.28}$$

where the 2 equalities come from Equation 3.10 and $\frac{1}{n-1}y^\top R_{\mathbf{1}}y$ is the empirical variance of y and therefore bounded.

By using the continuous mapping theorem (Mann and Wald, 1943) and Equation 3.25 to 3.28, the conditional distribution of the denominator converges in probability to $\sigma^2_{D\eta} + \sigma^2_\epsilon$.

We then show that the terms in the numerator ($\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^* \;\forall\; j \in 1, \dots, p-q$) converge to normal distributions. We compute its conditional expectation and variance:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y\right] = \frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}H_{\mathbf{1}}y = 0, \tag{3.29}$$

and

$$\begin{aligned}\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y\right] &= \frac{1}{(p-q)(n-1)}y^\top R_{\mathbf{1}}Q^\top_{R_DX:[j]}R_{\mathbf{1}}Q_{R_DX:[j]}y \\ &= \frac{1}{(p-q)(n-1)}y^\top R_{\mathbf{1}}y \to \frac{1}{p-q}\left(\sigma^2_{D\eta}+\sigma^2_\epsilon\right), \end{aligned} \tag{3.30}$$

where the third equation uses the property $Q^\top_{R_DX:[j]}R_{\mathbf{1}}Q_{R_DX:[j]} = 1$ if $\mathbf{1} \in D$.

Similarly to Pauly et al. (2015), $\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y$ is a weighted sum of $Y^*|y$ and it therefore has the same conditional distribution as $\frac{1}{\sqrt{p-q}}Q^{*\top}_{R_DX:[j]}Y|y$, where $Q^*_{R_DX:[j]}$ is defined as a random variable following the distribution by permutation of the vector $Q_{R_DX:[j]}$ (similarly to Equation 3.8). More precisely, each equiprobable outcome of $\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y$ corresponds to one outcome of $\frac{1}{\sqrt{p-q}}(Q^*_{R_DX:[j]})^\top Y|y$ such that:

$$\frac{1}{\sqrt{p-q}}\left(P^\top Q_{R_DX:[j]}\right)^\top Y|y = \frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}PY|y \;\forall\; P \in \mathcal{P}. \tag{3.31}$$

It follows the equality in distribution of the two notations:

$$\frac{1}{\sqrt{p-q}}\left(Q^*_{R_DX:[j]}\right)^\top y \overset{d}{=} \frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y. \tag{3.32}$$

Moreover, adapting the equations 8.3 to 8.7 of the Supplementary Material of Pauly et al. (2015) (applying himself theorems in Pauly (2011) and Janssen (2005)) , the asymptotic normality of Equation 3.32 is proven if 5 conditions (from Equation 3.33 to 3.37) are met. Like Pauly et al. (2015), we show that they directly result from the model assumptions presented in Equations 3.13, 3.14 and 3.15. The 5 conditions are written:

$$\max_{i\in 1,\dots,n}\left|\frac{1}{\sqrt{n}}\left(R_{\mathbf{1}}y\right)_i\right| \to 0, \tag{3.33}$$

$$\frac{1}{n}y^\top R_{\mathbf{1}}y \to \sigma^2_{D\eta}+\sigma^2_{\epsilon}, \tag{3.34}$$

$$\max_{i\in 1,\dots,n}\left|\left(Q_{R_DX:[j]}\right)_i\right| \to 0, \tag{3.35}$$

$$Q^\top_{R_DX:[j]}Q_{R_DX:[j]} = 1, \tag{3.36}$$

and, finally,

$$\sqrt{n}\left(Q^*_{R_DX:[j]}\right)_i \xrightarrow{D} (0,1)\ \forall\ i. \tag{3.37}$$

To verify the condition in Equation 3.33, we apply the results in Appendix C.2 within each cell to prove the convergence:

$$\max_{i\ \text{s.t.}\ g(i)=g}\left|\frac{1}{\sqrt{\hat{\kappa}_g n}}\left(y\right)_i\right| \to 0\ \forall\ g\ \in\ 1,\dots,G. \tag{3.38}$$

It follows that

$$\max_{g\in 1,\dots,G}\left(\max_{i\ \text{s.t.}\ g(i)=g}\left|\frac{1}{\sqrt{\hat{\kappa}_g n}}\left(y\right)_i\right|\right) \to 0, \tag{3.39}$$

which consequently verifies the condition in Equation 3.33.

Moreover, the condition in Equation 3.34 stems from the convergences defined in Equations 3.21 and 3.22.

The condition in Equation 3.35 depends directly on the behaviour of the diagonal of the "hat" matrix $H_{D,X}$. Using Equation 3.18, we find the following inequality: $\left|\left(Q_{R_DX:[j]}\right)_i\right| \leq \sqrt{H_{D,X:[ii]}}$ for all $i \in 1,\dots,n$. In a factorial design, we recall that the diagonal elements of the "hat" matrix depends on the proportion of observations in each group such that $\sqrt{H_{D,X:[ii]}} = (n\hat{\kappa}_{g(i)})^{-1/2}$. In addition, the condition in Equation 3.14 assumes the convergence to a finite value of $\hat{\kappa}_{g(i)} = \hat{\kappa}_g \to \kappa_g\ \forall\ g \in 1,\dots,G$ when $n\to\infty$ which implies condition in Equation 3.35.

The condition in Equation 3.36 is satisfied as it depends directly on the construction of the QR decomposition.

Finally, for a finite sample size, the distribution by permutation of $Q^*_{R_DX:[j]}$ is such that, $\mathrm{E}_{\mathcal{P}}\left[Q^*_{R_DX:[j]}\right] = H_{\mathbf{1}}Q_{R_DX:[j]} = 0$ and $\mathrm{Var}_{\mathcal{P}}\left[Q^*_{R_DX:[j]}\right] = \frac{1}{n-1}R_{\mathbf{1}}Q^\top_{R_DX:[j]}R_{\mathbf{1}}Q_{R_DX:[j]} = \frac{1}{n-1}R_{\mathbf{1}}$, as $Q_{R_DX:[j]}$ is orthogonal to $\mathbf{1}$ which implies the convergence in Equation 3.37.

Altogether, results in Pauly et al. (2015) hold for the terms of the numerator $\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y$ and proves their asymptotic normality ($\mathcal{N}\left(0,\frac{1}{p-q}(\sigma^2_{D\eta}+\sigma^2_{\epsilon})\right)$). As $Q^\top_{R_DX:[j]}Q_{R_DX:[k]} = 0\ \forall\ j\neq k$, the $p-q$ asymptotic distributions of $\frac{1}{\sqrt{p-q}}Q^\top_{R_DX:[j]}Y^*|y$ for $j\in 1,\dots,p-q$ are independent which proves Theorem 1.

3.5 Conclusion

In Chapter 3, we show how to compute both small samples and asymptotic properties of the F statistic under several permutation methods. The small samples properties are computed using expectation over all permutations of matrices or vectors. We show in Appendix C.1.2 and C.1.3 that similar results are obtained using shuffling matrices (which are used under heteroscedasticity) or ”bootstrap” matrices. Our new approach may lead to finite samples properties using these transformations too.

Note that we investigate finite samples and asymptotic properties in the univariate case only. Using the conditional distribution by permutation, we may find interesting properties in the multivariate cases. Indeed, we should be able to compute the conditional correlation (or covariance) between two conditional distributions of the response and extend these results to the conditional correlation (or covariance) between tests. In addition, the multiple comparisons procedures greatly benefit from the correlation between the tests, hence, understanding the conditional correlation between tests is directly related to the problem of EEG data analysis.

Finally, the asymptotic conditional distribution is proven only for a factorial design. However, the use of the QR decomposition allows to handle both regression and factorial design until the late stage of the proof. The results in (Jayakumar and Sulthan, 2014) which gives the distribution of the diagonal elements of the ”hat” matrix under multivariate normality of $[D\ X]$ or other appropriate assumptions on the design matrix $[D\ X]$ (e.g. compact support) could help to prove Theorem 1 for a general regression setting.

Chapter 4

The Correlation Structure of Cross-Random Effects Mixed-Effects Models

The following Chapter is the main part of an article submitted to the journal Psychological Methods (Frossard and Renaud, 2019).

Abstract. The design of experiments in psychology can often be summarized to participants reacting to stimuli. For such an experiment, the mixed effects model with crossed random effects is usually the appropriate tool to analyse the data because it considers the sampling of both participants and stimuli. However, these models let to users several choices when analysing data and this practice may be disruptive for researchers trained to a set of standardized analysis such as ANOVA. In the present article, we are focusing on the choice of the correlation structure of the data, because it is both subtle and influential on the results of the analysis. We provide an overview of several correlation structures used in the literature and we propose a new one that is the natural extension of the repeated measures ANOVA. A large simulation study shows that correlation structures that are either too simple or too complex fail to deliver credible results, even for designs with only three variables. We also show how the design of the experiment influences the correlation structure of the data. Moreover, we provide `R` code to estimate all the correlation structures presented in this article, as well as functions implemented in an `R` package to compute our new proposal.

4.1 Introduction

The statistical practice in psychology is dominated by the ANOVA. It has been a standardize tool to analyse various experiments or randomized control trials. ANOVA and particularly repeated measures ANOVA (rANOVA) are useful to consider the variability induced by the sampling of participants in the experiment. The complexity of the experiment tends to increase and there is a need for more complex statistical tools (Boisgontier and Cheval, 2016). The experiments often are designed by crossing a sample of participants and a sample of stimuli (e.g. images or words). To take into account the induced variability of both the sampling of participants and stimuli, methodologists suggest using crossed random effects mixed effects models (CRE-MEM) (Clark, 1973; Baayen, 2008; Lachaud and Renaud, 2011; Judd et al., 2012). These models are part of the family of mixed effects models (MEM) and sometimes called crossed random effects models. They have been introduced by Baayen (2008), Lachaud and Renaud (2011) or Judd et al. (2012) to psychologists, discussed by Barr et al. (2013) and Bates et al. (2015), and efficiently implemented by Bates et al. (2015) in the R programming language.

The CRE-MEM are used to test factors or fixed effects in experiments but need a correlation structure of the response variable as a set of parameters specified by users. The correlation structure is a set of assumptions on the distribution of the response variable, more specifically its covariance matrix. It models the full consequences that the variability induced by the sampling of participants and stimuli have on the response variables. For instance, if the responses of the same participant are correlated seems reasonable, or if the responses to the same stimuli are correlated seems also reasonable. This correlation structure is complexified by assuming that the responses of the same subject in the same condition are even more correlated, etc. In the literature, CRE-MEM are used assuming very simple correlation structures to very complex ones. In this paper, we discuss the effect of the choice of correlation structure on the tests of fixed effects. We will see that this discussion was already relevant in the rANOVA framework and that some of its conclusion should be transposed to CRE-MEM.

In Section 4.2, we explain the main differences and similarities between the CRE-MEM and rANOVA. We discuss the default choices that are made in the rANOVA framework and open choices in CRE-MEM, and we focus on the correlation structure of the data. In Section 4.3, we present the statistical models of rANOVA and CRE-MEM with published examples. In Section 4.4, we discuss how the design of the experiment influences the correlation structure and propose a classification of variables that is general for any experiment using both participants and stimuli. In Section 4.5, we present the main correlation structures used in the literature, discuss their properties and propose a new one that is the natural candidate to generalize the rANOVA. In Section 4.6, we present a simulation study that shows consequences of the choice of the correlation structure on the type I error of test of fixed effects. This simulation study allows us to advise again some correlation structures that may the inflate type I error. In Appendix, readers will find ready to use R codes for all correlation structures presented in this article as well as many extensions designed to understand CRE-MEM and their correlation structures.

4.2 ”All models are wrong, ...”, but why?

When analysing an experiment, we are mostly interested in testing a few hypotheses. However, this cannot be achieved without (explicitly or implicitly) building a statistical model, choosing a test statistic and computing its associated p value or other decision rules. All those steps require to choose settings among several options. For classical analyses, those options are often hidden to users (e.g. the model underlying rANOVA cannot be changed in many software although several models are conceivable) and even the distinction between model and test is fuzzy. For instance, the ANOVA refers both to tests statistics (test of the differences of means with a F statistics), or to a model (linear model with factors and all their interactions) (Gelman, 2005). We can hypothesise that this framework was created not only intentionally by methodologists, but also by tradition and by the use of default settings in software. By contrast, for CRE-MEM, no consensus exists for the choice of the statistic of the tests of fixed effects, neither for the choice of the model (and its correlation structure) and software usually let users tune each setting. Coming from the rANOVA framework, users must make choices that they are not used to, and, each of them will have consequences on the results of the data analysis (in particular on the p-value). The most influential choice is the statistical model which is a set of assumptions on the mathematical relationship between the response variable and the design and the measurements of an experiment. It is mathematically summarized by probability distributions with unknown parameters. Hence, to perform a good data analysis, researchers must choose a reasonably good model. But there is no unique answer to what a good data analysis is, hence, no unique choice for the statistical model.

Many interesting thoughts have already been written about the relationship between the statistical model and the real phenomena we are interesting in. And as Box and Draper (1987) warn, “Essentially, all models are wrong, but some are useful”. Indeed, even for the simplest experiment that compares two groups, nobody can guarantee that the data are Gaussian, with the same variance for the two groups, independent (and randomly sampled from the population), which are necessary assumptions of the t test or one-way ANOVA. If these assumptions are not met, one cannot know how misleading the p-value obtained by a t-test will be. There is no “right model” for this simplest experiment, and therefore no sure answer to the research question. For complex experiments with many variables, participants and stimuli, there is no “right model” either. However, some models will be more useful in the sense that they will deliver an answer to the research hypotheses that is more valuable. And we can measure the value of a model depending on the goal of the data analysis: it can be the prediction power, or the replicability of the findings or finding a model close to the real phenomena. So, the model and consequently the correlation structure in case of a CRE-MEM should be assumed in order to best fulfil the goal of the data analysis.

In this article, we concentrate on frequentist approach and evaluate the effect of the model on the p-value. However, all the argument developed here are also relevant in a Bayesian framework, as the choice of the model influences the results equally.

In the following sections, we recall the choices user faces when analysing data with a CRE-MEM, and establish parallels with more usual models like regression and ANOVA.

4.2.1 The Predictors in a Linear Model

Regression, 2 samples t test and factorial ANOVA (without repeated measures) are subsets of the same model: the linear model which is sometimes called general lineal model. The main feature of linear models is to assume a linear relationship between one or several predictor(s) (continuous variables, factors or interactions) and the mean of the response variable (also called dependent variable or outcome). Those 3 methods are mainly different in the choice of the predictors in the model. The 2 samples t test refers to a linear model with one factor with only two levels, the choice of the predictor in this model is entirely defined by the hypothesis. A factorial ANOVA refers to a linear model with several factors (with possibly more than 2 levels) and all their interactions. In that case, users have more options, as they may include more predictors than constraints by their research hypothesis. More importantly, the ANOVA tradition imposes to select all interactions although this might not be necessary. In contrast, for regression, we do not assume a model with constraints on the choice of the predictors, and usually users have to choose themselves the appropriate predictors and do not add interactions except if it represents a hypothesis.

It is noteworthy that the regression and ANOVA traditions are so different although they are based on the same (general) model. However, none of these approaches guarantees that the “right model” is used as, once again, all models are wrong.

In CRE-MEM, software let users choose which predictors and which interactions to include in the model, as in the regression tradition. However, to analyse data from experiments, the ANOVA tradition is still influential, and researchers tend to include all interactions of the selected factors.

Although very important as these choices will have an impact on the inference (p-value) of one’s hypothesis (e.g. the main effect of a precise predictor) in regression, ANOVA and in CRE-MEM, they will not be discussed further as we will concentrate on the correlation structure.

4.2.2 The Error Terms and the Correlation Structure

Concerning classical models, ANOVA (without repeated measures), regression and 2 samples t-test assume an error term which is (1) normally distributed, (2) independent, and (3) homoscedastic. When one or several of those assumptions is/are not true, users should perform other statistical analysis. The first strategy is to use other tests statistics and perform, for instance, quantile regression for skewed residuals (Koenker and Bassett, 1978), robust regression for heavy tails distribution of the errors (Heritier et al., 2009), Wilcoxon test (Wilcoxon, 1945) or Kruskal-Wallis test (Kruskal and Wallis, 1952) when the errors are not normally distributed. Moreover when the errors are heteroscedastic, we can also change the statistics and use Welch’s statistics (Welch, 1951, 1947) or Satterthwaite’s approximation (Brown and Forsythe, 1974). A second strategy is to use re-sampling methods like bootstrap (Efron and Tibshirani, 1994) or permutation tests which allow to compute distribution of statistics when the errors do not satisfy the default assumptions of normality. However, re-sampling method are still influenced by outliers and, in that case, it is still recommended to use robust estimators even within re-sampling (Salibian-Barrera and Zamar, 2002).

When some responses are linked, correlated with others, it implies that the errors may not be independent and homoscedastic, and it violates the assumption of the linear model. In that case, we must use a more complex model as we need to include the correlation structure of the data in its assumptions. This question arises as soon as several measures are made on the same sampling unit (e.g. a participant). This is the case when performing a one-way rANOVA. Its underlying model may be represented as a mixed-effects model (MEM) and it decomposes the error term into an error term per observation and a random effect (unique for each participant). We call the latter a random intercept and it considers that some participants are better/worst (have higher/lower response values) on average over all experimental conditions than others. For the MEM underlying a rANOVA, we must make assumptions on the distribution of both the error term (independent and homoscedastic), and the random effects (here also independent and homoscedastic) to fully define the model with its correlation structure.

In less complex models with only one measure per participant (like a 2 samples t tests or an ANOVA), the assumption of random intercepts per participant may be relevant in the sense that, in the true phenomena we are observing, some participants are likely to be better than others. But, with only one measure per participant, this random effect (its variability) will not be estimable because it will be included/confounded with the error term. Hence, a less complex design allows less random effects to be estimable. The corollary is that a more complex design allows more random effects to be estimable. And, with a complex experimental design, we may be able to estimate not only random intercepts per participant but also random slopes. By analogy to the regression, the literature defines a random slope as a random effect that describes the variation of a subject given the value of a covariate or a factor and represents the change of slope of this subject with respect to one of the population; and we can generalize this concept to random interactions which represents the change of interaction effect with respect to the population for each subject.

In a complex MEM, the choice of the correlation structure not only includes which random effects we assume in the model (random intercepts, slopes and interactions) but also their full (joint) distribution. If each participant has more than one random effect, all those random effects create together a multivariate random effect and we have to make assumptions on its full multivariate distribution. For instance, we can assume correlation between random effects or constraints like the equality of variance of several random effects. This correlation is for example relevant in learning processes, where it seems natural to assume that someone with a lower score at the beginning will learn more than someone with a higher starting score. This feature can be implemented in the model by assuming that the random intercept (for a participant, his score’s difference with respect to the average) is correlated with the random slope (for a participant, the learning’s difference with respect to the average learning); in this example, we therefore expect a negative correlation between the random intercept and the random slope. As a second example, the equality of some variances is used to implement spherical random effects and it is mainly used for factors. We could specify a different variability for each level of a factor, but this assumption is often neither useful nor rooted with strong information on the correlation structure. As a result, it is reasonable to assume spherical random effects which means that, in each level, random effects will have the same variability. Here, the sphericity assumption reduces the number of free parameters and increases the parsimony of the model.

rANOVA is said to control the type I error rate at its nominal level. However, this good property is mathematically derived assuming a specific model with its specific correlation structure. Specifically, the correlation structure underlying the rANOVA is assumed to be saturated which means that the random intercepts, random slopes and random interactions are all included (up to the last interaction which is confounded with the error term). Moreover, all effects are independent and spherical within each factor or interaction. The model underlying the rANOVA is a special case of MEM and researchers could use the MEM framework to analyse this type of data. It is however appropriate only if its assumptions are tenable.

When the response variable is the result of the crossing between the samples of 2 units (typically participants and stimuli/pictures), we model the response using a CRE-MEM. The correlation structure becomes much more complex because random effects can be associated with the participants and the stimuli but also with their interactions. To define the correlation of a CRE-MEM, we need to select which random intercepts, slopes or interactions and which multivariate distributions we assume for the participants and the stimuli. Moreover, if it is relevant to assume that some participants may be better/worst with some stimuli, we specify this feature by including in the correlation structure random effects associated to the interaction between the participants and stimuli. As for any MEM, the estimable random effects are also constrained by the design, and by which fixed variable is included in the model. And the set of random effects estimable which depends on the design becomes more complex with two sampling units rather than only one. To have the full picture of the estimable effects, we develop a classification of variables in Section 4.4.

4.2.3 The Test Statistics

The model is specified in Section 4.2.1 and Section 4.2.2 and for the sake of completeness, the testing procedure, which includes the test statistic, its distribution and the associated p value, have to be specified. For the regression or ANOVA model, the common solution is to use t or F statistics. If the model is well specified and the assumptions of independence, homoscedasticity and normality of the errors are met, they provide tests that are exact and powerful (low type II error rate). As we saw in Section 4.2.2, for those simple models, there exists alternative solutions when the model is not well specified.

However, for CRE-MEM, no exact solution exists even if the assumptions of the model are met. Methodologists still debate the advantages of 4 test statistics: the quasi-F statistic (Clark, 1973; Raaijmakers et al., 1999), the likelihood ratio statistics (and its chi-square asymptotic distribution), the Satterthwaite's approximation (Schaalje et al., 2002) and the Kenward-Roger approximation (Kenward and Roger, 1997). The quasi-F statistic is limited to balanced design using only factors and simulation studies tend to show that the likelihood ratio test does not control type I error rate very well. By contrast, the Satterthwaite and the Kenward-Roger approximations seem to be closer to the nominal level when the model is well specified and its assumptions are met (Luke, 2017). Simulations show that the latter seems to perform slightly better but at a higher computation cost.

Only the quasi-F statistics imposes condition on the correlation structure. As for rANOVA, the statistic is based on sum of squares and the distribution on the quasi-F statistic approximates a F distribution only by assuming spherical and uncorrelated random effects.

4.2.4 The Design-Driven vs Data-Driven Approach

In Section 4.2.1, 4.2.2 and 4.2.3, we described the main choices a researcher have to do when analysing data. Once again, no choice can guarantee that the right model is used, but there are two "families" of strategies that are often implemented to attempt to obtain this useful model. The two approaches are well described by Barr et al. (2013). Moreover, Shmueli (2010) found a similar dichotomy when discussing the goals of the data analysis which are either explanatory modelling with methods related to the design-driven approach or predictive modelling related to the data-driven approach. Here we recall and enhance their descriptions.

The first approach is called design-driven and is prominent in the ANOVA setting. It is usually performed to analyse data issued from experiments in which all variables are carefully chosen in advanced, and where we expect to find a causal relationship between variables. The creation of the model is instrumental to reporting a test of hypothesis and all the choices of the analysis could be made before even collecting the data. The main worry of the analyst is to control the type I error rate and the replicability of the findings.

The second approach, called data-driven, plays a prominent role in regression, which has many extensions in model selection like the Lasso (Tibshirani, 1996). The focus of the analysis is not to test hypotheses, but to find a good model in the specific sense that it provides good prediction capacity and is parsimonious. The chosen model should be the closest to the true underlying model or, at least, should produce similar predictions. In this context, external assumptions are kept minimal and ideally all the choices are merely based on the data. The main worry of the analyst is over-fitting.

This typology is clearly a caricature and real data analyses always lie somewhere between these two extremes. For experimental designs, the first approach seems more relevant. This is the case for the ANOVA/rANOVA framework which has standardized the data analysis. Regardless of the scientific domain, researchers perform tests of main effects, interactions, contrasts, simple effects or post-hoc analyses using the same construction of fixed effects and, for rANOVA, the same correlation structure. By contrast, no standard solution is as well established to analyse CRE-MEM: if researchers tend to include all fixed effects in the linear predictor, the random structure that should be assumed is still debated. Data-driven (Bates et al., 2015) or design-driven (Barr et al., 2013) solutions have been proposed, and our solution (referred latter to gANOVA in Section 4.5.5) is a parsimonious design-driven approach which generalizes the correlation structure of rANOVA to several random units.

4.3 rANOVA and CRE-MEM

In order to understand the (sometimes hidden) correlation structure of the models behind usual analyses, we first investigate the model of rANOVA, highlight some of its properties and show how this model can be extended to CRE-MEM. To illustrate with a research example, we present the publication of Erickson et al. (2011) who use a rANOVA to analyse their experiment. They are interested in the effect of exercise on the volume of the brain of elderly participants. They split the sample into an aerobic training group (AT) and a stretching control group (SC) and measured their brain volume using magnetic resonance images at the baseline (BL), after 6 months (6M) and after one year (1Y) of training. This experiment is analysed using a generic rANOVA with one between-participant variable (the group with two levels: AT and SC), and one within-participant variable (the time with 3 levels: BL, 6M and 1Y). We note that the between-participant variable is a feature of the participants because the participants are not allowed a change of levels during the experiment. And the within-participant variable indicates a feature of the experimental manipulation: the time of data collection, which is defined by the experimenter. As for many models, each response, is decomposed into fixed effects and random effects. Following the notation of e.g. Howell (2012), we write the underlying model using the equation:

$$\begin{aligned} y_{ijk} = \mu + \alpha_j + \psi_k + (\alpha\psi)_{jk} \\ + \pi_i + (\pi\psi)_{ik} + \epsilon_{ijk}, \end{aligned} \tag{4.1}$$

where y_{ijk} is the response variable, here the brain volume of the i^{th} participant, assigned to group j on the k^{th} occasion of measure. The fixed part of the equation is decomposed into the between-participants effects and within-participants effects. The between-participant effects are α_j for $j \in \{1, \dots, n_j\}$, which corresponds to the main effect of the groups. The within effects are ψ_k for $k \in \{1, \dots, n_k\}$, which correspond to the effect of the time and all interactions that contain it. Here $(\alpha\psi)_{jk}$ is the effect of the interaction group:time in our example. The random part is composed of random intercepts π_i for $i \in \{1, \dots, n_i\}$ which correspond to the participant's average brain volume relative to the overall mean brain volume, and random slopes, $(\pi\psi)_{ik}$, which is interpreted as the difference of the time's effect of each participant to the population time's effect on brain volume. Finally, the error term $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ capture everything that cannot be captured by the previous terms. The correlation structure used in the rANOVA model is implied by the assumed distribution of the π_i's and $(\pi\psi)_{ik}$'s. To have an exact test for the fixed effect in rANOVA, we must assume that the random effects ψ_k, $(\pi\psi)_{ik}$, follows each a normal homoscedastic distribution. Therefore, the assumed correlation structure needs 2 variance parameters in addition to the error term[1]. Except for this last random effect, each term in Equation 4.1 is associated to a sum of squares as produced by many statistical software. This model is today's standard in statistical software and in published researches, but it is the result of a debate dating from the 70's. We provide in Appendix D.1, a small summary of the discussion and arguments that lead to what we now today as a rANOVA.

The model represented by Equation 4.1, should only be used when we assume only one sampling unit, the participants. In many studies, in addition to the sampling of participants, researchers sample stimuli, and the value of the responses will depend on the crossing of a sample of participants and a sample of stimuli. In that setting, we also must consider the variability induced by this second sampling, which means using a CRE-MEM. To illustrate with a research example, Wilson et al. (2017) performed several experiments that highlight racial biases in judgement of physical size and weight and analyse them with a CRE-MEM. They showed to a sample of male participants a sample of stimuli (images of Black men) and asked the participants to evaluate the size of the people shown in the stimuli. As each participant viewed each stimulus, we say that the estimation of size was made by crossing participants and stimuli. Wilson et al. (2017) recorded also features of the participants, their race: Black (B) or White (W), and features of the stimuli, the actual size of the people in the images. Note that each stimulus is evaluated in each level of the factor race, which means they are evaluated by Black and by White participants. And each participant sees stimuli of different actual sizes. It may have been interesting to also change the presentation of the stimuli, where stimuli are shown in a neutral environment (N) or with a context (C) where the size is easier to evaluate [2]. We will assume that each participant sees each stimulus two times: in the neutral environment and with a context. This third factor is then a feature of the experimental manipulation each pair participant-stimulus are recorded in both contexts. To analyse this experiment, we use a CRE-MEM which equation is written:

$$\begin{aligned} y_{imk} =& \mu + \alpha_j + \psi_k + \phi_l + (\alpha\psi)_{jk} + (\alpha\psi)_{jk} + (\psi\phi)_{kl} + (\alpha\psi\phi)_{jkl} \\ &+ \pi_i + (\pi\psi)_{ik} + (\pi\phi)_{il} + (\pi\psi\phi)_{ikl} \\ &+ \omega_m + (\omega\psi)_{mk} + (\omega\alpha)_{mj} + (\omega\psi\alpha)_{mkj} \\ &+ (\pi\omega)_{im} + (\pi\omega\psi)_{imk} + \epsilon_{imk}, \end{aligned} \tag{4.2}$$

with the response y_{imk} (here the estimated size) and with a fixed part written in the first line of Equation 4.2 composed of α_j (features of participants: here the race), ψ_k (features of stimuli: here the actual size), ϕ_l (features of experimental manipulation: here the presentation of stimuli), and their interactions, $(\alpha\psi)_{jk}$, $(\alpha\psi)_{jk}$, $(\psi\phi)_{kl}$ and $(\alpha\psi\phi)_{jkl}$. The decomposition of the random part has then effects associated to the participants and their interactions with some factors (second line of Equation 4.2), π_i, $(\pi\psi)_{ik}$, $(\pi\phi)_{il}$ and $(\pi\psi\phi)_{ikl}$; effects associated to the stimuli and their interactions with some factors

[1] Actually, rANOVA models are always overparametrized as the highest interaction, here $(\pi\psi)_{ik}$, and the errors ϵ_{ik} are confounded. This has no harmful consequence as we do not estimate each parameter

[2] We add this third factor (presentation of stimuli) which was not in the original study not as a criticism but to highlight the effect of the design on the correlation structure of a CRE-MEM.

Table 4.1: Link between random units and the type of variables. A cross means that a random interaction (i.e. an interaction between the random effect define by the row and the fixed effect defined by the column) is estimable or allowed in a CRE-MEM with a fully balanced design.

	intercept	V_P	V_S	V_{PS}	V_M	V_O
Participants	X		X	X	X	X
Stimuli	X	X		X	X	X
Participants:Stimuli	X				X	X

(third line), ω_m, $(\omega\psi)_{mk}$, $(\omega\alpha)_{mj}$ and $(\omega\psi\alpha)_{mkj}$; and effects associated to the participants-stimuli interactions and to the error term (fourth line), $(\pi\omega)_{im}$ and $(\pi\omega\psi)_{imj}$. Note that the interaction between sampling units and a fixed effect is only feasible if the sampling units are evaluated in several levels of the fixed effect. The correlation structure of this CRE-MEM is implied by the multivariate distribution of the random effects (defined in the three last lines of Equation 4.2). The correlation structure is de facto more complex than the one for rANOVA and the different choices and assumptions are discussed in Section 4.5.

Note that the effects on the fourth line are seldom if ever presented or used in the context of CRE-MEM and correspond to the random effect associated to the interaction participant-stimuli and model the assumption that some participant will have a better response with some particular stimuli (or vice versa). With CRE-MEM, the dichotomy of “between-participant” variables and “within-participant” variables used in rANOVA is insufficient and, in Section 4.4, we present a classification of 5 types of variables for the CRE-MEM.

4.4 Classification of Variables for the CRE-MEM

In the ANOVA framework, explanatory variables are split into “within-participant” variables and “between-participant” variables. Often between-participant variables (that we will call V_P) represent a feature of the participants, like their sex, and the participants can be or is measured in only one level of the between-participant variables. The within-participant variables often represent the feature of the experimental manipulations (V_M) which means that the participants can be measured in multiple levels of a within-participant variables. This classification is feasible since there is only one sampling unit, the participants. Based on this dichotomy, we know that only within-participant variables may interact with the sampling unit to create random effects.

As viewed in the previous Section, many experiments in psychology are more complex as they cross one random sample of participants and one random sample of stimuli (and therefore must be analysed with CRE-MEM). In that setting, 3 random units are actually present: the participants, the stimuli, and their interactions. In order to know which models are at least feasible (or more precisely, which random effects can be included in the model), we have to know which explanatory variables may interact with which random units. The aim of this section is to provide a classification of variables that answers this question and to illustrate it with a few examples.

In summary, for CRE-MEM, as in rANOVA, some variables are either V_P or V_M as they specify a feature of the participants or of the manipulation. By symmetry with V_P, there might be variables that specify a feature of the stimuli, called V_S, and variables that specify a feature of the interaction of participants and stimuli, that will be called V_{PS}. Finally, some variables designate a feature of the specific occurrence, or observation, and will be called V_O. In more detail, here is the list of potential types of variables:

1. V_P: variables that specify a feature of the participants. The typical example is the sex of the participants. It can also happen that for experimental reasons, participants are (randomly) assigned to a single experimental condition, like a learning method in education or given a specific instruction in social psychology. Experimentally, this condition becomes a feature of the participant and thus the corresponding variable is also classified as V_P. In Equation 4.2, the effects α_j come from a V_P variable and is the race in the experiment of Wilson et al. (2017). This type of variables reduces to a between-participant variable in the rANOVA setting.

2. V_S: variables that specify a feature of stimuli, in the same way as V_P specify features of participants. The typical example is the valence of an image or the characteristic (frequency, type, ...) of a word. In Equation 4.2, the effects ϕ_l come from a V_S variable and correspond to the actual size of stimuli.

3. V_M: variables that specify a feature of the experimental manipulation. The experimenter has usually the ability to manipulate it independently of the participants and of the stimuli, showing several different conditions to the same pair participants-stimuli. Examples are the hemifield of presentation of a target, the lightning or surrounding sound conditions. In Equation 4.2, the effects ψ_k come from a V_M variable and correspond to the presentation of the stimuli. This type of variable reduces to the within-participant variables in the rANOVA.

4. V_{PS}: variables that specify a feature of the interaction between a participant and a stimulus, and therefore that cannot vary for a given pair participant-stimulus. Often, this type of variables is the results of constraints on the experimental design: if the same participant cannot see the same stimulus in several conditions, this factor is then specific for each pair of participant and stimulus and become a variable V_{PS}. For example, if only the high-frequency or only the low-frequency of an image is shown, for a given image, half of the participants will see its high-frequency version and the other half its low-frequency version, and conversely for another image. As an additional example in linguistics, in a novel word experiment design half the participants learn half the words with spelling and the other half with only the spoken input, and the association between one word and a level is balanced across participants. In the experiment of Wilson et al. (2017), if the participants were asked, afterward, to evaluate a characteristic of the stimuli, like the level of masculinity, and use this measure as a predictor, this variable would be of type V_{PS}.

5. V_O: variables that specify a feature of the specific occurrence or observation. It may be a physiological or physical measure taken at the precise time of the measures of the response (for a given participant subject to a given stimulus). The position of the trial in the experiment, or the RT to the previous stimulus fall also in this category.

As a rule, for any factors or variables, if the random unit is measured in several of its levels, then a random interaction between this variable and the random unit is estimable,

Table 4.2: List of 5 typical experimental designs involving participants and stimuli that will be exemplified in this article (column Use: ”Formu” meaning that its `R` formulas are given in the Appendix D.5 for all correlation structures, and ”Simul” means that it is used in the simulation study of Section 4.6). The five different types of variables (V_P, V_S, V_M, V_{PS}, and V_O) are defined in Section 4.4 and the number of levels are given within parentheses.

Model	Variables	Use
M1	Vp(2), Vs(2), Vm(2)	Formu/Simul
M2	Vp(3), Vs(3), Vm(3)	Simul
M3	Vp(3), Vs(3), Vm(3), Vm(2)	Formu
M4	Vp(3), Vs(3), Vm(3), Vps(2)	Simul
M5	Vp(3), Vs(3), Vm(3), Vm(2), Vps(2), Vo(2)	Formu

i.e. it can be included in the CRE-MEM. This random interaction is interpreted as a random slope. As a result, the random slope of a V_p variable is only estimable for the stimuli, the random slope of a V_M variable is estimable for the participants, the stimuli and their interaction, etc. Table 4.1 gives a summary of the estimable random slopes for the 5 types of variables. In this section, we discussed the cases where the variables are factors. When dealing with one or more continuous variables, the classification of variables and its consequences on the correlation structure of the data are the same. However, especially in the presence of interaction, a special care is needed in the interpretation of the results (e.g. depending if the variables are centred or not). Moreover, we believe that this classification and the approach described above can be extended for cases with more than 2 random units.

Finally, note that in rANOVA, the “wide” format of the data makes a clear distinction between the representation of the within-participant and between-participant variables. In Appendix D.2, we extend this representation for the 5 types of variables of the CRE-MEM.

4.5 Several Families of Correlation Structures

In this section, we describe the correlation structures that are discussed in the literature and propose a new one in Section 4.5.5. Several goals are pursued. First, to list the major models and to give them a name, second to link them with `R` code of the `lme4` package (Bates et al., 2015), third to explain their assumptions and fourth to compare them theoretically. In Section 4.6, we will compare them based on simulations so that some guidelines can be learned.

Note first that all correlation structures discussed in the literature assume independence between the random effects associated with (a) participants, (b) stimuli and (c) their interactions. For the model of Equation (4.2), it implies that for all following proposals, the random effects on the second line are independent with the random effects on the third line and the fourth line. This is a minor assumption if the interaction between participants and stimuli is included but might be questionable if not.

Second, each of the proposal may include effects coming from the interaction participants:stimuli even if the authors who originally described these correlation structures

Table 4.3: Number of parameters for the correlation structure for the five models described in Table 4.2 and for all the random structures defined in Section 4.5. A plus sign describe a random structure that includes the interaction participants:stimuli.

	M1	M2	M3	M4	M5
without interaction participants:stimuli					
RI	2	2	2	2	2
RI-L	8	8	16	16	64
MAX	20	90	342	342	5256
ZCP	8	18	36	36	144
gANOVA	8	8	16	16	64
with interaction participants:stimuli					
RI+	3	3	3	3	3
RI-L+	9	9	19	17	71
MAX+	21	91	352	343	5311
ZCP+	9	19	40	37	154
gANOVA+	9	9	19	17	71

did not include them. Those interactions terms model if some participants are especially good/bad with a particular stimulus. Below, a "+" sign in the name of a correlation structure indicates its inclusion.

In CRE-MEM, the optimization process is defined for parameters which are function of the elements of the correlation structure (Bates et al., 2015). So, having more free parameters in the correlation structure imply a more difficult optimization process and more convergence errors of the algorithm. And, as for any statistical model, including additional parameters makes the model "less wrong" (in the sense of the goodness of fit) at the price of reducing its parsimony. In practice, for CRE-MEM, the usual trade-off between parsimony and goodness of fit is disturbed by the convergence error of the algorithm. For the five models presented in Table 4.2, we show in Table 4.3 the number of parameters for each correlation structure presented next. It clarifies the huge difference between the proposed correlation structures.

Moreover, the replicability of the findings has become a major worry in many fields. The choices carried out when using CRE-MEM should reflect this tendency. For that purpose, the correlation structure should have good properties: specially to have the expected results of the analysis reproducible through experiments, be robust to some misspecification of the model and have a high rate of convergence. If the model is used for testing in a frequentist approach, a good choice will exhibit a type I error rate close to the nominal level under the null hypothesis and, at the same time, a high power under the alternative.

4.5.1 The Correlation Structure with Random Intercepts (RI)

In this simplest case, the correlation structure has only a random intercept for the participants and a random intercept for stimuli. These intercepts are not correlated, which means that only 1 variance parameter per random unit is estimated regardless of the number of fixed effects, hence the value of 2 in the first line and 3 in the sixth line of Table 4.3. All the interaction terms in the second, third and fourth lines of Equation (4.2) are removed, or equivalently, their variances are set to zero.

For two factors `f1` and `f2` and the participant and stimulus identifier `PT` and `ST`, the typical formula of RI using the `lme4` package is:

```
R> lmer(y ~ f1*f2 + (1|PT) + (1|SM), data = mydata)
```

and for th RI+ structure:

```
R> lmer(y ~ f1*f2 + (1|PT) + (1|SM) + (1|PT:SM), data = mydata)
```

Full examples are provided in Appendix D.5. For space reason and in order to focus on the part of interest, in the main text we will summarize the formulas. For the RI and the RI+, it becomes:

```
R> lmer(y ~ [...] + (1|PT)  [...] )
```

In lay language, this correlation structure just suppose that some participants are better than others on each measurement, and that some stimuli are more difficult than others for all participants alike. Although this correlation structure is used in the literature, in most cases, the true correlation structure will most probably be more complex that only random intercepts and more random effects are estimable. Choosing this correlation structure will reduce the reproducibility of the results with a gain in parsimony we do not really need. RI is probably too simple for most applications and does not provide credible inference (Barr et al., 2013).

4.5.2 The Correlation Structure with Random Intercepts at each Level (RI-L)

One way to view the rANOVA model (as in Equation (4.1)) is to think that it incorporates all interactions between the random effect of the participant (π_i) and the within-participant fixed effects (only μ and α_j here), to produce all possible random effects (here π_i and $(\pi\psi)_{ik}$). Bates et al. (2015) suggest following the same idea for CRE-MEM, starting with the random effect of both participants (π_i) and stimuli (ω_m). This random structure corresponds to random intercepts and slopes that are IID and spherical ($\pi_i \sim \mathcal{N}(0, \sigma^2_\pi)$, $(\pi\psi)_{ik} \sim \mathcal{N}(0, \sigma^2_{\pi\psi})$, $(\pi\phi)_{il} \sim \mathcal{N}(0, \sigma^2_{\pi\phi})$ and so on), and independence between them, for all the elements in the 2nd, 3rd and 4th lines of Equation (4.2). The typical formula of RI-L using the `lme4` package is:

```
R> lmer(y ~  [...] + (1 | PT) + (1 | PT:f1) + (1 | PT:f2)
        + (1 | PT:f1:f2)  [...] )
```

The RI-L correlation structure keeps a relatively low number of parameters which does not increase with respect to the number of levels of the factors (see Table 4.3).

This may seem the natural extension of rANOVA, however with the same number of parameters, gANOVA includes all the correlation structures that can be obtained with RI-L and strictly more. Therefore, the likelihood (and criteria like AIC and BIC) will always be better or equal when using gANOVA instead of RI-L. More details on the difference between the two correlation structures are given in Section 4.5.5. The numerical optimisation is also easier for gANOVA compared to RI-L, as exemplified in Appendix D.4.

4.5.3 The ”Maximal” Correlation Structure (MAX)

The “maximal” correlation structure is suggested by Barr et al. (2013) and it is defined by including all possible random effects associated with the participants on one side and all possible random effects associated with the stimuli on the other side. Moreover, Barr et al. (2013) let also a maximal correlation structure between random effects, which means that all random effects can correlate with each other (within the same random unit), or said differently, the covariance matrix of the random effects is full and unstructured. The typical formula of MAX using the lme4 package is:

```
R> lmer(y ~  [...]  + (f1*f2 | PT)  [...] )
```

This correlation structure may seem the appropriate choice without prior information on the correlation structure, but the problem is that it is not parsimonious enough except for the smallest models, see Table 4.3. Note that the authors did not specified explicitly how to handle factors with more than two levels. Moreover, the actual optimization algorithms often do not converge even for small models (see Table 4.4). It might therefore be used when the design has only one IV but is probably not suited for experiments with two variables or more.

4.5.4 The Zero-Correlation Parameter Correlation Structure (ZCP)

The ZCP (Bates et al., 2015) also includes all the random effects associated with stimuli and with participants. Unlike the MAX model, the ZCP model does not include correlations between the random effects. This means that one variance parameter is estimated for each effect, but no correlation is assumed. When one or more factors have 3 or more levels, there is a twist and the number of variance parameters that are estimated for each random part will be equal to the number of degree-of-freedom of the corresponding fixed factor (or interaction of factors). Said differently, variance parameters are attached to contrasts of the factor or interaction of factors and not to the factors themselves.

For two factors f1 and f2 and the participant identifier PT, one first transforms the factors f1 (e.g. with 4 levels) and f2 (e.g. with 3 levels) into coding variables x1a, x1b and x1c, respectively x2a and x2b (more information about the necessity of transformation into coding variable is in the Appendix D.5). Then the typical formula using the lme4 package is, for ZCP:

```
R> lmer(y ~  [...] + ((x1a + x1b + x1c)*(x2a + x2b) || PT)  [...] )
```

This correlation structure is relatively parsimonious when all factors have exactly two levels, but the number of parameters will increase with respect to the number of levels of the factors (see Table 4.3). It has the drawback to be dependent on the choice of the coding of the factors. This correlation structure can be viewed as a workaround to force lme4 not to add correlations between random effects, but that this workaround does not give the expected results with factors that have 3 or more levels. Although, we do not expect a huge difference in practice, the maximum likelihood and the inference will depend on the choice of the coding variables (or contrasts), even when they are forced to be orthonormal. In many applications, the choice of the coding variable does not correspond to any hypothesis and is therefore arbitrary. Moreover, this may be an obstacle for the reproducibility of the data analysis because it will be challenging to report the all coding variables of the factors and their interactions.

4.5.5 The Random Structure of the Generalized ANOVA (gANOVA)

In order to generalize rANOVA to CRE-MEM, gANOVA first assumes a saturated model with all random effects, the one associated to participants, with stimuli and with their interaction. As in the experimental design literature, the covariance structure of random effects is assumed to be minimal, i.e. each random effect is independent from the others, and spherical. A correlation structure with spherical random effects will have random effects that share the same variance for each level of the same factor; which means that the number of parameters will not increase with respect to the number of levels (see the lines gANOVA in Table 4.3 for models M1 vs M2). The model behind gANOVA is the same as RI-L suggested in Bates et al. (2015), as exemplified in Equation 4.2, and the number of parameters is also identical (compare the lines RI-L and gANOVA in Table 4.3). The difference with RI-L is that some constraints are assumed on the random effects. In Equation 4.2, those constraints are written: $\sum_k (\pi\psi)_{ik} = 0\ \forall i$, $\sum_l (\pi\phi)_{il} = 0\ \forall\ i$, $\sum_k (\pi\psi\phi)_{ikl} = 0\ \forall\ i, l$, and $\sum_l (\pi\psi\phi)_{ikl} = 0\ \forall\ i, k$. For the algorithm, those constraints are simply implemented by transforming the factors into orthonormal coding variables and forcing them to share the same variance parameter (within each factor).

It is not possible to use lme4 to estimate the gANOVA model, as it needs to satisfy both the constraints of equality of variances for each level and to use coding variables for the random interactions. However, a simple modification of the lmer function implemented in the gANOVA package (https://github.com/jaromilfrossard/gANOVA) allows to perform this optimization. For two factors f1 and f2 and the participant identifier PT, the gANOVA is performed using the gANOVA package and the formula:

```
R> gANOVA(y ~  [...]  + (1 | PT | f1*f2) [...] )
```

The justification for this correlation structure is that it is much more in line with the tradition of experimental design. Indeed, it is exactly as defined by Cornfield and Tukey (1956) for ANOVA including one or several random effects (see Appendix D.1). Moreover, one of the first tools to obtain p-values for balanced experiments where there is crossing of random samples of participants and stimuli was the quasi-F statistic (Winer, 1962). This statistic is based on sums of squares which are easy to compute after averaging over the participants and over the stimuli. The quasi-F statistic follows an approximative F distribution under some assumptions (Clark, 1973). These assumptions are identical to the one made in rANOVA (independence and homoscedasticity of the random effects). For balanced data, the model and the implied correlation structure are the same for quasi-F and CRE-MEM based on gANOVA (but the statistic, t-value and p-value are computed differently) and we expect to obtain quite similar results (very close p-values). However, gANOVA generalizes naturally to non-balanced designs since it is a mixed effect model.

Secondly, as mentioned in Section 4.5.2, the possible correlations between all responses assumed by RI-L are only a (strict) subset of the ones with gANOVA. For some data, both methods lead to the same variance-covariance matrix of the response, but its decomposition into variances of random effects are different. Which is similar to say that, in Equation 4.2, all the variances and covariances of the responses y_{imk} are the same for gANOVA and RI-L, but its decomposition into variances of random effects π_i, ..., $(\pi\omega\psi)_{imk}$ is different, due to the sum-to-zero constraints in gANOVA. In lay terms, without these constraints, higher order interaction random effects put restrictions on the variance of the random effects of lower interaction. Pehaps surprisingly, this imply that the possible variances and covariances of the responses are *reduced* in RI-L (compared to gANOVA), and for some data the variance-covariance matrix of the response will be different between the two methods. It often implies a solution at the boundary of the domain of definition (one or several variances set to zero) in RI-L during the optimization. In those cases, RI-L and gANOVA do not share the same solution and gANOVA has always a smaller deviance (and AIC, BIC, ...) which suggests a better fit.

The equations below show the relationship between the variance parameters of both parametrization for a model with one variable:

$$\begin{aligned}\sigma^2_{RIL;i} &= \sigma^2_{gANOVA;i} - a\sigma^2_{gANOVA;F}\\ \sigma^2_{RIL;F} &= \sigma^2_{gANOVA;F}\\ \sigma^2_{RIL;\epsilon} &= \sigma^2_{gANOVA;\epsilon},\end{aligned}$$

where $\sigma_{RIL;i}$ and $\sigma_{gANOVA;i}$ are the standard deviation of the random intercepts for both parametrizations, $\sigma_{RIL;F}$ and $\sigma_{gANOVA;F}$ are the standard of the random slopes (participant:F), and $\sigma_{RIL;\epsilon}$ is the standard deviation of the error term. a is positive constant that depends on the number of levels of the factor F: $a = 1 - 1/(\#\text{ levels})$. See Appendix D.4 for the full derivation of this example.

Several comments can be made. First, one has to be aware that the interpretation is different between the two covariance structures. Second, if $\sigma^2_{gANOVA;i} - a\sigma^2_{gANOVA;F}$ is positive, RI-L and gANOVA will produce the same variance-covariance of the response and the same deviance. The fit is exactly the same. However, there will cases where this term is negative. In that case, RI-L cannot attain the optimum and is forced to set a variance to zero (and to adjust the two other ones), leading to a poorer fit compared to the solution of gANOVA. This leads to the conclusion that gANOVA is strictly better than RI-L.

Concerning now the comparison between gANOVA and ZCP, they are the same model for designs that have factors with exactly two levels. But it is not the case when at least one factor has 3 levels or more. For this type of factors and when they interact with random effects, gANOVA has the assumption of sphericity which imposes the same variance parameters for each coding variables of the factors. On the other hand, ZCP will have a new variance parameter for each new coding variable and adding these new parameters has two drawbacks. First, the number of parameters to estimate increases which implies more variable estimations (see Table 4.3). Moreover, these new parameters are usually not dictated by theoretical ground but more by the convenience of an existing `R` formula. Secondly, the random structure, the maximum likelihood and the inference depend on this arbitrary choice of the coding variable, even when they are forced to be orthonormal. There are infinitely many groups of coding variables that may be used for a single dataset and for each of which ZCP will give a different p-value. This arbitrariness in a model that is precisely design-driven is not desirable. On the other hand, the sphericity assumption in gANOVA keeps one variance parameter for all coding variables of a given factor (or interaction). This will reduce the number of parameters and the arbitrariness of the coding of factors by being independent to the choice of the coding of the factors.

Finally, the constraints used in gANOVA (almost) orthogonalize the random effects (they would be orthogonal if fixed) such that the parameters have a small mutual influence in comparison with RI-L. Figure D.3 shows an example of the likelihood within the space of the parameters. For the same data (one sampling unit, one V_M variable and replications), and fitting random intercepts and random slopes, we see that the two ridges defining the two profile likelihoods cross almost at 90° at the optimum in the gANOVA case but is far more inclined for RI-L. This suggests less dependency between the parameters and a better optimization process for gANOVA. In higher dimension, it is known that all optimization suffers from the curse of dimensionality, and the better independence of gANOVA parameters is clearly an asset.

4.5.6 The Correlation Structure Based on PCA (CS-PCA)

Bates et al. (2015) proposed a heuristic to find the appropriate correlation structure. This correlation structure has a data-driven approach and therefore changes even between two experiments sharing the same design. To compare it to the previous methods, we summarize the proposition of Bates et al. (2015) by a fully defined algorithm in Algorithm 4. The idea behind this method is to use PCA on the estimated maximal correlation structure (from the MAX model) to estimate the dimensionality of the random effects; by assuming that the true correlation structure is of lower dimension than the one defined by the MAX model, we restrict to a subspace in which it is hoped that most of the variability of the random effect lives. Then by deleting random effects of the model (suppressing the higher-level interaction first), we match the correlation structure to the estimated dimensionality. Then based on the new maximal dimensionality, Bates et al. (2015) proposed to reduce the number of parameters based on test or goodness of fit; first we decide whether to drop the covariances between random effects and select the random structure based on test, then we decide whether to drop random effects one by one beginning with the higher interaction levels. We stop this procedure when it does not improve the model anymore. The selected random structure is then compared to a last one by adding or subtracting the covariance between random effects.

Algorithm 4 Correlation structure based on PCA

```
 1: Choose a model selection procedure $\mathscr{P}$.
 2: Estimate the model based on MAX $\mathscr{S}_{max}$.
 3: for participants and stimuli do
 4:     Perform PCA to find the dimensionality $r_{\mathscr{S}_{max}}$ of the random effects.
 5:     Drop random effects with higher interaction levels to match $r_{\mathscr{S}_{max}}$.
 6: Define the new random structure $\mathscr{S}^{+}_{PCA}$.
 7: Drop covariance between random effects and define this random structure $\mathscr{S}^{-}_{PCA}$.
 8: Choose between $\mathscr{S}^{+}_{PCA}$ and $\mathscr{S}^{-}_{PCA}$ using $\mathscr{P}$. The chosen random structure is called
    $\mathscr{S}_{reduced}$.
 9: while $\mathscr{P}$ suggests the smaller correlation structure do
10:     $\mathscr{S}^{0}_{reduced}$ is defined by dropping from $\mathscr{S}_{reduced}$ the random effect of the higher
    interaction levels.
11:     Choose between $\mathscr{S}_{reduced}$ and $\mathscr{S}^{0}_{reduced}$ using $\mathscr{P}$.
12:     Update $\mathscr{S}_{reduced}$ by the previous choice.
13: Given the choice made in 8, add or drop covariance to $\mathscr{S}_{reduced}$ to create $\mathscr{S}^{1}_{reduced}$.
14: Choose between $\mathscr{S}_{reduced}$ and $\mathscr{S}^{1}_{reduced}$ using $\mathscr{P}$.
```

This algorithm will choose a correlation structure that is a subset of the correlation structure defined by MAX. However, it is possible to imagine new algorithms that choose a correlation structure that is a subset of ZCP, RI-L or gANOVA correlation structure. Moreover, being based on a MAX correlation structure, CS-PCA will have problems in complex designs.

Table 4.4: Percentage of convergence error for all simulations ($N_{sim} = 4000$) under the null hypothesis. Results are split by rows according to the simulation settings based on (1) the sample size for stimuli, (2) the true correlation between random effects, (3) the presence/absence of random effects associated with the participants:stimuli interaction and (4) the size of the design. The columns represent the type of estimation: all 7 correlation structures are assumed with (+) and without (-) the interaction participants:stimuli. The dash ”-” indicates settings without simulations. MAX and to a lesser extent CS-PCA present problems of convergence.

				RI		RI-L		MAX		ZCP-sum		ZCP-poly		gANOVA		CS-PCA	
				-	+	-	+	-	+	-	+	-	+	-	+	-	+
18	spheric.	no PT:SM	M1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	12.7	15.5	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.7
			M4	0.0	0.0	0.0	0.0	-	-	0.0	0.0	0.0	0.0	0.0	0.0	-	-
		PT:SM	M1	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	13.4	9.4	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.4
			M4	0.0	0.0	0.0	0.0	-	-	0.0	0.0	0.0	0.0	0.0	0.0	-	-
	corr.	no PT:SM	M1	0.0	0.0	0.0	0.0	4.4	8.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	29.7	35.0	0.0	0.0	0.0	0.0	0.0	0.0	6.1	8.2
			M4	0.0	0.0	0.0	0.0	-	-	0.0	0.0	0.0	0.0	0.0	0.0	-	-
		PT:SM	M1	0.0	0.0	0.0	0.0	8.8	9.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	39.1	34.0	0.0	0.1	0.0	0.0	0.0	0.0	10.0	9.3
			M4	0.0	0.0	0.0	0.0	-	-	0.0	0.0	0.0	0.0	0.0	0.0	-	-
36	spheric.	no PT:SM	M1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	15.8	19.2	0.1	0.0	0.0	0.0	0.0	0.0	1.0	2.9
		PT:SM	M1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.1	14.9	15.4	0.1	0.0	0.0	0.0	0.0	0.0	2.4	2.5
	corr.	no PT:SM	M1	0.0	0.0	0.0	0.0	3.7	11.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	30.0	41.5	0.2	0.0	0.0	0.0	0.0	0.0	7.2	14.2
		PT:SM	M1	0.0	0.0	0.0	0.0	10.4	10.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			M2	0.0	0.0	0.0	0.0	41.0	34.6	0.0	0.1	0.0	0.0	0.0	0.0	16.4	14.8

Note that even if the goal of the algorithm is to match the data, some design-driven components persist, like reducing the higher interaction levels first or, keeping the dimension of the random effects based on the design even after a PCA.

It is not possible to produce meaningful theoretical comparison with the other correlation structures discussed above and we will compare it through simulations in Section 4.6. However, the correlation structure CS-PCA, being mainly data-driven, may seem to come from a different family than the design-driven random structure RI-L, ZCP, MAX or gANOVA. However, all correlation structures can be summarized as a function or algorithm of the design and of the data. The main difference is that the procedures RI-L, ZCP, MAX and gANOVA will mostly use information about the design to select the correlation structure and CS-PCA will also use information from the data. And, again, they will all be false, and the goal is to select the most useful one.

4.6 Simulation Study

This simulation study is designed to compare the above correlation structures when performing tests on the fixed effects, which is the principal interest of researchers. Our focus are the type I error rate and the convergence rate of the methods. The type I error rate is the average number of rejected null hypotheses per simulation settings. It should be close to the nominal level, that is set here to $\alpha = 5\%$; a lower value indicates a conservative method and a higher value indicates a liberal one. Moreover, to evaluate the power of

Table 4.5: Type I error rate for three variables of the common model (M2 of in Table 4.3). Correct methods should be close the the nominal level $\alpha = .050$. The first column indicates the true correlation between random effects (homoscedastic or correlated). The second one indicates if the true model is generated with the interaction participants-stimuli. The third column indicates if model is estimated assuming the interaction participants:stimuli (+) or not (-). The RI and CS-PCA correlation structures show huge deviations from the nominal level. Confidence intervals are computed using Agresti and Coull (1998). Bold font corresponds to nominal level (5%) within the confidence interval, red font corresponds to confidence interval above the nominal level and italic font corresponds to confidence interval below the nominal level.

			RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vs									
spheric.	no PT:SM	-	.136 [.126;.148]	**.051** **[.044;.058]**	**.051** **[.044;.059]**	**.049** **[.043;.056]**	**.052** **[.045;.059]**	**.050** **[.044;.058]**	**.052** **[.045;.059]**
		+	.136 [.126;.148]	**.053** **[.047;.061]**	**.053** **[.046;.061]**	**.053** **[.046;.060]**	**.054** **[.048;.062]**	**.053** **[.047;.061]**	**.055** **[.048;.062]**
	PT:SM	-	.139 [.129;.150]	**.054** **[.047;.061]**	.059 [.052;.068]	**.053** **[.046;.060]**	.058 [.051;.065]	**.054** **[.047;.061]**	.058 [.051;.066]
		+	.139 [.129;.150]	**.054** **[.047;.061]**	.059 [.051;.067]	**.053** **[.047;.060]**	.058 [.051;.065]	**.054** **[.047;.061]**	.058 [.051;.066]
corr.	no PT:SM	-	.134 [.124;.145]	**.050** **[.044;.058]**	**.056** **[.048;.065]**	**.046** **[.040;.053]**	**.050** **[.043;.057]**	**.050** **[.044;.058]**	**.050** **[.043;.057]**
		+	.134 [.124;.145]	**.052** **[.046;.060]**	**.050** **[.042;.060]**	**.048** **[.042;.055]**	**.052** **[.046;.060]**	**.052** **[.046;.060]**	**.052** **[.045;.060]**
	PT:SM	-	.140 [.129;.151]	**.050** **[.044;.058]**	**.055** **[.047;.065]**	**.044** **[.039;.051]**	**.053** **[.046;.060]**	**.050** **[.044;.058]**	**.056** **[.049;.065]**
		+	.140 [.129;.151]	**.051** **[.044;.058]**	**.051** **[.043;.060]**	**.045** **[.039;.052]**	**.053** **[.046;.060]**	**.050** **[.044;.058]**	.057 [.050;.066]
Vp:Vm									
spheric.	no PT:SM	-	.723 [.709;.737]	.059 [.052;.066]	.065 [.057;.074]	.077 [.069;.086]	.069 [.061;.077]	.058 [.051;.066]	.077 [.069;.085]
		+	.809 [.797;.821]	**.055** **[.048;.062]**	.064 [.056;.073]	.075 [.067;.083]	.063 [.056;.071]	**.054** **[.048;.062]**	.068 [.060;.076]
	PT:SM	-	.782 [.769;.795]	**.054** **[.048;.062]**	.059 [.052;.068]	.072 [.064;.081]	.064 [.057;.072]	**.054** **[.048;.062]**	.066 [.059;.074]
		+	.824 [.812;.836]	**.054** **[.048;.062]**	.059 [.052;.068]	.072 [.064;.080]	.064 [.057;.072]	**.054** **[.048;.062]**	.067 [.060;.076]
corr.	no PT:SM	-	.698 [.684;.712]	.059 [.052;.067]	**.055** **[.047;.064]**	.075 [.068;.084]	.067 [.060;.075]	.059 [.052;.067]	.077 [.068;.086]
		+	.794 [.781;.806]	**.053** **[.047;.061]**	**.053** **[.044;.062]**	.072 [.064;.080]	.063 [.056;.071]	**.053** **[.047;.061]**	.075 [.067;.084]
	PT:SM	-	.770 [.757;.783]	**.054** **[.048;.062]**	**.058** **[.049;.068]**	.070 [.062;.078]	.064 [.057;.072]	**.054** **[.048;.062]**	.073 [.065;.083]
		+	.809 [.797;.821]	**.054** **[.047;.061]**	**.058** **[.049;.067]**	.069 [.062;.078]	.064 [.056;.072]	**.054** **[.047;.061]**	.073 [.065;.083]
Vp:Vs:Vm									
spheric.	no PT:SM	-	.250 [.237;.264]	.075 [.067;.083]	*.036* *[.030;.043]*	.103 [.094;.112]	.087 [.079;.096]	.075 [.067;.083]	.409 [.394;.425]
		+	.446 [.431;.462]	**.051** **[.045;.059]**	*.040* *[.034;.047]*	.092 [.083;.101]	.070 [.063;.079]	**.051** **[.045;.059]**	.249 [.235;.263]
	PT:SM	-	.360 [.345;.375]	**.049** **[.043;.056]**	*.039* *[.033;.046]*	.085 [.077;.094]	.060 [.053;.068]	**.049** **[.043;.056]**	.195 [.183;.208]
		+	.463 [.448;.479]	**.049** **[.042;.056]**	*.040* *[.034;.047]*	.085 [.076;.094]	.059 [.052;.067]	**.049** **[.042;.056]**	.189 [.177;.202]
corr.	no PT:SM	-	.245 [.232;.259]	.083 [.075;.092]	*.026* *[.020;.032]*	.113 [.104;.124]	.086 [.078;.096]	.083 [.075;.092]	.457 [.441;.473]
		+	.420 [.405;.436]	.060 [.053;.068]	*.028* *[.022;.035]*	.103 [.094;.113]	.069 [.062;.077]	.060 [.053;.068]	.494 [.478;.511]
	PT:SM	-	.351 [.337;.366]	.058 [.051;.065]	*.028* *[.022;.036]*	.106 [.097;.116]	.068 [.061;.077]	.058 [.051;.065]	.502 [.485;.519]
		+	.448 [.433;.463]	.057 [.050;.064]	*.027* *[.022;.034]*	.106 [.097;.116]	.067 [.059;.075]	.056 [.050;.064]	.493 [.477;.510]

the tests, we recorded, under the alternative hypothesis, the average number of true positive (the empirical power). A higher number of true positive indicates a more powerful method.

Table 4.6: Type I error rate for three variables of the common model (M1 of in Table 4.3) with 36 stimuli. Correct methods should be close the the nominal level $\alpha = .050$. The data are generated without random intercepts. In this setting, the type I error rates of RI-L deviate strongly from the nominal level, whereas gANOVA stay close to it. Confidence intervals are computed using Agresti and Coull (1998). Bold font corresponds to nominal level (5%) within the confidence interval, red font corresponds to confidence interval above the nominal level and italic font corresponds to confidence interval below the nominal level.

			RI-L	RI-L+	gANOVA	gANOVA+
Vp	corr.	no PT:SM	*.005 [.003;.008]*	*.005 [.003;.008]*	**.046 [.040;.054]**	**.046 [.040;.054]**
		PT:SM	*.006 [.004;.008]*	*.006 [.004;.008]*	**.049 [.043;.056]**	**.049 [.043;.056]**
	spheric.	no PT:SM	*.006 [.004;.008]*	*.006 [.004;.008]*	**.051 [.045;.058]**	**.051 [.045;.058]**
		PT:SM	*.005 [.003;.008]*	*.005 [.003;.008]*	*.039 [.033;.045]*	*.039 [.033;.045]*
Vm	corr.	no PT:SM	.119 [.109;.129]	.119 [.109;.129]	**.047 [.041;.054]**	**.047 [.041;.054]**
		PT:SM	.118 [.108;.128]	.118 [.108;.128]	**.050 [.043;.057]**	**.050 [.043;.057]**
	spheric.	no PT:SM	.117 [.107;.127]	.117 [.107;.127]	**.046 [.040;.054]**	**.046 [.040;.054]**
		PT:SM	.107 [.098;.117]	.107 [.098;.117]	**.051 [.044;.058]**	**.051 [.044;.058]**
Vp:Vs	corr.	no PT:SM	.114 [.105;.125]	.114 [.105;.125]	**.052 [.045;.059]**	**.052 [.045;.059]**
		PT:SM	.114 [.104;.124]	.113 [.104;.123]	**.050 [.044;.058]**	**.050 [.044;.058]**
	spheric.	no PT:SM	.113 [.103;.123]	.113 [.103;.123]	**.047 [.041;.054]**	**.047 [.041;.054]**
		PT:SM	.111 [.101;.121]	.110 [.101;.121]	**.050 [.044;.058]**	**.050 [.044;.058]**

4.6.1 Simulating the Datasets

We choose several simulation settings in order to match likely experimental designs and 4000 samples were simulated in order to have small confidence interval of our metrics. The settings vary according to 3 different designs, 2 different sample sizes, 2 different correlations of random effects, and the fact that random effects for the interaction participants:stimuli are included or not.

The 3 experimental designs are: a small design with only 2 levels per factor (M1 in Table 4.3), a rather common design with 3 levels per factor (M2 in Table 4.3) and a larger design with more variables (M4 in Table 4.3). The designs M1 and M2 have variables of type V_P,V_S, V_M and M4 has an additional variable of type V_{PS}.

Two correlations between random effects are used for the generation of the data. In the first case, the random effects are spherical (spheric.) and in the second case, the random effects are fully correlated (corr.); the fully correlated covariance matrix is such that random effects spanned a space of half of the dimension of the random effects (but not in the canonical directions). In order to give more importance to the main effects, the standard deviations of random effects are halved when increasing an order (or degree) of interaction. Moreover, all standard deviations of the random effects associated to stimulus and interaction participants:stimuli are shrunken by 0.9, respectively 0.8. Each design is simulated with random effects associated to the interaction participants:stimuli (PT:SM) and without (no PT:SM). Moreover, the small (M1) and common (M2) designs are simulated with 2 different sample sizes: 18 participants and 18 stimuli, and 18 participants and 36 stimuli. The large model (M4) was only simulated using 18 participants and 18 stimuli to reduce computation time.

Because decreasing variability as the interaction level increases favours RI-L correlation structure, we also produce another simulation based on design M1. We change the variance of random effects to highlight the difference of gANOVA and RI-L on the type I error rate. In that case, we simulate data without random intercepts while all others standard

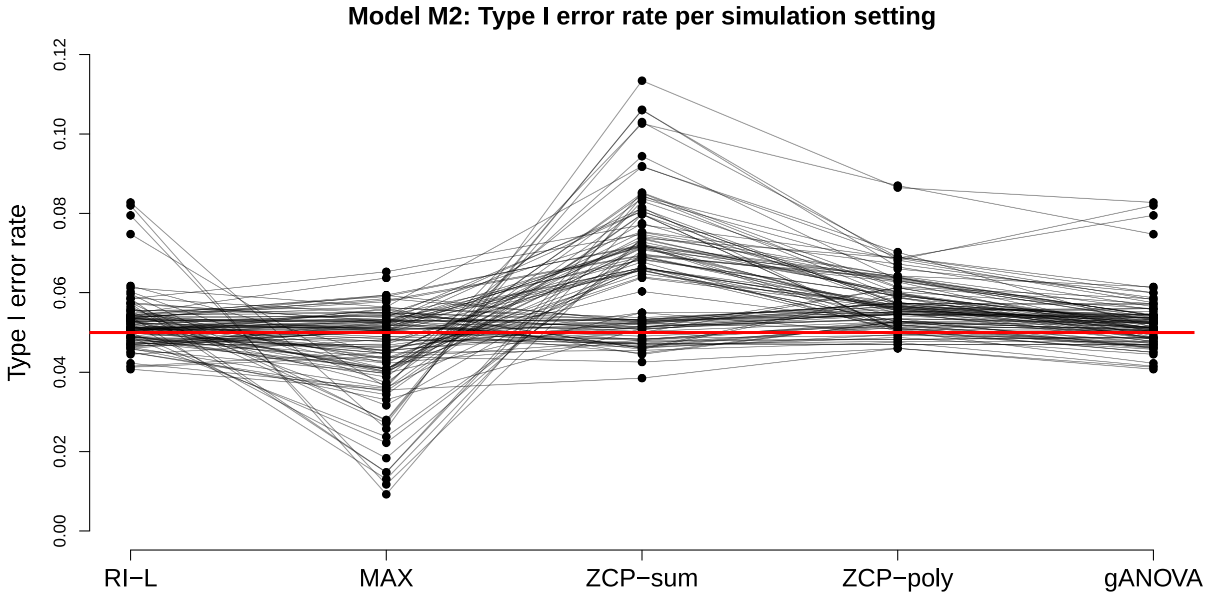

Figure 4.1: Display of type I error rates for all simulations setting of the model M2 (2 sample sizes × 2 correlations of random effects × 2 interactions in simulation × 2 interactions in estimation × 7 effects = 112 settings). The spherical correlation structures (RI-L and gANOVA) produce results closer to the nominal level $\alpha = .050$ represented by a red line.

deviations of random effects are kept the same.

Finally, we produce a power analysis by increasing the fixed effect parameters. We first estimated a maximum value for the parameter such that the empirical power exceeds 90% for each effect. Then, each effect is tested by multiplying this value by .2, .4, .6, .8 and 1. For factors with more than 3 levels, we let all fixed parameters increase simultaneously. Moreover, to reduce the computation time, we increased all fixed effects simultaneously (all main effects and all interactions).

4.6.2 Fitting of the Data

The randomly generated data are fitted using the correlation structures presented in Section 4.5: the random intercepts (RI), the random intercepts at each level (RI-L), the maximal (MAX), the zero-correlation parameter (ZCP), the generalized ANOVA (gANOVA) and correlation structure based on PCA (CS-PCA). ZCP is computed once with the default (non-orthonormal) “sum” coding (ZCP-sum) and once with a “polynomial” (and orthonormal) coding (ZCP-poly) on the random effects. Each model is estimated with (+) and without (-) assuming the random effects associated to the interaction participant-stimulus. Moreover, the significance is evaluated using the type III test with Satterthwaite’s approximation of the degrees of freedom using the `lmerTest` package (Kuznetsova et al., 2017) and the restricted maximum likelihood (REML) estimation (Bates et al., 2015). The larger model (M4) was only estimated using RI, ZCP and gANOVA to reduce computation time.

To reduce the convergence error, each model is first optimized using the default `BOBYQA` optimizer (Powell, 2009), then the Nelder-Mead optimizer (Nelder and Mead, 1965), then from the `optimx` package (Nash and Varadhan, 2011) the `nlminb` optimizer and the `L-BFGS-B` optimizer. We stop the procedure when a solution is found without convergence

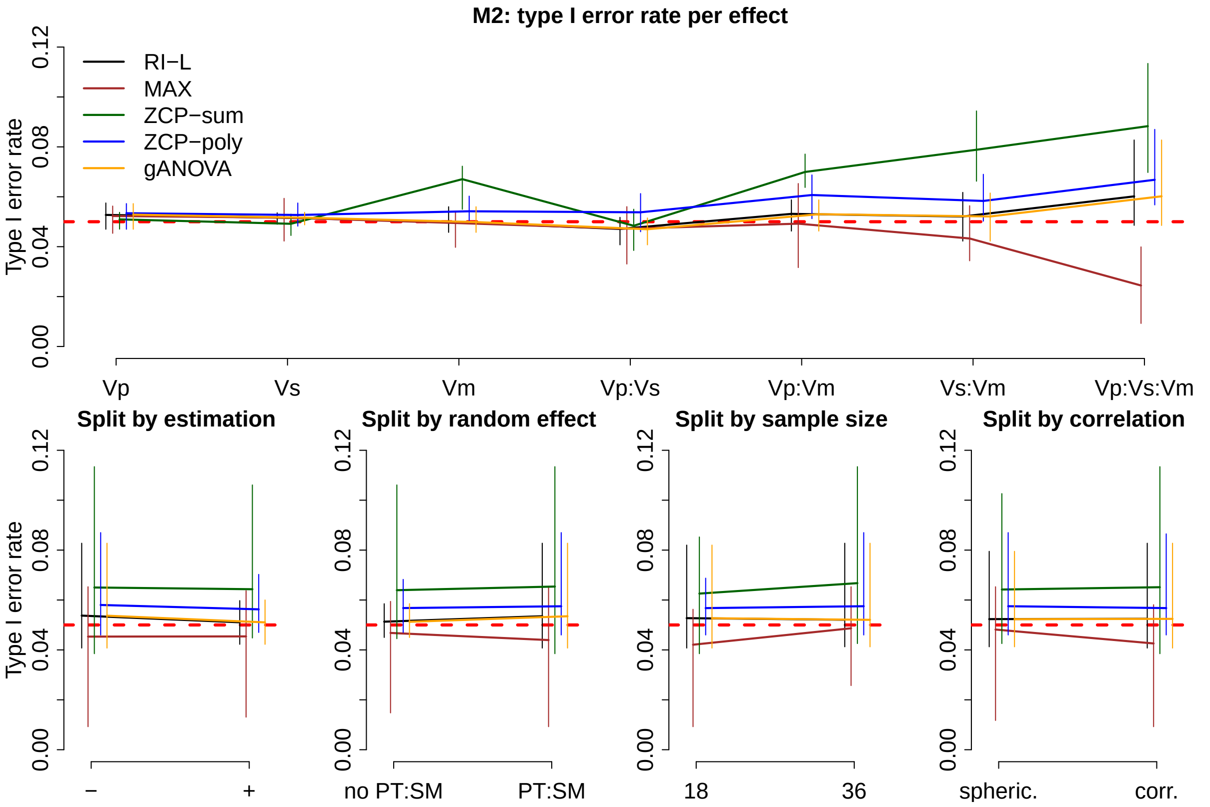

Figure 4.2: Display of type I error rates of the model M2 split given the simulations settings. The vertical lines indicate the range of all simulations within the condition. RI-L and gANOVA are the closest to the nominal level $\alpha = .050$ represented by a red dashed line. The variable V_M and its interaction produce higher deviation from the nominal level across all correlation structures. No other simulation setting tends to have an effect on the type I error rate.

error. If all optimizers fail, we declare a failure of convergence for that sample.

4.6.3 Evaluation of Simulation

Table 4.4 shows the percentage of samples with convergence error based on 4000 simulated samples for all simulation settings (designs M1, M2 and M4 in Table 4.3). We deduce that MAX is not scalable to even moderately sized designs because with only 3 levels per factor we recorded up to 40% of convergence error. Moreover, our implementation of the CS-PCA by Algorithm 4.5.6 did not reach a low number of convergence error. For the other correlation structure, we achieve a high convergence rate. This means that using several optimizers seems a good practice to reduce convergence error.

Table 4.5 shows estimated type I error rates with their confidence intervals (Agresti and Coull, 1998) for the common model (M2). The liberal type I error rates are shown in red and the conservative ones in italic. The rates are computed using only the samples without convergence error which might bias the results for MAX. One sees that RI and CS-PCA are globally too liberal as their type I error rates show huge deviations from the nominal level. The second observation is that including the interaction participants:stimuli does not influence the number of convergence error (see "+" versus "-" in Table 4.4) nor does it increase the type I error rate. Interestingly, the ZCP-poly and ZCP-sum exhibit

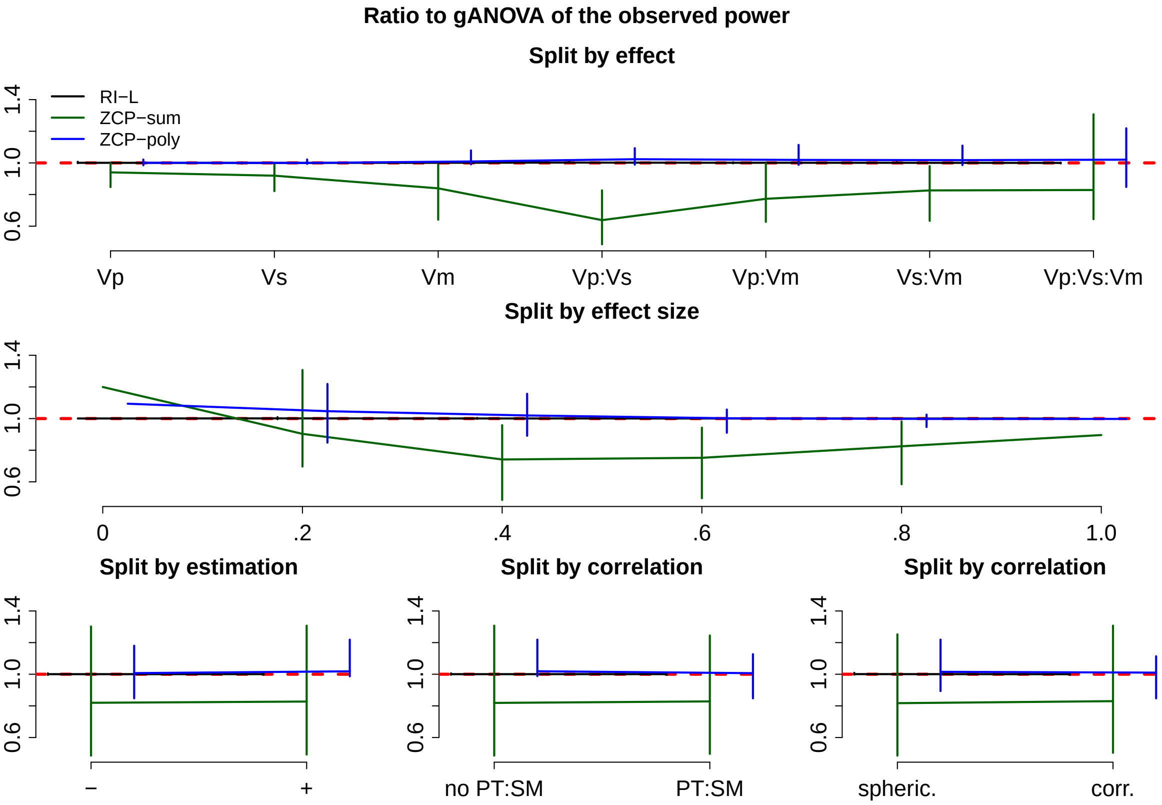

Figure 4.3: Ratio of uncorrected average observed powers of RI-L, ZCP-sum and ZCP-poly compared to the one of gANOVA. A method has larger power than gANOVA if the ratio is bigger than one. The vertical lines indicate the range of all simulations within the condition. ZCP-poly is liberal which explains that for lower effects size, gANOVA has a slightly lower power than ZCP-poly, but the deviation reduces for higher effect sizes.

differences in their type I error rate. It implies that the choice of the coding variable of the random effects will influence the results of the tests, which is not a desirable property. The orthonormal coding (polynomial) has a better control on the type I error rate.

The type I error rate seems reasonably close to the nominal level for gANOVA, ZCP-poly, and RI-L. To show the results more graphically, Figure 4.1 plots the type I error rates of Table 4.5 for the five best correlation structures: RI-L, MAX, ZCP-sum, ZCP-poly, and gANOVA. Best methods are those with points close to the red line. The superiority of ZCP-poly compared to the ZCP-sum is noticeable. Moreover, gANOVA and RI-L seem superior to ZCP-poly. By aggregating the results with respect to the simulation settings (see Figure 4.2), we see that neither the type of effects, the sample size, the third interaction in the data generation or estimation show influence on the type I error rate. And, on average, gANOVA and RI-L perform better than ZCP-poly. On the other side, the type of variable is influential on the type I error rate. We see that the variable V_M (or the interactions with this type of variable) induce type I error rates that deviate more from the nominal level.

In the supplementary material, the results for all variables and the three models M1, M2 and M4 are displayed. The above findings, including that gANOVA and RI-L are closer to the nominal level than ZCP-poly and ZCP-sum, that the lack of influence of the simulation setting and the higher deviation of the nominal level from test of variable V_M (and its interaction), are similar for all simulations, including for the larger design M4.

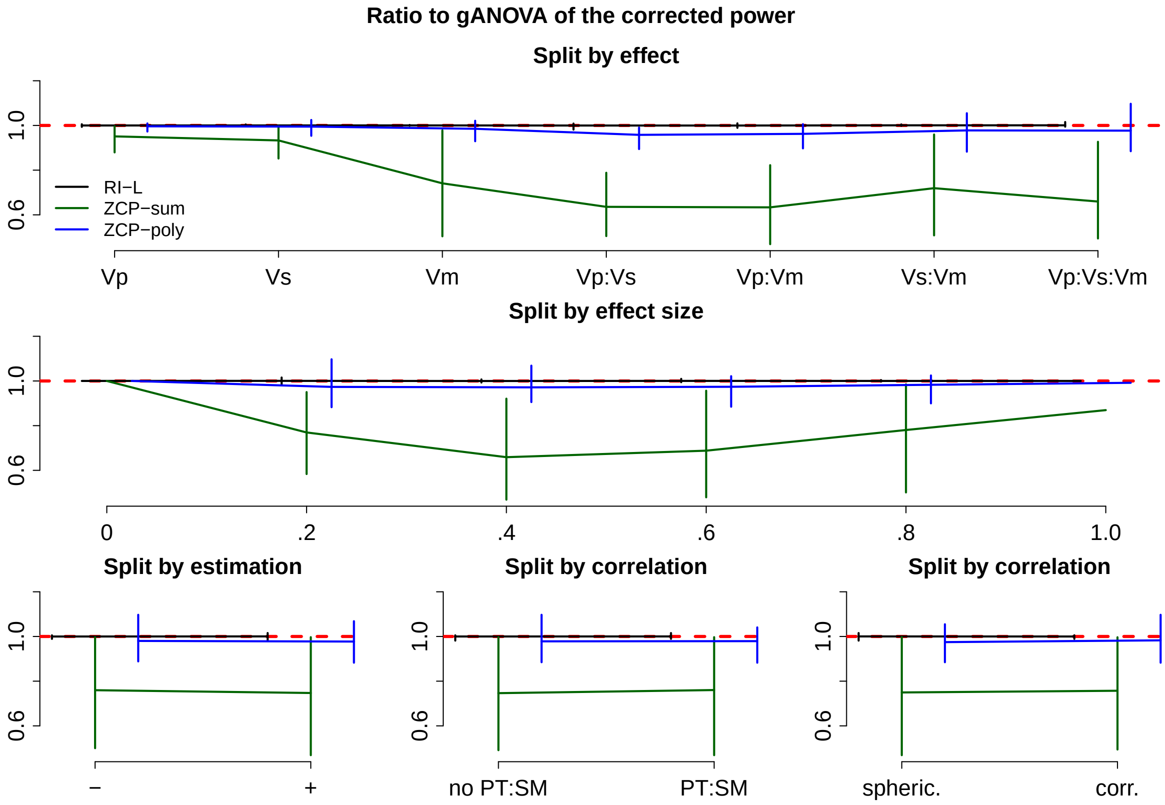

Figure 4.4: Ratio of corrected average observed powers of RI-L, ZCP-sum and ZCP-poly compared to the one of gANOVA. A method has larger power than gANOVA if the ratio is bigger than one. The vertical lines indicate the range of all simulations within the condition. In all simulation settings, no correlation structure performs better than gANOVA.

The difference between gANOVA and RI-L is visible in Table 4.6 which shows simulations with data from model M2 with a null standard deviation for the random intercepts. For these datasets, gANOVA stay close to the nominal level but RI-L shows large deviations (liberal or conservative). In addition to the theoretical remarks on the difference between RI-L and gANOVA given in Section 4.5 and Appendix D.4, these simulations show that gANOVA is strictly better than RI-L for reporting tests of fixed effects.

Moreover, the parsimony of gANOVA seems to endow it an advantage for the power of the test. All the power results are provided in the supplementary material. Figure 4.3 summarizes the findings by computing the ratio of uncorrected average observed powers of RI-L, ZCP-sum and ZCP-poly compared to the one of gANOVA. A method has larger power than gANOVA if the ratio is bigger than one and smaller power if < 1. One sees that ZCP-poly and gANOVA perform clearly better than ZCP-sum and that RI-L is close to gANOVA due to the choice of the variance parameters of the random effects. As seen in Figure 4.1, ZCP-poly has a higher type I error rate for most of the simulation settings under the null hypothesis. However, by increasing the effect size, this deviation decreases. This means that the better control of the type I error rate of gANOVA is not achieved at the expense of the power of the test. Moreover, when the power is corrected by using a critical value set at the nominal level (Figure 4.4), gANOVA performs better than ZCP-poly. No other simulation settings have a big influence on the average difference of power between the methods.

4.7 Conclusion

Using CRE-MEM in psychology is a growing practice and there is a diversity of correlation structures that are available and that are used. These correlation structures directly depend on the design of experiment and we develop a classification of variables in order to help users in the planning of the experiment and its analysis. All correlation structures do not share the same advantages and there is no predefined tools to help researchers select the appropriate correlation structure given the experiment. Depending on the goal of the analysis some properties are more important to control than others.

In the case of hypothesis testing, we have shown that the gANOVA correlation structure has many desirable properties. Simulations show that it controls the type I error rate without a loss in power even with misspecifications of the model. It also provides a high rate of convergence and is scalable to complex datasets. Moreover, it is in line with the experimental design tradition represented by the ANOVA/rANOVA framework, which is helpful for the interpretation of the results for researchers familiar with the ANOVA.

Acknowledgement

We are particularly grateful for the assistance given by Audrey Bürki that greatly improved this manuscript. She provided many comments coming from her extended reading of this paper and her expertise with CRE-MEM; although any errors are our own.

Chapter 5

A General Re-sampling Method for Multiple Tests in CRE-MEM

5.1 Introduction

In Chapter 4 we explain that experiments in psychology are often the results of the crossing of a sample of participants and a sample of stimuli. For this type of design, CRE-MEM is the general class of model that should be used as it considers both the variability induced by the sampling of participants and stimuli. Moreover, crossing participants and stimuli is also performed when recording the brain activity with EEG or fMRI. However, in order to analyse EEG or fMRI data, we must perform thousands of tests, and a multiple comparisons procedure should also be used to control the number of type I errors (Bullmore et al., 1999). As presented in Chapter 1, permutation tests with multiple comparisons procedures like the cluster-mass test (Maris and Oostenveld, 2007) or the threshold-free cluster-enhancement (Smith and Nichols, 2009) on this type of signals are very powerful and control the FWER as they wisely use correlation between tests. These procedures are used with permutation but no general permutation method that considers the effect of stimuli exists. In Bürki et al. (2018), we show that ignoring the effect of the stimuli drastically increases the type I error rate in the univariate case and also increases the FWER for comparison of signals and we proposed a re-sampling approach based on the quasi-F statistic (Clark, 1973) using sums of squares. In the present Chapter, we propose a general framework for re-sampling tests used for multiple comparisons procedure in CRE-MEM that includes the approach presented in Bürki et al. (2018) and we recall our simulation findings.

Several re-sampling methods exist for mixed model and Modugno and Giannerini (2015) gives a nice overview of bootstrapping procedures. They consider a parametric bootstrap, a case bootstrap and a residuals bootstrap as proposed by Carpenter et al. (2003). However, for multiples comparisons using the same design, only re-sampling procedures that conserve the links between tests are useful. Hence, separated generations of data using parameter estimates, like the parametric bootstrap, are proscribed as it destroys these temporal correlations (without an additional model of the correlations). Moreover, the case bootstrap cannot be applied when crossing 2 samples as there is no independent cases or sampling units. Hence, the following propositions are based on re-sampling of the predictions of the random effects.

In the next Sections we propose a general framework that allows multiple comparisons procedures with re-sampling method in CRE-MEM.

5.2 The CRE-MEM Equation

In order to simplify the notation, we describe the model and the equations for only one test (i.e. for one time-point) and the CRE-MEM equation is written as:

$$Y = D\eta + X\beta + \sum_{j=1}^{J} \sum_{k=1}^{K(j)} Z_{jk} G_{jk}^{-} G_{jk} \gamma_{jk} + \epsilon, \tag{5.1}$$

where Y is the response, β are the parameters of interest associated to the design X and the nuisance variables are D associated to η. For each J random unit (e.g. $j = 1$ for participants, $j = 2$ for stimuli, $j = 3$ their interaction or more), we decompose the random part into $K(j)$ sets of independent random effects $\gamma_{jk} \sim \mathcal{N}(0, \Sigma_{jk})$. The Z_{jk} are matrices with 0 and 1 coding which random effect influences which observation. The dummy coding of the Z_{jk} matrices needs the contrasts G_{jk}^{-} depending on the assumed correlation structure. The matrices G_{jk} and G_{jk}^{-} are either identity matrices (for RI-L and the random intercepts) or orthonormal contrasts (for the random interactions in gANOVA) such that $G_{jk}^{-} G_{jk} = I_{n_j} \otimes R_{\mathbf{1}}$ and $G_{jk} G_{jk}^{-} = I$, where n_j is the the sample size of the j random unit. Finally, we add the error term $\epsilon \sim \mathcal{N}(0, I\sigma^2)$.

Equation 4.2 can be expressed in the form of Equation 5.1. The lines 2 to 4 of Equation 4.2 are the random effects associated to the participants (line 2 in Equation 4.2 and $j = 1$ in Equation 5.1) , to the stimuli (line 3 in Equation 4.2 and $j = 2$ in Equation 5.1) and finally to their interaction (line 4 in Equation 4.2 and $j = 3$ in Equation 5.1). Moreover, the π_i in Equation 4.2 code the random intercept associated to the participants of the ith observation which correspond to the ith element of $Z_{1,1}\gamma_{1,1}$ where $j = 1$, $k = 1$ in Equation 5.1 as G_{ij} is usually the indentity matrix for random intercept. Moreover, the random interaction between fixed effects and the participants are also represented in both equations as they are $(\pi\psi)_{ik}$, $(\pi\phi)_{il}$ and $(\pi\psi\phi)_{ikl}$ in Equation 4.2 and $Z_{1,k} G_{1k}^{-} G_{1k} \gamma_{1,k}$, $j = 1$, $k = 2, 3, 4$ in Equation 5.1. The same relationship between the two equations exists for the stimuli and the interaction.

Then, the variance of the response is:

$$\mathrm{Var}\left[Y\right] = \Omega = I\sigma^2 + \sum_{j=1}^{J} \sum_{k=1}^{K(j)} Z_{jk} G_{jk}^{-} G_{jk} \Sigma_{jk} G_{jk}^{-} G_{jk} Z_{jk}^{\top}. \tag{5.2}$$

The correlation structures, RI, RI-L, gANOVA or ZCP, assume a simple form for the correlation matrices of random effects such that $\Sigma_{jk} = I\sigma_{jk}^2$.

In the model represented in Equation 5.1, we are interested by testing the null hypothesis:

$$H_0 : \beta = 0. \tag{5.3}$$

5.3 The Estimation

Given the observation of y, an estimation of the full model produces the estimates $\hat{\beta}$, $\hat{\eta}$ and predictors $\hat{\gamma}_{jk}$ and $\hat{\epsilon}$ such that:

$$y = D\hat{\eta} + X\hat{\beta} + \sum_{j=1}^{J} \sum_{k=1}^{K(j)} Z_{jk} G_{jk}^{-} G_{jk} \hat{\gamma}_{jk} + \hat{\epsilon}. \tag{5.4}$$

A first approach of the estimation can be performed assuming γ_{jk} as random and using the mixed linear models framework. One needs to specify the correlation structure which defines the Z_{jk}'s, G_{jk}'s and Σ_{jk}'s. We saw in Chapter 4 that ZCP or gANOVA are good choices when testing fixed effects in experiments with cross-random effects. The estimation may be performed with the `lme4` package for maximum likelihood (ML) or restricted maximum likelihood (REML) (Bates and DebRoy, 2004; Bates et al., 2015) or `gANOVA` (https://github.com/jaromilfrossard/gANOVA). The $\hat{\gamma}_{jk}$ are usually referred to as best linear unbiased predictor (BLUP). The predictions of the BLUP are shrunk towards zero such that their empirical variance is typically smaller than σ^2_{jk}. They must therefore be modified if used in a re-sampling procedure (Morris, 2002).

Another approach is to estimate η, β and γ_{jk} as fixed effects using OLS. Contrary to BLUPs, the OLS does not creates shrinkage of random effects (Efron and Morris, 1977). Moreover, finding the OLS estimates is much less time consuming than the optimization of a CRE-MEM.

In both cases, the higher interaction term of the random effects ($(\pi\omega\phi)_{imk}$ in Equation 4.2) is usually confounded with the error term. It implies that its variance cannot be estimated. It happens when the same setting (same participant, same stimulus and same manipulation) is not repeated during the experiment and only one observation is collected. In that case, this last interaction term is omitted for the estimation.

5.4 Re-Sampling Methods

Given the estimates and predictions $\hat{\eta}$, $\hat{\beta}$, $\hat{\gamma}_{jk}$ and $\hat{\epsilon}$ (produced either from the mixed model framework or from OLS), we compute the re-samples under the alternative hypothesis using:

$$Y_1^* = D\hat{\eta} + X\hat{\beta} + \sum_{j=1}^{J}\sum_{k=1}^{K(j)} \kappa_{jk} Z_{jk} P_{jk} G^{-}_{jk} G_{jk} \hat{\gamma}_{jk} + P_\epsilon \hat{\epsilon}, \tag{5.5}$$

where P_{jk} and P_ϵ, are shuffling or bootstrapping matrices (instead of permutation matrices). Shuffling matrices are diagonal matrices with elements -1 or 1 in the diagonal. A bootstrapping matrix is a square matrix B with elements 1 or 0 such that $\mathbf{1}B = \mathbf{1}$; on the other side, any permutation matrix P must satisfied both equality, $\mathbf{1}^\top P = \mathbf{1}^\top$ and $P\mathbf{1} = \mathbf{1}$ in addition to have elements 0 or 1. For each new re-sample Y_1^* we randomly select new P_{jk} and new P_ϵ. The inflation factors κ_{jk} are added in order to correct for the variance of the predictors of the random effects if estimated in the mixed model framework. This parameter is inspired by the bootstrap developed by Carpenter et al. (2003) and correct for the shrinkage of the predictions of random effect induced by the estimation using mixed models. We use for instance $\kappa_{jk} = \frac{\hat{\sigma}^2_{jk}}{\text{sd}(\hat{\gamma}_{jk})}$, where $\text{sd}(\cdot)$ is the sample standard deviation.

The contrast matrices G_{jk} are especially useful for the re-sampling of the gANOVA correlation structure or when using the OLS estimation. In these cases the estimation only uses the "constraint" matrices $Z_{jk}G^{-}_{jk}$ and software produces a prediction of the contrasts $G_{jk}\gamma_{jk}$. However, shuffling/bootstrapping/permuting directly these predictions $G_{jk}\gamma_{jk}$ does not change the average of the re-sampled random effects. Indeed, for all shuffling, bootstrap or permutation matrix P_{jk}, we have the following equality $\mathbf{1}G^{-}_{jk}P'_{jk}G_{jk}\hat{\gamma}_{jk} = 0P'_{jk}G_{jk}\hat{\gamma}_{jk} = 0$ when G_{jk} is a contrast a therefore orthogonal to $\mathbf{1}$. This implies that the sum of squares associated with this effect would stay constant over all re-samples which is contrary to the requirement that re-samples should be like new samples. It is not the case when shuffling/bootstrapping $P_{jk}G^{-}_{jk}G_{jk}\hat{\gamma}_{jk}$ which is our new proposal. Note that for the correlation structure for which $G_{jk} = G^{-}_{jk} = I$ this difference disappears. Moreover, permuting random effects are also proscribed as $\mathbf{1}^{\top}P_{jk}G^{-}_{jk}G_{jk}\hat{\gamma}_{jk} = 0$ when P_{jk} is a permutation matrix.

Using re-samples under the alternative hypothesis in Equation 5.5, we easily produce re-samples under the null hypothesis inspired by the `terBraak` method (ter Braak, 1992):

$$Y^{*}_{tB} = Y^{*}_{1} - X\hat{\beta}. \tag{5.6}$$

Another option inspired by the `dekker` method (Dekker et al., 2007) is computing the re-sampled statistic using Y^{*}_{1} as the response variable and the permuted fixed part of the design $[D\ P_X R_D X]$ where P_X is a permutation matrix. Permuting the design breaks any links between the re-sampled response under the alternative and the effect of interest β.

In summary, we have two re-sampling propositions which are summarized by transformations of the observed data. The `terBraak`-like method is the transformation of the observed data $\{y,\ D,\ X,\ Z\} \ \rightarrow\ \{Y^{*}_{tB},\ D,\ X,\ Z\}$ and the `dekker`-like method is $\{y,\ D,\ X,\ Z\} \ \rightarrow\ \{Y^{*}_{1},\ D,\ PR_DX,\ Z\}$.

5.5 Test Statistics, Parametric and Re-sampled *p*-values

In order to test hypothesis in Equation 5.3, four statistics are commonly used in the parametric setting. The quasi-F statistics (Clark, 1973) is fastest to compute because it is based on sums of squares but can be used with a balanced designed without missing value. Then, based on the mixed model framework, the Satterthwaite's approximation (Schaalje et al., 2002) and the Kenward-Roger approximation (Kenward and Roger, 1997) propose approximation of the degree of freedom. They produce good and similar results in terms of type I error rate in the parametric setting. However, the Kenward-Roger approximation is computationally intensive which makes it not adapted for re-sampling tests using the present algorithms. Finally, we can also use in the mixed model framework the likelihood ratio test statistic which asymptotic is a χ^2 distribution. However, it shows some divergence between the type I error rate and the nominal level in the parametric setting.

Note that the quasi-F statistic cannot be used in combination with the `dekker`-like method as the statistic relies on the orthogonality of the design matrix in balanced design. The `dekker`-like method permutes the design which breaks the orthogonality in the design and invalidates the use of the statistic.

The quasi-F statistic, the Satterthwaite's approximation, and the Kenward-Roger approximation rely on approximations of the degree of freedom to compute the p-values. Unlike F statistic in ANOVA, the distribution from which the p-value is computed also depends on the observed response. In a re-sampling test, it implies that computing the p-value using re-sampled statistics or their transformations into the probability scale produces different results. For the simple model like ANOVA, running a permutation test using the distribution by permutation of the statistic (e.g. F_y) or of its corresponding p-value (e.g. $1 - F^{-1}_{df_1,df_2}(F_y)$ when df_1 and df_2 are degrees of freedom corresponding on the parametric test) leads to the same results. It is only true when $F^{-1}_{df_1,df_2}$ is a strictly increasing function of F_y and df_1 and df_2 are fixed. However, this equivalence does not hold

Table 5.1: All possible combinations of estimations, re-sampling methods and test statistics presented in Chapter 5.

Methods	Statistics	Drawbacks	
Estimation by OLS			
terBraak	quasi-F	Balanced Design	Simulation study
	Satterwhaite	Slow	Univariate
	LRT	Slow	Univariate
	Kenward-Roger	Very slow	
dekker	quasi-F	Not feasible	
	Satterwhaite	Slow	Univariate
	LRT	Slow	Univariate
	Kenward-Roger	Very slow	
Estimation by MLM			
terBraak	quasi-F	Balanced Design	Signals
	Satterwhaite	Slow	Univariate
	LRT	Slow	Univariate
	Kenward-Roger	Very slow	
dekker	quasi-F	Not feasible	
	Satterwhaite	Slow	Univariate
	LRT	Slow	Univariate
	Kenward-Roger	Very slow	

for methods relying on approximation of degrees of freedom, like the quasi-F statistic. If the degrees of freedom are different through re-samples, using p-values instead of the statistics seems a more valid re-sampling approach to compute the p-value as it compares values on the same scale.

Moreover, this approach has a second advantage in the multiple comparisons problems as the degrees of freedom are also different through tests. For instance, the cluster-mass test procedure relies on summation of statistics to compute the cluster-mass. It relies on the assumption that the distributions under the null hypothesis are the same through time-points and using different degrees of freedom acknowledge that these distributions are indeed different. Transforming the statistics into p-values (more precisely into the logarithm of their inverse) insures that each test has the same scale when computing the cluster-mass. Moreover, using the logarithm of the p-values insures that the smallest p-value has an important effects on the cluster-mass in comparison the relatively small one; in lay terms, the logarithm of the p-value is a more ”natural” scale (Fisher, 1948), especially when combining statistics. In summary, from a re-sampled response Y^*, we produce a re-sampled statistics T^* and re-sampled degrees of freedom df_1^* and df_2^*, which are transformed into a p-value using $p^* = 1 - F^{-1}_{df_1^*,df_2^*}(T^*)$, where $F^{-1}_{df_1^*,df_2^*}$ is the inverse of the theoretical distribution function of the statistic (e.g. a F distribution). Then, the re-sampled statistics in the probability scale are transformed using logarithm: $T^*_{\ln p} = -\ln(p^*)$. Finally, performing this for all time-points allows us to use the cluster-mass procedure by combining statistics on the same scale. The usual threshold set at the 95 percentile of the parametric distribution becomes $\tau = -\ln 0.05$.

Table 5.1 shows all possible combinations of estimations, re-sampling method and test statistics described in the previous sections. Section 5.6 shows the combination of the OLS estimation, with the `terBraak` re-sampling method and the quasi-F statistic. It is the fastest case which is feasible for a simulation study in both the univariate case and the comparison of signals. Moreover, the first estimation by MLM (also with the quasi-F and `terBraak` method) could also be tested in a simulation study with a reasonably low computing time for the comparison of signals as only one MLM optimization per time-points must be performed. Moreover, the computing time may be reduced by taking the optimal parameters at time-point $t-1$ as the first guess of the parameters at time t as the signals show high temporal correlations. As explained previously, the `dekker` method disturbs the design which makes it not compatible with the quasi-F statistic. Finally, using the Satterthwaite's approximation, or likelihood ratio test as test statistics implies an optimisation of MLM for each re-sample. These procedures are obviously time consuming. However, some computing time may be gained by setting, through re-samples, the first guess of the parameters as the average (or median) of the previous optimal parameters. Indeed, the optimal parameters of all re-samples may be close to a central value through re-samples. However, even with these small optimizations, it is actually not testable in a simulation study for comparison of signals in a reasonable low computing time. Nevertheless, Winkler et al. (2016) propose to use matrix completion as a method to predict permuted test statistics without performing all optimizations. Indeed, the multiple comparisons procedures like cluster-mass test are performed using a large matrix of test statistics sorted by re-samples in rows (around 5000) and time-points (around 1000) in columns. In CRE-MEM, finding each 5 millions elements of this matrix is a time consuming task and Winkler et al. (2016) show that this full matrix can be imputed by matrix completion using only a subset of the re-sampled statistical values.

5.6 Simulation Study

In Bürki et al. (2018), we presented simulation studies implementing the re-sampling test for the quasi-F in CRE-MEM in the univariate case and for the comparison of signals with the cluster-mass test. In both cases, OLS is used to estimate the $\hat{\beta}$, $\hat{\eta}_{jk}$ and $\hat{\epsilon}$, the matrices P_{jk} and P_{ϵ} are shuffling matrices, the re-sampling method is the `terBraak` method and the scale of re-sampled random effects is not changed as we set $\kappa_{jk}=1$.

5.6.1 The Univariate Case

For the univariate case, the design crosses 20 participants with 18 or 36 stimuli. It assumes one variable of type V_S with 2 levels (e.g. sex), one variable of type V_S with 3 levels (e.g. emotions of the stimuli) and one variable of type V_M with 2 levels (e.g. test and retest of the same settings); see Section 4.4 for the typology of variables in CRE-MEM. These simulation settings produce 720 observations for the case with 18 stimuli and 1440 for the case with 36 stimuli.

We simulated the data assuming the variances of random effects decreasing with respect to the interaction levels ($sd = 1, 0.5, 0.3$) but a higher variability for the error term ($sd = 2$). All random effects and the error term are simulated from normal distributions.

In the univariate case, 5 different tests are compared. First, the F1 ANOVA (rANOVA after averaging the data over the stimuli) with a parametric p-value and p-value by permutation (Kherad-Pajouh and Renaud, 2015) which model does not assume random effects

Table 5.2: Type I error rate of the tests of main effects of 3 variables (see Section 4.4 for the typology). The F1 ANOVA is not a reliable test when crossing stimuli and items as it does not control the type I error rate. Moreover, using CRE-MEM or quasi-F statistic we achieve type I error rate close to the nominal level for both the parametric and the re-sampling tests. Confidence intervals are computed using Agresti and Coull (1998). Bold font corresponds to nominal level (5%) within the confidence interval, red font corresponds to confidence interval above the nominal level and italic font corresponds to confidence interval below the nominal level. Table presented in Bürki et al. (2018).

	V_P	V_S	V_M
18 Stimuli			
ANOVA F1 (parametric)	.064 [.057;.072]	.628 [.614;.644]	.120 [.111;.131]
ANOVA F1 (permut.)	.064 [.057;.072]	.626 [.611;.641]	.120 [.111;.131]
CRE-MEM (Satterthwaite)	**.054 [.046;.063]**	**.053 [.045;.062]**	**.050 [.042;.059]**
Quasi-F (parametric)	**.048 [.042;.056]**	**.051 [.045;.058]**	*.036 [.031;.043]*
Quasi-F (permut.)	**.052 [.045;.059]**	**.054 [.048;.062]**	**.043 [.038;.050]**
Quasi-F (permut. log-p)	**.052 [.046;.060]**	**.052 [.046;.060]**	**.045 [.039;.052]**
36 Stimuli			
ANOVA F1 (parametric)	**.056 [.049;.063]**	.555 [.540;.571]	.093 [.084;.102]
ANOVA F1 (permut.)	**.055 [.048;.062]**	.554 [.539;.570]	.093 [.084;.102]
CRE-MEM (Satterthwaite)	**.049 [.042;.056]**	**.052 [.045;.059]**	**.048 [.041;.055]**
Quasi-F (parametric)	**.046 [.040;.053]**	**.048 [.042;.055]**	*.038 [.033;.044]*
Quasi-F (permut.)	**.050 [.043;.057]**	**.052 [.046;.060]**	**.043 [.038;.050]**
Quasi-F (permut. log-p)	**.050 [.044;.057]**	**.051 [.045;.059]**	*.043 [.037;.050]*

associated to stimuli. Then, we estimated the CRE-MEM with a RI-L correlation structure and tested the effect using the Satterwhaite's approximation. Finally, we compared 3 different types of quasi-F statistic: the parametric one, the re-sampled test with a p-value computed on the raw statistics and the re-sampled test with the transformation to the logarithm of the p-value (see Section 5.5).

Table 5.2 shows the results of estimated type I error rate based on 4000 simulations with their confidence intervals. We first see that the F1 ANOVA does not control the type I error rate for both parametric and permutation tests. Moreover, the discrepancy is high as it produces up to 60% of false positive (here for 18 stimuli and variable V_S). However, the test using CRE-MEM (with the Satterwhaite's approximation) and the 3 types of quasi-F statistics produce type I error rates close to the nominal as they are designed to take into account the variability induced by the sampling of stimuli. The Satterwhaite's approximation performs better but the 3 quasi-F statistic are still reliable tests for controlling the type I error rate as they may only be too conservative. Finally, computing p-values on the raw re-sampled quasi-F statistics or on their transformations to the logarithm of the p-values does not seem to influence the type I error rate.

5.6.2 Comparison of Signals

Table 5.3: FWER of the comparison of signal of 3 main effects. The F1 ANOVA are not reliable method to control the FWER rate. The quasi-F statistic produces type I error rate below 10% with the cluster-mass test and conservative one with the control of the FDR (Benjamini and Hochberg, 1995). Confidence intervals are computed using Agresti and Coull (1998). Bold font corresponds to nominal level (5%) within the confidence interval, red font corresponds to confidence interval above the nominal level and italic font corresponds to confidence interval below the nominal level.

	VP	VS	VM
9 stimuli, 10 participants			
ANOVA F1 (cluster-mass)	.128 [.118;.139]	.998 [.997;1]	.153 [.142;.164]
ANOVA F1, log-p (cluster-mass)	.151 [.141;.163]	.998 [.997;.999]	.146 [.135;.157]
ANOVA F1, param. (B.-H.)	**.043 [.038;.050]**	.998 [.997;1]	.095 [.087;.105]
Quasi-F, log-p (cluster-mass)	.090 [.082;.099]	.058 [.052;.066]	.066 [.058;.074]
Quasi-F, param. (B.-H.)	*.016 [.013;.021]*	.077 [.069;.086]	*.006 [.004;.010]*
18 stimuli, 20 participants			
ANOVA F1 (cluster-mass)	.117 [.107;.127]	1 [1;1]	.242 [.230;.256]
ANOVA F1, log-p (cluster-mass)	.126 [.116;.136]	1 [1;1]	.216 [.204;.229]
ANOVA F1, param. (B.-H.)	**.049 [.043;.056]**	1 [1;1]	.212 [.200;.225]
Quasi-F, log-p (cluster-mass)	.082 [.073;.090]	.066 [.059;.074]	.091 [.083;.100]
Quasi-F, param. (B.-H.)	*.024 [.020;.029]*	*.034 [.029;.040]*	*.010 [.008;.014]*
36 stimuli, 20 participants			
ANOVA F1 (cluster-mass)	.079 [.071;.088]	.998 [.996;.999]	.133 [.122;.143]
ANOVA F1, log-p (cluster-mass)	.083 [.075;.092]	.997 [.995;.999]	.122 [.112;.132]
ANOVA F1, param. (B.-H.)	*.038 [.033;.045]*	1 [1;1]	.120 [.111;.131]
Quasi-F, log-p (cluster-mass)	.069 [.062;.078]	.079 [.071;.088]	.091 [.083;.100]
Quasi-F, param. (B.-H.)	*.028 [.024;.034]*	*.028 [.023;.033]*	*.011 [.009;.015]*
36 stimuli, 40 participants			
ANOVA F1 (cluster-mass)	.096 [.088;.106]	1 [1;1]	.314 [.300;.329]
ANOVA F1, log-p (cluster-mass)	.098 [.089;.108]	1 [1;1]	.283 [.270;.298]
ANOVA F1, param. (B.-H.)	**.047 [.041;.054]**	1 [1;1]	.337 [.322;.352]
Quasi-F, log-p (cluster-mass)	.062 [.055;.070]	.066 [.058;.074]	.085 [.077;.094]
Quasi-F, param. (B.-H.)	*.022 [.018;.028]*	*.027 [.022;.033]*	*.012 [.009;.016]*

For the comparison of signals, we assume 4 different designs: 9 stimuli and 10 participants, 18 stimuli and 20 participants, 36 stimuli and 20 participants, and finally, 36 stimuli and 20 participants. As for the univariate case, we assume one variable of type V_S (2 levels), one variable of type V_S (3 levels) and one variable of type V_M (2 levels).

As for the univariate case, we simulated the response variable with normal random effects and their variances decreasing with respect to the interaction levels ($sd = 1, .5, .3$) but a higher variability for the error term ($sd = 2$). To simulate time correlations of random effects and errors terms, we use a Gaussian correlation function $\rho(\tau) = \exp(-3\tau^2/R)$ (Abrahamsen, 1997), with $R = 60$ for the random effects associated to the participants, $R = 40$ for the random effects associate to the stimuli and $R = 20$ for the random effects associate to the error terms.

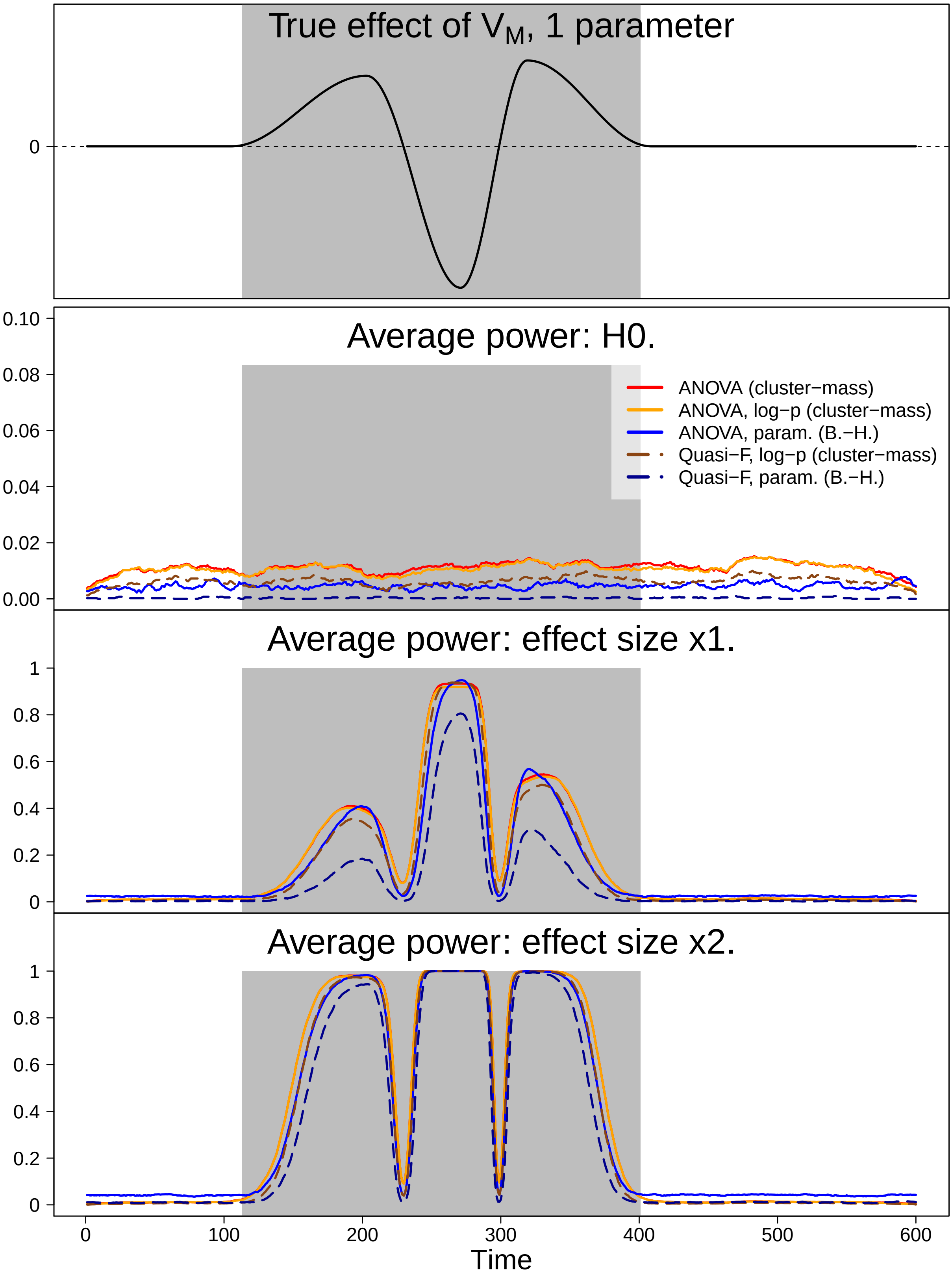

Figure 5.1: Average power of the test of V_M variable for the model with 36 stimuli and 20 participants. The cluster-mass test with the quasi-F statistic still have a high power. However, the control of the FDR results in less powerful test.

Five different methods are evaluated and their FWER are shown in Table 5.3: the permuted F1 ANOVA with the cluster-mass test, the permuted F1 ANOVA with the cluster-mass test on the logarithm of p-values, the parametric F1 ANOVA controlling the false discovery rate (FDR) (Benjamini and Hochberg, 1995), the re-sampled quasi-F with the cluster-mass on the logarithm of p-values and finally the parametric quasi-F controlling the false discovery rate.

As expected, the 3 methods based on the F1 ANOVA are not reliable to control the FWER and may produce up to $FWER = 100\%$ (variable V_S with 18 stimuli and 20 participants). On the contrary, the cluster-mass test using the quasi-F statistics keeps the nominal level under 10%, which is satisfying in comparison to the alternative of the F1 ANOVA. The parametric quasi-F with the control of the false discovery rate using Benjamini and Hochberg (1995) produces conservative FWER (with a minimal at 0.6% for the variable V_M on the smallest design.). As shown in Figure 5.1, the average power of the quasi-F statistic is still high when using the cluster-mass test while the same statistics with the control of the FDR using Benjamini and Hochberg (1995) is reduced.

5.7 Conclusion

In Chapter 5, we propose a general re-sampling approach for testing fixed effects in CRE-MEM. Moreover, Bürki et al. (2018) shows that not taking to account the stimuli effect in EEG may lead to an increase of the FWER. However, this procedure is actually based on the quasi-F statistic and has some drawbacks for EEG data. It is the fastest statistic but is too stringent for real data analysis as it must be performed using a balanced design. In EEG data analysis, even if the design planned by the experimenters is balanced, we usually delete numerous trials that shows recording errors. In most cases, we finally have an unbalanced dataset to analyse. A first solution may arise by slightly modifying the quasi-F statistic. Using the appropriate orthogonalization of the projections matrices may result in a statistic for unbalanced design. A second solution may be the completion of the missing values. Indeed, the recoding error of the signals may only appear in a small time windows of the signal but the full signal must be deleted for the tests. The missing part of the signal may be imputed using matrix completion. In addition, if the full signal is corrupted taking the average of the signal in the cell may be an appropriate solution for restoring a balance design. Research has still to be carry on to provide the best approach to analyse EEG data coming from experiments.

Conclusion

More technical conclusions have already been proposed at the end of each Chapter. To end this thesis, I first propose some thoughts on my work as a statistician. I feel lucky to have been able to work on three kinds of field related to statistics: the application of existing methods on real data analyses, the development of new theoretical results and the creation of computer software. My experience shows that these three domains are complementary.

First, the creation of `permuco` proved to be helpful to prove the asymptotic distribution presented in Chapter 3. Indeed, to produce a software fast enough to compute permutation tests on signals I had to use QR decomposition and learn some basic properties of this method. It turns out to be a key ingredient to prove the asymptotic of the distribution by permutation. Moreover, it allows me to propose the most recent permutation method available to analyse the data in Cheval et al. (2018) as I ”just” needed to generalize `permuco` from one electrode to the full scalp.

Secondly, working with real data applications allows to collaborate with people in other fields. You learn from their data analysis practice which helps you to formulate methodological problems. When working for the manuscript in Cheval et al. (2018), it was not clear which covariates should be used to adjust the daily physical activity of the participants. The covariates were strongly correlated, and the final results were not impacted by this choice. Moreover, one measure was more consensual but using it may still be refused by reviewers which implies missing an opportunity to publish our findings. The ideal solution would be performing one test per covariate and using multiple comparisons procedures to control for the FWER. In that case, it seems that permutation tests may be a good procedure as they maintain correlation between the tests. This problem must be investigated from a methodological perspective and may lead to useful applications.

Finally, developing new methods allows you to have a wide range of knowledge. It helps to develop abstract thought and a deep understanding of the statistical problems. Together it is helpful for the creation of software as you can structure them for easier maintenance. In addition, it helps to collaborate for real data applications as you can propose the best statistical practices.

Among the natural development of this works is the application of permutation tests and cluster-mass tests using robust statistics. The question is how to use robust statistic based on M-estimator in combination with permutation methods. First, note that optimizing a robust estimator is a computing intensive task and it would be time consuming to optimize it for each permutation and each time-point of EEG signals. Even in a simple design, without nuisance variables, it would be an issue to optimize a robust estimator on the raw permutations. One solution may come from taking advantage of the weights produced on the robust estimation of the observed data set. It may produce valid test for simple models but becomes more complicated when introducing nuisance variables in the models. Indeed, the permutation methods we saw in Chapter 1 are usually based on residual matrices which orthogonalize the interest variables and the response variables to the nuisance variables. In that case, the transformation is not robust and will be influenced by outlying observations. The permutation methods of Chapter 1 cannot be directly used and must be transformed to have robust properties. Again, one solution may come from the weights but projections using weights are oblique. It results that all useful properties linked to the orthogonalization to the nuisance variables are lost when using oblique projections. Finally, even if we find a satisfying permutation method using a robust statistic in the univariate case, its application on signals raises also multiple questions. For comparing signals, the weights may be different for each time-point, for each observation (or, more precisely, complete signal) or for both. All three cases should be considered in terms of the re-sampling methods, application to real data problems or to the interpretation of the results using cluster statistics.

Moreover, we saw that bootstrap or shuffling (”coin-flip”) may be used instead of permutation with some of the permutation methods presented in Section 1.2.2. Moreover, Langsrud (2005) shows that rotation is also a valid transformation under spherical distributions to produce a null distribution by re-sampling. When shuffling residuals, we preserve some heteroscedasticity but assume the first moment set to zero and the symmetry of the distribution. However, the difference between bootstrap and permutation is not obvious. When using permutation methods, the permuted residuals are not fully exchangeable, and we have no guarantee of an exact test (except for the `huh_jhun` method under sphericity (Kherad Pajouh and Renaud, 2010)). For the bootstrap, there is no exact property in finite sample size even for the simple cases without nuisance variables. To our knowledge, we don't actually have any proof for the best choice between permutation rather than bootstrap when re-sampling some sorts of residuals. ter Braak (1992) suggests that: ”Permutation may have some advantage here because there is maximum balance in each random permutation”. This argument suggests that permutation samples are more typical instances of the observed sample rather than bootstrap samples. Indeed, bootstrapping allows, with a very low probability, no variability in the re-samples. In more concrete cases, the problem was also noted by Salibian-Barrera and Zamar (2002) for the robust bootstrap as a small proportion of bootstrap samples may be composed fully of outliers, which implies some numerical instability.

Finally, we saw in Section 1.4.5 that cluster-based methods produce an inference at the cluster level. A time-point inference is then proscribed. EEG allows a precise temporal recording and it is disappointing that timing the brain activity is not advisable using cluster-based procedures. The precision of the tools is then lost in the data analysis although some smoothing of the raw data already biases the exact timing of the events. Answering the question, ”When does the effect occur ?”, may lead to errors using cluster-based procedures. However, the `troendle` procedure (presented in Chapter 1) does not have this problem and also controls the FWER. It is based on both a min-p procedure and a step-wise argument. Moreover, if the signals are smooth enough, the individual tests are sufficiently correlated to produce powerful tests using `troendle`. It may be a good solution to have a proper timing of the effects. Moreover, the cluster-mass test controls only the weak FWER which considers the special case when all hypotheses are under the null. When there are true effects on some parts of the signals, the cluster-mass test does not give a guarantee of the control of the FWER for the remaining null hypotheses. In that case, we should prefer multiple comparisons methods that control the strong FWER (like `troendle`) which considers all possible combinations of null and alternative hypotheses on the signal. We can hypothesize that there might be a relationship between the strong control of the FWER and the ability to interpret the timing of the effects but the question remains: is the strong control of the FWER a necessary condition to interpret the timing of the effects in EEG? Finally, it may be a fruitful topic to investigate: the use of the `troendle` method in combination of tests on the slopes (inspired by the works in Section 2.3) to improve the timing of the effects in EEG.

List of Tables

List of Figures

Appendix A

Supplementary Material for Chapter 1

A.1 Comparisons of Existing Packages

A.1.1 ANOVA and ANCOVA

```
R> install.packages("lmPerm")
R> install.packages("flip")
R> install.packages("GFD")
R> library("lmPerm")
R> library("flip")
R> library("GFD")

R> emergencycost$LOSc <- scale(emergencycost$LOS, scale = FALSE)
R> contrasts(emergencycost$sex) <- contr.sum
R> contrasts(emergencycost$insurance) <- contr.sum
R>
R> X <- model.matrix( ~ sex+insurance, data = emergencycost)[, -1]
R> colnames(X) <- c("sex_num", "insurance_num")
R> emergencycost <- data.frame(emergencycost,X)
R>
R> anova_permuco <- aovperm(cost ~ sex * insurance, data = emergencycost)
R> anova_GFD <- GFD(cost ~ sex * insurance, data = emergencycost,
R>                  CI.method = "perm", nperm  = 5000)
R>
R>
R> ancova_permuco <- aovperm(cost ~ LOSc * sex * insurance, data = emergencycost,
R>                           method = "huh_jhun")
R> ancova_flip <- flip(cost ~1, X = ~ sex_num, Z = ~ LOSc * insurance_num * sex_num
R>                    - sex_num, data = emergencycost, statTest = "ANOVA",
R>                    perms = 5000)
R> ancova_lmPerm <- aovp(cost ~ LOS * sex * insurance, data = emergencycost,
R>                      seqs = FALSE, nCycle = 1)

R> anova_permuco
```

```
Anova Table
Permutation test using freedman_lane to handle nuisance variables and
 5000 permutations.

                       SS  df      F parametric P(>F)
sex              60470803   1 0.7193           0.3975
insurance       598973609   1 7.1249           0.0083
sex:insurance   334349436   1 3.9771           0.0477
Residuals     14459666504 172
              permutation P(>F)
sex                      0.3978
insurance                0.0120
sex:insurance            0.0508
Residuals

R> anova_GFD

Call:
cost ~ sex * insurance

Wald-Type Statistic (WTS):
              Test statistic df    p-value p-value WTPS
sex                0.6397413  1 0.42380448       0.4662
insurance          6.3367469  1 0.01182616       0.0584
sex:insurance      3.5371972  1 0.06000678       0.0730

ANOVA-Type Statistic (ATS):
              Test statistic df1      df2   p-value
sex                0.6397413   1 5.743756 0.4556003
insurance          6.3367469   1 5.743756 0.0471947
sex:insurance      3.5371972   1 5.743756 0.1112178

R> ancova_permuco

Anova Table
Permutation test using huh_jhun to handle nuisance variables and
 5000, 5000, 5000, 5000, 5000, 5000, 5000 permutations.

                           SS  df        F parametric P(>F)
LOSc               2162110751   1 483.4422           0.0000
sex                  14630732   1   3.2714           0.0723
insurance              618366   1   0.1383           0.7105
LOSc:sex              8241073   1   1.8427           0.1765
LOSc:insurance       29107536   1   6.5084           0.0116
sex:insurance          123892   1   0.0277           0.8680
LOSc:sex:insurance   13457877   1   3.0091           0.0846
Residuals           751350616 168
                   permutation P(>F)
LOSc                          0.0002
```

```
sex                          0.0736
insurance                    0.7224
LOSc:sex                     0.1756
LOSc:insurance               0.0102
sex:insurance                0.8704
LOSc:sex:insurance           0.0820
Residuals

R> summary(ancova_lmPerm)

Component 1 :
                   Df    R Sum Sq  R Mean Sq Iter            Pr(Prob)
LOS                 1 2162110751 2162110751 5000 <0.0000000000000002
sex                 1   14630732   14630732 4159              0.0236
LOS:sex             1    8241073    8241073 1525              0.0616
insurance           1     618366     618366   94              0.5213
LOS:insurance       1   29107536   29107536 5000              0.0010
sex:insurance       1     123892     123892   80              0.5625
LOS:sex:insurance   1   13457877   13457877 2238              0.0429
Residuals         168  751350616    4472325

LOS               ***
sex               *
LOS:sex           .
insurance
LOS:insurance     ***
sex:insurance
LOS:sex:insurance *
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R> ancova_flip


     Test  Stat tail p-value
cost    F 3.271    >  0.0724
```

A.1.2 Repeated Measures ANOVA

```
R> jpah2016$id = as.factor(jpah2016$id)
R> jpah2016$bmic = scale(jpah2016$bmi,scale = FALSE)
R>
R> rancova_permuco <- aovperm(iapa ~ bmic * condition * time + Error(id/(time)),
R>                             data = jpah2016)
R> rancova_lmPerm <- aovp(iapa ~ bmic * condition * time + Error(id/(time)),
R>                           data = jpah2016, nCycle = 1, seqs = FALSE)

R> rancova_permuco
```

```
Permutation test using Rd_kheradPajouh_renaud to handle nuisance
variables and 5000 permutations.

                           SSn dfn      SSd dfd       MSEn      MSEd
bmic                   18.6817   1 106883.5  13    18.6817  8221.808
condition           27878.1976   2 106883.5  13 13939.0988  8221.808
bmic:condition      89238.4780   2 106883.5  13 44619.2390  8221.808
time                  268.8368   1 167304.9  13   268.8368 12869.607
bmic:time             366.4888   1 167304.9  13   366.4888 12869.607
condition:time      21159.7735   2 167304.9  13 10579.8867 12869.607
bmic:condition:time 29145.7201   2 167304.9  13 14572.8601 12869.607
                         F parametric P(>F) permutation P(>F)
bmic                0.0023           0.9627            0.9660
condition           1.6954           0.2217            0.2180
bmic:condition      5.4269           0.0193            0.0248
time                0.0209           0.8873            0.8856
bmic:time           0.0285           0.8686            0.8666
condition:time      0.8221           0.4611            0.4392
bmic:condition:time 1.1323           0.3521            0.3528
```

```
R> summary(rancova_lmPerm)


Error: id
Component 1 :
               Df R Sum Sq R Mean Sq Iter Pr(Prob)
bmic            1     3270      3270   51   0.8824
condition       2    20000     10000  801   0.3009
bmic:condition  2    89238     44619 4863   0.0255 *
Residuals      13   106884      8222
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1


Error: id:time
Component 1 :
                    Df R Sum Sq R Mean Sq Iter Pr(Prob)
time                 1     1047    1047.4   51   0.9412
bmic:time            1       31      31.5   51   0.8039
condition:time       2    29793   14896.4  320   0.3875
bmic:condition:time  2    29146   14572.9  419   0.3914
Residuals           13   167305   12869.6
```

Appendix B

Supplementary Material for Chapter 2

B.1 Analysis of Full-Scalp EEG Data

In this appendix you fill find R script to reproduce the full-scalp analysis of Cheval et al. (2018). We will need the following R packages (and the devtools and plyr package must also be installed):

```
R> library(edf)
R> library(abind)
R> library(xlsx)
R> library(igraph)
R> devtools::install_github("jaromilfrossard/clustergraph")
R> library(permuco)
R> library(clustergraph)
R> library(Matrix)
R> library(tidyr)
R> library(dplyr)
```

The EEG data are freely available in the ZENODO repository. We download it and decompress it using:

```
R> rep <- "https://zenodo.org/record/1169140/files/"
R> file <- "ERP_by_subject_by_condition_data.zip"
R> download.file(paste0(rep, file), "eeg.zip")
R> unzip("eeg.zip")
```

It creates a folder named raw_data which contains one ".edf" file for each experimental condition for each subject. These files contains signals that we need to download in R and store in a 3D array. The first dimension of this 3D array stores the signal with respect to the design (experimental conditions × participants), the second stores it with respect to the time and the third stores the data with respect to the electrodes. First, we create the design data.frame with the participants and the within-participants factors: the type of stimuli, and the task (here action).

```
R> design <- expand.grid(subject = c(111, 113, 115, 116, 117, 118, 120, 122,
R>                 123, 124, 126, 127, 128, 130, 131, 132, 134, 135, 137, 138,
R>                139, 140, 141, 142, 143, 144,  145, 146, 147),
R>                stimuli = c("AP", "SED"), action = c("Av_App", "Av_Ev"))
```

To produce the array containing the signals, we download in R for each row of the design the appropriate signals. Then, we take the difference with the neutral condition.

```
R> edf_filname <- list.files("raw_data")
R>
R> signal <- list()
R> for(i in 1:nrow(design)){
R>   # Select the experimental condition
R>   nid <- grepl(design$subject[i], edf_filname)
R>   nstim <- grepl(design$stimuli[i], edf_filname)
R>   naction <- grepl(design$action[i], edf_filname)
R>
R>   # Select the neutral VS Not neutral
R>   nneutr <- grepl("Rond", edf_filname)
R>   ntask <- grepl("Task", edf_filname)
R>
R>   # Download the data
R>   data_task <- read.edf(paste("raw_data/",
R>        edf_filname[nid&nstim&naction&ntask], sep = ""))
R>   data_neutr <- read.edf(paste("raw_data/",
R>        edf_filname[nid&naction&nneutr], sep = ""))
R>
R>   # Store the signals
R>   data_task <- data_task$signal[sort(names(data_task$signal))]
R>   data_task <- t(sapply(data_task, function(x)x$data))
R>   data_neutr <- data_neutr$signal[sort(names(data_neutr$signal))]
R>   data_neutr <- t(sapply(data_neutr, function(x)x$data))
R>
R>   # Store the signals relative to neutral
R>   signal[[i]] <- data_task - data_neutr
R> }
R>
R> # Create the 3D array
R> signal <- abind(signal, along = 3)
R> signal <- aperm(signal, c(3, 2, 1))
R>
R> # Select usefull time windows (from 0ms to 800ms, with a frequency of 512hz)
R> signal <- signal[,102:512,]
```

Then, we add the between-participants variable mvpa describing the usual physical activity of each participant. We use for the test the centred variable mvpac.

```
R> file <- "data_self_report_R_subset_zen.csv"
R> data_sr <- read.csv(paste0(rep,file),sep=";")
R> design <- left_join(design, data_sr, by = "subject")
R>
```

```
R> # Reshape the design
R> design$stimuli <- plyr::revalue(design$stimuli,
R>       c("AP" = "pa","SED" = "sed"))
R> design$action <- plyr::revalue(design$action,
R>       c("Av_App" = "appr","Av_Ev" = "avoid"))
R> design$mvpac <- as.numeric(scale(design$MVPA, scale = F))
```

Then, we create a matrix of adjacency using the theoretical position of the electrodes which are found on the website (www.biosemi.com).

```
R> # Download the Position of electrodes
R> download.file("https://www.biosemi.com/download/Cap_coords_all.xls",
R>        "Cap_coords_all.xls",mode="wb")
R> coord <-  read.xlsx(file="Cap_coords_all.xls", sheetIndex = 3,
R>        header =T,startRow=34)
R>
R> # Clean the coordinate data
R> coord <- coord[1:64,c(1,4:6)]
R> colnames(coord) <- c("electrode","x","y","z")
R>
R> coord$electrode <- plyr::revalue(coord$electrode, c("T7 (T3)" = "T7",
R>                  "Iz (inion)" = "Iz", "T8 (T4)" = "T8", "Afz" = "AFz"))
R> coord$electrode <-  as.character(coord$electrode)
```

From the position of the electrodes, we compute all pairs of Euclidean distances. After some trials, we find that 35mm is the minimal distance such that the graph of adjacency is connected.

```
R> # Create adjacency matrix
R> distance_matrix <- dist(coord[, 2:4])
R> adjacency_matrix <- as.matrix(distance_matrix) < 35
R> diag(adjacency_matrix) <- FALSE
R>
R> dimnames(adjacency_matrix) = list(coord[,1], coord[,1])
```

We convert the adjacency matrix into a `igraph` object and we add the position of electrodes as attributes for each vertex.

```
R> graph <- graph_from_adjacency_matrix(adjacency_matrix, mode = "undirected")
R> graph <- delete_vertices(graph,
R>           V(graph)[!get.vertex.attribute(graph, "name")%in%(coord[,1])])
R>
R> graph <- set_vertex_attr(graph,"x",
R>           value = coord[match(vertex_attr(graph,"name"),coord[,1]),2])
R> graph <- set_vertex_attr(graph,"y",
R>           value = coord[match(vertex_attr(graph,"name"),coord[,1]),3])
R> graph <- set_vertex_attr(graph,"z",
R>           value = coord[match(vertex_attr(graph,"name"),coord[,1]),4])
```

We set the parameters for the cluster-mass test (number of permutation, mass function, number of processors, the design, and the coding of factors.). The threshold is set by default at the 95 percentile of the F statistic.

```
R> np <- 4000
R> aggr_FUN <- sum
R> ncores <- 5
R> formula <- ~ mvpac*stimuli*action + Error(subject/(stimuli*action))
R> contr <- contr.sum
R>
R> pmat <- Pmat(np = np, n = nrow(design))
```

Then, we perform the 7 tests of the rANOVA. The following function computes a total of 7 effects $\times 4000$ permutations $\times 411$ time-points $\times 64$ electrodes $= 73651200$ F statistics using 5 processors. The computation takes several minutes:

```
R> model <- clustergraph_rnd(formula = formula, data = design,
R>          signal = signal, graph = graph, aggr_FUN = aggr_FUN,
R>          method = "Rd_kheradPajouh_renaud", contr = contr,
R>          return_distribution = F, threshold = NULL, ncores = ncores,
R>          P = pmat)
```

The results are plotted for the interaction `stimuli:action` (the 6th effect) using the `image` method. Printing the object `model` gives more information on the clusters for each effect.

```
R> image(model, effect = 6)
```

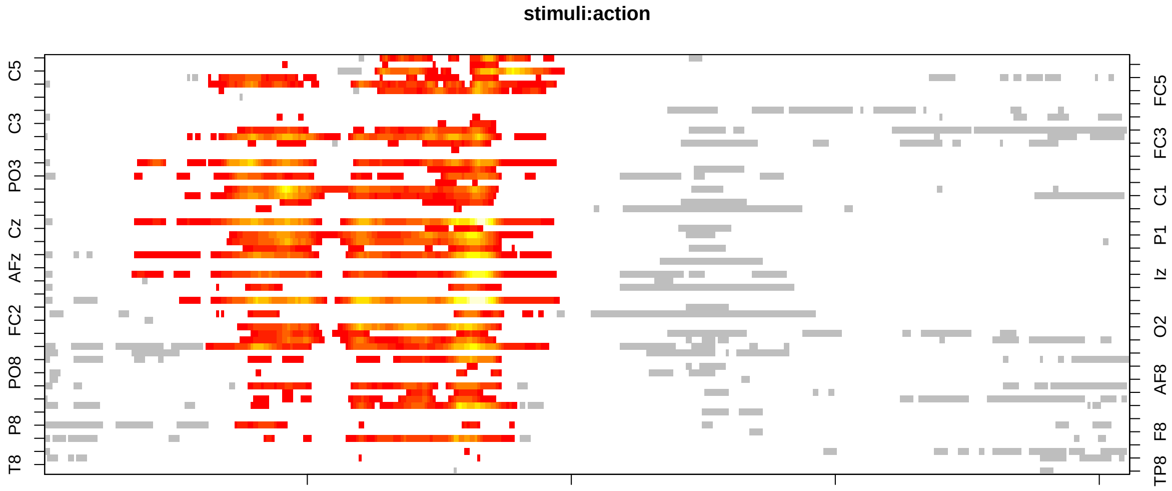

Figure B.1: Results of the full-scalp cluster-mass test for the interaction stimuli × action. In the X-axis are displayed the time measures (at 512Hz) and in the Y-axis are displayed the electrodes. The coloured cluster is a significant effect and the grey one are non significant (but above the threshold).

Appendix C

Supplementary Material for Chapter 3

C.1 Average over Transformation of a Square Matrix

C.1.1 Average over all Permutations

When permuting a square matrix, all diagonal elements are rearranged in the diagonal and all off-diagonal elements are rearranged in the upper and lower triangles. The average over all permutations of a square matrix A of size n is simply computed by averaging diagonal and off-diagonal elements separately:

$$\begin{aligned}
\frac{1}{n!}\sum_{P\in\mathcal{P}} P^\top A P &= I\frac{1}{n}\sum_{i=1}^{n} A_{i,i} + \left(\mathbf{1}\mathbf{1}^\top - I\right)\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} A_{i,j} \\
&= I\left[\left(\frac{1}{n}\sum_{i=1}^{n} A_{i,i}\right) - \left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} A_{i,j}\right)\right] + \mathbf{1}\mathbf{1}^\top\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} A_{i,j} \\
&= \frac{1}{n}\mathrm{tr}(A)I + \frac{1}{n(n-1)}\left(\mathbf{1}^\top A\mathbf{1} - \mathrm{tr}(A)\right)\left(\mathbf{1}\mathbf{1}^\top - I\right) \\
&= \frac{1}{n(n-1)}\left(n\mathrm{tr}(A) - \mathbf{1}^\top A\mathbf{1}\right)I + \frac{1}{n(n-1)}\left(\mathbf{1}^\top A\mathbf{1} - \mathrm{tr}(A)\right)\mathbf{1}\mathbf{1}^\top \\
&= \frac{n\mathrm{tr}(A) - \mathbf{1}^\top A\mathbf{1}}{n(n-1)}I - \frac{\mathrm{tr}(A) - \mathbf{1}^\top A\mathbf{1}}{n(n-1)}\mathbf{1}\mathbf{1}^\top.
\end{aligned} \tag{C.1}$$

For the special case when $A = yy^\top$:

$$\frac{1}{n!}\sum_{P\in\mathcal{P}} P^\top yy^\top P = \frac{ny^\top y - (\mathbf{1}^\top y)^2}{n(n-1)}I - \frac{y^\top y - (\mathbf{1}^\top y)^2}{n(n-1)}\mathbf{1}\mathbf{1}^\top. \tag{C.2}$$

For the special case when $\mathbf{1}^\top A_0\mathbf{1} = 0$ (e.g. for any projection matrices orthogonal to $\mathbf{1}$):

$$\frac{1}{n!}\sum_{P\in\mathcal{P}} P^\top A_0 P = \frac{1}{n-1}\mathrm{tr}(A_0)R_{\mathbf{1}}. \tag{C.3}$$

For the special case when $\mathbf{1}^\top A_1\mathbf{1} = n$ (e.g. for any projection matrices into a space containing $\mathbf{1}$):

$$\frac{1}{n!}\sum_{P\in\mathcal{P}} P^\top A_1 P = \frac{\mathrm{tr}(A_1) - 1}{(n-1)}I - \frac{\mathrm{tr}(A_1) - n}{(n-1)}H_{\mathbf{1}}. \tag{C.4}$$

C.1.2 Average over all Distinct Bootstrap Samples

Similarly to average over all permutations, the average over all bootstrap samples of the vector y is computed using the set of all "bootstrap matrices" $\mathcal{B} = \{B_1, \dots, B_{n^n}\}$:

$$\frac{1}{n^n} \sum_{B \in \mathcal{B}} By = H_{\mathbf{1}} y. \tag{C.5}$$

We conjecture that the average over all bootstrap samples of the matrix A is:

$$\begin{aligned} \frac{1}{n^n} \sum_{B \in \mathcal{B}} B^\top A B &= \left(\frac{1}{n} \mathrm{tr}(A) + \frac{1}{n^2} \left(\mathbf{1}^\top A \mathbf{1} - \mathrm{tr}(A) \right) \right) I + \frac{1}{n^2} \left(\mathbf{1}^\top A \mathbf{1} - \mathrm{tr}(A) \right) \left(\mathbf{1}\mathbf{1}^\top - I \right) \\ &= \frac{1}{n} \mathrm{tr}(A) I + \frac{1}{n} \left(\mathbf{1}^\top A \mathbf{1} - \mathrm{tr}(A) \right) H_{\mathbf{1}}. \end{aligned} \tag{C.6}$$

C.1.3 Average over all Distinct Shuffle Samples

Similarly to average over all permutations or bootstrap sample, the average over all shuffle samples of the vector y is computed using the set of all "shuffle matrices" $\mathcal{S} = \{S_1, \dots, S_{2^n}\}$.

$$\frac{1}{2^n} \sum_{S \in \mathcal{S}} Sy = 0. \tag{C.7}$$

When shuffling a square matrix, the diagonal elements stay the same and the off diagonal change sign depending on the shuffle, which leads to:

$$\frac{1}{2^n} \sum_{S \in \mathcal{S}} S^\top A S = \mathrm{diag}(A), \tag{C.8}$$

where $\mathrm{diag}(A)$ is a diagonal matrix with the elements of A.

C.2 Asymptotic of Maximum

Consider the random variable $Y_i \sim F_Y$ and its absolute value $|Y_i| \sim F_{|Y|}$. Then, the probability of the maximum of n independent instance of Y_i scaled by $\sqrt{n}$ to be greater that the value m is simply:

$$\begin{aligned} \Pr\left(\max_{i \in 1,\dots,n} \left| \frac{1}{\sqrt{n}} Y_i \right| > m \right) &= 1 - \Pr\left(\max_{i \in 1,\dots,n} \left| \frac{1}{\sqrt{n}} Y \right| < m \right) \\ &\overset{iid}{=} 1 - \left(\Pr\left(|Y_i| < \sqrt{n} m \right) \right)^n \\ &= 1 - \left(F_{|Y|}(n^{1/2} m) \right)^n. \end{aligned} \tag{C.9}$$

Assuming a polynomial decrease of the tail of the distribution such that $1 - F_{|Y|}(y) \sim y^{-p}$ then it density is $f_{|Y|}(y) \sim y^{-p-1}$ and we have:

$$\begin{aligned} 1 - \left(F_{|Y|}(n^{1/2} m) \right)^n &= 1 - \left(1 - (n^{1/2} m)^{-p} \right)^n \\ &= 1 - \left(1 - m^{-p} n^{-p/2} \right)^n \to 0 \text{ iff } p > 2, \end{aligned} \tag{C.10}$$

and, with $p = 2$, it converges to $1 - \exp m^{-2}$ and with $p < 2$ it converges to ∞. Moreover, a polynomial decrease of the tails imply also the second moment of Y_i such that:

$$\begin{aligned} \mathrm{E}\left[Y_i^2\right] &\leq C \int_0^\infty y^2 y^{-p-1} \mathrm{d}y \\ &= C \int_0^\infty y^{-p+1} \mathrm{d}y < \infty \text{ iff } p > 2. \end{aligned} \tag{C.11}$$

Together, Equation C.9 and C.11 implies $\mathrm{Var}\left[Y\right] < \infty \leftrightarrow \max_{i \in 1,\ldots,n} \left|\frac{1}{\sqrt{n}} Y_i\right| \to 0$.

C.3 Additional Asymptotic Results

Assuming a regression model as Equation 3.1 and using the convergences defined in Equations 3.20, 3.21 and 3.22, we find under the null hypothesis ($\beta = 0$):

$$\frac{1}{n-p} y^\top y = \frac{1}{n-p} \left(\eta^\top D^\top D\eta + 2e^\top D\eta + e^\top e\right) \to \mu_{D\eta}^2 + \sigma_{D\eta}^2 + \sigma_\epsilon^2, \tag{C.12}$$

as,

$$\frac{1}{n-p} \eta^\top D^\top D\eta = \frac{1}{n-p} \eta^\top H_{\mathbf{1}} D^\top D\eta + \frac{1}{n-p} \eta^\top R_{\mathbf{1}} D^\top D\eta \to \mu_{D\eta}^2 + \sigma_{D\eta}^2, \tag{C.13}$$

$$\frac{2}{n-p} e^\top D\eta \leq \frac{2}{n-p} e^\top \mathbf{1} \max\left(D\eta\right) \to 0, \tag{C.14}$$

$$\frac{1}{n-p} e^\top e \to \sigma_\epsilon^2. \tag{C.15}$$

Moreover, we also deduce:

$$\frac{1}{n} y^\top R_D y = \frac{1}{n} e^\top R_D e \to \sigma_\epsilon^2, \tag{C.16}$$

and

$$\frac{1}{n} y^\top H_D y = \frac{1}{n} y^\top y - \frac{1}{n} y^\top R_D y \to \mu_{D\eta}^2 + \sigma_{D\eta}^2. \tag{C.17}$$

C.4 Finite Sample Size and Asymptotic Results under Several Permutation Methods

C.4.1 Properties of the F Statistic with the `kennedy` Permutation Method

The `kennedy` permutation method is defined by the transformation $\{y, D, X\} \to \{PR_D y, -, R_D X\}$. The conditional distribution by permutation of the response under the `kennedy` method is then:

$$\Pr\left((Y_K^* | y) = PR_D Y\right) = \frac{1}{n!} \ \forall\ P \ \in\ \mathcal{P}, \tag{C.18}$$

where $\mathcal{P}$ is the set of all $n!$ permutation matrices of size $n \times n$. Its conditional expectation is defined as:

$$\mathrm{E}_{\mathcal{P}}\left[Y_K^* | y\right] = 0, \tag{C.19}$$

and its conditional variance as:

$$\mathrm{Var}_{\mathcal{P}}\left[Y_K^*|y\right] = \mathrm{E}_{\mathcal{P}}\left[Y_K^* Y_K^{*\ \top}|y\right] = \frac{1}{n-1} R_{\mathbf{1}} y^\top R_D y. \tag{C.20}$$

The permuted statistic is then:

$$F_{Y_K^*} = \frac{Y_K^{*\ \top} H_{R_D X} Y_K^* \ / \ (p-q)}{Y_K^{*\ \top} R_{R_D X} Y_K^* \ / \ (n-p)}. \tag{C.21}$$

Note that the projection matrix in the denominator is slightly different from Equation 3.3 due to the transformation of the design: the `kennedy` method transforms $[D\ X]$ into $[\ R_D X]$.

For a finite sample size, the conditional expectation of the numerator under all permutations is simply:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{p-q} Y_K^{*\top} H_{R_D X} Y_K^* | y\right] = \frac{1}{n-1} y^\top R_D y, \tag{C.22}$$

and for the denominator:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{n-p} Y_K^{*\top} R_{R_D X} Y_K^* | y\right] = \frac{1}{n-1} \frac{n-p+q}{n-p} y^\top R_D y. \tag{C.23}$$

For the `kennedy` method, the conditional expectation of the numerator and denominator of the F statistic are different. This suggests that we can correct the distribution of permuted F statistic using by the factor of $\frac{n-p+q}{n-p}$ to insure equal moments as proposed in Section 3.3.2. This difference is the results of the reduction of the rank of the design from p to $p-k$.

To prove asymptotic properties of Equation C.21 we first decompose its numerator:

$$\frac{1}{p-q} Y_K^{*\ \top} H_{R_D X} Y_K^* = \sum_{i=1}^{p-q} \left(\frac{1}{\sqrt{p-q}} Q_{R_D X:[i]}^\top Y_K^* \right)^2, \tag{C.24}$$

and the denominator:

$$\frac{1}{n-p} Y_K^{*\ \top} R_{R_D X} Y_K^* = \frac{1}{n-p} Y_K^{*\ \top} Y_K^* - \sum_{i=1}^{p} \left(\frac{1}{\sqrt{n-p}} (Q_{R_D X:[i]}^\top) Y_K^{*\ \top} \right)^2. \tag{C.25}$$

Moreover, assuming the conditions of the linear model defined in Equation 3.13, 3.14 and 3.15, we have the following asymptotic properties for the first part of the denominator:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{n-p} Y_K^{*\ \top} Y_K^* | y\right] = \frac{1}{n-p} y^\top R_D y \to \sigma_\epsilon^2, \tag{C.26}$$

and

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{n-p} Y_K^{*\ \top} Y_K^* | y\right] = 0. \tag{C.27}$$

For the second part, we compute:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}} (Q_{R_D X:[i]}^\top) Y_K^{*\ \top} | y\right] = 0, \tag{C.28}$$

and

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}}(Q^{\top}_{R_D X:[i]})Y_K^{*\ \top}|y\right] = \frac{1}{n-p}\frac{p-q}{n-1}y^\top R_D y \to 0. \tag{C.29}$$

Using the continuous mapping theorem, the denominator converges in probability to σ^2_ϵ.

For the numerator, we have the asymptotic conditional expectation:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{p-q}}(Q^{\top}_{R_D X:[i]})Y_K^{*\ \top}|y\right] = 0, \tag{C.30}$$

and the asymptotic variance:

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{p-q}(Q^{\top}_{R_D X:[i]})Y_K^{*\ \top}|y\right] = \frac{1}{p-q}\frac{p-q}{n-1}y^\top R_D y \to \sigma^2_\epsilon. \tag{C.31}$$

Then, it is easy to show that the numerator is equal in distribution to:

$$\frac{1}{p-q}(Q^{\top}_{R_D X:[i]})Y_K^{*\ \top}|y \overset{d}{=} \frac{1}{p-q}(Q^{*\top}_{R_D X:[i]})R_D y. \tag{C.32}$$

To prove asymptotic normality of Equation C.32, we need to verify similar conditions as Equations 3.33, 3.34, 3.35, 3.36 and 3.37. Only conditions involving the observations y will be different. This means that condition 3.33 becomes:

$$\max_{i\in 1,\dots,n}\left|\frac{1}{\sqrt{n}}\left(R_{\mathbf{1}}R_D y\right)_i\right| \to 0, \tag{C.33}$$

and condition 3.34 becomes:

$$\frac{1}{n}y^\top R_D y \to \sigma^2_\epsilon. \tag{C.34}$$

Finally, the conditions in Equations 3.35, 3.36 and 3.37 stay unchanged.

Moreover, using the convergence in Equation 3.39, we deduce the convergence in Equation C.33.

The condition in Equation C.34 is true as depends on the model assumptions.

Altogether, it shows that Theorem 1 holds when using the `kennedy` method.

C.4.2 Properties of the F Statistic with the `freedman_lane` Permutation Method

The `freedman_lane` permutation method is defined by the transformation $\{y, D, X\} \to \{(H_D + PR_D)y, D, X\}$. The conditional distribution by permutation of the response under the `freedman_lane` method is then:

$$\Pr\left((Y^*_{FL}|y) = (H_D + PR_D)y\right) = \frac{1}{n!}\ \forall\ P\ \in\ \mathcal{P}, \tag{C.35}$$

where $\mathcal{P}$ is the set of all $n!$ permutation matrices of size $n \times n$. Moreover, the conditional distribution of the `freedman_lane` method is a function of the conditional distribution of the `kennedy` method as:

$$Y^*_{FL}|y = \left(H_D Y + Y^*_K\right)|y. \tag{C.36}$$

Its conditional expectation is defined as:

$$\mathrm{E}_{\mathcal{P}}\left[Y^*_{FL}|y\right] = H_D y, \tag{C.37}$$

and its conditional variance as:

$$\mathrm{Var}_{\mathcal{P}}\left[Y^*_{FL}|y\right] = \mathrm{Var}_{\mathcal{P}}\left[Y^*_K Y^{*\ \top}_K|y\right] = \frac{1}{n-1} R_{\mathbf{1}} y^\top R_D y. \tag{C.38}$$

The permuted statistic under the `freedman_lane` permutation method is rewritten as:

$$F_{Y^*_{FL}} = \frac{Y^{*\top}_{FL} H_{R_D X} Y^*_{FL}\ /\ (p-q)}{Y^{*\top}_{FL} R_{D,X} Y^*_{FL}\ /\ (n-p)} = \frac{Y^{*\ \top}_K H_{R_D X} Y^*_K\ /\ (p-q)}{Y^{*\ \top}_K R_{D,X} Y^*_K\ /\ (n-p)}. \tag{C.39}$$

Only the denominator is different from the `kennedy` method (Equation C.21). For the numerator, the results of the `kennedy` method in Appendix C.4.1 hold using `freedman_lane`. Then, the conditional expectation of the denominator is simply:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{n-p} Y^{*\top}_{FL} R_{D,X} Y^*_{FL}|y\right] = \frac{1}{n-1} y^\top R_D y. \tag{C.40}$$

Moreover, we decompose the denominator using:

$$\frac{1}{n-p} Y^{*\ \top}_K R_{D,X} Y^*_K = \frac{1}{n-p} Y^{*\ \top}_K Y^*_K - \sum_{i=1}^{p} \left(\frac{1}{\sqrt{n-p}} (Q^\top_{D,X:[j]}) Y^{*\ \top}_K\right)^2. \tag{C.41}$$

Concerning the asymptotic results, in the denominator, for the first term ($\frac{1}{n-p} Y^{*\ \top}_K Y^*_K$), results from Equation C.26 and C.27 and implies its convergence in probability to σ^2_ϵ. Concerning the second term of the denominator, we find:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}} (Q^\top_{D,X:[j]}) Y^{*\ \top}_K|y\right] = 0, \tag{C.42}$$

and

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}} (Q^\top_{D,X:[j]}) Y^{*\ \top}_K|y\right] = \frac{1}{n-p} \frac{p-1}{n-1} y^\top R_D y \to 0. \tag{C.43}$$

Again, using the continuous mapping theorem, we find that the denominator converges in probability to σ^2_ϵ.

Finally, for the numerator, the results of the `kennedy` method in Appendix C.4.1 hold which implies Theorem 1 for the `freedman_lane` method.

C.4.3 Properties of the F Statistic with the `terBraak` Permutation Method

The `terBraak` method is the transformation $\{y, D, X\} \to \{(H_{D,X} + PR_{D,X})y, D, X\}$ and we compute the test statistic using another null hypothesis, $H_0 : \beta = \hat{\beta}|y = (X^\top R_D X) X^\top R_D Y|y$. This is equivalent than using the response $\left((H_{D,X} + PR_{D,X})Y - \hat{\beta}\right)|y$ and using the null hypothesis $H_0 : \beta = 0$. We need this second notation in this chapter to simplify the computations.

We then define the conditional distribution by permutation of the `terBraak` method:

$$\Pr\left((Y^*_{tB}|y) = \left(H_{R,D} - X(X^\top R_D X)^{-1} X^\top R_D + P R_{D,X}\right) y\right) = \frac{1}{n!} \ \forall\ P\ \in\ \mathcal{P}, \tag{C.44}$$

where $\mathcal{P}$ is the set of all $n!$ permutation matrices of size $n \times n$. Y^*_{tB} is defined using another conditional distribution as the `freedman_lane` method:
$Y^*_{tB}|y = \left(H_{R,D}Y - X(X^\top R_D X)^{-1} X^\top R_D Y + Y^*_{tB-}\right)|y$, where

$$\Pr\left((Y^*_{tB-}|y) = P R_{D,X} y\right) = \frac{1}{n!} \ \forall\ P\ \in\ \mathcal{P}. \tag{C.45}$$

The conditional expectations of $Y^*_{tB}|y$ and $Y^*_{tB-}|y$ over all permutation are:

$$\mathrm{E}_{\mathcal{P}}\left[Y^*_{tB}|y\right] = H_{R,D}Y - X(X^\top R_D X)^{-1} X^\top R_D Y, \tag{C.46}$$

and

$$\mathrm{E}_{\mathcal{P}}\left[Y^*_{tB-}|y\right] = 0. \tag{C.47}$$

Their conditional variances are:

$$\mathrm{Var}_{\mathcal{P}}\left[Y^*_{tB}|y\right] = \mathrm{Var}_{\mathcal{P}}\left[Y^*_{tB-}|y\right] = \frac{1}{n-1} R_{\mathbf{1}} y^\top R_{X,D} y. \tag{C.48}$$

The permuted statistic is written as a function of Y^*_{tB} or Y^*_{tB-}:

$$F_{Y^*_{tB}} = \frac{{Y^*_{tB}}^\top H_{R_D X} Y^*_{tB}\ /\ (p-q)}{{Y^*_{tB}}^\top R_{D,X} Y^*_{tB}\ /\ (n-p)} = \frac{{Y^*_{tB-}}^\top H_{R_D X} Y^*_{tB-}\ /\ (p-q)}{{Y^*_{tB-}}^\top R_{D,X} Y^*_{tB-}\ /\ (n-p)}. \tag{C.49}$$

The conditional expectation under all permutation of the numerator is:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{p-q} Y^{*\top}_{tB-} H_{R_D X} Y^*_{tB-}|y\right] = \frac{1}{n-1} y^\top R_{D,X} y. \tag{C.50}$$

For the denominator, we have:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{n-p} Y^{*\top}_{tB-} R_{D,X} Y^*_{tB-}|y\right] = \frac{1}{n-1} y^\top R_{D,X} y. \tag{C.51}$$

As the `freedman_lane` method, the `terBraak` method have the same conditional expectation of the denominator and numerator in small sample size.

To prove asymptotic properties of Equation C.49, we rewrite its numerator:

$$\frac{1}{p-q} {Y^*_{tB-}}^\top H_{R_D X} Y^*_{tB-} = \sum_{i=1}^{p-q} \left(\frac{1}{\sqrt{p-q}} Q^\top_{R_D X:[j]} Y^*_{tB-}\right)^2. \tag{C.52}$$

Similarly, for the denominator, we have:

$$\frac{1}{n-p} {Y^*_{tB-}}^\top R_{D,X} Y^*_{tB-} = \frac{1}{n-p} {Y^*_{tB-}}^\top Y^*_{tB-} - \sum_{i=1}^{p} \left(\frac{1}{\sqrt{n-p}} (Q^\top_{X,D:[j]}) {Y^*_{tB-}}^\top\right)^2. \tag{C.53}$$

Under the null hypothesis, the denominator converges in probability as, for the first part, its asymptotic conditional exception and variance are:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{n-p}Y^{*\top}_{tB-}Y^{*}_{tB-}|y\right] = \frac{1}{n-p}y^\top R_{D,X}y \to \sigma^2_\epsilon, \tag{C.54}$$

and,

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{n-p}Y^{*\top}_{tB-}Y^{*}_{tB-}|y\right] = 0. \tag{C.55}$$

For the second part of the denominator, we also find convergence in probability as its asymptotic conditional exception and variance are:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}}(Q^\top_{D,X:[j]})Y^{*\top}_{tB-}|y\right] = 0, \tag{C.56}$$

and,

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{\sqrt{n-p}}(Q^\top_{D,X:[j]})Y^{*\top}_{tB-}|y\right] = \frac{1}{n-p}\frac{p-1}{n-1}y^\top R_{D,X}y \to 0. \tag{C.57}$$

Once again, using the continuous mapping theorem, we deduce that the denominator converges in probability to σ^2_ϵ.

For the numerator, we compute its asymptotic expectation:

$$\mathrm{E}_{\mathcal{P}}\left[\frac{1}{\sqrt{p-q}}(Q^\top_{R_D X:[j]})Y^{*\top}_{tB-}|y\right] = 0, \tag{C.58}$$

and its asymptotic variance:

$$\mathrm{Var}_{\mathcal{P}}\left[\frac{1}{\sqrt{p-q}}(Q^\top_{R_D X:[j]})Y^{*\top}_{tB-}|y\right] = \frac{1}{(p-q)(n-1)}y^\top R_{D,X}y \to \frac{1}{p-q}\sigma^2_\epsilon. \tag{C.59}$$

Then, it is easy to show that the terms in the numerator are equal in distribution to:

$$\frac{1}{p-q}(Q^\top_{R_D X:[i]})Y^{*\top}_{-tB}|y \overset{d}{=} \frac{1}{p-q}(Q^{*\top}_{R_D X:[i]})R_{D,X}y. \tag{C.60}$$

To prove asymptotic normality of the term in Equation C.60, we need to verify similar conditions as Equations 3.33, 3.34, 3.35, 3.36 and 3.37. Only conditions involving the observation y are different which means that condition in Equation 3.33 becomes:

$$\max_{i\in 1,\dots,n}\left|\frac{1}{\sqrt{n}}\left(R_{\mathbf{1}}R_{D,X}y\right)_i\right| \to 0. \tag{C.61}$$

The condition in Equation 3.34 becomes:

$$\frac{1}{n}y^\top R_{D,X}y \to \sigma^2_\epsilon. \tag{C.62}$$

Finally, the conditions in Equations 3.35, 3.36 and 3.37 stay unchanged.

Assuming a factorial design, the convergence in Equation C.61 is proven using the results in Appendix C.2 within each cell.

The condition in Equation C.62 is true as depends on the model assumptions.

Altogether, it shows that Theorem 1 holds when using the `terBraak` method.

Appendix D

Supplementary Material for Chapter 4

D.1 Correlation Structure of the rANOVA

In the earlier times when the repeated measures ANOVA model was first discussed, an abundant literature on the choice of the model and especially on the choice of the correlation structure was written. At that time, the necessity to compute everything by hand gave constraints on the analysis and all the available test statistics were based on sum of squares. Under specific assumptions, these statistics possess an exact F distribution. Let's first mention that Huynh and Feldt (1970) showed that the sphericity (or circularity) of the covariance structure was a necessary and sufficient condition for an exact test. Box (1954) and Huynh and Feldt (1976) proposed modifications in the degrees of freedom – called ϵ-correction – when this condition is not fulfilled. Note that even earlier, a discussion was engaged whether to include in the correlation structure the interactions between the participant and the fixed effects ($(\pi\psi)_{ik}$ in Equation 4.1), which are now called the random slopes. Rouanet and Lepine (1970), for example, compare the models with and without them. Ultimately, the model that includes all random slopes became the reference and statistical software, like SPSS or Statistica, use it as if it was the only possible random structure.

On a more technical side, for the fixed effects, some constraints have to be chosen since the model is otherwise overparametrized and no estimation or test can be obtained (Cardinal and Aitken, 2013). In order to keep the interpretation of main effects as in the ANOVA tradition, we use the sigma-restricted parametrizations, which corresponds in Equation 4.1 to the following constraints: $\sum_j \alpha_j = 0$, $\sum_j (\alpha\psi)_{jk} = 0 \; \forall k$ and $\sum_k (\alpha\psi)_{jk} = 0 \; \forall j$, and so on. Concerning random slopes, which are technically interactions between a fixed and a random unit, it is worth asking if this type of constraints should also be also applied. In a very influential paper, Cornfield and Tukey (1956) , with a construction called "pigeonhole" that includes both fixed and random factors as special cases, show that the distributional behaviour for factorial designs lead to constraints only for one margin: $\sum_k (\pi\psi)_{ik} = 0 \; \forall i$, see e.g. Montgomery (2017). If these constraints are not included e.g. when simulating data, some variances will be inflated.

The discussion in the previous paragraphs shows that there was a debate over several decades on the best model for rANOVA, both for the fixed part and for the correlation structure. It is therefore not surprising that concerning the more recently proposed CREMEM, a similar debate is ongoing. It is noteworthy that it concerns exactly the same questions on the best correlation structure and the correlation structures for mixed effect models that are presented in Section 4.5 take root in the above-mentioned literature on rANOVA.

D.2 The Database Representation of the 5 Types of Variables

As a complement to Section 4.4, we propose here a generalisation of the “wide” format for the 5 types of variables. In the rANOVA framework, software like SPSS or Statistica use data in the “wide” format. This format makes explicit the difference between within-participant variables (V_P) and between-participant variables (V_M), see Figure D.1. Each line of the data represents one participant and the “data” column are split into between-participant variables, and columns which values are the response recorded in one level of the within-participant variables. So, the responses are stored in a two-dimensional (2D) array which entries, by rows, correspond to the participants and by column to the within-participants levels. This 2D array is the results of the crossing of one table storing the participant and V_P variables and one table storing the experimental manipulations and the V_M variables.

The “wide” format shows clearly the fundamental difference between within-participant and between-participant variables. This representation can be extended to CRE-MEM. In that setting, we cross 3 tables (instead of 2): one for the participants and the variables V_P (Table P), one for the stimuli and the variables V_S (Table S) and one for the experimental manipulations and the variables V_M (Table M). The crossing of those 3 tables creates a 3D array (like a building, of dimension $[n_P,\ n_S,\ n_M]$) in which the responses can be stored. This construction is represented in the Figure D.2. In this example, each participant is represented by one row of the Table P and each stimulus is represented by one row (here rotated by 90°) of the Table S. Then, the responses of one participant are a stored in one “floor” (of dimension $[1,\ n_S,\ n_M]$), the responses of one stimulus are stored in one “vertical slide” from forefront to back (of dimension $[n_P,\ 1,\ n_M]$) and the responses in one experimental manipulation is one “slice” (of dimension $[n_P,\ n_S,\ 1]$) of the 3D array. The responses associated to one pair participant-stimulus is consequently a “pile” (of dimension $[1,\ 1,\ n_M]$) of the 3D array. Note that to simplify the representation the variables V_{PS} and V_O are not represented in Figure D.2. However, they could be associated both to their own table (Table PS and Table O) and each entry of the Table PS would be associated with one “pile” of the 3D array. Finally, each entry of the Table O would be associated with one cell of the 3D array.

This representation is a tool to understand which interactions between fixed effects and random units are allowed in a model. Each “floor” (the responses by participants) crosses all levels of the V_S and V_M variables, which implies that participants are measured in all levels of V_S and V_M. Moreover each “floor” is composed of multiple “piles” and multiple “cells” which means that a participant will be measured in multiple levels of the variables V_{PS} and V_O. All the random interactions participant:V_S, participant:V_M, participant:V_{PS}, participant:V_O are then allowed for CRE-MEM. The same rational is applied for stimuli: each stimulus is represented in by one “vertical slide” and will be measured in all levels of the V_P and V_M, then the random interaction stimuli:V_P, stimuli:V_M, stimuli:V_{PS}, stimuli:V_O are feasible. To understand which random slopes associated to the interaction participants:stimuli can be included in the model we apply the same strategy for the “pile” and each interaction crosses all levels of the variables V_M. And each “pile”” is composed of multiple cells (which are associated to variables V_O). Which means that the random interactions between participants:stimuli:V_M and participants:stimuli:V_O are feasible. These findings are summarized in Table 4.1.

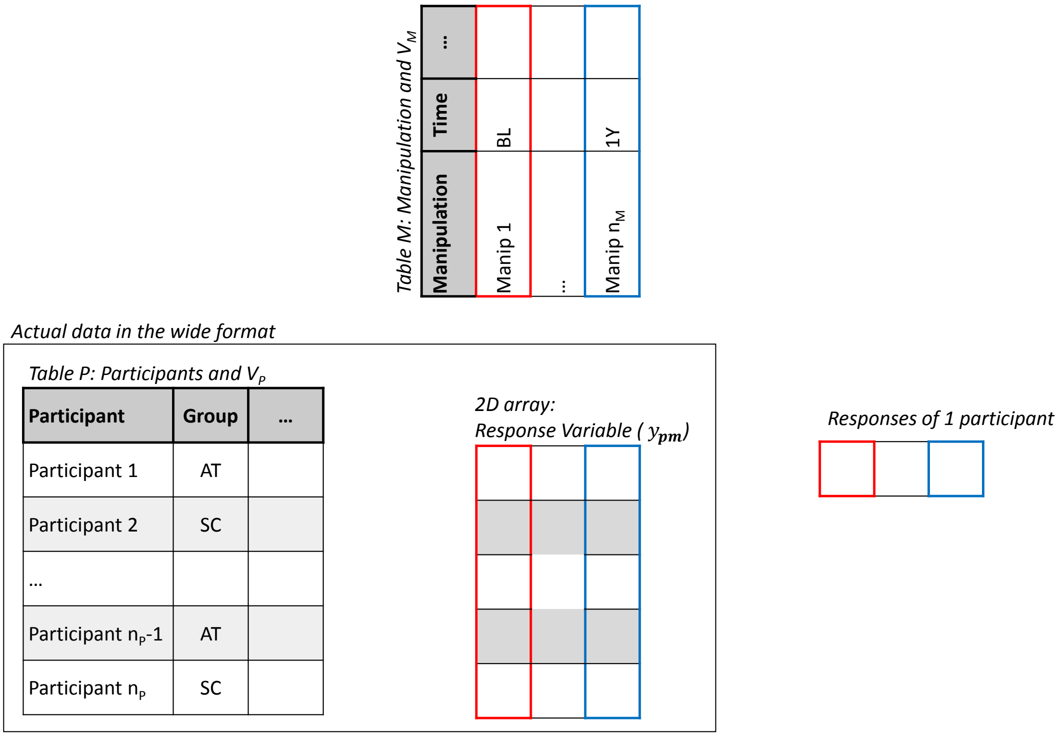
Table M: Manipulation and V_M
Actual data in the wide format
Table P: Participants and V_P
2D array:
Response Variable (y_{pm})
Responses of 1 participant

Manipulation	Time	...
Manip 1	BL	
...		
Manip n_M	1Y	

Participant	Group	...
Participant 1	AT	
Participant 2	SC	
...		
Participant n_P-1	AT	
Participant n_P	SC	

Figure D.1: Representation of the variables and the responses of a rANOVA in the ”wide” format. The responses are stored in a 2D array that stem from the crossing of 2 tables. Table P stores the participants and their features V_P (or between-participant variables) and Table M stores the features of the experimental manipulations V_M (or within-participant variables).

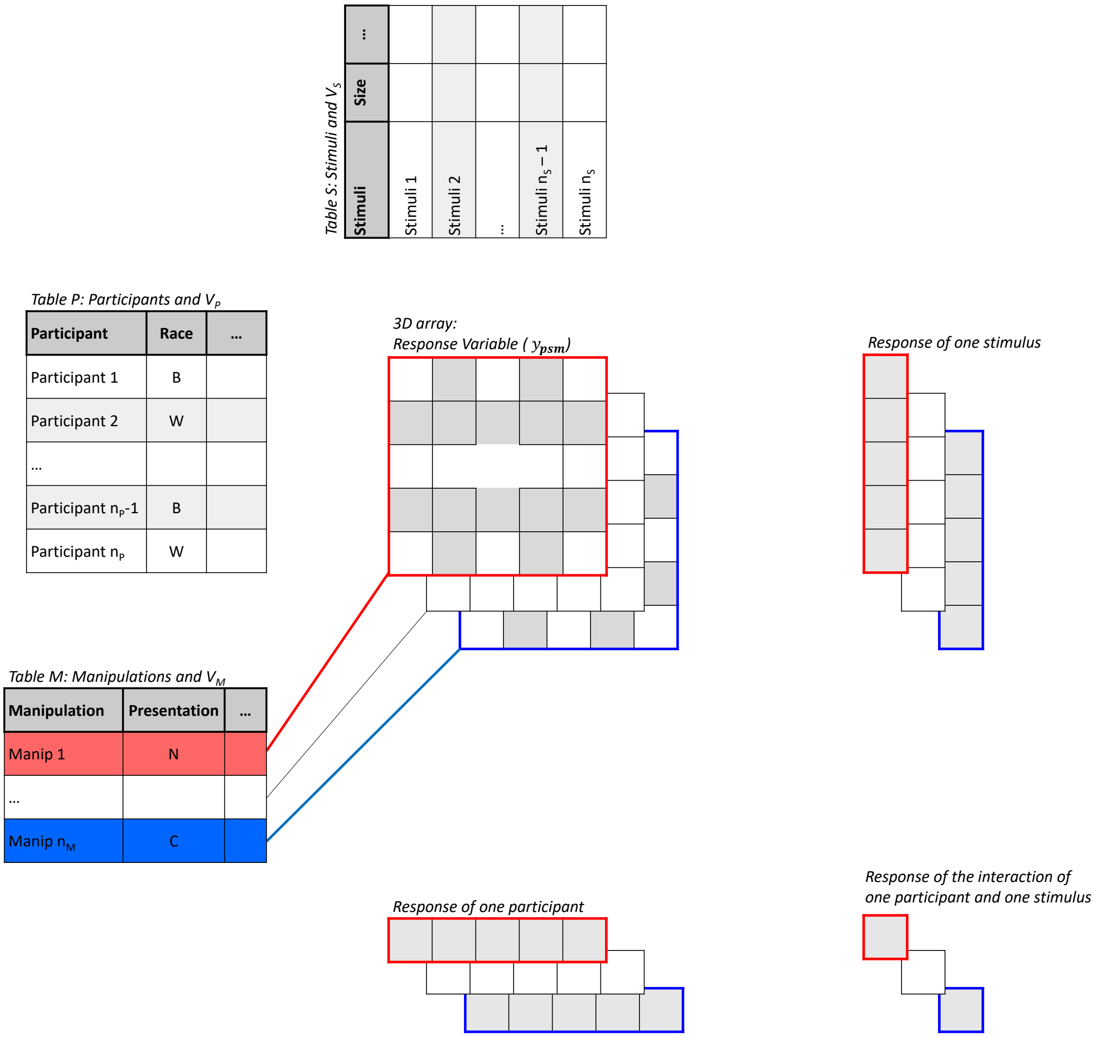

Table S: Stimuli and V_S

Stimuli	Size	…
Stimuli 1		
Stimuli 2		
…		
Stimuli $n_S - 1$		
Stimuli n_S		

Table P: Participants and V_P

Participant	Race	…
Participant 1	B	
Participant 2	W	
…		
Participant n_P-1	B	
Participant n_P	W	

Table M: Manipulations and V_M

Manipulation	Presentation	…
Manip 1	N	
…		
Manip n_M	C	

3D array:
Response Variable (y_{psm})
Response of one stimulus
Response of one participant
Response of the interaction of one participant and one stimulus

Figure D.2: Representation of the variables and the responses of CRE-MEM. The responses are stored in a 3D array that is the result of the crossing of Table P, Table S and Table M. Participants are associated with the Table P and the levels of variable V_P will be identical for each "floor". Stimuli are associated with the Table S and variables V_S will takes the same value for each "vertical slide". And the experimental manipulations are associated with the Table M and each "slice" of the 3D array.

D.3 Matrix Formulation of the CRE-MEM

D.3.1 General Notation for Mixed Models

Following Bates et al. (2015), we define the mixed linear model:

$$y = X\beta + Z\gamma + \epsilon \tag{D.1}$$

where y is the response, the fixed part of the design is X and the random part is Z. The fixed parameters are β, the random effects are $\gamma \sim (0, \Sigma)$ and the error terms are $\epsilon \sim$

$(0, \sigma^2 I)$. For a CRE-MEM, we split the random effects into G independent components $\gamma = \left[\gamma_1^\top | \dots | \gamma_g^\top | \dots | \gamma_G^\top\right]^\top$ and their associated design matrices $Z = [Z_1 | \dots | Z_g | \dots | Z_G]$ which led to the decomposition of the covariance matrix of the random effects into $\Sigma = \text{diag}(\Sigma_1, \dots, \Sigma_g, \dots, \Sigma_G)$. The covariance matrix of the response variable y can then be written as $\Omega = Z\Sigma Z^\top + I\sigma^2 = Z_1\Sigma_1 Z_1^\top + \dots + Z_g\Sigma_g Z_g^\top + \dots + Z_G\Sigma_G Z_G^\top + I\sigma^2$.

D.3.2 Generalized ANOVA and RI-L for the CRE-MEM

The gANOVA is written following the Equation D.1 with specific constraints on the random effects. Independence between random components is specified using $\gamma = \left[\gamma_P^\top | \gamma_S^\top | \gamma_{P:S}^\top\right]^\top$ and by defining a covariance matrix of observations split into the parts relative to participants, stimuli and their interactions:

$$\Omega = Z\Sigma Z^\top + I\sigma^2 = Z_P\Sigma_P Z_P^\top + Z_S\Sigma_S Z_S^\top + Z_{P:S}\Sigma_{P:S} Z_{P:S}^\top + I\sigma^2, \tag{D.2}$$

where $Z_P = \left(X_{parti}^\top * [\mathbf{1}|X_S|X_{PS}|X_M|X_O]^\top\right)^\top$, $Z_S = \left(X_{stimulus}^\top * [\mathbf{1}|X_P|X_{PS}|X_M|X_O]^\top\right)^\top$, $Z_{P:S} = \left((X_{parti}^\top * X_{stimulus}^\top) * [\mathbf{1}|X_M|X_O]^\top\right)^\top$ and the Khatri-Rao product (Khatri and Rao, 1968) is written using $*$. X_{parti} and $X_{stimulus}$ are matrices dummy coding xxoli est ce bien des 0/1 for the participants and stimuli, and Σ_P, Σ_S, $\Sigma_{P:S}$ are covariance matrices.

If the dataset has no missing value and after the appropriate permutation (P_P,P_S and P_{PS}), the covariance matrices are written as block diagonal matrices of the individual covariance structure: $\Sigma_P = P_P(I_P \otimes \Sigma_P^0)P_P^\top$, $\Sigma_S = P_S(I_S \otimes \Sigma_S^0)P_S^\top$ and $\Sigma_{P:S} = P_{P:S}(I_{P:S} \otimes \Sigma_{P:S}^0)P_{P:S}^\top$. gANOVA assumes that the matrices Σ_P^0, Σ_S^0 and Σ_{PS}^0 are diagonal matrices with the same value for the random effects (i.e. contrasts) associated to the same factor.

For gANOVA, the $X_.$ matrices X_S, X_P, X_{PS}, X_M and X_O are written using orthonormal contrasts $C_.$ (e.g.: `contr.poly`) and overparametrized dummy-coded design matrices $X_.^0$ such that $X_. = X_.^0 C_.^-$ where $C_.^-$ is the generalized inverse of $C_.$. If $C_.$ is orthonormal gives the properties $C_.C_.^- = I$ and $C_.^-C_.^{-\top} = I - a\mathbf{1}\mathbf{1}^\top$ where a is a positive value that depends on the dimension of $C_.$.

Concerning the RI-L model, the only difference with gANOVA is that there is no constraints on the random effects. Its matrix formulation is therefore the same as gANOVA except that the contrasts $C_.$ are not used (or are replaced by an identity matrix I). Note that the $X_.^0$ matrices are used to construct the fixed part of the design and are usually associated with contrasts. Depending on the hypothesis, these contrasts may be represented by non-orthonormal matrices (e.g.: using `contr.sum`).

D.4 Comparison of the gANOVA and RI-L Model

In this appendix, we provide evidence that gANOVA has a better decomposition of the correlation structure than RI-L. For the sake of the argument, we focus the comparison to a model with only one sampling unit (the participants), balanced design and replications (one variable V_M), which means that there will be 3 variances parameters: one for random intercepts, one for the random slopes and one for the residuals. The replications will make the variance of the random slopes estimable. The gANOVA and RI-L correlation structure differ by the constraints on the random effects and those constraints are represented by contrast matrices. We will call constraint (c) the design of gANOVA and unconstraint (uc) the one of RI-L.

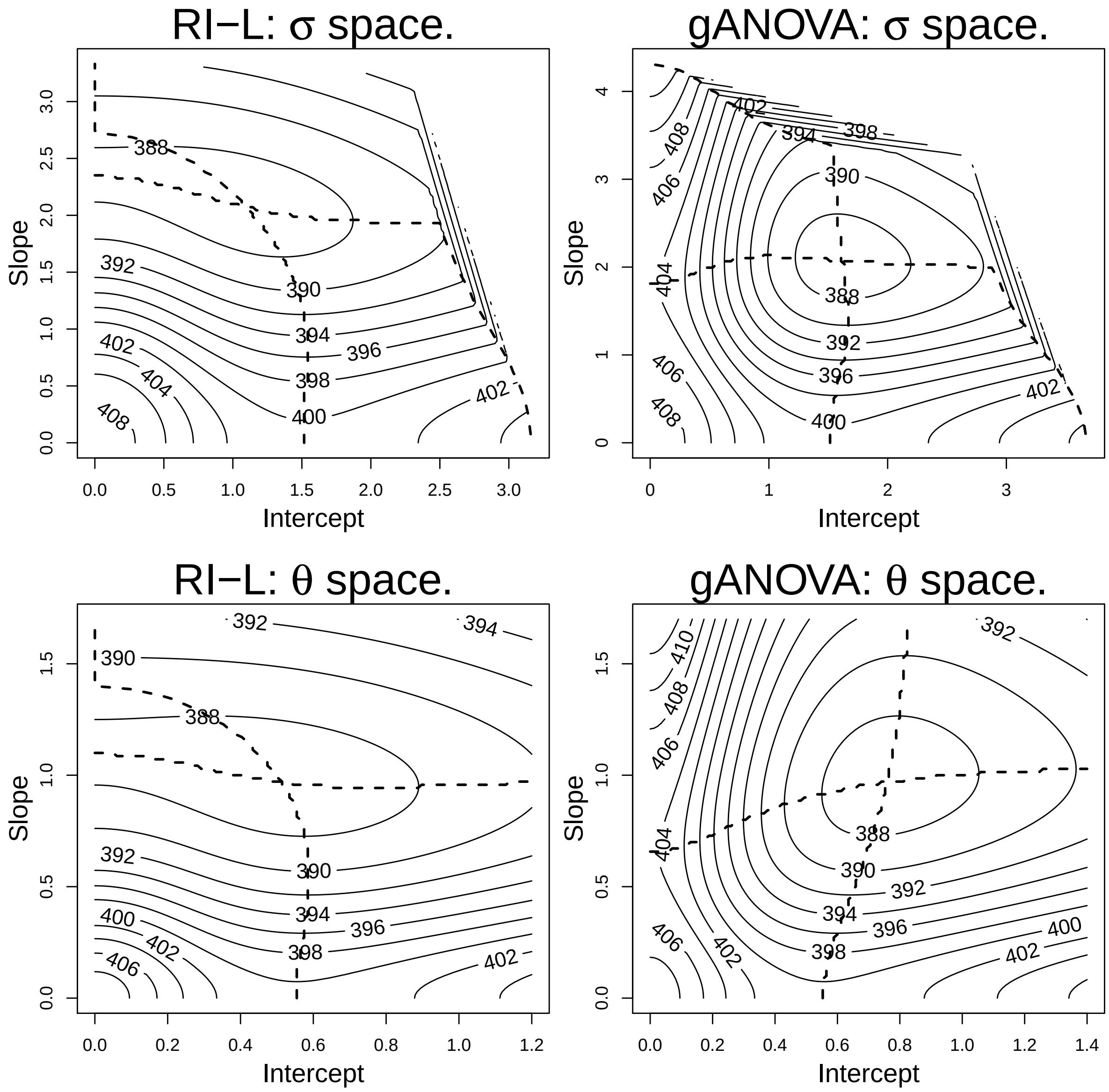

Figure D.3: Likelihood of RI-L and gANOVA for a model with one sampling units and 1 variable V_M with replications and assuming random intercepts and random slopes. The two top figures represent the likelihood in the σ parameters space (standard deviation of random effects), and two bottom one represents the θ parameters space (space of optimization parameters; see Equation (4) in Bates et al. (2015) for an exact definition of the θ's.). The dotted lines are the two ridges defining the profile likelihoods. We see in this example that the gANOVA tends to orthogonalize the profile likelihoods as they cross almost at a 90° angle, which suggest less dependency of the two parameters and an easier optimization process.

In that setting, the RI-L model is written:

$$y = X\beta + Z_{uc}\gamma_{uc} + \epsilon,$$

where $\gamma_{uc} \sim (0, I_{N_P} \otimes \Sigma_{uc})$, $\epsilon = (0, I\sigma^2_{uc;\epsilon})$ for N_P the number of participants. Σ_{uc} is a diagonal covariance matrix of dimension $N_M + 1$: $\Sigma_{uc} = \text{diag}(\sigma^2_{uc;i}, I_{N_M}\sigma^2_{uc;T})$.

The design matrix of the random effects is written $Z_{uc} = \left(Z^\top * \left[\mathbf{1} \quad X_{uc}\right]^\top\right)^\top$ and $*$ denotes the column-wise Khatri-Rao product (Khatri and Rao, 1968). Assuming a balanced design, we write:

$$\begin{aligned} Z_{uc} &= \left((\mathbf{1}_{N_P} \otimes I_{N_M+1})^\top * \left[\mathbf{1} \quad \mathbf{1}_{N_P} \otimes X_{uc;p}\right]^\top\right)^\top \\ &= \left(I_{N_P} \otimes \left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]\right), \end{aligned}$$

where $X_{uc;p}$ is a $N_M N_R \times N_M$ matrix representing the overparametrized design of one participant for a V_M factor of N_M levels and assuming N_R replications in each cell. The covariance matrix of the response y is:

$$\begin{aligned} Z_{uc}(I_{N_P} \otimes \Sigma_{uc})Z_{uc}^\top + I\sigma^2_{uc;\epsilon} &= \left(I_{N_P} \otimes \left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]\right)(I_{N_P} \otimes \Sigma_{uc})\left(I_{N_P} \otimes \left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]^\top\right) + I\sigma^2_{uc;\epsilon} \\ &= \left(I_{N_P} \otimes \left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]\right)\left(I_{N_P} \otimes \Sigma_{uc}\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]^\top\right) + I\sigma^2_{uc;\epsilon} \\ &= I_{N_P} \otimes (\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]\Sigma_{uc}\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]^\top) + I\sigma^2_{uc;\epsilon} \\ &= I_{N_P} \otimes (\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]\Sigma_{uc}\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]^\top + I_{N_M}\sigma^2_{uc;\epsilon}). \end{aligned}$$

The covariance matrix of the response is a block diagonal matrix with block-elements of the form:

$$\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]\Sigma_{uc}\left[\mathbf{1}_{N_M} \quad X_{uc;p}\right]^\top + I_{N_M}\sigma^2_{uc;\epsilon} = \mathbf{1}_{N_M}\mathbf{1}_{N_M}^\top\sigma^2_{uc;i} + X_{uc;p}X_{uc;p}^\top\sigma_{uc;F} + I_{N_M}\sigma^2_{uc;\epsilon}.$$

Similarly, gANOVA is written:

$$y = X\beta + Z_c\gamma_c + \epsilon,$$

where $\gamma_c \sim (0, I_{N_P} \otimes \Sigma_c)$, $\epsilon = (0, I\sigma^2_{c;\epsilon})$. Σ_c is a diagonal matrix covariance matrix of dimension N_M : $\Sigma_c = \text{diag}(\sigma^2_{c;i}, I_{N_M-1}\sigma^2_{c;T})$.

The design matrix of the random effects is written using the orthonormal contrast C: $Z_c = \left(Z^\top * \left[\mathbf{1} \quad X_{uc}C^-\right]^\top\right)^\top$ and assuming that it is balanced, it becomes: $Z_c = \left(I_{N_P} \otimes \left[\mathbf{1} \quad X_{uc;p}C^-\right]\right)$. The covariance matrix of the response is then:

$$\begin{aligned} Z_c(I_{N_P} \otimes \Sigma_c)Z_c^\top + I\sigma^2_{c;\epsilon} &= I_{N_P} \otimes \left(\left[\mathbf{1}_{N_M} \quad X_{c;p}\right]\Sigma_c\left[\mathbf{1}_{N_M} \quad X_{c;p}\right]^\top + I_{N_M}\sigma^2_{c;\epsilon}\right) \\ &= I_{N_P} \otimes \left(\left[\mathbf{1}_{N_M} \quad X_{uc;p}C^-\right]\Sigma_c\left[\mathbf{1}_{N_M} \quad X_{uc;p}C^-\right]^\top + I_{N_M}\sigma^2_{c;\epsilon}\right). \end{aligned}$$

Using the properties of the orthonormal contrasts, the block-elements of the covariance matrix simplify:

$$\begin{aligned} \left[\mathbf{1}_{N_M} \quad X_{uc;p}C^-\right]\Sigma_c\left[\mathbf{1}_{N_M} \quad X_{uc;p}C^-\right]^\top + I_{N_M}\sigma^2_{c;\epsilon} &= \mathbf{1}_{N_M}\mathbf{1}_{N_M}^\top\ \sigma^2_{c;i} + X_{uc;p}C^-C^{-\top}X_{uc;p}^\top\sigma_{c;F} + I\sigma^2_{c;\epsilon} \\ &= \mathbf{1}_{N_M}\mathbf{1}_{N_M}^\top\sigma^2_{c;i} + X_{uc;p}(I - \mathbf{1}\mathbf{1}^\top a)X_{uc;p}^\top\sigma_{c;F} + I\sigma^2_{c;\epsilon} \\ &= \mathbf{1}_{N_M}\mathbf{1}_{N_M}^\top(\sigma^2_{c;i} - a\sigma^2_{c;F}) + X_{uc;p}X_{uc;p}^\top\sigma_{c;F} + I\sigma^2_{c;\epsilon}, \end{aligned}$$

with positive value a defined by the number of levels of V_M: $a = 1-1/N_M$. The covariance matrices of the 2 models are equal if and only if:

$$\begin{aligned}\sigma^2_{uc;i} &= \sigma^2_{c;i} - a\sigma^2_{c;F}\\ \sigma^2_{uc;F} &= \sigma^2_{c;F}\\ \sigma^2_{uc;\epsilon} &= \sigma^2_{c;\epsilon}.\end{aligned}$$

These equalities show us that the 2 models are equal for values of the variances of random effects. But, the first equality tells us that the 2 models are not equal when $\sigma^2_{c;i} < a\sigma^2_{c;F}$. Which means that adding the constraint of the contrasts C increases the possible covariance matrices of the observations. With more factors in the model, higher level interactions put similar conditions on lower level interactions. The RI-L model will produce more estimates equal to 0 with maximal values of the optimal function at the boundary of the parameter space. And, when RI-L and gANOVA are not equal, gANOVA will always have a better likelihood which suggests a better fit of the model.

D.5 Examples of lme4 Formulas for CRE-MEM with Different Correlation Structures

In this appendix, some examples of `R` formula from simple to more complex designs (see Table 4.2 for the selected designs). The type of variable is explained in the Section 4.4. We use the notation `PT` and `SM` for the identifier variables of the participants and stimuli respectively and `y` for the response variable. All those variables and the design will be variables of the `mydata dataframe`. To interpret correctly main effects as in the ANOVA framework, it is extremely important to use contrasts that sum to zero (like contr.sum or contr.poly) and not the contr.treat default in `R`). The fixed part will be assumed to be a full factorial design in each case. For saturated design, we drop the last interaction term because we assume no replication of the same observations (multiple observations associated to the same cell in the design) and the last interaction terms would not be estimable because it is confounded with the error terms.

Using factors in the formula will not produce always the model that the users expect. `R` assigns to factors the maximum degree of freedom (some sort of contrasts) available which means that, for the factors `A` and `B` with 2 levels, using the formula `~A*B` and `~A:B` will assign respectively 1 and 3 degrees of freedom for the interaction `A:B`; one other example is `~A` and `~0 + A` which assign 1 and 2 degrees of freedom for the effect of `A`. In `lme4`, this feature happens when `R` compute separately the effects; for instance, `~(1|PT) + (A|PT)` will assign 2 variances (+ 1 covariance) for the levels of A and `~(A*B||PT)` will assign 4 variances (+6 covariances) for the interaction `A:B`. In order to produce the correlation structures described in the Section 4.5, the solution is to convert factors into numeric variables, we use the `model.matrix()` function for that purpose.

D.5.1 A Simple Case, Variables $V_P(2)$, $V_S(2)$ and $V_M(2)$

This design corresponds to model M1 in Table 4.3. In that simple case, the saturated correlation structure will not have the last term of interaction between the variable V_M and the random units `participant:stimulus`. With only 2 levels per factor, the “sum” coding will produce similar results to the polynomial coding variable.

RI and RI+

```
R> lmer(y ~ Vp*Vs*Vm + (1|PT) + (1|SM) + (1|PT:SM), data = mydata)
```

RI-L and RI-L+

```
R> lmer(y ~ Vp*Vs*Vm + (1|PT) + (1|PT:Vs) + (1|PT:Vm) + (1|PT:Vs:Vm)
R>   + (1|SM) + (1|SM:Vp) + (1|SM:Vm) + (1|SM:Vp:Vm)
R>   + (1|PT:SM), data = mydata)
```

MAX and MAX+

```
R> lmer(y ~ Vp*Vs*Vm + (Vs*Vm|PT) + (Vp*Vm|SM) + (1|PT:SM), data = mydata)
```

ZCP and ZCP+

```
R> mydata$Xs <- model.matrix(~ Vs, data = mydata)[, -1]
R> mydata$Xp <- model.matrix(~ Vp, data = mydata)[, -1]
R> mydata$Xm <- model.matrix(~ Vm, data = mydata)[, -1]
R> lmer(y ~ Vp*Vs*Vm + (Xs*Xm||PT) + (Xp*Xm||SM) + (1|PT:SM), data = mydata)
```

gANOVA and gANOVA+

The gANOVA package use the same notation as the RI-L formula (`+ (1|PT) + (1|PT:f1) + (1|PT:f2) + (1|PT:f1:f2)`) and puts orthonormal coding from the second factors of the right part of the random effect. Doing so, the order of the factors matters and the first variable to the right of the bar must be the random unit. Note that it implies that the interaction participant-stimulus should be written as only one variable. Moreover, the multiple terms in the notation can be reduced to the formula `+ (1|PT|f1*f2)`. Then gANOVA is specified using:

```
R> mydata$PTSM <- interaction(mydata$PT, mydata$SM)
R>
R> gANOVA(y ~ Vp*Vs*Vm + (1|PT) + (1|PT:Vs) + (1|PT:Vm) + (1|PT:Vs:Vm)
R>          + (1|SM) + (1|SM:Vp) + (1|SM:Vm)+ (1|SM:Vp:Vm)
R>          + (1|PTSM), data = mydata)
```

or equivalently:

```
R> mydata$PTSM <- interaction(mydata$PT, mydata$SM)
R>
R> gANOVA(y ~ Vp*Vs*Vm + (1|PT|Vs*Vm)+ (1|SM|Vp*Vm) + (1|PTSM),
R>       data = mydata)
```

D.5.2 A Common Case, Variables $V_P(3)$, $V_S(3)$, $V_M(3)$ and $V_M(2)$

This design corresponds to model M3 in Table 4.3. The last interaction terms is confounded with the error term is the interaction between the 2 variables V_{M1}, V_{M2} and the `participant:stimulus` random unit.

RI and RI+

```
R> lmer(y ~ Vp*Vs*Vm1*Vm2 + (1|PT) + (1|SM) + (1|PT:SM), data = mydata)
```

RI-L and RI-L+

```
R> lmer(y ~ Vp*Vs*Vm1*Vm2 + (1|PT) + (1|PT:Vs) + (1|PT:Vm1) + (1|PT:Vm2)
R>   + (1|PT:Vs:Vm1) + (1|PT:Vs:Vm2) + (1|PT:Vm1:Vm2) + (1|PT:Vs:Vm1:Vm2)
R>   + (1|SM) + (1|SM:Vp) + (1|SM:Vm1) + (1|SM:Vm2) + (1|SM:Vp:Vm1)
R>   + (1|SM:Vp:Vm2) + (1|SM:Vm1:Vm2) + (1|SM:Vp:Vm1:Vm2)
R>   + (1|PT:SM) + (1|PT:SM:Vm1) + (1|PT:SM:Vm2), data = mydata)
```

MAX and MAX+

With more than two levels, it is not advisable to use the `sum` coding for the random part and we must create new factors with the appropriate coding variable. Moreover, `lmer` does not handle different type of coding for the fixed part and the random part. We suggest creating new variables with the appropriate coding. Here we choose the orthonormal `contr.poly` coding:

```
R> mydata$VpPoly <- mydata$Vp; contrasts(mydata$VpPoly) <- contr.poly
R> mydata$VsPoly <- mydata$Vs; contrasts(mydata$VsPoly) <- contr.poly
R> mydata$Vm1Poly <- mydata$Vm1; contrasts(mydata$Vm1Poly) <- contr.poly
R> mydata$Vm2Poly <- mydata$Vm2; contrasts(mydata$Vm2Poly) <- contr.poly
R>
R> lmer(y ~ Vp*Vs*Vm1*Vm2 + (VsPoly*Vm1Poly*Vm2Poly|PT)
R>      + (VpPoly*Vm1Poly*Vm2Poly|SM)
R>      + (Vm1Poly + Vm2Poly|PT:SM), data = mydata)
```

ZCP and ZCP+

For ZCP, we transform the factors into orthonormal coding variable. We first need to set the coding using the procedure described in the previous section. Then, we transform the factors into numeric variables using the `model.matrix()` function.

```
R> dataVp <- data.frame(model.matrix( ~ VpPoly, data = mydata)[,-1])
R> colnames(dataVp) <- c("Xpa", "Xpb")
R>
R> dataVs <- data.frame(model.matrix( ~ VsPoly, data = mydata)[,-1])
R> colnames(dataVs) <- c("Xsa", "Xsb")
R>
R> dataVm1 <- data.frame(model.matrix( ~ Vm1Poly, data = mydata)[,-1])
R> colnames(dataVm1) <- c("Xm1a", "Xm1b")
R>
R> dataVm2 <- data.frame(model.matrix( ~ Vm2Poly, data = mydata)[,-1])
R> colnames(dataVm2) <- c("Xm2a")
R>
R> mydata <- cbind(mydata, dataVp, dataVs, dataVm1, dataVm2)
R>
R> lmer(y ~ Vp*Vs*Vm1*Vm2 + ((Xsa + Xsb)*(Xm1a + Xm1b)*Xm2a||PT)
R>      + ((Xpa + Xpb)*(Xm1a + Xm1b)*Xm2a||SM)
R>      + (Xm1a + Xm1b + Xm2a||PT:SM), data = mydata)
```

gANOVA and gANOVA+

As explained previously, the interaction participant:stimuli should be written as only one variable when we can run the gANOVA function:

```
R> mydata$PTSM <- interaction(mydata$PT, mydata$SM)
R>
R> gANOVA(y ~ Vp*Vs*Vm1*Vm2 + (1|PT|Vs*Vm1*Vm2) + (1|SM|Vp*Vm1*Vm2)
R>        + (1|PTSM|Vm1+Vm2), data = mydata)
```

D.5.3 A Complex Case, Variables $V_P(3)$, $V_S(3)$, $V_M(3)$, $V_M(2)$, $V_{PS}(2)$, $V_O(2)$

This design corresponds to model M5 of Table 4.3. The last interaction term confounded with the error term is the interaction between the 3 variables V_{M1}, V_{M2}, V_O and the `participant:stimulus` random unit.

RI and RI+

```
R> lmer(y ~ Vp*Vs*Vm1*Vm2*Vps*Vo + (1|PT) + (1|SM) + (1|PT:SM), data = mydata)
```

RI-L and RI-L+

```
R> lmer(y ~ Vp*Vs*Vm1*Vm2*Vps*Vo + (1|PT) + (1|PT:Vs) + (1|PT:Vm1) + (1|PT:Vm2)
R>     + (1|PT:Vps) + (1|PT:Vo) + (1|PT:Vs:Vm1) + (1|PT:Vs:Vm2) + (1|PT:Vs:Vmsi)
R>     + (1|PT:Vs:Vo) + (1|PT:Vm1:Vm2) + (1|PT:Vm1:Vps) + (1|PT:Vm1:Vo)
R>     + (1|PT:Vm2:Vps) + (1|PT:Vm2:Vo) + (1|PT:Vps:Vo) + (1|PT:Vs:Vm1:Vm2)
R>     + (1|PT:Vs:Vm1:Vps) + (1|PT:Vs:Vm1:Vo) + (1|PT:Vs:Vm2:Vmsi)
R>     + (1|PT:Vs:Vps:Vo)
R>     + (1|PT:Vm1:Vm2:Vps) + (1|PT:Vm2:Vps:Vo) + (1|PT:Vs:Vm1:Vm2:Vps)
R>     + (1|PT:Vs:Vm1:Vm2:Vo) + (1|PT:Vs:Vm1:Vps:Vo) + (1|PT:Vs:Vm2:Vps:Vo)
R>     + (1|PT:Vm1:Vm2:Vps:Vo) + (1|PT:Vs:Vm1:Vm2:Vps:Vo)
R>     + (1|SM) + (1|SM:Vp) + (1|SM:Vm1) + (1|SM:Vm2)
R>     + (1|SM:Vps) + (1|SM:Vo) + (1|SM:Vp:Vm1) + (1|SM:Vp:Vm2) + (1|SM:Vp:Vmsi)
R>     + (1|SM:Vp:Vo) + (1|SM:Vm1:Vm2) + (1|SM:Vm1:Vps) + (1|SM:Vm1:Vo)
R>     + (1|SM:Vm2:Vps) + (1|SM:Vm2:Vo) + (1|SM:Vps:Vo) + (1|SM:Vp:Vm1:Vm2)
R>     + (1|SM:Vp:Vm1:Vps) + (1|SM:Vp:Vm1:Vo) + (1|SM:Vp:Vm2:Vmsi)
R>     + (1|SM:Vp:Vps:Vo)
R>     + (1|SM:Vm1:Vm2:Vps) + (1|SM:Vm2:Vps:Vo) + (1|SM:Vp:Vm1:Vm2:Vps)
R>     + (1|SM:Vp:Vm1:Vm2:Vo) + (1|SM:Vp:Vm1:Vps:Vo)+ (1|SM:Vp:Vm2:Vps:Vo)
R>     + (1|SM:Vm1:Vm2:Vps:Vo) + (1|SM:Vp:Vm1:Vm2:Vps:Vo)
R>     + (1|PT:SM) + (1|PT:SM:Vm1) + (1|PT:SM:Vm2) + (1|PT:SM:Vo)
R>     + (1|PT:SM:Vm1:Vm2) + (1|PT:SM:Vm1:Vo) + (1|PT:SM:Vm2:Vo),
R>     data = mydata)
```

MAX and MAX+

We set the orthonormal coding using the following functions. Note that the variables with 2 levels do not need to change from coding of type "sum" to coding of type "polynomial".

```
R> mydata$VpPoly <- mydata$Vp; contrasts(mydata$VpPoly) <- contr.poly
R> mydata$VsPoly <- mydata$Vs; contrasts(mydata$VsPoly) <- contr.poly
R> mydata$Vm1Poly <- mydata$Vm1; contrasts(mydata$Vm1Poly) <- contr.poly
R> mydata$Vm2Poly <- mydata$Vm2; contrasts(mydata$Vm2Poly) <- contr.poly
R> mydata$VpsPoly <- mydata$Vps; contrasts(mydata$VpsPoly) <- contr.poly
R> mydata$VoPoly <- mydata$Vo; contrasts(mydata$VoPoly) <- contr.poly
```

```
R> lmer(y ~ Vp*Vs*Vm1*Vm2*Vps*Vo + (VsPoly*Vm1Poly*Vm2Poly*VpsPoly*VoPoly|PT)
R>   + (VpPoly*Vm1Poly*Vm2Poly*VpsPoly*VoPoly|SM)
R>   + (Vm1Poly + Vm2Poly + VoPoly + Vm1Poly:PolyVm2 + Vm1Poly:VoPoly
R>   + Vm2Poly:VoPoly|PT:SM),
R>   data = mydata)
```

ZCP and ZCP+

See the MAX model to change the coding of the factors.

```
R> dataVp <- data.frame(model.matrix( ~ VpPoly, data = mydata)[,-1])
R> colnames(dataVp) <- c("Xpa", "Xpb")
R>
R> dataVs <- data.frame(model.matrix( ~ VsPoly, data = mydata)[,-1])
R> colnames(dataVs) <- c("Xsa", "Xsb")
R>
R> dataVm1 <- data.frame(model.matrix( ~ Vm1Poly, data = mydata)[,-1])
R> colnames(dataVm1) <- c("Xm1a", "Xm1b")
R>
R> mydata$Xm2 <- model.matrix( ~ Vm2Poly, data = mydata)[, -1]
R> mydata$Xps <- model.matrix( ~ VpsPoly, data = mydata)[, -1]
R> mydata$Xo <- model.matrix( ~ VoPoly, data = mydata)[, -1]
R>
R> mydata <- cbind(mydata,dataVp,dataVs,dataVm1)
```

```
R> lmer(y ~ Vp*Vs*Vm1*Vm2*Vps*Vo + ((Xsa+Xsb)*(Xm1a+Xm1b)*Xm2*Xps*Xo||PT)
R>  + ((Xpa+Xpb)*(Xm1a+Xm1b)*Xm2*Xps*Xo||SM)
R>  + ((Xm1a+Xm1b)+Xm2+Xo + (Xm1a+Xm1b):Xm2 + (Xm1a+Xm1b):Xo + Xm2:Xo||PT:SM),
R>  data = mydata)
```

gANOVA and gANOVA+

```
R> mydata$PTSM <- interaction(mydata$PT, mydata$SM)
R>
R> gANOVA(y ~ Vp*Vs*Vm1*Vm2*Vps*Vo + (1|PT|Vs*Vm1*Vm2*Vps*Vo)
R>        + (1|SM|Vp*Vm1*Vm2*Vps*Vo)
R>        + (1|PTSM|Vm1 + Vm2 + Vo + Vm1:Vm2 + Vm1:Vo + Vm2:Vo),
R>        data = mydata)
```

Appendix E

Supplementary Simulation Results for Chapter 4

E.1 Results of simulation: type I error rate

Table E.1: List of 5 typical experimental designs.

Model	Variables	Use
M1	Vp(2), Vs(2), Vm(2)	F/S
M2	Vp(3), Vs(3), Vm(3)	S
M3	Vp(3), Vs(3), Vm(3), Vm(2)	F
M4	Vp(3), Vs(3), Vm(3), Vps(2)	S
M5	Vp(3), Vs(3), Vm(3), Vm(2), Vps(2), Vo(2)	F

E.1.1 Design M1

Table E.2: Type I error rate of the design M1 (see Table E.1): The data are simulated using spherical random effects and 18 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.082	**.050**	**.050**	**.050**	**.050**	**.050**	**.050**
		[.074;.091]	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**
	PT:SM	.082	**.050**	**.050**	**.050**	**.050**	**.050**	**.050**
		[.074;.091]	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**
Vs	no PT:SM	.086	**.049**	**.049**	**.049**	**.049**	**.049**	**.049**
		[.078;.095]	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**
	PT:SM	.088	**.049**	**.049**	**.049**	**.049**	**.049**	**.049**
		[.080;.098]	**[.043;.056]**	**[.043;.057]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**
Vm	no PT:SM	.378	**.050**	**.050**	**.050**	**.050**	**.050**	**.056**
		[.363;.393]	**[.043;.057]**	**[.043;.057]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.049;.064]**
	PT:SM	.387	**.050**	**.050**	**.050**	**.050**	**.050**	**.053**
		[.372;.403]	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.046;.060]**
Vp:Vs	no PT:SM	.408	**.049**	**.048**	**.048**	**.048**	**.048**	**.048**
		[.393;.423]	**[.043;.056]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**
	PT:SM	.461	**.049**	**.048**	**.049**	**.048**	**.048**	**.049**
		[.446;.476]	**[.043;.056]**	**[.042;.055]**	**[.042;.056]**	**[.042;.056]**	**[.042;.056]**	**[.043;.056]**
Vp:Vm	no PT:SM	.308	**.050**	**.048**	**.049**	**.049**	**.049**	.076
		[.294;.322]	**[.044;.058]**	**[.042;.055]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	[.069;.085]
	PT:SM	.319	**.050**	**.048**	**.049**	**.049**	**.049**	.070
		[.305;.334]	**[.043;.057]**	**[.042;.055]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	[.062;.078]
Vs:Vm	no PT:SM	.271	**.048**	**.046**	**.047**	**.047**	**.047**	.070
		[.258;.285]	**[.042;.055]**	**[.040;.053]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	[.063;.079]
	PT:SM	.283	**.047**	**.046**	**.046**	**.046**	**.046**	.067
		[.270;.298]	**[.041;.054]**	**[.040;.053]**	**[.040;.053]**	**[.040;.053]**	**[.040;.053]**	[.060;.075]
Vp:Vs:Vm	no PT:SM	.162	**.048**	**.044**	**.048**	**.048**	**.048**	.167
		[.151;.174]	**[.042;.055]**	**[.038;.051]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	[.156;.179]
	PT:SM	.171	**.047**	**.045**	**.048**	**.048**	**.048**	.162
		[.159;.183]	**[.041;.054]**	**[.039;.052]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	[.151;.174]

Table E.3: Estimated type I error rate of the model M1 (see Table E.1): The data are simulated using correlated random effects and 18 stimuli and its represents the subset of model estimated without assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vp	no PT:SM	.076	**.050**	**.049**	**.051**	**.050**	**.050**	**.051**
		[.068;.084]	**[.043;.057]**	**[.043;.057]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.045;.059]**
	PT:SM	.076	**.047**	**.045**	**.047**	**.047**	**.047**	**.048**
		[.068;.085]	**[.041;.054]**	**[.039;.053]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.042;.055]**
Vs	no PT:SM	.087	**.046**	**.044**	**.046**	**.046**	**.046**	**.047**
		[.078;.096]	**[.040;.053]**	**[.038;.051]**	**[.040;.053]**	**[.040;.053]**	**[.040;.053]**	**[.041;.054]**
	PT:SM	.093	**.049**	**.048**	**.050**	**.050**	**.050**	**.051**
		[.084;.102]	**[.043;.056]**	**[.041;.055]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.044;.058]**
Vm	no PT:SM	.336	**.053**	**.046**	**.053**	**.053**	**.053**	.281
		[.321;.350]	**[.047;.061]**	**[.040;.053]**	**[.047;.060]**	**[.046;.060]**	**[.046;.060]**	[.267;.295]
	PT:SM	.378	**.050**	**.044**	**.050**	**.050**	**.050**	.212
		[.363;.393]	**[.044;.057]**	**[.038;.051]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	[.200;.226]
Vp:Vs	no PT:SM	.444	**.050**	**.047**	**.049**	**.049**	**.049**	**.051**
		[.429;.460]	**[.044;.058]**	**[.041;.054]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.045;.058]**
	PT:SM	.494	**.051**	**.048**	**.051**	**.050**	**.050**	**.052**
		[.479;.510]	**[.045;.058]**	**[.042;.056]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.046;.059]**
Vp:Vm	no PT:SM	.250	**.052**	*.040*	**.052**	**.052**	**.052**	.246
		[.237;.264]	**[.046;.060]**	*[.034;.047]*	**[.045;.059]**	**[.046;.060]**	**[.046;.060]**	[.233;.259]
	PT:SM	.286	**.049**	*.041*	**.049**	**.049**	**.049**	.233
		[.273;.301]	**[.043;.056]**	*[.035;.048]*	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	[.220;.246]
Vs:Vm	no PT:SM	.237	**.052**	*.039*	**.052**	**.052**	**.052**	.237
		[.224;.250]	**[.046;.060]**	*[.033;.046]*	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**	[.224;.250]
	PT:SM	.275	**.053**	**.046**	**.053**	**.053**	**.053**	.222
		[.261;.289]	**[.047;.061]**	**[.039;.053]**	**[.047;.061]**	**[.047;.061]**	**[.047;.061]**	[.210;.236]
Vp:Vs:Vm	no PT:SM	.121	**.049**	*.027*	**.047**	**.047**	**.047**	.160
		[.111;.132]	**[.043;.056]**	*[.023;.033]*	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	[.149;.172]
	PT:SM	.148	**.047**	*.036*	**.047**	**.046**	**.046**	.199
		[.138;.160]	**[.041;.054]**	*[.030;.042]*	**[.040;.054]**	**[.040;.054]**	**[.040;.054]**	[.187;.212]

Table E.4: Type I error rate of the model M1 (see Table E.1): The data are simulated using correlated random effects and 18 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.076	**.051**	**.050**	**.052**	**.052**	**.052**	**.052**
		[.068;.084]	**[.045;.058]**	**[.044;.058]**	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**
	PT:SM	.076	**.047**	**.047**	**.047**	**.047**	**.047**	**.048**
		[.068;.085]	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.042;.055]**
Vs	no PT:SM	.087	**.047**	**.044**	**.047**	**.047**	**.047**	**.047**
		[.078;.096]	**[.041;.054]**	**[.038;.051]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**
	PT:SM	.093	**.050**	**.048**	**.050**	**.050**	**.050**	**.050**
		[.084;.102]	**[.043;.057]**	**[.042;.056]**	**[.044;.057]**	**[.044;.057]**	**[.044;.057]**	**[.044;.058]**
Vm	no PT:SM	.376	**.052**	**.048**	**.051**	**.051**	**.051**	.240
		[.361;.391]	**[.045;.059]**	**[.042;.055]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	[.228;.254]
	PT:SM	.394	**.050**	**.044**	**.050**	**.050**	**.050**	.205
		[.380;.410]	**[.044;.057]**	**[.038;.051]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	[.193;.218]
Vp:Vs	no PT:SM	.402	**.052**	**.047**	**.050**	**.050**	**.050**	**.052**
		[.387;.417]	**[.045;.059]**	**[.041;.055]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.046;.060]**
	PT:SM	.480	**.051**	**.050**	**.051**	**.051**	**.051**	**.052**
		[.464;.495]	**[.045;.059]**	**[.043;.058]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.046;.059]**
Vp:Vm	no PT:SM	.290	**.048**	*.041*	**.048**	**.048**	**.048**	.231
		[.277;.305]	**[.042;.055]**	*[.035;.048]*	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	[.219;.245]
	PT:SM	.300	**.049**	*.043*	**.048**	**.048**	**.048**	.229
		[.286;.315]	**[.043;.056]**	*[.037;.050]*	**[.042;.056]**	**[.042;.056]**	**[.042;.056]**	[.216;.242]
Vs:Vm	no PT:SM	.283	**.046**	*.038*	**.046**	**.046**	**.046**	.226
		[.269;.297]	**[.040;.054]**	*[.033;.045]*	**[.040;.054]**	**[.040;.054]**	**[.040;.054]**	[.213;.239]
	PT:SM	.289	**.052**	**.048**	**.053**	**.053**	**.053**	.217
		[.275;.303]	**[.046;.060]**	**[.041;.055]**	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**	[.205;.230]
Vp:Vs:Vm	no PT:SM	.163	**.044**	*.029*	*.043*	*.043*	*.043*	.180
		[.152;.175]	**[.039;.051]**	*[.024;.035]*	*[.037;.050]*	*[.037;.050]*	*[.037;.050]*	[.169;.193]
	PT:SM	.159	**.046**	*.035*	**.046**	**.046**	**.046**	.200
		[.148;.171]	**[.040;.053]**	*[.030;.042]*	**[.040;.053]**	**[.040;.053]**	**[.040;.053]**	[.187;.212]

Table E.5: Type I error rate of the model M1 (see Table E.1): The data are simulated using spherical random effects and 36 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vp	no PT:SM	.068	**.050**	**.051**	**.051**	**.051**	**.051**	**.051**
		[.061;.077]	**[.043;.057]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**
	PT:SM	.070	**.051**	**.052**	**.052**	**.052**	**.052**	**.052**
		[.062;.078]	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**
Vs	no PT:SM	.144	**.053**	**.053**	**.053**	**.053**	**.053**	**.053**
		[.134;.156]	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**	**[.046;.060]**
	PT:SM	.147	**.054**	**.054**	**.054**	**.054**	**.054**	**.054**
		[.137;.159]	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**
Vm	no PT:SM	.419	**.051**	**.051**	**.050**	**.051**	**.050**	**.051**
		[.404;.435]	**[.045;.058]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	**[.045;.058]**
	PT:SM	.462	**.049**	**.049**	**.049**	**.049**	**.049**	**.049**
		[.447;.478]	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**
Vp:Vs	no PT:SM	.540	**.051**	**.050**	**.050**	**.050**	**.050**	**.050**
		[.525;.556]	**[.045;.058]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**
	PT:SM	.571	**.052**	**.051**	**.051**	**.051**	**.051**	**.052**
		[.556;.587]	**[.046;.060]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**
Vp:Vm	no PT:SM	.380	**.053**	**.052**	**.052**	**.052**	**.052**	.057
		[.365;.395]	**[.046;.060]**	**[.045;.059]**	**[.046;.059]**	**[.046;.059]**	**[.046;.059]**	[.050;.065]
	PT:SM	.424	**.050**	**.050**	**.050**	**.050**	**.050**	**.052**
		[.408;.439]	**[.044;.058]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.045;.059]**
Vs:Vm	no PT:SM	.278	**.056**	**.055**	**.055**	**.055**	**.055**	.084
		[.264;.292]	**[.049;.063]**	**[.048;.062]**	**[.048;.063]**	**[.048;.063]**	**[.048;.063]**	[.076;.093]
	PT:SM	.323	**.044**	**.044**	**.044**	**.044**	**.044**	.064
		[.309;.338]	**[.038;.051]**	**[.038;.051]**	**[.038;.051]**	**[.038;.051]**	**[.038;.051]**	[.057;.072]
Vp:Vs:Vm	no PT:SM	.177	.058	**.056**	.057	.057	.057	.148
		[.166;.189]	[.051;.065]	**[.049;.063]**	[.050;.065]	[.050;.065]	[.050;.065]	[.138;.160]
	PT:SM	.216	**.052**	**.050**	**.051**	**.051**	**.051**	.109
		[.203;.229]	**[.045;.059]**	**[.044;.058]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**	[.100;.119]

Table E.6: Type I error rate of the model M1 (see Table E.1): The data are simulated using spherical random effects and 36 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.068	**.050**	**.051**	**.051**	**.051**	**.051**	**.051**
		[.061;.077]	**[.044;.058]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**
	PT:SM	.070	**.051**	**.052**	**.052**	**.052**	**.052**	**.052**
		[.062;.078]	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**
Vs	no PT:SM	.144	**.055**	**.055**	**.055**	**.055**	**.055**	**.055**
		[.134;.156]	**[.048;.063]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**
	PT:SM	.147	**.054**	**.054**	**.054**	**.054**	**.054**	**.054**
		[.137;.159]	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**	**[.048;.062]**
Vm	no PT:SM	.463	**.049**	**.048**	**.048**	**.048**	**.048**	**.048**
		[.448;.479]	**[.043;.056]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	**[.042;.055]**	**[.042;.056]**
	PT:SM	.472	**.049**	**.049**	**.049**	**.049**	**.049**	**.049**
		[.457;.488]	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**	**[.043;.056]**
Vp:Vs	no PT:SM	.497	**.052**	**.051**	**.051**	**.051**	**.051**	**.051**
		[.482;.513]	**[.046;.059]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**	**[.045;.058]**
	PT:SM	.561	**.052**	**.051**	**.051**	**.051**	**.051**	**.052**
		[.546;.577]	**[.046;.060]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**
Vp:Vm	no PT:SM	.424	**.051**	**.050**	**.050**	**.050**	**.050**	**.052**
		[.409;.439]	**[.044;.058]**	**[.044;.057]**	**[.044;.057]**	**[.044;.057]**	**[.044;.057]**	**[.046;.060]**
	PT:SM	.433	**.050**	**.050**	**.050**	**.050**	**.050**	**.051**
		[.418;.449]	**[.044;.058]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.043;.057]**	**[.045;.058]**
Vs:Vm	no PT:SM	.322	**.050**	**.050**	**.050**	**.050**	**.050**	.072
		[.308;.337]	**[.043;.057]**	**[.044;.057]**	**[.044;.057]**	**[.044;.057]**	**[.044;.057]**	[.064;.080]
	PT:SM	.337	**.044**	**.044**	**.044**	**.044**	**.044**	.064
		[.323;.352]	**[.038;.051]**	**[.038;.051]**	**[.038;.051]**	**[.038;.051]**	**[.038;.051]**	[.056;.072]
Vp:Vs:Vm	no PT:SM	.218	**.052**	**.052**	**.052**	**.052**	**.052**	.115
		[.206;.231]	**[.046;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	**[.045;.059]**	[.106;.125]
	PT:SM	.227	**.051**	**.050**	**.050**	**.050**	**.050**	.107
		[.214;.240]	**[.045;.058]**	**[.043;.057]**	**[.044;.058]**	**[.044;.058]**	**[.044;.058]**	[.098;.117]

Table E.7: Type I error rate of the model M1 (see Table E.1): The data are simulated using correlated random effects and 36 stimuli and its represents the subset of model estimated without assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vp	no PT:SM	.065 [.058;.073]	**.051** **[.044;.058]**	**.049** **[.043;.056]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.052** **[.045;.059]**
	PT:SM	.065 [.058;.073]	**.051** **[.045;.058]**	**.051** **[.044;.059]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.052** **[.046;.059]**
Vs	no PT:SM	.136 [.126;.147]	**.054** **[.047;.061]**	**.052** **[.045;.059]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	**.054** **[.048;.062]**
	PT:SM	.136 [.125;.147]	**.054** **[.047;.061]**	**.055** **[.048;.063]**	**.053** **[.046;.060]**	**.053** **[.046;.060]**	**.053** **[.046;.060]**	**.054** **[.047;.061]**
Vm	no PT:SM	.409 [.394;.425]	**.047** **[.041;.054]**	*.042* *[.036;.049]*	**.047** **[.041;.054]**	**.047** **[.041;.054]**	**.047** **[.041;.054]**	.373 [.358;.388]
	PT:SM	.481 [.466;.496]	**.052** **[.045;.059]**	**.051** **[.044;.058]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	.322 [.308;.337]
Vp:Vs	no PT:SM	.534 [.519;.549]	**.054** **[.048;.062]**	**.052** **[.045;.059]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**
	PT:SM	.551 [.536;.566]	**.052** **[.045;.059]**	**.050** **[.043;.058]**	**.051** **[.044;.058]**	**.051** **[.044;.058]**	**.051** **[.044;.058]**	**.052** **[.045;.059]**
Vp:Vm	no PT:SM	.377 [.362;.392]	**.054** **[.048;.062]**	**.046** **[.040;.053]**	**.054** **[.048;.062]**	**.054** **[.048;.062]**	**.054** **[.048;.062]**	.378 [.363;.393]
	PT:SM	.427 [.412;.442]	.057 [.050;.064]	**.051** **[.044;.059]**	.056 [.050;.064]	.056 [.050;.064]	.056 [.050;.064]	.366 [.351;.381]
Vs:Vm	no PT:SM	.280 [.267;.295]	.058 [.051;.065]	**.045** **[.039;.052]**	.058 [.051;.065]	.058 [.051;.065]	.058 [.051;.065]	.289 [.275;.303]
	PT:SM	.330 [.315;.344]	**.055** **[.048;.062]**	**.050** **[.043;.058]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**	.264 [.251;.279]
Vp:Vs:Vm	no PT:SM	.174 [.162;.186]	**.051** **[.045;.058]**	*.035* *[.029;.041]*	**.051** **[.044;.058]**	**.051** **[.044;.058]**	**.051** **[.044;.058]**	.213 [.200;.226]
	PT:SM	.230 [.217;.243]	**.053** **[.047;.061]**	**.045** **[.038;.052]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	.286 [.272;.300]

Table E.8: Type I error rate of the model M1 (see Table E.1): The data are simulated using correlated random effects and 36 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.065 [.058;.073]	**.051** **[.045;.058]**	**.049** **[.043;.057]**	**.052** **[.046;.059]**	**.052** **[.046;.059]**	**.052** **[.046;.059]**	**.052** **[.046;.059]**
	PT:SM	.065 [.058;.073]	**.051** **[.045;.058]**	**.049** **[.043;.057]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.052** **[.046;.059]**
Vs	no PT:SM	.136 [.126;.147]	**.055** **[.048;.063]**	**.052** **[.045;.060]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**
	PT:SM	.136 [.125;.147]	**.054** **[.047;.061]**	**.054** **[.047;.062]**	**.053** **[.047;.061]**	**.053** **[.047;.061]**	**.053** **[.047;.061]**	**.054** **[.047;.061]**
Vm	no PT:SM	.450 [.435;.466]	*.043* *[.037;.050]*	*.040* *[.034;.047]*	*.043* *[.037;.050]*	*.043* *[.037;.050]*	*.043* *[.037;.050]*	.320 [.306;.335]
	PT:SM	.491 [.476;.506]	**.051** **[.045;.059]**	**.049** **[.043;.057]**	**.051** **[.045;.058]**	**.051** **[.045;.058]**	**.051** **[.045;.058]**	.312 [.298;.327]
Vp:Vs	no PT:SM	.499 [.484;.515]	**.056** **[.049;.063]**	**.052** **[.045;.059]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**	**.055** **[.048;.062]**	**.054** **[.048;.062]**
	PT:SM	.540 [.525;.556]	**.052** **[.045;.059]**	**.050** **[.044;.058]**	**.051** **[.044;.058]**	**.051** **[.044;.058]**	**.051** **[.044;.058]**	**.052** **[.045;.059]**
Vp:Vm	no PT:SM	.418 [.402;.433]	**.052** **[.045;.059]**	**.047** **[.041;.055]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	**.051** **[.045;.059]**	.358 [.343;.373]
	PT:SM	.438 [.423;.454]	.056 [.050;.064]	**.050** **[.044;.058]**	**.056** **[.049;.063]**	**.056** **[.049;.063]**	**.056** **[.049;.063]**	.362 [.348;.378]
Vs:Vm	no PT:SM	.331 [.317;.346]	**.052** **[.045;.059]**	**.045** **[.038;.052]**	**.052** **[.045;.059]**	**.052** **[.045;.059]**	**.052** **[.045;.059]**	.258 [.245;.272]
	PT:SM	.345 [.331;.360]	**.054** **[.047;.061]**	**.050** **[.043;.057]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	**.054** **[.047;.061]**	.257 [.244;.271]
Vp:Vs:Vm	no PT:SM	.214 [.201;.227]	**.048** **[.042;.055]**	*.038* *[.032;.045]*	**.048** **[.042;.055]**	**.048** **[.042;.055]**	**.048** **[.042;.055]**	.228 [.216;.242]
	PT:SM	.242 [.229;.255]	**.053** **[.046;.060]**	**.046** **[.040;.054]**	**.053** **[.047;.061]**	**.053** **[.047;.061]**	**.053** **[.047;.061]**	.286 [.272;.300]

E.1.2 Design M2

Table E.9: Type I error rate of the design M2 (see Table E.1): The data are simulated using spherical random effects and 18 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.074 [.067;.083]	**.054** **[.047;.061]**	**.056** **[.049;.065]**	**.054** **[.047;.061]**	.056 [.050;.064]	**.054** **[.048;.062]**	**.056** **[.049;.064]**
	PT:SM	.075 [.067;.084]	**.056** **[.049;.063]**	**.055** **[.048;.063]**	**.053** **[.047;.061]**	**.056** **[.049;.063]**	**.056** **[.049;.064]**	**.056** **[.049;.064]**
Vs	no PT:SM	.086 [.078;.096]	**.052** **[.046;.060]**	**.051** **[.044;.059]**	**.052** **[.045;.059]**	**.054** **[.047;.061]**	**.052** **[.046;.060]**	**.053** **[.046;.060]**
	PT:SM	.088 [.079;.097]	**.053** **[.046;.060]**	**.053** **[.046;.061]**	**.052** **[.045;.059]**	**.053** **[.047;.061]**	**.053** **[.046;.060]**	**.053** **[.046;.060]**
Vm	no PT:SM	.550 [.535;.566]	**.047** **[.041;.054]**	**.047** **[.040;.055]**	.068 [.060;.076]	**.050** **[.044;.058]**	**.047** **[.041;.054]**	**.053** **[.046;.060]**
	PT:SM	.560 [.545;.576]	**.046** **[.040;.053]**	**.051** **[.044;.058]**	.072 [.064;.080]	**.051** **[.045;.058]**	**.046** **[.040;.053]**	**.054** **[.048;.062]**
Vp:Vs	no PT:SM	.673 [.658;.687]	**.048** **[.042;.055]**	**.047** **[.040;.054]**	**.055** **[.048;.063]**	**.055** **[.048;.063]**	**.048** **[.042;.055]**	**.054** **[.048;.062]**
	PT:SM	.777 [.764;.790]	**.049** **[.043;.056]**	**.056** **[.049;.064]**	**.046** **[.040;.053]**	.060 [.053;.068]	**.048** **[.042;.056]**	.060 [.053;.068]
Vp:Vm	no PT:SM	.608 [.593;.623]	**.048** **[.042;.056]**	*.041* *[.035;.048]*	.069 [.062;.077]	.057 [.050;.064]	**.048** **[.042;.055]**	.089 [.080;.098]
	PT:SM	.623 [.608;.638]	**.048** **[.042;.055]**	*.041* *[.035;.048]*	.064 [.057;.072]	**.056** **[.049;.064]**	**.048** **[.042;.055]**	.079 [.071;.087]
Vs:Vm	no PT:SM	.543 [.528;.559]	**.054** **[.047;.061]**	**.043** **[.037;.051]**	.066 [.059;.074]	.058 [.051;.065]	**.054** **[.047;.061]**	.090 [.081;.099]
	PT:SM	.560 [.545;.576]	**.052** **[.046;.060]**	*.041* *[.035;.048]*	.071 [.064;.080]	**.056** **[.049;.063]**	**.052** **[.046;.060]**	.084 [.076;.093]
Vp:Vs:Vm	no PT:SM	.261 [.247;.275]	**.052** **[.046;.060]**	*.018* *[.014;.023]*	.074 [.066;.082]	.060 [.053;.067]	**.052** **[.046;.060]**	.367 [.353;.383]
	PT:SM	.271 [.257;.285]	**.049** **[.043;.056]**	*.024* *[.019;.029]*	.070 [.062;.078]	.063 [.056;.071]	**.049** **[.043;.056]**	.365 [.350;.380]

Table E.10: Estimated type I error rate of the model M2 (see Table E.1): The data are simulated using correlated random effects and 18 stimuli and its represents the subset of model estimated without assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vp	no PT:SM	.077 [.069;.085]	**.051** **[.045;.059]**	**.052** **[.044;.061]**	**.048** **[.042;.055]**	**.053** **[.046;.060]**	**.052** **[.045;.059]**	**.049** **[.043;.056]**
	PT:SM	.078 [.070;.087]	.057 [.050;.065]	**.053** **[.045;.063]**	**.053** **[.047;.061]**	.057 [.050;.065]	.057 [.050;.065]	**.056** **[.049;.064]**
Vs	no PT:SM	.085 [.077;.094]	**.051** **[.044;.058]**	**.050** **[.043;.059]**	**.048** **[.042;.055]**	**.052** **[.046;.060]**	**.051** **[.044;.058]**	**.052** **[.045;.060]**
	PT:SM	.086 [.078;.096]	**.049** **[.043;.056]**	**.042** **[.035;.051]**	**.048** **[.042;.055]**	**.048** **[.042;.055]**	**.049** **[.043;.056]**	**.049** **[.042;.056]**
Vm	no PT:SM	.443 [.428;.459]	**.056** **[.049;.064]**	*.042* *[.035;.050]*	.072 [.064;.080]	.060 [.053;.068]	**.056** **[.049;.064]**	**.056** **[.049;.064]**
	PT:SM	.535 [.520;.550]	**.053** **[.046;.060]**	**.050** **[.042;.060]**	.068 [.061;.077]	.059 [.052;.067]	**.053** **[.046;.060]**	**.056** **[.049;.064]**
Vp:Vs	no PT:SM	.816 [.804;.828]	*.041* *[.035;.047]*	*.036* *[.029;.043]*	*.038* *[.033;.045]*	**.046** **[.040;.053]**	*.041* *[.035;.047]*	*.042* *[.036;.049]*
	PT:SM	.832 [.820;.844]	**.051** **[.045;.059]**	**.046** **[.038;.055]**	**.050** **[.044;.057]**	.061 [.054;.069]	**.051** **[.045;.058]**	**.051** **[.044;.058]**
Vp:Vm	no PT:SM	.472 [.457;.487]	**.055** **[.049;.063]**	*.032* *[.026;.039]*	.069 [.062;.078]	.056 [.050;.064]	**.055** **[.049;.063]**	.079 [.071;.088]
	PT:SM	.552 [.537;.568]	**.053** **[.046;.060]**	*.040* *[.033;.049]*	.066 [.059;.074]	.057 [.050;.064]	**.052** **[.046;.060]**	.079 [.071;.088]
Vs:Vm	no PT:SM	.411 [.396;.426]	.060 [.053;.068]	*.037* *[.031;.045]*	.084 [.076;.093]	.066 [.059;.074]	.060 [.053;.068]	.089 [.080;.098]
	PT:SM	.496 [.481;.512]	**.050** **[.044;.058]**	**.045** **[.037;.054]**	.074 [.066;.082]	.060 [.053;.068]	**.051** **[.045;.058]**	.084 [.076;.094]
Vp:Vs:Vm	no PT:SM	.113 [.104;.123]	.082 [.074;.091]	*.009* *[.006;.014]*	.084 [.076;.093]	.068 [.061;.077]	.082 [.074;.091]	.304 [.289;.319]
	PT:SM	.190 [.178;.203]	.058 [.052;.066]	*.015* *[.011;.020]*	.085 [.077;.094]	.062 [.054;.069]	.058 [.052;.066]	.451 [.435;.468]

Table E.11: Type I error rate of the model M2 (see Table E.1): The data are simulated using correlated random effects and 18 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.077 [.069;.085]	**.054** **[.047;.061]**	**.045** **[.038;.054]**	**.052** **[.045;.059]**	**.055** **[.048;.062]**	**.054** **[.047;.061]**	**.051** **[.045;.059]**
	PT:SM	.078 [.070;.087]	.058 [.051;.065]	**.053** **[.045;.062]**	**.053** **[.047;.061]**	.057 [.050;.065]	.057 [.050;.065]	**.056** **[.049;.064]**
Vs	no PT:SM	.085 [.077;.094]	**.054** **[.047;.061]**	**.051** **[.043;.060]**	**.051** **[.045;.059]**	**.055** **[.048;.063]**	**.054** **[.047;.061]**	**.053** **[.046;.060]**
	PT:SM	.086 [.078;.096]	**.049** **[.043;.056]**	**.044** **[.036;.052]**	**.048** **[.042;.055]**	**.048** **[.042;.055]**	**.049** **[.043;.056]**	**.046** **[.040;.054]**
Vm	no PT:SM	.537 [.522;.553]	**.051** **[.045;.059]**	**.048** **[.040;.057]**	.066 [.059;.075]	**.054** **[.048;.062]**	**.051** **[.045;.058]**	**.053** **[.046;.060]**
	PT:SM	.580 [.565;.596]	**.053** **[.046;.060]**	**.054** **[.046;.064]**	.068 [.061;.077]	.059 [.052;.067]	**.052** **[.046;.060]**	**.055** **[.048;.063]**
Vp:Vs	no PT:SM	.671 [.657;.686]	**.046** **[.040;.053]**	*.033* *[.027;.041]*	**.051** **[.045;.058]**	**.054** **[.047;.061]**	**.046** **[.040;.053]**	*.042* *[.036;.050]*
	PT:SM	.762 [.749;.776]	**.052** **[.045;.059]**	**.049** **[.041;.058]**	**.051** **[.045;.058]**	.061 [.054;.069]	**.052** **[.045;.059]**	**.050** **[.043;.058]**
Vp:Vm	no PT:SM	.596 [.581;.612]	**.046** **[.040;.053]**	*.036* *[.030;.044]*	.066 [.059;.075]	**.051** **[.045;.059]**	**.046** **[.040;.053]**	.075 [.067;.084]
	PT:SM	.607 [.592;.622]	**.053** **[.046;.060]**	**.044** **[.036;.052]**	.065 [.058;.073]	.056 [.050;.064]	**.052** **[.046;.060]**	.079 [.070;.088]
Vs:Vm	no PT:SM	.543 [.528;.559]	**.052** **[.045;.059]**	*.041* *[.034;.049]*	.078 [.070;.086]	.059 [.052;.067]	**.052** **[.046;.059]**	.089 [.081;.099]
	PT:SM	.568 [.553;.584]	**.050** **[.044;.057]**	**.045** **[.037;.053]**	.073 [.065;.081]	.060 [.053;.068]	**.050** **[.044;.058]**	.086 [.077;.095]
Vp:Vs:Vm	no PT:SM	.257 [.244;.271]	**.054** **[.047;.061]**	*.013* *[.009;.018]*	.080 [.072;.089]	.057 [.050;.064]	**.054** **[.048;.062]**	.402 [.386;.418]
	PT:SM	.272 [.258;.286]	.057 [.050;.065]	*.015* *[.011;.020]*	.083 [.075;.092]	.061 [.054;.069]	.057 [.050;.065]	.452 [.436;.468]

Table E.12: Type I error rate of the model M2 (see Table E.1): The data are simulated using spherical random effects and 36 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vp	no PT:SM	.062	**.054**	**.054**	**.051**	**.054**	**.054**	**.055**
		[.055;.070]	**[.047;.061]**	**[.047;.063]**	**[.044;.058]**	**[.048;.062]**	**[.047;.061]**	**[.048;.062]**
	PT:SM	.062	**.054**	**.054**	**.052**	**.054**	**.054**	**.054**
		[.055;.070]	**[.047;.061]**	**[.047;.062]**	**[.046;.059]**	**[.048;.062]**	**[.048;.062]**	**[.047;.061]**
Vs	no PT:SM	.136	**.051**	**.051**	**.049**	**.052**	**.050**	**.052**
		[.126;.148]	**[.044;.058]**	**[.044;.059]**	**[.043;.056]**	**[.045;.059]**	**[.044;.058]**	**[.045;.059]**
	PT:SM	.139	**.054**	.059	**.053**	.058	**.054**	.058
		[.129;.150]	**[.047;.061]**	[.052;.068]	**[.046;.060]**	[.051;.065]	**[.047;.061]**	[.051;.066]
Vm	no PT:SM	.584	**.054**	**.053**	.069	**.056**	**.053**	**.056**
		[.569;.599]	**[.047;.061]**	**[.046;.061]**	[.062;.077]	**[.049;.063]**	**[.047;.061]**	**[.049;.064]**
	PT:SM	.633	**.051**	**.053**	.066	.056	**.051**	.057
		[.618;.648]	**[.045;.058]**	**[.046;.061]**	[.059;.075]	[.050;.064]	**[.044;.058]**	[.050;.065]
Vp:Vs	no PT:SM	.906	*.041*	**.044**	*.043*	**.046**	*.041*	**.046**
		[.896;.915]	*[.036;.048]*	**[.038;.051]**	*[.037;.049]*	**[.040;.053]**	*[.036;.048]*	**[.040;.053]**
	PT:SM	.926	**.046**	**.052**	**.046**	**.052**	**.046**	**.053**
		[.918;.934]	**[.040;.053]**	**[.045;.060]**	**[.040;.053]**	**[.045;.059]**	**[.040;.053]**	**[.046;.061]**
Vp:Vm	no PT:SM	.723	.059	.065	.077	.069	.058	.077
		[.709;.737]	[.052;.066]	[.057;.074]	[.069;.086]	[.061;.077]	[.051;.066]	[.069;.085]
	PT:SM	.782	**.054**	.059	.072	.064	**.054**	.066
		[.769;.795]	**[.048;.062]**	[.052;.068]	[.064;.081]	[.057;.072]	**[.048;.062]**	[.059;.074]
Vs:Vm	no PT:SM	.478	.061	**.056**	.092	.069	.061	.122
		[.463;.493]	[.054;.069]	**[.049;.065]**	[.083;.101]	[.062;.077]	[.054;.069]	[.112;.133]
	PT:SM	.565	**.051**	**.054**	.081	.057	**.051**	.081
		[.550;.581]	**[.045;.059]**	**[.047;.062]**	[.073;.089]	[.050;.065]	**[.045;.059]**	[.073;.090]
Vp:Vs:Vm	no PT:SM	.250	.075	*.036*	.103	.087	.075	.409
		[.237;.264]	[.067;.083]	*[.030;.043]*	[.094;.112]	[.079;.096]	[.067;.083]	[.394;.425]
	PT:SM	.360	**.049**	*.039*	.085	.060	**.049**	.195
		[.345;.375]	**[.043;.056]**	*[.033;.046]*	[.077;.094]	[.053;.068]	**[.043;.056]**	[.183;.208]

Table E.13: Type I error rate of the model M2 (see Table E.1): The data are simulated using spherical random effects and 36 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.062 [.055;.070]	**.054** **[.048;.062]**	**.051** **[.044;.059]**	**.053** **[.046;.060]**	**.055** **[.049;.063]**	**.054** **[.048;.062]**	**.055** **[.048;.062]**
	PT:SM	.062 [.055;.070]	**.054** **[.048;.062]**	**.055** **[.048;.063]**	**.053** **[.046;.060]**	**.054** **[.048;.062]**	**.055** **[.048;.062]**	**.053** **[.047;.061]**
Vs	no PT:SM	.136 [.126;.148]	**.053** **[.047;.061]**	**.053** **[.046;.061]**	**.053** **[.046;.060]**	**.054** **[.048;.062]**	**.053** **[.047;.061]**	**.055** **[.048;.062]**
	PT:SM	.139 [.129;.150]	**.054** **[.047;.061]**	.059 [.051;.067]	**.053** **[.047;.060]**	.058 [.051;.065]	**.054** **[.047;.061]**	.058 [.051;.066]
Vm	no PT:SM	.660 [.645;.674]	**.051** **[.044;.058]**	**.052** **[.045;.061]**	.066 [.058;.074]	**.051** **[.044;.058]**	**.050** **[.044;.058]**	**.052** **[.045;.059]**
	PT:SM	.666 [.652;.681]	**.051** **[.044;.058]**	**.053** **[.046;.061]**	.066 [.059;.074]	**.056** **[.049;.064]**	**.050** **[.044;.057]**	.057 [.050;.065]
Vp:Vs	no PT:SM	.817 [.805;.829]	**.047** **[.041;.054]**	**.046** **[.040;.054]**	**.052** **[.046;.060]**	**.051** **[.045;.058]**	**.047** **[.041;.054]**	**.052** **[.045;.059]**
	PT:SM	.885 [.875;.895]	**.047** **[.041;.054]**	**.053** **[.046;.061]**	**.047** **[.040;.054]**	**.052** **[.046;.059]**	**.047** **[.041;.054]**	**.053** **[.046;.061]**
Vp:Vm	no PT:SM	.809 [.797;.821]	**.055** **[.048;.062]**	.064 [.056;.073]	.075 [.067;.083]	.063 [.056;.071]	**.054** **[.048;.062]**	.068 [.060;.076]
	PT:SM	.824 [.812;.836]	**.054** **[.048;.062]**	.059 [.052;.068]	.072 [.064;.080]	.064 [.057;.072]	**.054** **[.048;.062]**	.067 [.060;.076]
Vs:Vm	no PT:SM	.606 [.591;.621]	**.051** **[.045;.059]**	**.053** **[.045;.061]**	.081 [.073;.090]	.058 [.051;.065]	**.051** **[.045;.059]**	.086 [.077;.095]
	PT:SM	.622 [.607;.637]	**.051** **[.045;.058]**	**.051** **[.045;.059]**	.081 [.073;.089]	.057 [.050;.065]	**.051** **[.045;.058]**	.080 [.072;.089]
Vp:Vs:Vm	no PT:SM	.446 [.431;.462]	**.051** **[.045;.059]**	*.040* *[.034;.047]*	.092 [.083;.101]	.070 [.063;.079]	**.051** **[.045;.059]**	.249 [.235;.263]
	PT:SM	.463 [.448;.479]	**.049** **[.042;.056]**	*.040* *[.034;.047]*	.085 [.076;.094]	.059 [.052;.067]	**.049** **[.042;.056]**	.189 [.177;.202]

Table E.14: Type I error rate of the model M2 (see Table E.1): The data are simulated using correlated random effects and 36 stimuli and its represents the subset of model estimated without assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	MAX	ZCP-sum	ZCP-poly	gANOVA	CS-PCA
Vp	no PT:SM	.058	**.048**	**.050**	**.047**	**.048**	**.048**	**.049**
		[.051;.065]	**[.042;.055]**	**[.043;.059]**	**[.041;.054]**	**[.042;.055]**	**[.042;.055]**	**[.043;.057]**
	PT:SM	.059	**.047**	**.050**	**.047**	**.047**	**.047**	**.050**
		[.052;.067]	**[.041;.054]**	**[.042;.060]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.043;.058]**
Vs	no PT:SM	.134	**.050**	**.056**	**.046**	**.050**	**.050**	**.050**
		[.124;.145]	**[.044;.058]**	**[.048;.065]**	**[.040;.053]**	**[.043;.057]**	**[.044;.058]**	**[.043;.057]**
	PT:SM	.140	**.050**	**.055**	**.044**	**.053**	**.050**	**.056**
		[.129;.151]	**[.044;.058]**	**[.047;.065]**	**[.039;.051]**	**[.046;.060]**	**[.044;.058]**	**[.049;.065]**
Vm	no PT:SM	.564	**.050**	**.045**	.060	**.052**	**.050**	**.050**
		[.549;.580]	**[.044;.057]**	**[.038;.054]**	[.053;.068]	**[.046;.060]**	**[.044;.057]**	**[.044;.058]**
	PT:SM	.635	**.048**	**.050**	.066	**.053**	**.048**	**.053**
		[.621;.650]	**[.041;.055]**	**[.042;.060]**	[.058;.074]	**[.046;.060]**	**[.041;.055]**	**[.046;.061]**
Vp:Vs	no PT:SM	.908	**.044**	**.048**	**.046**	**.049**	**.044**	**.045**
		[.899;.917]	**[.039;.051]**	**[.041;.057]**	**[.040;.053]**	**[.043;.056]**	**[.039;.051]**	**[.039;.053]**
	PT:SM	.923	**.051**	**.052**	**.052**	.056	**.051**	**.053**
		[.915;.932]	**[.044;.058]**	**[.044;.062]**	**[.046;.060]**	[.050;.064]	**[.044;.058]**	**[.046;.061]**
Vp:Vm	no PT:SM	.698	.059	**.055**	.075	.067	.059	.077
		[.684;.712]	[.052;.067]	**[.047;.064]**	[.068;.084]	[.060;.075]	[.052;.067]	[.068;.086]
	PT:SM	.770	**.054**	**.058**	.070	.064	**.054**	.073
		[.757;.783]	**[.048;.062]**	**[.049;.068]**	[.062;.078]	[.057;.072]	**[.048;.062]**	[.065;.083]
Vs:Vm	no PT:SM	.478	**.053**	*.037*	.094	.063	**.053**	.118
		[.463;.494]	**[.046;.060]**	*[.030;.044]*	[.086;.104]	[.056;.071]	**[.046;.060]**	[.108;.128]
	PT:SM	.540	**.045**	*.034*	.080	**.051**	**.045**	.117
		[.525;.556]	**[.039;.052]**	*[.028;.043]*	[.072;.089]	**[.044;.058]**	**[.039;.052]**	[.107;.129]
Vp:Vs:Vm	no PT:SM	.245	.083	*.026*	.113	.086	.083	.457
		[.232;.259]	[.075;.092]	*[.020;.032]*	[.104;.124]	[.078;.096]	[.075;.092]	[.441;.473]
	PT:SM	.351	.058	*.028*	.106	.068	.058	.502
		[.337;.366]	[.051;.065]	*[.022;.036]*	[.097;.116]	[.061;.077]	[.051;.065]	[.485;.519]

Table E.15: Type I error rate of the model M2 (see Table E.1): The data are simulated using correlated random effects and 36 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	MAX+	ZCP-sum+	ZCP-poly+	gANOVA+	CS-PCA+
Vp	no PT:SM	.058	**.049**	**.048**	**.048**	**.049**	**.048**	**.050**
		[.051;.065]	**[.043;.056]**	**[.040;.057]**	**[.042;.056]**	**[.043;.056]**	**[.042;.056]**	**[.043;.057]**
	PT:SM	.059	**.047**	**.049**	**.047**	**.047**	**.047**	**.048**
		[.052;.067]	**[.041;.054]**	**[.041;.058]**	**[.041;.054]**	**[.041;.054]**	**[.041;.054]**	**[.042;.056]**
Vs	no PT:SM	.134	**.052**	**.050**	**.048**	**.052**	**.052**	**.052**
		[.124;.145]	**[.046;.060]**	**[.042;.060]**	**[.042;.055]**	**[.046;.060]**	**[.046;.060]**	**[.045;.060]**
	PT:SM	.140	**.051**	**.051**	**.045**	**.053**	**.050**	.057
		[.129;.151]	**[.044;.058]**	**[.043;.060]**	**[.039;.052]**	**[.046;.060]**	**[.044;.058]**	[.050;.066]
Vm	no PT:SM	.642	**.046**	*.040*	**.055**	**.050**	**.046**	**.048**
		[.627;.657]	**[.040;.053]**	*[.033;.049]*	**[.048;.063]**	**[.043;.057]**	**[.040;.053]**	**[.041;.056]**
	PT:SM	.672	**.047**	**.052**	.066	**.053**	**.047**	**.054**
		[.658;.687]	**[.041;.054]**	**[.044;.061]**	[.058;.074]	**[.046;.060]**	**[.041;.054]**	**[.047;.062]**
Vp:Vs	no PT:SM	.813	**.049**	**.046**	**.053**	**.055**	**.049**	**.044**
		[.801;.825]	**[.043;.056]**	**[.038;.056]**	**[.046;.060]**	**[.048;.062]**	**[.043;.056]**	**[.038;.052]**
	PT:SM	.881	**.051**	**.050**	**.052**	.057	**.051**	**.050**
		[.871;.891]	**[.045;.058]**	**[.042;.059]**	**[.046;.060]**	[.050;.064]	**[.045;.058]**	**[.044;.058]**
Vp:Vm	no PT:SM	.794	**.053**	**.053**	.072	.063	**.053**	.075
		[.781;.806]	**[.047;.061]**	**[.044;.062]**	[.064;.080]	[.056;.071]	**[.047;.061]**	[.067;.084]
	PT:SM	.809	**.054**	**.058**	.069	.064	**.054**	.073
		[.797;.821]	**[.047;.061]**	**[.049;.067]**	[.062;.078]	[.056;.072]	**[.047;.061]**	[.065;.083]
Vs:Vm	no PT:SM	.610	*.042*	*.036*	.082	**.050**	*.042*	.113
		[.595;.625]	*[.036;.049]*	*[.029;.044]*	[.073;.090]	**[.044;.058]**	*[.036;.049]*	[.103;.124]
	PT:SM	.609	**.045**	*.035*	.080	**.051**	**.045**	.116
		[.594;.625]	**[.039;.052]**	*[.029;.043]*	[.072;.089]	**[.044;.058]**	**[.039;.052]**	[.106;.127]
Vp:Vs:Vm	no PT:SM	.420	.060	*.028*	.103	.069	.060	.494
		[.405;.436]	[.053;.068]	*[.022;.035]*	[.094;.113]	[.062;.077]	[.053;.068]	[.478;.511]
	PT:SM	.448	.057	*.027*	.106	.067	.056	.493
		[.433;.463]	[.050;.064]	*[.022;.034]*	[.097;.116]	[.059;.075]	[.050;.064]	[.477;.510]

E.1.3 Design M4

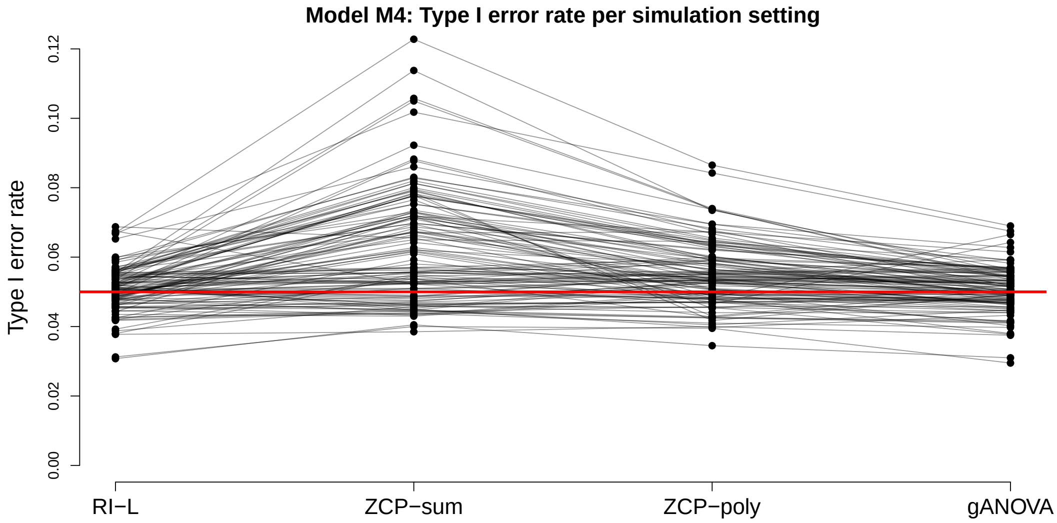

Figure E.1: Type I error rate of the model M4 for all simulations setting (1 sample sizes × 2 correlations of random effects × 2 interactions in simulation × 2 interactions in estimation × 15 effects = 120 settings). The spherical correlation structures (RI-L and gANOVA) produce results closer to the nominal level.

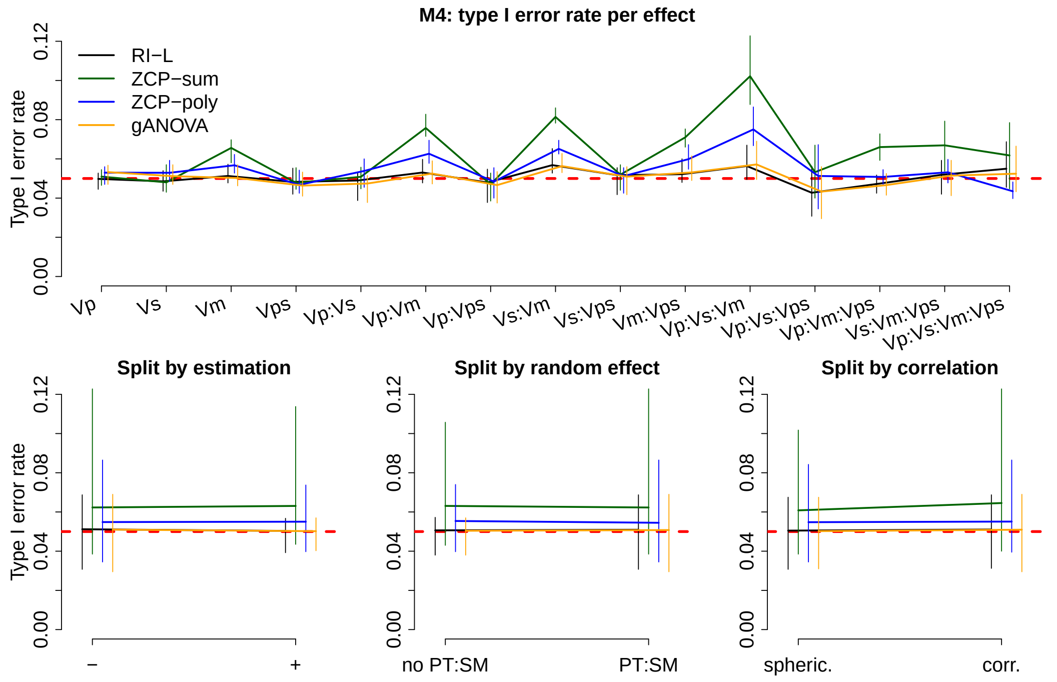

Figure E.2: Type I error rate of the model M4 split given the simulations settings. The vertical lines indicate the range of all simulations within the condition. No simulation setting tend to have an effect on the type I error rate. The variable V_M and its interaction have also a higher deviation from the nominal level.

Table E.16: Type I error rate of the model M4 (see Table E.1): The data are simulated using spherical random effects and 18 stimuli and its represents the subset of model estimated without assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	ZCP-sum	ZCP-poly	gANOVA
Vp	no PT:SM	.078 [.070;.087]	**.050 [.043;.057]**	**.052 [.046;.060]**	**.054 [.047;.061]**	**.055 [.048;.062]**
	PT:SM	.080 [.072;.089]	**.053 [.046;.060]**	**.052 [.046;.060]**	**.055 [.049;.063]**	.056 [.050;.064]
Vs	no PT:SM	.091 [.082;.100]	**.046 [.040;.053]**	*.043* [.037;.050]	**.049 [.043;.056]**	**.048 [.042;.055]**
	PT:SM	.094 [.086;.104]	*.044* [.038;.050]	*.043* [.037;.050]	**.048 [.042;.056]**	**.047 [.041;.054]**
Vm	no PT:SM	.507 [.492;.523]	**.052 [.045;.059]**	.067 [.060;.075]	.056 [.050;.064]	**.051 [.044;.058]**
	PT:SM	.540 [.525;.555]	**.049 [.043;.056]**	.067 [.060;.075]	**.053 [.046;.060]**	**.047 [.041;.054]**
Vps	no PT:SM	.417 [.402;.433]	**.051 [.044;.058]**	**.049 [.043;.056]**	**.048 [.042;.055]**	**.048 [.042;.055]**
	PT:SM	.422 [.407;.438]	**.051 [.045;.058]**	**.046 [.040;.053]**	**.047 [.041;.054]**	**.048 [.042;.055]**
Vp:Vs	no PT:SM	.789 [.776;.802]	*.039* [.033;.045]	**.046 [.040;.053]**	**.046 [.039;.052]**	*.038* [.032;.044]
	PT:SM	.820 [.808;.832]	**.053 [.046;.060]**	**.056 [.049;.063]**	.059 [.052;.067]	**.051 [.044;.058]**
Vp:Vm	no PT:SM	.545 [.530;.561]	.060 [.053;.068]	.083 [.075;.092]	.070 [.062;.078]	.059 [.052;.067]
	PT:SM	.582 [.566;.597]	**.053 [.046;.060]**	.078 [.070;.087]	.064 [.057;.072]	**.052 [.046;.059]**
Vp:Vps	no PT:SM	.502 [.487;.518]	*.038* [.032;.044]	*.039* [.033;.045]	*.040* [.034;.047]	*.038* [.032;.044]
	PT:SM	.519 [.504;.535]	**.055 [.048;.062]**	**.053 [.046;.060]**	**.054 [.047;.061]**	**.053 [.046;.060]**
Vs:Vm	no PT:SM	.498 [.482;.513]	.059 [.052;.067]	.083 [.075;.092]	.068 [.061;.077]	.058 [.051;.066]
	PT:SM	.545 [.530;.560]	**.056 [.049;.063]**	.080 [.072;.089]	.064 [.056;.072]	**.056 [.049;.063]**
Vs:Vps	no PT:SM	.456 [.441;.471]	*.042* [.036;.048]	**.044 [.038;.051]**	*.042* [.037;.049]	*.042* [.036;.048]
	PT:SM	.476 [.461;.492]	**.054 [.047;.061]**	**.052 [.046;.060]**	**.054 [.047;.061]**	**.052 [.045;.059]**
Vm:Vps	no PT:SM	.162 [.151;.174]	.060 [.053;.068]	.072 [.064;.080]	.067 [.060;.075]	.062 [.054;.069]
	PT:SM	.190 [.178;.203]	**.048 [.042;.055]**	.074 [.066;.082]	.060 [.053;.068]	**.050 [.044;.057]**
Vp:Vs:Vm	no PT:SM	.190 [.179;.203]	.067 [.059;.075]	.102 [.093;.112]	.084 [.076;.093]	.068 [.060;.076]
	PT:SM	.264 [.251;.279]	**.051 [.045;.058]**	.088 [.080;.097]	.068 [.060;.076]	**.052 [.045;.059]**
Vp:Vs:Vps	no PT:SM	.378 [.364;.394]	*.031* [.026;.037]	*.041* [.035;.047]	*.034* [.029;.041]	*.031* [.026;.037]
	PT:SM	.385 [.370;.401]	**.054 [.047;.061]**	.065 [.058;.073]	.064 [.057;.073]	**.054 [.048;.062]**
Vp:Vm:Vps	no PT:SM	.104 [.095;.114]	**.050 [.043;.057]**	.062 [.055;.070]	**.050 [.044;.058]**	**.049 [.043;.056]**
	PT:SM	.156 [.145;.167]	**.049 [.043;.056]**	.063 [.056;.071]	**.053 [.046;.060]**	**.047 [.041;.054]**
Vs:Vm:Vps	no PT:SM	.104 [.095;.114]	**.055 [.048;.062]**	.057 [.050;.064]	**.052 [.046;.060]**	**.053 [.047;.061]**
	PT:SM	.148 [.137;.159]	**.049 [.043;.056]**	.062 [.054;.069]	**.049 [.043;.056]**	**.049 [.043;.056]**
Vp:Vs:Vm:Vps	no PT:SM	*.038* [.032;.044]	.068 [.060;.076]	**.052 [.046;.060]**	**.046 [.040;.054]**	.064 [.057;.072]
	PT:SM	**.055 [.048;.063]**	**.047 [.041;.054]**	**.045 [.039;.052]**	*.040* [.035;.047]	**.045 [.039;.052]**

Table E.17: Type I error rate of the model M4 (see Table E.1): The data are simulated using sphercial random effect and 18 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	ZCP-sum+	ZCP-poly+	gANOVA+
Vp	no PT:SM	.079 [.071;.088]	**.052 [.045;.059]**	**.054 [.048;.062]**	**.056 [.049;.064]**	**.056 [.049;.063]**
	PT:SM	.081 [.073;.090]	**.053 [.046;.060]**	**.053 [.046;.060]**	**.056 [.049;.063]**	.057 [.050;.064]
Vs	no PT:SM	.091 [.083;.100]	**.048 [.042;.055]**	**.047 [.041;.054]**	**.053 [.046;.060]**	**.050 [.044;.058]**
	PT:SM	.094 [.086;.104]	**.044 [.039;.051]**	*.044* [.038;.050]	**.050 [.043;.057]**	**.047 [.041;.054]**
Vm	no PT:SM	.594 [.578;.609]	**.049 [.043;.056]**	.064 [.057;.072]	**.056 [.049;.063]**	**.048 [.042;.055]**
	PT:SM	.601 [.586;.617]	**.048 [.042;.056]**	.066 [.059;.075]	**.053 [.046;.060]**	**.046 [.040;.054]**
Vps	no PT:SM	.311 [.297;.326]	**.055 [.049;.063]**	**.056 [.049;.063]**	**.054 [.048;.062]**	**.053 [.047;.061]**
	PT:SM	.350 [.336;.366]	**.051 [.044;.058]**	**.046 [.040;.053]**	**.047 [.041;.054]**	**.048 [.041;.055]**
Vp:Vs	no PT:SM	.624 [.609;.639]	**.047 [.041;.054]**	**.056 [.049;.063]**	**.052 [.046;.060]**	**.046 [.039;.052]**
	PT:SM	.702 [.688;.717]	**.053 [.046;.060]**	**.056 [.049;.063]**	.060 [.053;.068]	**.052 [.045;.059]**
Vp:Vm	no PT:SM	.678 [.664;.693]	**.056 [.049;.063]**	.078 [.070;.086]	.065 [.058;.073]	**.055 [.048;.062]**
	PT:SM	.663 [.648;.678]	**.052 [.046;.060]**	.078 [.070;.087]	.063 [.056;.071]	**.052 [.045;.059]**
Vp:Vps	no PT:SM	.350 [.336;.365]	**.046 [.040;.053]**	**.047 [.041;.054]**	**.047 [.041;.054]**	**.044 [.039;.051]**
	PT:SM	.402 [.387;.417]	**.055 [.048;.062]**	**.052 [.046;.060]**	**.056 [.049;.063]**	**.054 [.047;.061]**
Vs:Vm	no PT:SM	.641 [.626;.656]	**.053 [.046;.060]**	.078 [.070;.087]	.062 [.055;.070]	**.053 [.046;.060]**
	PT:SM	.638 [.623;.653]	**.054 [.048;.062]**	.080 [.072;.088]	.063 [.056;.071]	**.054 [.048;.062]**
Vs:Vps	no PT:SM	.292 [.278;.306]	**.050 [.043;.057]**	**.052 [.046;.060]**	**.049 [.043;.056]**	**.050 [.043;.057]**
	PT:SM	.348 [.334;.363]	**.054 [.048;.062]**	**.054 [.047;.061]**	**.054 [.047;.061]**	**.052 [.046;.060]**
Vm:Vps	no PT:SM	.264 [.250;.278]	**.054 [.047;.061]**	.069 [.061;.077]	.060 [.053;.067]	**.054 [.047;.061]**
	PT:SM	.248 [.234;.261]	**.048 [.042;.055]**	.073 [.065;.082]	.059 [.052;.067]	**.049 [.043;.056]**
Vp:Vs:Vm	no PT:SM	.390 [.375;.406]	**.052 [.046;.060]**	.092 [.084;.102]	.074 [.066;.082]	**.054 [.047;.061]**
	PT:SM	.411 [.396;.426]	**.050 [.043;.057]**	.088 [.079;.097]	.067 [.059;.075]	**.050 [.043;.057]**
Vp:Vs:Vps	no PT:SM	.168 [.157;.181]	**.046 [.040;.053]**	**.051 [.045;.058]**	**.048 [.042;.055]**	**.049 [.043;.056]**
	PT:SM	.216 [.203;.229]	.056 [.050;.064]	.067 [.060;.075]	.067 [.060;.075]	**.056 [.049;.064]**
Vp:Vm:Vps	no PT:SM	.216 [.204;.230]	*.042* [.037;.049]	.059 [.052;.067]	**.047 [.041;.054]**	*.042* [.036;.048]
	PT:SM	.245 [.232;.259]	**.048 [.042;.055]**	.062 [.055;.070]	**.051 [.044;.058]**	**.045 [.039;.052]**
Vs:Vm:Vps	no PT:SM	.220 [.208;.233]	*.042* [.036;.049]	**.054 [.047;.061]**	**.048 [.042;.055]**	*.041* [.036;.048]
	PT:SM	.232 [.219;.245]	**.048 [.041;.055]**	.061 [.054;.069]	**.048 [.042;.055]**	**.048 [.041;.055]**
Vp:Vs:Vm:Vps	no PT:SM	.133 [.123;.144]	**.051 [.045;.059]**	**.054 [.047;.061]**	*.043* [.037;.050]	**.049 [.043;.056]**
	PT:SM	.124 [.114;.134]	**.046 [.039;.052]**	**.045 [.039;.052]**	*.040* [.034;.046]	*.043* [.037;.050]

Table E.18: Type I error rate of the model M4 (see Table E.1): The data are simulated using correlated random effects and 18 stimuli and its represents the subset of model estimated without assuming random effects associated to the interaction participants:stimuli.

		RI	RI-L	ZCP-sum	ZCP-poly	gANOVA
Vp	no PT:SM	.077 [.069;.086]	**.044 [.039;.051]**	**.047 [.040;.054]**	**.047 [.041;.054]**	**.047 [.041;.054]**
	PT:SM	.080 [.072;.089]	**.050 [.044;.058]**	**.051 [.044;.058]**	**.053 [.047;.061]**	**.052 [.046;.060]**
Vs	no PT:SM	.102 [.094;.112]	**.051 [.045;.059]**	**.053 [.046;.060]**	**.053 [.047;.061]**	**.054 [.048;.062]**
	PT:SM	.096 [.088;.106]	**.050 [.044;.058]**	**.048 [.042;.056]**	**.055 [.048;.063]**	**.052 [.046;.060]**
Vm	no PT:SM	.504 [.489;.520]	**.050 [.044;.058]**	.062 [.055;.070]	.058 [.051;.066]	**.050 [.043;.057]**
	PT:SM	.532 [.517;.548]	.057 [.050;.065]	.070 [.062;.078]	.062 [.055;.070]	.057 [.050;.065]
Vps	no PT:SM	.410 [.395;.426]	**.045 [.039;.052]**	**.046 [.040;.053]**	**.044 [.038;.051]**	**.044 [.038;.051]**
	PT:SM	.420 [.405;.435]	*.042* [.036;.049]	**.044 [.039;.051]**	*.042* [.037;.049]	*.041* [.035;.048]
Vp:Vs	no PT:SM	.796 [.783;.808]	**.046 [.040;.054]**	**.049 [.042;.056]**	**.050 [.043;.057]**	**.045 [.039;.052]**
	PT:SM	.816 [.804;.828]	**.051 [.045;.059]**	**.045 [.039;.052]**	**.053 [.046;.060]**	**.050 [.043;.057]**
Vp:Vm	no PT:SM	.536 [.521;.552]	.056 [.050;.064]	.075 [.068;.084]	.064 [.056;.072]	.056 [.050;.064]
	PT:SM	.602 [.587;.617]	**.048 [.042;.055]**	.072 [.064;.080]	.058 [.051;.066]	**.048 [.042;.055]**
Vp:Vps	no PT:SM	.479 [.464;.494]	*.043* [.037;.050]	**.044 [.038;.051]**	*.041* [.035;.048]	*.040* [.034;.046]
	PT:SM	.508 [.493;.524]	**.050 [.044;.057]**	**.048 [.041;.055]**	**.052 [.045;.059]**	**.050 [.044;.057]**
Vs:Vm	no PT:SM	.473 [.458;.489]	.065 [.058;.073]	.086 [.078;.095]	.070 [.062;.078]	.063 [.056;.071]
	PT:SM	.550 [.535;.565]	.056 [.050;.064]	.082 [.074;.091]	.066 [.058;.074]	**.056 [.049;.063]**
Vs:Vps	no PT:SM	.459 [.444;.475]	**.048 [.042;.055]**	**.044 [.038;.051]**	**.046 [.040;.053]**	**.046 [.040;.053]**
	PT:SM	.472 [.457;.488]	**.055 [.048;.063]**	**.056 [.049;.063]**	**.055 [.048;.063]**	**.056 [.049;.063]**
Vm:Vps	no PT:SM	.146 [.135;.157]	.058 [.052;.066]	.075 [.068;.084]	.064 [.057;.072]	.059 [.052;.067]
	PT:SM	.188 [.177;.201]	**.049 [.043;.056]**	.068 [.060;.076]	.057 [.050;.065]	**.050 [.043;.057]**
Vp:Vs:Vm	no PT:SM	.200 [.187;.212]	.067 [.060;.075]	.123 [.113;.133]	.086 [.078;.096]	.069 [.062;.077]
	PT:SM	.265 [.252;.279]	**.055 [.048;.063]**	.106 [.097;.116]	.074 [.066;.083]	**.056 [.049;.063]**
Vp:Vs:Vps	no PT:SM	.384 [.370;.400]	*.031* [.026;.037]	*.040* [.034;.047]	*.040* [.034;.046]	*.030* [.025;.035]
	PT:SM	.362 [.347;.377]	*.038* [.033;.044]	**.054 [.048;.062]**	**.048 [.042;.055]**	*.038* [.033;.044]
Vp:Vm:Vps	no PT:SM	.114 [.104;.124]	**.052 [.045;.059]**	.073 [.065;.081]	**.054 [.048;.062]**	**.052 [.046;.060]**
	PT:SM	.147 [.137;.159]	**.046 [.040;.054]**	.070 [.062;.078]	**.051 [.045;.058]**	**.047 [.041;.054]**
Vs:Vm:Vps	no PT:SM	.104 [.094;.113]	.059 [.052;.067]	.073 [.065;.082]	**.056 [.049;.064]**	.059 [.052;.067]
	PT:SM	.150 [.140;.162]	.057 [.050;.065]	.079 [.071;.088]	.060 [.053;.068]	.057 [.050;.065]
Vp:Vs:Vm:Vps	no PT:SM	*.036* [.031;.043]	.069 [.061;.077]	.066 [.058;.074]	**.048 [.042;.055]**	.066 [.059;.075]
	PT:SM	.067 [.059;.075]	**.052 [.046;.060]**	.078 [.070;.087]	*.043* [.037;.050]	**.050 [.044;.058]**

Table E.19: Type I error rate of the model M4 (see Table E.1): The data are simulated using correlated random effects and 18 stimuli and its represents the subset of model estimated assuming random effects associated to the interaction participants:stimuli.

		RI+	RI-L+	ZCP-sum+	ZCP-poly+	gANOVA+
Vp	no PT:SM	.077 [.069;.085]	**.046 [.039;.052]**	**.047 [.041;.054]**	**.050 [.043;.057]**	**.050 [.044;.058]**
	PT:SM	.080 [.072;.089]	**.051 [.045;.058]**	**.051 [.045;.058]**	**.054 [.047;.061]**	**.052 [.046;.060]**
Vs	no PT:SM	.102 [.093;.112]	**.054 [.047;.061]**	.057 [.050;.065]	.059 [.052;.067]	.057 [.050;.065]
	PT:SM	.097 [.088;.107]	**.051 [.044;.058]**	**.049 [.043;.056]**	**.055 [.049;.063]**	**.053 [.047;.061]**
Vm	no PT:SM	.595 [.580;.611]	**.048 [.042;.055]**	.058 [.051;.066]	**.054 [.047;.061]**	**.046 [.040;.053]**
	PT:SM	.593 [.578;.608]	.057 [.050;.064]	.070 [.062;.078]	.062 [.055;.070]	.057 [.050;.064]
Vps	no PT:SM	.305 [.291;.319]	**.049 [.043;.056]**	**.050 [.044;.057]**	**.048 [.042;.055]**	**.048 [.042;.055]**
	PT:SM	.346 [.331;.361]	*.042* [.037;.049]	**.045 [.039;.052]**	*.043* [.037;.050]	*.041* [.036;.048]
Vp:Vs	no PT:SM	.623 [.608;.638]	**.051 [.044;.058]**	**.055 [.049;.063]**	.058 [.051;.066]	**.050 [.043;.057]**
	PT:SM	.698 [.684;.713]	**.052 [.046;.059]**	**.045 [.039;.052]**	**.053 [.047;.061]**	**.050 [.044;.057]**
Vp:Vm	no PT:SM	.670 [.656;.685]	**.052 [.046;.059]**	.072 [.064;.080]	.060 [.053;.068]	**.052 [.045;.059]**
	PT:SM	.686 [.671;.700]	**.048 [.042;.055]**	.072 [.064;.080]	.058 [.051;.065]	**.047 [.041;.054]**
Vp:Vps	no PT:SM	.318 [.304;.333]	**.048 [.042;.055]**	**.046 [.040;.053]**	**.046 [.039;.052]**	**.044 [.038;.051]**
	PT:SM	.390 [.375;.405]	**.050 [.044;.058]**	**.048 [.042;.055]**	**.052 [.046;.060]**	**.051 [.044;.058]**
Vs:Vm	no PT:SM	.626 [.611;.641]	.056 [.050;.064]	.081 [.073;.090]	.064 [.057;.072]	.056 [.050;.064]
	PT:SM	.636 [.622;.652]	**.055 [.048;.063]**	.081 [.073;.090]	.065 [.058;.073]	**.054 [.048;.062]**
Vs:Vps	no PT:SM	.303 [.289;.318]	**.056 [.049;.063]**	.056 [.050;.064]	**.054 [.048;.062]**	**.055 [.048;.062]**
	PT:SM	.361 [.346;.376]	**.056 [.049;.063]**	.057 [.050;.065]	**.056 [.049;.063]**	**.056 [.049;.064]**
Vm:Vps	no PT:SM	.238 [.225;.252]	**.052 [.045;.059]**	.072 [.064;.080]	**.055 [.048;.062]**	**.052 [.046;.059]**
	PT:SM	.260 [.247;.274]	**.049 [.043;.056]**	.066 [.059;.074]	.056 [.050;.064]	**.050 [.043;.057]**
Vp:Vs:Vm	no PT:SM	.396 [.381;.411]	**.056 [.049;.063]**	.114 [.104;.124]	.074 [.066;.082]	.056 [.050;.064]
	PT:SM	.409 [.394;.424]	**.054 [.047;.061]**	.105 [.096;.115]	.074 [.066;.082]	**.054 [.047;.061]**
Vp:Vs:Vps	no PT:SM	.181 [.169;.193]	**.046 [.040;.054]**	**.053 [.046;.060]**	.059 [.052;.067]	**.049 [.043;.056]**
	PT:SM	.200 [.188;.213]	*.039* [.034;.046]	**.055 [.048;.063]**	**.050 [.043;.057]**	*.040* [.035;.047]
Vp:Vm:Vps	no PT:SM	.238 [.225;.252]	**.044 [.038;.051]**	.072 [.064;.080]	**.049 [.043;.056]**	**.044 [.039;.051]**
	PT:SM	.236 [.224;.250]	**.046 [.040;.053]**	.068 [.061;.077]	**.051 [.044;.058]**	**.046 [.040;.054]**
Vs:Vm:Vps	no PT:SM	.219 [.207;.232]	**.050 [.043;.057]**	.071 [.063;.079]	**.054 [.047;.061]**	**.050 [.043;.057]**
	PT:SM	.232 [.220;.246]	**.056 [.049;.064]**	.079 [.071;.088]	.059 [.052;.067]	**.055 [.048;.062]**
Vp:Vs:Vm:Vps	no PT:SM	.134 [.124;.145]	**.056 [.049;.063]**	.077 [.069;.085]	**.045 [.039;.052]**	**.052 [.046;.059]**
	PT:SM	.152 [.142;.164]	**.052 [.045;.059]**	.078 [.071;.087]	*.042* [.036;.048]	**.049 [.043;.056]**

E.2 Results of Simulation: Power Analysis

Table E.20: Power analysis of model M2 (see Table E.1): The data are simulated using spherical random effects, without the interaction participants:stimuli. The models are estimated without assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.074 [.067;.083]	.146 [.136;.158]	.383 [.368;.399]	.713 [.699;.727]	.922 [.913;.930]	.989 [.986;.992]
	RI-L	.052 [.046;.059]	.112 [.103;.122]	.318 [.304;.333]	.658 [.643;.672]	.897 [.888;.906]	.984 [.980;.987]
	ZCP-sum	.052 [.045;.059]	.099 [.090;.109]	.280 [.266;.294]	.607 [.592;.622]	.872 [.862;.883]	.977 [.973;.982]
	ZCP-poly	.054 [.047;.061]	.112 [.103;.122]	.318 [.303;.332]	.659 [.644;.674]	.897 [.888;.907]	.983 [.980;.987]
	gANOVA	.052 [.046;.060]	.112 [.102;.122]	.318 [.304;.333]	.658 [.643;.672]	.897 [.888;.906]	.984 [.980;.987]
Vs	RI	.086 [.078;.096]	.183 [.172;.196]	.468 [.453;.484]	.784 [.771;.797]	.951 [.945;.958]	.996 [.994;.998]
	RI-L	.051 [.045;.059]	.122 [.113;.133]	.377 [.363;.393]	.712 [.699;.727]	.928 [.920;.936]	.992 [.990;.995]
	ZCP-sum	.047 [.041;.054]	.104 [.095;.114]	.324 [.310;.339]	.655 [.640;.670]	.903 [.894;.912]	.988 [.984;.991]
	ZCP-poly	.052 [.045;.059]	.124 [.114;.135]	.377 [.362;.392]	.710 [.696;.724]	.928 [.920;.936]	.992 [.990;.995]
	gANOVA	.051 [.045;.059]	.122 [.113;.133]	.377 [.363;.393]	.712 [.698;.726]	.928 [.920;.936]	.992 [.990;.995]
Vm	RI	.463 [.448;.479]	.632 [.617;.647]	.891 [.881;.901]	.986 [.983;.990]	1 [.999;1]	1 [1;1]
	RI-L	.051 [.045;.059]	.146 [.135;.157]	.451 [.436;.467]	.812 [.800;.824]	.971 [.966;.976]	.998 [.997;1]
	ZCP-sum	.075 [.067;.083]	.130 [.119;.140]	.328 [.313;.342]	.651 [.636;.666]	.903 [.894;.912]	.984 [.980;.987]
	ZCP-poly	.056 [.049;.063]	.154 [.143;.166]	.459 [.443;.474]	.809 [.797;.821]	.966 [.961;.972]	.998 [.996;.999]
	gANOVA	.051 [.045;.059]	.146 [.135;.157]	.450 [.435;.466]	.812 [.799;.824]	.971 [.966;.976]	.998 [.997;1]
Vp:Vs	RI	.814 [.803;.827]	.863 [.852;.874]	.942 [.935;.949]	.984 [.980;.988]	.998 [.997;.999]	1 [.999;1]
	RI-L	.042 [.036;.048]	.073 [.065;.081]	.194 [.182;.207]	.428 [.413;.444]	.704 [.690;.718]	.908 [.899;.917]
	ZCP-sum	.046 [.040;.053]	.056 [.049;.063]	.101 [.092;.111]	.212 [.200;.225]	.413 [.398;.428]	.658 [.643;.672]
	ZCP-poly	.046 [.040;.053]	.080 [.072;.088]	.197 [.185;.209]	.433 [.418;.448]	.701 [.687;.715]	.898 [.889;.908]
	gANOVA	.042 [.036;.048]	.073 [.065;.081]	.194 [.182;.207]	.427 [.412;.443]	.702 [.688;.716]	.906 [.898;.916]
Vp:Vm	RI	.467 [.452;.482]	.582 [.566;.597]	.788 [.775;.800]	.936 [.929;.944]	.994 [.991;.996]	1 [1;1]
	RI-L	.056 [.049;.063]	.100 [.091;.110]	.266 [.252;.280]	.545 [.530;.560]	.814 [.802;.826]	.954 [.947;.960]
	ZCP-sum	.076 [.068;.084]	.095 [.087;.105]	.186 [.174;.198]	.367 [.352;.382]	.612 [.597;.627]	.820 [.808;.832]
	ZCP-poly	.066 [.058;.074]	.104 [.095;.114]	.269 [.255;.283]	.537 [.522;.553]	.804 [.792;.816]	.948 [.942;.955]
	gANOVA	.056 [.049;.063]	.100 [.091;.110]	.265 [.251;.279]	.544 [.529;.560]	.814 [.802;.826]	.954 [.947;.960]
Vs:Vm	RI	.408 [.394;.424]	.572 [.557;.588]	.862 [.852;.873]	.984 [.980;.988]	.999 [.998;1]	1 [1;1]
	RI-L	.062 [.055;.070]	.126 [.116;.136]	.408 [.393;.424]	.785 [.772;.798]	.963 [.957;.969]	.998 [.997;1]
	ZCP-sum	.078 [.071;.087]	.120 [.111;.131]	.283 [.270;.298]	.575 [.560;.591]	.864 [.854;.875]	.975 [.970;.980]
	ZCP-poly	.068 [.061;.076]	.134 [.123;.145]	.414 [.399;.430]	.774 [.761;.787]	.958 [.952;.965]	.999 [.998;1]
	gANOVA	.062 [.054;.069]	.126 [.116;.136]	.407 [.392;.422]	.784 [.772;.797]	.963 [.957;.969]	.998 [.997;1]
Vp:Vs:Vm	RI	.109 [.100;.119]	.150 [.139;.161]	.257 [.244;.271]	.465 [.450;.480]	.705 [.691;.719]	.890 [.881;.900]
	RI-L	.080 [.072;.088]	.100 [.091;.110]	.185 [.173;.197]	.350 [.336;.366]	.581 [.566;.596]	.804 [.792;.816]
	ZCP-sum	.072 [.064;.080]	.083 [.075;.092]	.131 [.121;.142]	.241 [.228;.254]	.407 [.392;.422]	.610 [.595;.625]
	ZCP-poly	.070 [.063;.079]	.091 [.083;.100]	.165 [.154;.177]	.322 [.307;.336]	.556 [.541;.572]	.777 [.764;.790]
	gANOVA	.080 [.072;.088]	.100 [.092;.110]	.185 [.173;.197]	.350 [.336;.366]	.581 [.566;.596]	.804 [.792;.816]

Table E.21: Power analysis of model M2 (see Table E.1): The data are simulated using spherical random effects, without the interaction participants:stimuli. The models are estimated with assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.074 [.067;.083]	.146 [.136;.158]	.383 [.368;.399]	.713 [.699;.727]	.922 [.913;.930]	.989 [.986;.992]
	RI-L	.054 [.047;.061]	.116 [.106;.126]	.325 [.311;.340]	.662 [.648;.677]	.900 [.890;.909]	.985 [.981;.989]
	ZCP-sum	.056 [.049;.063]	.105 [.096;.115]	.296 [.282;.310]	.621 [.606;.636]	.881 [.871;.891]	.980 [.975;.984]
	ZCP-poly	.055 [.048;.063]	.114 [.105;.125]	.324 [.310;.339]	.662 [.648;.677]	.900 [.891;.910]	.985 [.981;.989]
	gANOVA	.054 [.048;.062]	.115 [.106;.126]	.324 [.310;.339]	.662 [.648;.677]	.900 [.890;.909]	.985 [.981;.989]
Vs	RI	.086 [.078;.096]	.183 [.172;.196]	.468 [.453;.484]	.784 [.771;.797]	.951 [.945;.958]	.996 [.994;.998]
	RI-L	.052 [.046;.060]	.126 [.116;.136]	.385 [.370;.400]	.718 [.704;.732]	.930 [.922;.938]	.993 [.990;.996]
	ZCP-sum	.052 [.045;.059]	.112 [.102;.122]	.339 [.325;.354]	.669 [.654;.683]	.908 [.899;.917]	.989 [.985;.992]
	ZCP-poly	.055 [.048;.063]	.128 [.118;.139]	.384 [.370;.400]	.716 [.703;.731]	.930 [.922;.937]	.993 [.990;.995]
	gANOVA	.052 [.046;.060]	.125 [.115;.136]	.385 [.370;.400]	.718 [.704;.732]	.930 [.922;.938]	.993 [.990;.996]
Vm	RI	.550 [.535;.566]	.702 [.688;.717]	.920 [.912;.929]	.991 [.988;.994]	1 [.999;1]	1 [1;1]
	RI-L	.047 [.041;.054]	.136 [.126;.147]	.437 [.422;.452]	.803 [.791;.816]	.968 [.963;.974]	.998 [.996;.999]
	ZCP-sum	.071 [.063;.079]	.117 [.107;.127]	.297 [.283;.311]	.612 [.597;.627]	.885 [.875;.895]	.980 [.975;.984]
	ZCP-poly	.051 [.045;.059]	.147 [.136;.158]	.447 [.432;.463]	.799 [.787;.812]	.965 [.959;.971]	.998 [.996;.999]
	gANOVA	.047 [.041;.054]	.136 [.126;.147]	.436 [.421;.452]	.802 [.790;.815]	.968 [.963;.974]	.998 [.996;.999]
Vp:Vs	RI	.673 [.658;.687]	.732 [.719;.746]	.871 [.860;.881]	.962 [.956;.968]	.993 [.991;.996]	1 [.999;1]
	RI-L	.048 [.042;.055]	.084 [.076;.093]	.212 [.200;.225]	.450 [.435;.466]	.723 [.710;.737]	.915 [.907;.924]
	ZCP-sum	.055 [.048;.063]	.069 [.062;.077]	.124 [.115;.135]	.261 [.247;.275]	.476 [.461;.492]	.715 [.701;.729]
	ZCP-poly	.054 [.047;.061]	.090 [.081;.099]	.218 [.205;.231]	.455 [.440;.470]	.723 [.709;.737]	.905 [.896;.914]
	gANOVA	.048 [.042;.055]	.084 [.075;.093]	.212 [.199;.225]	.448 [.433;.464]	.721 [.707;.735]	.915 [.906;.924]
Vp:Vm	RI	.608 [.593;.623]	.696 [.682;.710]	.857 [.846;.868]	.965 [.960;.971]	.998 [.997;.999]	1 [1;1]
	RI-L	.048 [.042;.056]	.091 [.083;.101]	.243 [.230;.257]	.518 [.503;.534]	.798 [.785;.810]	.946 [.939;.953]
	ZCP-sum	.072 [.064;.080]	.087 [.078;.096]	.165 [.154;.177]	.328 [.314;.343]	.569 [.554;.585]	.793 [.781;.806]
	ZCP-poly	.058 [.051;.066]	.093 [.085;.103]	.252 [.239;.266]	.521 [.506;.537]	.792 [.779;.804]	.944 [.937;.952]
	gANOVA	.048 [.042;.055]	.091 [.083;.101]	.242 [.229;.256]	.518 [.503;.534]	.797 [.785;.810]	.946 [.939;.953]
Vs:Vm	RI	.543 [.528;.559]	.697 [.683;.711]	.914 [.906;.923]	.991 [.988;.994]	1 [1;1]	1 [1;1]
	RI-L	.054 [.047;.061]	.110 [.100;.120]	.378 [.363;.393]	.760 [.747;.774]	.959 [.953;.965]	.998 [.997;.999]
	ZCP-sum	.072 [.064;.080]	.106 [.097;.116]	.245 [.232;.259]	.522 [.507;.538]	.830 [.818;.841]	.965 [.959;.970]
	ZCP-poly	.061 [.054;.069]	.122 [.112;.132]	.391 [.376;.406]	.754 [.741;.768]	.956 [.949;.962]	.998 [.996;.999]
	gANOVA	.054 [.047;.061]	.110 [.100;.120]	.378 [.363;.393]	.759 [.746;.773]	.959 [.953;.965]	.998 [.997;.999]
Vp:Vs:Vm	RI	.261 [.247;.275]	.310 [.296;.325]	.460 [.445;.476]	.666 [.652;.681]	.858 [.848;.869]	.954 [.948;.961]
	RI-L	.052 [.046;.060]	.069 [.062;.077]	.136 [.126;.147]	.283 [.270;.298]	.507 [.492;.523]	.741 [.728;.755]
	ZCP-sum	.071 [.063;.079]	.079 [.071;.088]	.115 [.106;.125]	.195 [.183;.207]	.329 [.314;.343]	.510 [.495;.526]
	ZCP-poly	.060 [.053;.068]	.078 [.070;.087]	.143 [.132;.154]	.292 [.278;.306]	.509 [.494;.525]	.739 [.726;.753]
	gANOVA	.052 [.046;.060]	.069 [.062;.077]	.136 [.126;.148]	.284 [.270;.298]	.507 [.492;.523]	.741 [.728;.755]

Table E.22: Power analysis of model M2 (see Table E.1): The data are simulated using spherical random effects, with the interaction participants:stimuli. The models are estimated without assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.075 [.067;.084]	.152 [.141;.164]	.410 [.395;.425]	.738 [.724;.751]	.935 [.928;.943]	.991 [.988;.994]
	RI-L	.056 [.049;.063]	.117 [.107;.127]	.350 [.335;.365]	.686 [.672;.701]	.918 [.910;.927]	.986 [.983;.990]
	ZCP-sum	.054 [.047;.061]	.109 [.100;.119]	.312 [.298;.327]	.642 [.627;.657]	.892 [.882;.901]	.984 [.980;.987]
	ZCP-poly	.055 [.048;.063]	.120 [.110;.130]	.350 [.335;.365]	.686 [.672;.701]	.918 [.910;.927]	.986 [.983;.990]
	gANOVA	.056 [.049;.064]	.117 [.107;.127]	.350 [.335;.365]	.686 [.672;.701]	.918 [.910;.927]	.986 [.983;.990]
Vs	RI	.088 [.079;.097]	.190 [.179;.203]	.478 [.463;.494]	.798 [.786;.811]	.962 [.956;.968]	.996 [.994;.998]
	RI-L	.053 [.046;.060]	.128 [.118;.139]	.394 [.379;.409]	.738 [.725;.752]	.943 [.936;.950]	.994 [.992;.996]
	ZCP-sum	.050 [.043;.057]	.111 [.102;.121]	.339 [.325;.354]	.675 [.661;.690]	.922 [.913;.930]	.990 [.987;.993]
	ZCP-poly	.052 [.046;.060]	.128 [.118;.139]	.394 [.379;.409]	.736 [.723;.750]	.942 [.935;.949]	.994 [.992;.996]
	gANOVA	.053 [.046;.060]	.128 [.118;.139]	.394 [.379;.409]	.738 [.725;.752]	.943 [.936;.950]	.994 [.992;.996]
Vm	RI	.515 [.500;.531]	.680 [.665;.694]	.909 [.900;.918]	.991 [.988;.994]	1 [.999;1]	1 [1;1]
	RI-L	.046 [.040;.054]	.141 [.131;.152]	.440 [.425;.456]	.803 [.791;.815]	.970 [.965;.976]	.999 [.998;1]
	ZCP-sum	.070 [.062;.078]	.108 [.099;.118]	.282 [.269;.297]	.607 [.592;.623]	.888 [.878;.898]	.983 [.979;.987]
	ZCP-poly	.053 [.046;.060]	.145 [.134;.156]	.441 [.426;.457]	.795 [.783;.808]	.970 [.964;.975]	.999 [.998;1]
	gANOVA	.046 [.040;.054]	.141 [.131;.152]	.440 [.425;.456]	.803 [.791;.815]	.970 [.965;.976]	.999 [.998;1]
Vp:Vs	RI	.843 [.832;.854]	.884 [.874;.894]	.948 [.941;.955]	.987 [.983;.990]	1 [.999;1]	1 [1;1]
	RI-L	.049 [.043;.056]	.087 [.079;.096]	.230 [.217;.243]	.484 [.469;.500]	.763 [.750;.776]	.929 [.921;.937]
	ZCP-sum	.047 [.041;.054]	.062 [.055;.070]	.111 [.102;.121]	.254 [.241;.268]	.467 [.452;.483]	.720 [.706;.734]
	ZCP-poly	.058 [.051;.065]	.094 [.086;.104]	.234 [.221;.248]	.494 [.479;.510]	.761 [.748;.774]	.928 [.919;.936]
	gANOVA	.048 [.042;.055]	.087 [.078;.096]	.229 [.217;.243]	.483 [.468;.498]	.762 [.749;.776]	.929 [.921;.937]
Vp:Vm	RI	.558 [.542;.573]	.662 [.648;.677]	.840 [.828;.851]	.958 [.951;.964]	.994 [.991;.996]	1 [.999;1]
	RI-L	.049 [.043;.056]	.095 [.087;.105]	.247 [.234;.260]	.522 [.507;.538]	.800 [.788;.813]	.949 [.942;.956]
	ZCP-sum	.064 [.056;.072]	.088 [.080;.098]	.162 [.151;.174]	.327 [.313;.342]	.572 [.557;.587]	.791 [.778;.803]
	ZCP-poly	.059 [.052;.067]	.099 [.090;.109]	.257 [.244;.271]	.532 [.517;.548]	.791 [.779;.804]	.944 [.937;.951]
	gANOVA	.049 [.043;.056]	.095 [.086;.105]	.246 [.234;.260]	.522 [.506;.537]	.800 [.787;.812]	.949 [.942;.956]
Vs:Vm	RI	.496 [.480;.511]	.646 [.632;.661]	.893 [.884;.903]	.989 [.985;.992]	1 [1;1]	1 [1;1]
	RI-L	.053 [.046;.060]	.121 [.111;.132]	.380 [.365;.395]	.757 [.744;.770]	.953 [.946;.959]	.997 [.995;.999]
	ZCP-sum	.067 [.060;.075]	.102 [.093;.112]	.244 [.231;.258]	.521 [.505;.536]	.813 [.801;.825]	.960 [.954;.966]
	ZCP-poly	.054 [.048;.062]	.132 [.122;.143]	.388 [.374;.404]	.756 [.743;.770]	.948 [.942;.955]	.998 [.996;.999]
	gANOVA	.053 [.046;.060]	.121 [.111;.131]	.380 [.365;.395]	.756 [.743;.769]	.952 [.946;.959]	.997 [.995;.999]
Vp:Vs:Vm	RI	.190 [.178;.203]	.235 [.222;.249]	.380 [.366;.396]	.596 [.580;.611]	.808 [.796;.821]	.941 [.934;.948]
	RI-L	.050 [.044;.057]	.071 [.063;.079]	.140 [.130;.151]	.292 [.279;.307]	.516 [.500;.531]	.754 [.741;.768]
	ZCP-sum	.072 [.064;.080]	.086 [.078;.095]	.122 [.112;.132]	.199 [.187;.212]	.334 [.320;.349]	.512 [.497;.528]
	ZCP-poly	.064 [.057;.072]	.084 [.075;.093]	.159 [.148;.171]	.308 [.294;.322]	.527 [.512;.543]	.754 [.741;.768]
	gANOVA	.050 [.044;.057]	.071 [.063;.079]	.140 [.130;.151]	.292 [.278;.307]	.516 [.501;.532]	.754 [.741;.767]

Table E.23: Power analysis of model M2 (see Table E.1): The data are simulated using spherical random effects, with the interaction participants:stimuli. The models are estimated with assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.075 [.067;.084]	.152 [.141;.164]	.410 [.395;.425]	.738 [.724;.751]	.935 [.928;.943]	.991 [.988;.994]
	RI-L	.056 [.049;.063]	.117 [.108;.127]	.350 [.336;.365]	.686 [.672;.701]	.918 [.910;.927]	.986 [.983;.990]
	ZCP-sum	.054 [.047;.061]	.109 [.100;.119]	.312 [.298;.327]	.644 [.629;.659]	.892 [.883;.902]	.984 [.980;.988]
	ZCP-poly	.056 [.049;.063]	.120 [.110;.130]	.351 [.336;.366]	.686 [.672;.701]	.918 [.910;.927]	.987 [.983;.990]
	gANOVA	.056 [.049;.064]	.117 [.107;.127]	.350 [.335;.365]	.686 [.672;.701]	.918 [.910;.927]	.986 [.983;.990]
Vs	RI	.088 [.079;.097]	.190 [.179;.203]	.478 [.463;.494]	.798 [.786;.811]	.962 [.956;.968]	.996 [.994;.998]
	RI-L	.053 [.047;.060]	.129 [.119;.140]	.395 [.380;.410]	.739 [.725;.753]	.943 [.936;.950]	.994 [.992;.996]
	ZCP-sum	.051 [.045;.058]	.112 [.102;.122]	.341 [.327;.356]	.676 [.662;.691]	.922 [.914;.931]	.990 [.987;.993]
	ZCP-poly	.053 [.046;.060]	.128 [.118;.139]	.394 [.379;.409]	.737 [.723;.751]	.942 [.935;.950]	.994 [.992;.996]
	gANOVA	.053 [.046;.060]	.129 [.119;.140]	.395 [.380;.410]	.739 [.725;.752]	.943 [.936;.950]	.994 [.992;.996]
Vm	RI	.560 [.545;.576]	.709 [.695;.723]	.924 [.916;.932]	.992 [.989;.995]	1 [.999;1]	1 [1;1]
	RI-L	.046 [.040;.053]	.140 [.130;.152]	.439 [.424;.455]	.801 [.789;.814]	.970 [.965;.975]	.999 [.998;1]
	ZCP-sum	.069 [.062;.078]	.108 [.098;.118]	.282 [.268;.296]	.606 [.591;.621]	.887 [.877;.897]	.983 [.979;.987]
	ZCP-poly	.052 [.046;.059]	.144 [.134;.156]	.441 [.426;.457]	.795 [.783;.808]	.970 [.964;.975]	.999 [.998;1]
	gANOVA	.046 [.040;.053]	.140 [.130;.151]	.439 [.424;.455]	.802 [.789;.814]	.970 [.965;.976]	.999 [.998;1]
Vp:Vs	RI	.777 [.764;.790]	.831 [.819;.843]	.926 [.918;.934]	.978 [.973;.982]	.997 [.996;.999]	1 [1;1]
	RI-L	.049 [.043;.056]	.088 [.079;.097]	.231 [.218;.244]	.485 [.470;.501]	.764 [.751;.777]	.930 [.922;.938]
	ZCP-sum	.047 [.041;.054]	.062 [.055;.070]	.113 [.104;.124]	.257 [.244;.271]	.472 [.457;.488]	.724 [.710;.738]
	ZCP-poly	.058 [.051;.066]	.095 [.087;.105]	.236 [.223;.250]	.496 [.480;.511]	.763 [.750;.776]	.928 [.920;.936]
	gANOVA	.048 [.042;.056]	.087 [.079;.096]	.230 [.217;.243]	.484 [.469;.499]	.763 [.750;.776]	.930 [.922;.938]
Vp:Vm	RI	.623 [.608;.638]	.714 [.700;.728]	.871 [.861;.882]	.967 [.962;.973]	.996 [.994;.998]	1 [.999;1]
	RI-L	.048 [.041;.055]	.095 [.086;.104]	.245 [.232;.259]	.521 [.506;.537]	.800 [.787;.812]	.949 [.942;.956]
	ZCP-sum	.063 [.056;.071]	.088 [.080;.098]	.162 [.150;.173]	.327 [.312;.341]	.570 [.555;.585]	.789 [.776;.802]
	ZCP-poly	.058 [.051;.065]	.099 [.090;.109]	.255 [.242;.269]	.530 [.515;.546]	.790 [.777;.803]	.944 [.937;.951]
	gANOVA	.048 [.041;.055]	.094 [.086;.104]	.245 [.232;.259]	.521 [.505;.536]	.799 [.787;.812]	.948 [.942;.955]
Vs:Vm	RI	.560 [.544;.575]	.702 [.688;.717]	.916 [.907;.924]	.993 [.990;.996]	1 [1;1]	1 [1;1]
	RI-L	.052 [.046;.060]	.120 [.111;.131]	.378 [.364;.394]	.755 [.742;.769]	.952 [.946;.959]	.997 [.995;.999]
	ZCP-sum	.067 [.060;.075]	.102 [.093;.112]	.241 [.228;.254]	.518 [.502;.533]	.811 [.799;.823]	.959 [.953;.965]
	ZCP-poly	.054 [.048;.062]	.131 [.121;.142]	.388 [.373;.404]	.755 [.742;.768]	.948 [.941;.955]	.998 [.996;.999]
	gANOVA	.052 [.046;.060]	.120 [.110;.130]	.378 [.364;.394]	.754 [.741;.768]	.952 [.945;.959]	.997 [.995;.999]
Vp:Vs:Vm	RI	.271 [.257;.285]	.327 [.313;.342]	.477 [.462;.493]	.690 [.676;.705]	.867 [.856;.877]	.964 [.958;.970]
	RI-L	.048 [.042;.055]	.069 [.061;.077]	.136 [.126;.147]	.288 [.274;.302]	.512 [.496;.527]	.748 [.735;.762]
	ZCP-sum	.071 [.063;.079]	.086 [.078;.095]	.120 [.111;.131]	.197 [.185;.210]	.329 [.315;.344]	.506 [.490;.521]
	ZCP-poly	.063 [.056;.071]	.084 [.075;.093]	.157 [.146;.169]	.304 [.290;.319]	.524 [.509;.540]	.751 [.737;.764]
	gANOVA	.048 [.042;.055]	.068 [.061;.077]	.136 [.126;.147]	.288 [.274;.302]	.511 [.496;.527]	.748 [.735;.762]

Table E.24: Power analysis of model M2 (see Table E.1): The data are simulated using correlated random effects, without the interaction participants:stimuli. The models are estimated without assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.076 [.069;.085]	.169 [.158;.181]	.390 [.375;.405]	.706 [.692;.720]	.910 [.901;.919]	.990 [.987;.993]
	RI-L	.051 [.045;.059]	.131 [.121;.142]	.340 [.325;.354]	.653 [.638;.668]	.887 [.878;.897]	.982 [.978;.986]
	ZCP-sum	.049 [.043;.056]	.110 [.101;.121]	.302 [.288;.317]	.604 [.590;.620]	.866 [.855;.876]	.974 [.969;.979]
	ZCP-poly	.052 [.046;.060]	.128 [.119;.139]	.338 [.324;.353]	.651 [.637;.666]	.886 [.876;.896]	.983 [.979;.987]
	gANOVA	.051 [.045;.059]	.130 [.120;.141]	.340 [.325;.355]	.653 [.638;.668]	.888 [.878;.897]	.982 [.978;.986]
Vs	RI	.085 [.077;.094]	.174 [.163;.187]	.472 [.457;.488]	.797 [.784;.809]	.955 [.949;.961]	.996 [.993;.998]
	RI-L	.049 [.043;.056]	.116 [.106;.126]	.379 [.364;.394]	.724 [.711;.738]	.938 [.931;.945]	.992 [.990;.995]
	ZCP-sum	.048 [.042;.055]	.098 [.089;.107]	.317 [.303;.332]	.654 [.640;.669]	.907 [.898;.916]	.986 [.983;.990]
	ZCP-poly	.051 [.045;.058]	.116 [.106;.126]	.378 [.364;.394]	.723 [.709;.737]	.938 [.930;.945]	.993 [.990;.995]
	gANOVA	.050 [.043;.057]	.116 [.107;.127]	.379 [.364;.394]	.724 [.711;.738]	.938 [.930;.945]	.992 [.990;.995]
Vm	RI	.443 [.428;.459]	.611 [.596;.626]	.878 [.868;.889]	.983 [.979;.987]	.999 [.998;1]	1 [1;1]
	RI-L	.056 [.049;.064]	.140 [.130;.151]	.444 [.429;.459]	.794 [.782;.807]	.962 [.957;.968]	.997 [.996;.999]
	ZCP-sum	.072 [.065;.081]	.122 [.113;.133]	.315 [.301;.329]	.649 [.634;.664]	.896 [.886;.905]	.988 [.984;.991]
	ZCP-poly	.061 [.054;.069]	.147 [.136;.158]	.443 [.428;.458]	.794 [.782;.807]	.960 [.954;.966]	.997 [.996;.999]
	gANOVA	.056 [.049;.064]	.140 [.130;.151]	.443 [.428;.459]	.794 [.782;.807]	.962 [.957;.968]	.997 [.996;.999]
Vp:Vs	RI	.815 [.803;.827]	.856 [.845;.867]	.936 [.928;.944]	.986 [.983;.990]	.999 [.998;1]	1 [1;1]
	RI-L	.040 [.034;.046]	.074 [.067;.083]	.190 [.178;.203]	.432 [.416;.447]	.701 [.687;.715]	.906 [.897;.915]
	ZCP-sum	.040 [.034;.046]	.053 [.046;.060]	.100 [.092;.110]	.217 [.204;.230]	.410 [.395;.425]	.651 [.637;.666]
	ZCP-poly	.047 [.041;.054]	.079 [.071;.088]	.200 [.188;.213]	.431 [.416;.446]	.702 [.687;.716]	.900 [.891;.910]
	gANOVA	.040 [.034;.046]	.074 [.066;.083]	.190 [.178;.202]	.430 [.415;.446]	.701 [.687;.715]	.906 [.897;.915]
Vp:Vm	RI	.471 [.456;.487]	.564 [.549;.579]	.766 [.753;.779]	.931 [.923;.939]	.987 [.983;.990]	1 [.999;1]
	RI-L	.056 [.049;.064]	.101 [.092;.111]	.261 [.247;.275]	.526 [.510;.541]	.798 [.786;.811]	.946 [.940;.953]
	ZCP-sum	.070 [.062;.078]	.092 [.084;.102]	.186 [.174;.198]	.363 [.348;.378]	.614 [.599;.630]	.831 [.819;.842]
	ZCP-poly	.058 [.051;.066]	.105 [.096;.115]	.272 [.259;.286]	.525 [.510;.540]	.788 [.775;.801]	.940 [.933;.947]
	gANOVA	.056 [.049;.064]	.101 [.092;.111]	.260 [.247;.274]	.525 [.510;.540]	.798 [.786;.811]	.946 [.940;.953]
Vs:Vm	RI	.411 [.396;.427]	.573 [.558;.589]	.850 [.839;.861]	.980 [.976;.984]	.999 [.998;1]	1 [1;1]
	RI-L	.061 [.054;.069]	.139 [.129;.150]	.403 [.388;.418]	.755 [.742;.769]	.957 [.951;.963]	.996 [.994;.998]
	ZCP-sum	.084 [.076;.093]	.128 [.119;.139]	.295 [.281;.309]	.591 [.576;.606]	.860 [.849;.871]	.976 [.972;.981]
	ZCP-poly	.068 [.061;.076]	.144 [.133;.155]	.405 [.390;.420]	.755 [.742;.769]	.955 [.949;.961]	.996 [.994;.998]
	gANOVA	.061 [.054;.069]	.139 [.129;.150]	.402 [.388;.418]	.754 [.741;.768]	.957 [.950;.963]	.996 [.994;.998]
Vp:Vs:Vm	RI	.114 [.104;.124]	.150 [.140;.162]	.261 [.247;.275]	.466 [.451;.482]	.696 [.681;.710]	.888 [.878;.898]
	RI-L	.082 [.073;.090]	.105 [.096;.115]	.193 [.181;.206]	.360 [.345;.375]	.582 [.567;.598]	.791 [.778;.803]
	ZCP-sum	.082 [.074;.091]	.100 [.091;.110]	.159 [.148;.170]	.263 [.250;.277]	.443 [.428;.458]	.634 [.620;.650]
	ZCP-poly	.069 [.062;.077]	.089 [.081;.098]	.172 [.161;.184]	.328 [.313;.342]	.551 [.536;.566]	.768 [.755;.781]
	gANOVA	.082 [.073;.090]	.105 [.096;.115]	.193 [.181;.206]	.360 [.345;.375]	.582 [.567;.597]	.791 [.778;.803]

Table E.25: Power analysis of model M2 (see Table E.1): The data are simulated using correlated random effects, without the interaction participants:stimuli. The models are estimated with assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.076 [.069;.085]	.169 [.158;.181]	.390 [.375;.405]	.706 [.692;.720]	.910 [.901;.919]	.990 [.987;.993]
	RI-L	.054 [.047;.061]	.134 [.123;.144]	.346 [.331;.361]	.658 [.644;.673]	.890 [.881;.900]	.984 [.980;.988]
	ZCP-sum	.052 [.045;.059]	.121 [.111;.131]	.315 [.301;.329]	.620 [.605;.635]	.871 [.860;.881]	.978 [.973;.983]
	ZCP-poly	.054 [.048;.062]	.134 [.123;.144]	.344 [.330;.359]	.656 [.641;.670]	.890 [.881;.900]	.984 [.980;.988]
	gANOVA	.054 [.047;.061]	.133 [.122;.143]	.346 [.331;.361]	.658 [.643;.673]	.890 [.881;.900]	.984 [.980;.988]
Vs	RI	.085 [.077;.094]	.174 [.163;.187]	.472 [.457;.488]	.797 [.784;.809]	.955 [.949;.961]	.996 [.993;.998]
	RI-L	.052 [.046;.060]	.121 [.112;.132]	.386 [.371;.401]	.730 [.717;.744]	.939 [.932;.947]	.993 [.990;.995]
	ZCP-sum	.051 [.045;.058]	.104 [.095;.114]	.334 [.319;.349]	.672 [.658;.687]	.914 [.905;.923]	.988 [.984;.991]
	ZCP-poly	.054 [.047;.061]	.121 [.111;.131]	.385 [.370;.401]	.730 [.717;.744]	.940 [.932;.947]	.993 [.990;.996]
	gANOVA	.053 [.046;.060]	.122 [.112;.132]	.387 [.372;.402]	.730 [.717;.744]	.939 [.932;.947]	.993 [.990;.995]
Vm	RI	.536 [.520;.551]	.691 [.677;.705]	.909 [.900;.918]	.989 [.986;.992]	1 [.999;1]	1 [1;1]
	RI-L	.052 [.045;.059]	.132 [.122;.143]	.428 [.413;.444]	.784 [.772;.797]	.962 [.956;.968]	.997 [.995;.999]
	ZCP-sum	.067 [.060;.075]	.108 [.099;.118]	.285 [.272;.300]	.621 [.606;.636]	.880 [.871;.891]	.983 [.979;.987]
	ZCP-poly	.056 [.049;.063]	.139 [.128;.150]	.430 [.415;.445]	.784 [.771;.797]	.958 [.951;.964]	.997 [.995;.999]
	gANOVA	.052 [.045;.059]	.133 [.122;.143]	.428 [.413;.444]	.784 [.772;.797]	.962 [.956;.968]	.997 [.995;.999]
Vp:Vs	RI	.670 [.656;.685]	.736 [.722;.749]	.866 [.856;.877]	.966 [.960;.971]	.995 [.993;.997]	1 [.999;1]
	RI-L	.046 [.040;.053]	.085 [.077;.094]	.205 [.193;.218]	.452 [.437;.467]	.718 [.704;.732]	.916 [.907;.924]
	ZCP-sum	.052 [.045;.059]	.068 [.061;.076]	.128 [.118;.138]	.260 [.247;.274]	.476 [.460;.491]	.714 [.700;.728]
	ZCP-poly	.055 [.048;.063]	.090 [.082;.100]	.216 [.204;.229]	.456 [.441;.472]	.724 [.710;.738]	.911 [.902;.920]
	gANOVA	.046 [.040;.053]	.084 [.076;.093]	.205 [.193;.218]	.451 [.435;.466]	.718 [.704;.732]	.915 [.906;.924]
Vp:Vm	RI	.597 [.582;.613]	.678 [.664;.693]	.840 [.829;.851]	.959 [.953;.965]	.993 [.990;.996]	1 [.999;1]
	RI-L	.046 [.040;.053]	.088 [.079;.097]	.238 [.226;.252]	.498 [.483;.514]	.783 [.770;.796]	.940 [.933;.948]
	ZCP-sum	.067 [.060;.075]	.088 [.079;.097]	.169 [.158;.181]	.333 [.319;.348]	.568 [.553;.584]	.799 [.787;.811]
	ZCP-poly	.054 [.047;.061]	.098 [.089;.107]	.254 [.240;.267]	.505 [.490;.521]	.776 [.763;.789]	.935 [.927;.943]
	gANOVA	.046 [.040;.053]	.088 [.079;.097]	.238 [.226;.252]	.497 [.482;.513]	.782 [.770;.795]	.939 [.932;.947]
Vs:Vm	RI	.543 [.528;.559]	.699 [.685;.713]	.908 [.899;.917]	.989 [.985;.992]	1 [.999;1]	1 [1;1]
	RI-L	.052 [.045;.059]	.121 [.111;.131]	.380 [.365;.395]	.736 [.722;.750]	.950 [.943;.957]	.996 [.994;.998]
	ZCP-sum	.078 [.070;.087]	.116 [.106;.126]	.258 [.245;.272]	.537 [.522;.553]	.822 [.810;.834]	.968 [.963;.974]
	ZCP-poly	.061 [.054;.069]	.130 [.120;.141]	.384 [.369;.399]	.741 [.728;.755]	.950 [.943;.957]	.995 [.993;.997]
	gANOVA	.052 [.045;.059]	.121 [.111;.131]	.380 [.365;.395]	.735 [.722;.749]	.949 [.943;.956]	.996 [.994;.998]
Vp:Vs:Vm	RI	.256 [.243;.270]	.298 [.284;.312]	.448 [.433;.464]	.664 [.650;.679]	.845 [.834;.856]	.959 [.953;.965]
	RI-L	.055 [.049;.063]	.074 [.066;.082]	.138 [.128;.149]	.288 [.274;.302]	.506 [.490;.521]	.740 [.727;.754]
	ZCP-sum	.080 [.072;.089]	.092 [.083;.101]	.132 [.122;.143]	.212 [.199;.225]	.352 [.338;.367]	.531 [.515;.546]
	ZCP-poly	.058 [.052;.066]	.080 [.072;.089]	.144 [.134;.156]	.289 [.276;.304]	.513 [.498;.529]	.730 [.716;.744]
	gANOVA	.055 [.049;.063]	.074 [.066;.082]	.138 [.128;.149]	.288 [.274;.302]	.506 [.491;.522]	.739 [.726;.753]

Table E.26: Power analysis of model M2 (see Table E.1): The data are simulated using correlated random effects, with the interaction participants:stimuli. The models are estimated without assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.080 [.072;.089]	.164 [.152;.175]	.408 [.393;.423]	.723 [.709;.737]	.918 [.909;.926]	.989 [.985;.992]
	RI-L	.058 [.051;.066]	.128 [.118;.139]	.352 [.338;.367]	.674 [.660;.689]	.900 [.890;.909]	.983 [.979;.987]
	ZCP-sum	.053 [.046;.060]	.114 [.105;.124]	.313 [.299;.328]	.634 [.619;.649]	.883 [.873;.893]	.976 [.972;.981]
	ZCP-poly	.059 [.052;.066]	.128 [.118;.139]	.353 [.338;.368]	.673 [.659;.688]	.900 [.890;.909]	.984 [.980;.988]
	gANOVA	.058 [.051;.066]	.128 [.118;.139]	.352 [.338;.367]	.675 [.660;.689]	.900 [.890;.909]	.983 [.979;.987]
Vs	RI	.086 [.077;.095]	.183 [.171;.195]	.476 [.461;.492]	.802 [.789;.814]	.961 [.955;.967]	.996 [.993;.998]
	RI-L	.048 [.042;.055]	.127 [.117;.137]	.386 [.371;.401]	.739 [.726;.753]	.943 [.936;.950]	.993 [.991;.996]
	ZCP-sum	.047 [.041;.054]	.104 [.095;.114]	.326 [.312;.341]	.677 [.663;.692]	.921 [.913;.929]	.989 [.986;.992]
	ZCP-poly	.049 [.042;.056]	.126 [.116;.137]	.389 [.374;.404]	.737 [.723;.750]	.941 [.933;.948]	.993 [.991;.996]
	gANOVA	.048 [.042;.055]	.127 [.117;.137]	.386 [.371;.402]	.740 [.726;.753]	.943 [.936;.950]	.993 [.991;.996]
Vm	RI	.528 [.513;.544]	.679 [.664;.693]	.899 [.890;.908]	.987 [.983;.991]	.998 [.997;1]	1 [1;1]
	RI-L	.049 [.043;.056]	.147 [.137;.159]	.446 [.431;.462]	.794 [.782;.807]	.963 [.957;.969]	.996 [.995;.998]
	ZCP-sum	.068 [.060;.076]	.119 [.109;.129]	.315 [.301;.329]	.629 [.614;.644]	.890 [.880;.899]	.982 [.978;.986]
	ZCP-poly	.054 [.047;.061]	.151 [.140;.162]	.447 [.432;.463]	.796 [.784;.809]	.964 [.959;.970]	.996 [.995;.998]
	gANOVA	.048 [.042;.056]	.147 [.136;.158]	.446 [.430;.461]	.794 [.782;.807]	.963 [.957;.969]	.996 [.995;.998]
Vp:Vs	RI	.834 [.823;.846]	.877 [.867;.887]	.957 [.950;.963]	.990 [.987;.993]	1 [1;1]	1 [1;1]
	RI-L	.050 [.044;.058]	.092 [.084;.102]	.234 [.221;.248]	.490 [.475;.505]	.774 [.761;.787]	.936 [.928;.943]
	ZCP-sum	.049 [.043;.056]	.064 [.057;.072]	.123 [.113;.134]	.253 [.240;.267]	.472 [.457;.487]	.719 [.705;.733]
	ZCP-poly	.059 [.052;.067]	.096 [.087;.106]	.245 [.232;.259]	.492 [.477;.508]	.768 [.755;.781]	.934 [.927;.942]
	gANOVA	.050 [.044;.058]	.092 [.084;.102]	.234 [.221;.247]	.489 [.474;.505]	.774 [.761;.787]	.935 [.928;.943]
Vp:Vm	RI	.555 [.540;.571]	.647 [.633;.662]	.833 [.822;.845]	.958 [.952;.964]	.996 [.994;.998]	1 [1;1]
	RI-L	.053 [.046;.060]	.092 [.084;.102]	.248 [.235;.262]	.519 [.504;.535]	.787 [.775;.800]	.948 [.941;.955]
	ZCP-sum	.069 [.061;.077]	.087 [.078;.096]	.166 [.155;.178]	.346 [.331;.361]	.573 [.558;.589]	.809 [.797;.821]
	ZCP-poly	.058 [.051;.066]	.102 [.093;.112]	.254 [.241;.268]	.519 [.503;.534]	.782 [.770;.795]	.945 [.938;.952]
	gANOVA	.052 [.046;.060]	.093 [.084;.102]	.248 [.235;.262]	.519 [.503;.534]	.786 [.774;.799]	.946 [.940;.953]
Vs:Vm	RI	.500 [.485;.516]	.655 [.640;.670]	.882 [.872;.892]	.985 [.982;.989]	1 [.999;1]	1 [1;1]
	RI-L	.048 [.042;.055]	.122 [.112;.133]	.395 [.380;.410]	.745 [.731;.758]	.948 [.941;.954]	.996 [.994;.998]
	ZCP-sum	.066 [.059;.074]	.107 [.098;.117]	.251 [.238;.265]	.543 [.528;.559]	.820 [.808;.832]	.960 [.954;.966]
	ZCP-poly	.057 [.050;.064]	.130 [.120;.141]	.401 [.386;.417]	.740 [.726;.754]	.943 [.936;.950]	.995 [.993;.997]
	gANOVA	.048 [.042;.055]	.122 [.112;.133]	.395 [.380;.410]	.744 [.731;.758]	.947 [.940;.954]	.996 [.994;.998]
Vp:Vs:Vm	RI	.192 [.181;.205]	.240 [.227;.253]	.371 [.356;.386]	.583 [.568;.599]	.804 [.792;.816]	.939 [.932;.946]
	RI-L	.058 [.051;.066]	.076 [.068;.085]	.150 [.140;.162]	.296 [.283;.311]	.516 [.501;.532]	.749 [.735;.762]
	ZCP-sum	.087 [.078;.096]	.099 [.090;.109]	.141 [.131;.152]	.229 [.216;.242]	.353 [.338;.368]	.528 [.513;.543]
	ZCP-poly	.063 [.056;.071]	.083 [.075;.092]	.159 [.148;.171]	.306 [.292;.321]	.525 [.510;.541]	.752 [.739;.766]
	gANOVA	.058 [.052;.066]	.076 [.068;.085]	.150 [.140;.162]	.296 [.283;.311]	.516 [.501;.532]	.748 [.735;.762]

Table E.27: Power analysis of model M2 (see Table E.1): The data are simulated using correlated random effects, with the interaction participants:stimuli. The models are estimated with assuming random effects associated to the interaction participants:stimuli.

variable	model	H0	0.2	0.4	0.6	0.8	1.0
Vp	RI	.080 [.072;.089]	.164 [.152;.175]	.408 [.393;.423]	.723 [.709;.737]	.918 [.909;.926]	.989 [.985;.992]
	RI-L	.059 [.052;.066]	.128 [.118;.139]	.353 [.338;.368]	.675 [.660;.689]	.900 [.890;.909]	.983 [.979;.987]
	ZCP-sum	.053 [.047;.061]	.114 [.105;.125]	.314 [.300;.328]	.635 [.620;.650]	.883 [.873;.893]	.976 [.972;.981]
	ZCP-poly	.058 [.052;.066]	.129 [.119;.139]	.353 [.339;.368]	.673 [.659;.688]	.900 [.890;.909]	.984 [.980;.988]
	gANOVA	.058 [.051;.066]	.128 [.118;.139]	.353 [.338;.368]	.675 [.661;.690]	.900 [.890;.909]	.983 [.979;.987]
Vs	RI	.086 [.077;.095]	.183 [.171;.195]	.476 [.461;.492]	.802 [.789;.814]	.961 [.955;.967]	.996 [.993;.998]
	RI-L	.048 [.042;.055]	.127 [.117;.138]	.386 [.372;.402]	.740 [.727;.754]	.943 [.936;.950]	.993 [.991;.996]
	ZCP-sum	.047 [.041;.054]	.104 [.095;.114]	.326 [.312;.341]	.679 [.664;.693]	.922 [.914;.930]	.989 [.986;.992]
	ZCP-poly	.049 [.043;.056]	.126 [.116;.137]	.390 [.375;.405]	.737 [.724;.751]	.941 [.934;.948]	.994 [.991;.996]
	gANOVA	.048 [.042;.055]	.127 [.117;.138]	.387 [.372;.402]	.740 [.727;.754]	.943 [.936;.950]	.993 [.991;.996]
Vm	RI	.575 [.560;.591]	.710 [.696;.724]	.914 [.906;.923]	.989 [.986;.992]	.999 [.998;1]	1 [1;1]
	RI-L	.048 [.042;.055]	.147 [.137;.159]	.445 [.430;.461]	.794 [.781;.806]	.963 [.957;.969]	.996 [.995;.998]
	ZCP-sum	.068 [.060;.076]	.118 [.108;.128]	.314 [.300;.328]	.628 [.613;.643]	.888 [.879;.898]	.982 [.978;.986]
	ZCP-poly	.053 [.047;.061]	.150 [.140;.162]	.446 [.431;.462]	.796 [.784;.809]	.964 [.959;.970]	.996 [.995;.998]
	gANOVA	.048 [.042;.055]	.147 [.136;.158]	.444 [.429;.460]	.794 [.781;.806]	.963 [.957;.969]	.996 [.995;.998]
Vp:Vs	RI	.762 [.749;.775]	.826 [.814;.837]	.932 [.924;.940]	.982 [.978;.986]	.999 [.998;1]	1 [1;1]
	RI-L	.051 [.044;.058]	.093 [.084;.102]	.235 [.223;.249]	.492 [.477;.508]	.775 [.762;.788]	.935 [.928;.943]
	ZCP-sum	.050 [.043;.057]	.065 [.058;.073]	.124 [.114;.135]	.255 [.241;.268]	.474 [.459;.490]	.721 [.707;.735]
	ZCP-poly	.059 [.052;.067]	.097 [.088;.106]	.247 [.234;.261]	.494 [.478;.509]	.770 [.757;.783]	.935 [.928;.943]
	gANOVA	.051 [.044;.058]	.093 [.084;.102]	.235 [.222;.249]	.492 [.477;.507]	.775 [.762;.788]	.935 [.928;.943]
Vp:Vm	RI	.609 [.594;.625]	.696 [.682;.711]	.866 [.856;.877]	.968 [.963;.974]	.996 [.995;.998]	1 [1;1]
	RI-L	.052 [.046;.059]	.093 [.084;.102]	.247 [.234;.260]	.517 [.502;.533]	.786 [.773;.798]	.948 [.941;.955]
	ZCP-sum	.068 [.060;.076]	.087 [.078;.096]	.165 [.154;.177]	.343 [.329;.358]	.570 [.555;.586]	.807 [.795;.819]
	ZCP-poly	.057 [.050;.065]	.101 [.092;.111]	.253 [.240;.267]	.517 [.502;.533]	.780 [.768;.793]	.944 [.937;.951]
	gANOVA	.051 [.045;.059]	.092 [.084;.102]	.246 [.234;.260]	.517 [.501;.532]	.785 [.772;.798]	.946 [.940;.953]
Vs:Vm	RI	.574 [.558;.589]	.703 [.689;.717]	.905 [.896;.914]	.988 [.985;.991]	1 [.999;1]	1 [1;1]
	RI-L	.048 [.041;.055]	.121 [.111;.131]	.394 [.379;.409]	.744 [.731;.758]	.947 [.940;.954]	.996 [.994;.998]
	ZCP-sum	.065 [.058;.073]	.106 [.097;.116]	.249 [.236;.263]	.541 [.526;.557]	.817 [.805;.829]	.958 [.952;.964]
	ZCP-poly	.056 [.050;.064]	.129 [.119;.140]	.400 [.385;.415]	.739 [.725;.752]	.942 [.935;.949]	.995 [.993;.997]
	gANOVA	.048 [.042;.055]	.120 [.111;.131]	.393 [.378;.409]	.744 [.731;.758]	.946 [.940;.953]	.996 [.994;.998]
Vp:Vs:Vm	RI	.270 [.257;.285]	.323 [.309;.338]	.462 [.447;.477]	.679 [.664;.693]	.863 [.853;.874]	.962 [.956;.968]
	RI-L	.057 [.050;.065]	.075 [.067;.083]	.148 [.138;.160]	.293 [.280;.308]	.512 [.497;.528]	.746 [.733;.760]
	ZCP-sum	.084 [.076;.093]	.098 [.089;.108]	.139 [.129;.150]	.225 [.213;.239]	.350 [.335;.365]	.520 [.505;.536]
	ZCP-poly	.062 [.055;.070]	.082 [.074;.091]	.156 [.146;.168]	.302 [.288;.317]	.520 [.505;.536]	.749 [.735;.762]
	gANOVA	.057 [.050;.065]	.075 [.067;.084]	.148 [.138;.160]	.293 [.279;.307]	.512 [.497;.528]	.746 [.733;.760]

E.3 Results of Simulation: gANOVA vs RI-L

Table E.28: Type I error rate of design M1 where the data are generated without random intercepts. Subset of data simulated using 18 stimuli.

			RI-L	RI-L+	gANOVA	gANOVA+
Vp	corr.	no PT:SM	*.023 [.019;.028]*	*.023 [.019;.028]*	**.045 [.039;.052]**	**.045 [.039;.052]**
		PT:SM	*.026 [.021;.031]*	*.026 [.021;.031]*	**.051 [.045;.058]**	**.051 [.045;.059]**
	spheric.	no PT:SM	*.018 [.015;.023]*	*.018 [.015;.023]*	**.048 [.042;.055]**	**.048 [.042;.055]**
		PT:SM	*.021 [.017;.026]*	*.021 [.017;.026]*	**.044 [.038;.051]**	**.044 [.038;.051]**
Vs	corr.	no PT:SM	*.034 [.029;.040]*	*.034 [.029;.040]*	**.051 [.045;.059]**	**.051 [.045;.059]**
		PT:SM	*.040 [.035;.047]*	*.040 [.034;.047]*	**.051 [.045;.059]**	**.051 [.045;.059]**
	spheric.	no PT:SM	*.036 [.030;.042]*	*.036 [.030;.042]*	**.052 [.045;.059]**	**.052 [.045;.059]**
		PT:SM	*.042 [.036;.048]*	*.041 [.036;.048]*	**.050 [.043;.057]**	**.050 [.043;.057]**
Vm	corr.	no PT:SM	.120 [.110;.131]	.120 [.110;.131]	**.050 [.044;.058]**	**.050 [.044;.058]**
		PT:SM	.115 [.106;.126]	.115 [.106;.125]	**.051 [.044;.058]**	**.050 [.044;.058]**
	spheric.	no PT:SM	.113 [.104;.124]	.113 [.104;.124]	**.050 [.044;.058]**	**.050 [.044;.058]**
		PT:SM	.105 [.096;.115]	.105 [.096;.115]	**.051 [.045;.059]**	**.051 [.045;.058]**
Vp:Vs	corr.	no PT:SM	.112 [.102;.122]	.112 [.102;.122]	**.048 [.042;.055]**	**.048 [.042;.055]**
		PT:SM	.106 [.097;.116]	.106 [.097;.116]	**.044 [.039;.051]**	**.044 [.039;.051]**
	spheric.	no PT:SM	.109 [.100;.119]	.109 [.100;.119]	**.049 [.043;.056]**	**.049 [.043;.056]**
		PT:SM	.110 [.100;.120]	.109 [.100;.119]	**.050 [.043;.057]**	**.050 [.043;.057]**
Vp:Vm	corr.	no PT:SM	.082 [.074;.091]	.082 [.074;.091]	**.051 [.045;.058]**	**.051 [.045;.058]**
		PT:SM	.079 [.071;.088]	.079 [.071;.088]	**.050 [.044;.058]**	**.050 [.044;.057]**
	spheric.	no PT:SM	.079 [.071;.088]	.079 [.071;.088]	**.052 [.046;.059]**	**.052 [.046;.059]**
		PT:SM	.080 [.072;.088]	.079 [.071;.088]	**.050 [.044;.058]**	**.050 [.044;.057]**
Vs:Vm	corr.	no PT:SM	.066 [.059;.074]	.066 [.059;.074]	**.044 [.038;.051]**	**.044 [.038;.051]**
		PT:SM	.068 [.061;.076]	.068 [.061;.076]	**.047 [.041;.054]**	**.047 [.041;.054]**
	spheric.	no PT:SM	.083 [.075;.092]	.083 [.075;.092]	.060 [.053;.068]	.060 [.053;.068]
		PT:SM	.083 [.075;.092]	.083 [.075;.092]	.060 [.053;.067]	.060 [.053;.067]
Vp:Vs:Vm	corr.	no PT:SM	*.036 [.031;.043]*	*.036 [.031;.043]*	*.042 [.037;.049]*	*.042 [.037;.049]*
		PT:SM	*.038 [.033;.045]*	*.038 [.033;.045]*	**.051 [.045;.059]**	**.051 [.045;.058]**
	spheric.	no PT:SM	*.035 [.030;.041]*	*.035 [.030;.041]*	*.044 [.038;.050]*	*.044 [.038;.050]*
		PT:SM	*.035 [.030;.041]*	*.035 [.030;.041]*	**.045 [.039;.052]**	**.045 [.039;.052]**

Table E.29: Type I error rate of design M1 where the data are generated without random intercepts. Subset of data simulated using 36 stimuli.

			RI-L	RI-L+	gANOVA	gANOVA+
Vp	corr.	no PT:SM	*.005 [.003;.008]*	*.005 [.003;.008]*	**.046 [.040;.054]**	**.046 [.040;.054]**
		PT:SM	*.006 [.004;.008]*	*.006 [.004;.008]*	**.049 [.043;.056]**	**.049 [.043;.056]**
	spheric.	no PT:SM	*.006 [.004;.008]*	*.006 [.004;.008]*	**.051 [.045;.058]**	**.051 [.045;.058]**
		PT:SM	*.005 [.003;.008]*	*.005 [.003;.008]*	*.039 [.033;.045]*	*.039 [.033;.045]*
Vs	corr.	no PT:SM	.064 [.057;.072]	.064 [.057;.072]	**.048 [.042;.056]**	**.048 [.042;.056]**
		PT:SM	.066 [.059;.074]	.066 [.058;.074]	**.050 [.044;.057]**	**.050 [.044;.058]**
	spheric.	no PT:SM	.061 [.054;.069]	.062 [.054;.069]	**.048 [.041;.055]**	**.048 [.041;.055]**
		PT:SM	.063 [.056;.071]	.062 [.055;.070]	**.047 [.041;.054]**	**.047 [.041;.054]**
Vm	corr.	no PT:SM	.119 [.109;.129]	.119 [.109;.129]	**.047 [.041;.054]**	**.047 [.041;.054]**
		PT:SM	.118 [.108;.128]	.118 [.108;.128]	**.050 [.043;.057]**	**.050 [.043;.057]**
	spheric.	no PT:SM	.117 [.107;.127]	.117 [.107;.127]	**.046 [.040;.054]**	**.046 [.040;.054]**
		PT:SM	.107 [.098;.117]	.107 [.098;.117]	**.051 [.044;.058]**	**.051 [.044;.058]**
Vp:Vs	corr.	no PT:SM	.114 [.105;.125]	.114 [.105;.125]	**.052 [.045;.059]**	**.052 [.045;.059]**
		PT:SM	.114 [.104;.124]	.113 [.104;.123]	**.050 [.044;.058]**	**.050 [.044;.058]**
	spheric.	no PT:SM	.113 [.103;.123]	.113 [.103;.123]	**.047 [.041;.054]**	**.047 [.041;.054]**
		PT:SM	.111 [.101;.121]	.110 [.101;.121]	**.050 [.044;.058]**	**.050 [.044;.058]**
Vp:Vm	corr.	no PT:SM	.095 [.086;.104]	.095 [.086;.104]	**.051 [.045;.058]**	**.051 [.045;.058]**
		PT:SM	.099 [.090;.109]	.099 [.090;.108]	**.051 [.045;.059]**	**.051 [.045;.059]**
	spheric.	no PT:SM	.092 [.083;.101]	.092 [.083;.101]	**.052 [.045;.059]**	**.052 [.045;.059]**
		PT:SM	.089 [.081;.099]	.089 [.081;.099]	**.052 [.046;.060]**	**.052 [.046;.060]**
Vs:Vm	corr.	no PT:SM	.056 [.050;.064]	.056 [.050;.064]	**.048 [.042;.055]**	**.048 [.042;.055]**
		PT:SM	**.055 [.049;.063]**	**.055 [.049;.063]**	**.048 [.041;.055]**	**.048 [.041;.055]**
	spheric.	no PT:SM	.059 [.052;.067]	.059 [.052;.067]	**.048 [.042;.055]**	**.048 [.042;.055]**
		PT:SM	.058 [.052;.066]	.058 [.051;.066]	**.051 [.045;.058]**	**.051 [.045;.058]**
Vp:Vs:Vm	corr.	no PT:SM	*.040 [.035;.047]*	*.040 [.034;.047]*	**.052 [.045;.059]**	**.052 [.045;.059]**
		PT:SM	*.037 [.032;.043]*	*.037 [.032;.043]*	**.051 [.045;.059]**	**.051 [.045;.059]**
	spheric.	no PT:SM	*.033 [.028;.039]*	*.033 [.028;.039]*	**.047 [.041;.054]**	**.047 [.041;.054]**
		PT:SM	*.032 [.027;.038]*	*.032 [.027;.038]*	**.048 [.042;.056]**	**.048 [.042;.055]**

Appendix F

Supplementary Simulation Results for Chapter 5

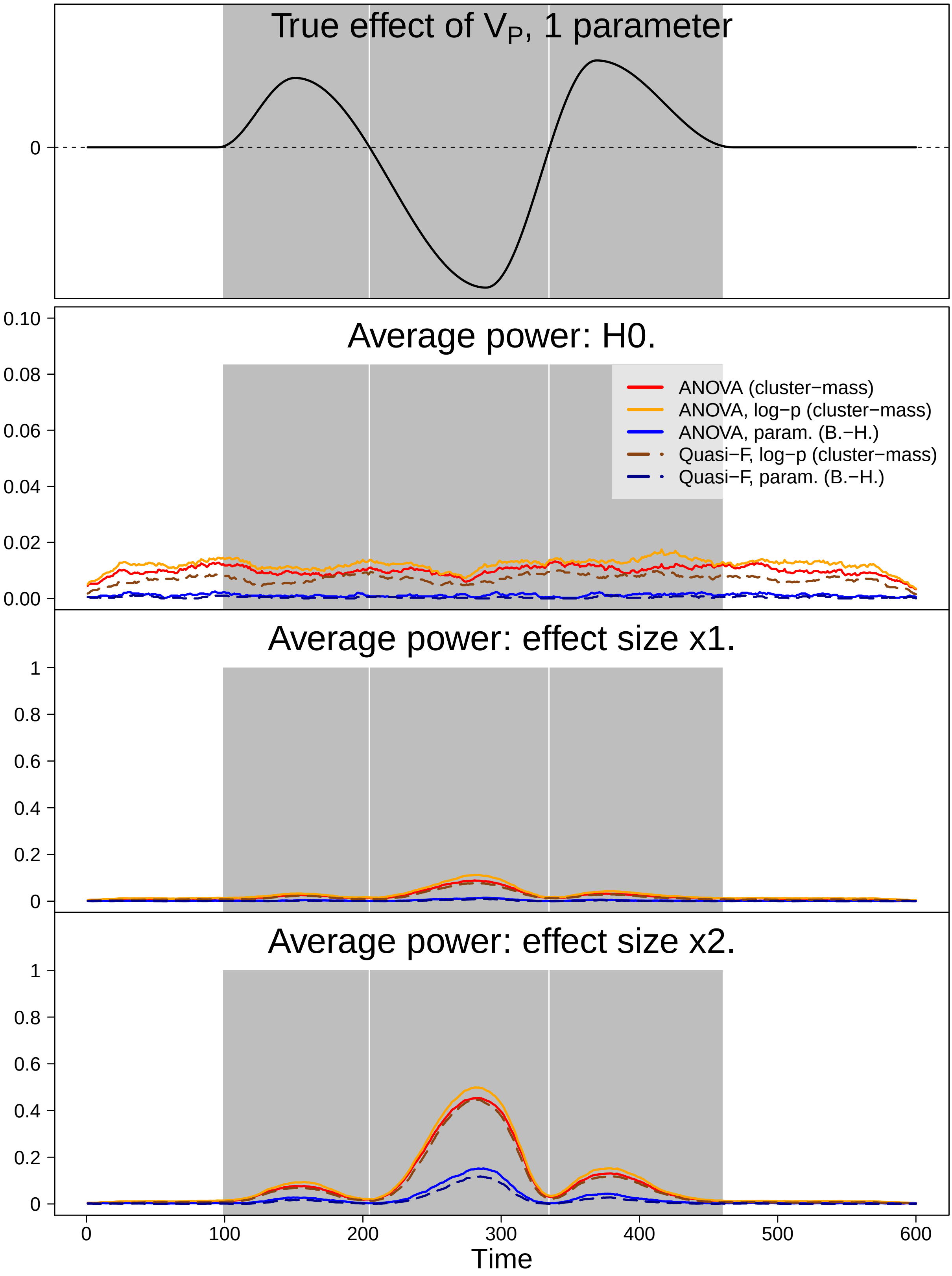

Figure F.1: Average power of the test of V_M for the model with 9 stimuli and 21 participants.

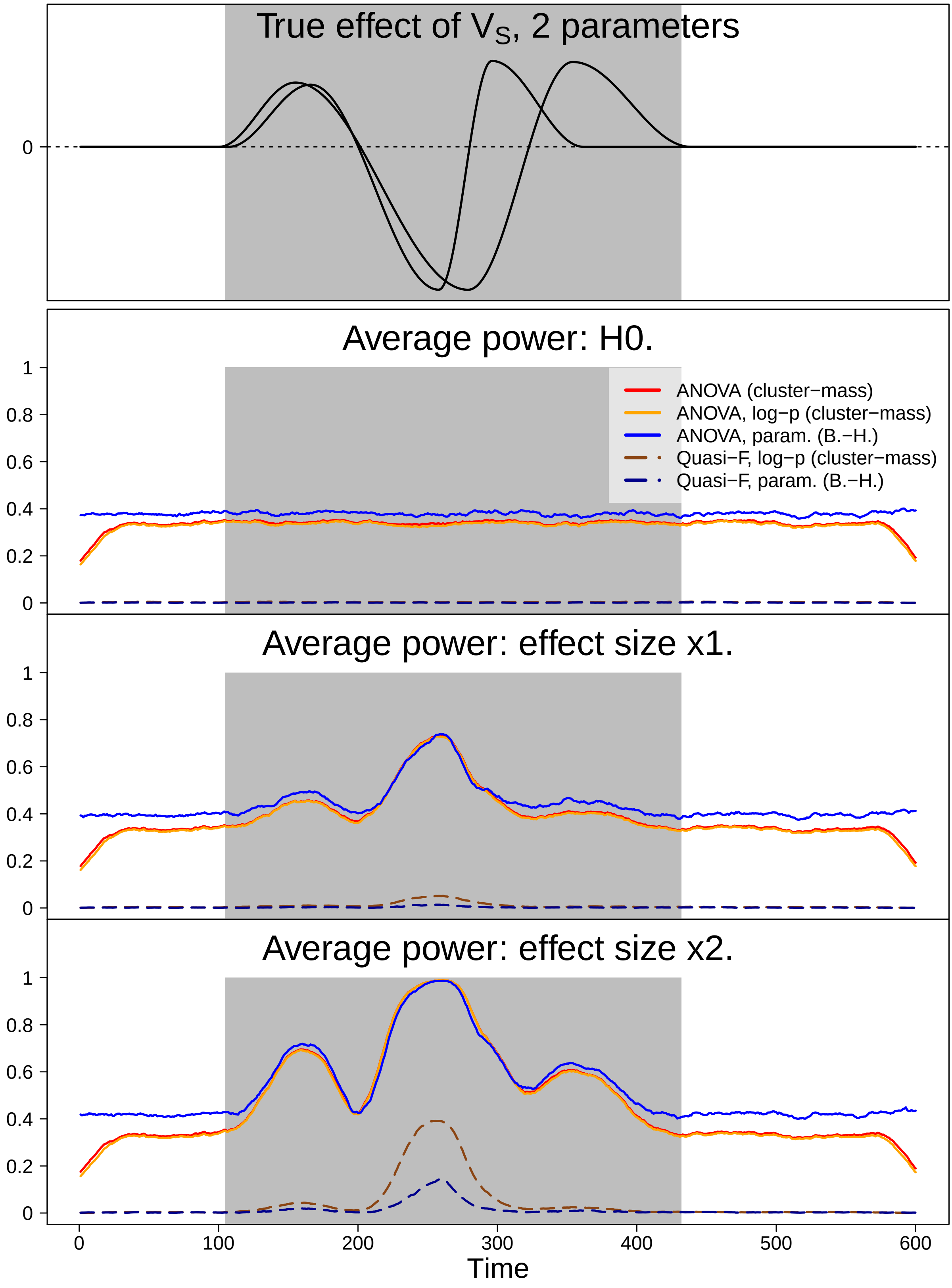

Figure F.2: Average power of the test of V_M for the model with 9 stimuli and 21 participants.

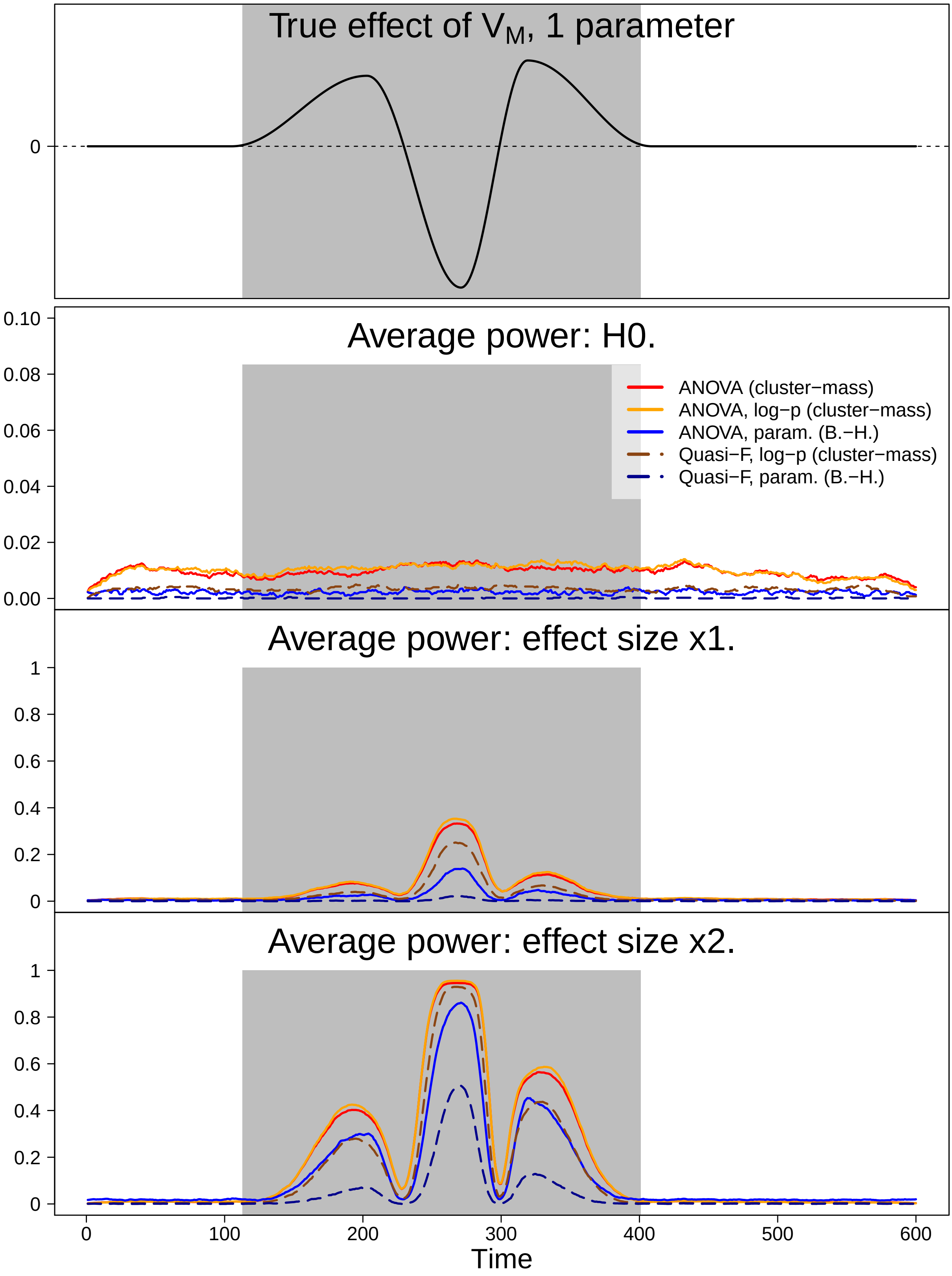

Figure F.3: Average power of the test of V_M for the model with 9 stimuli and 21 participants.

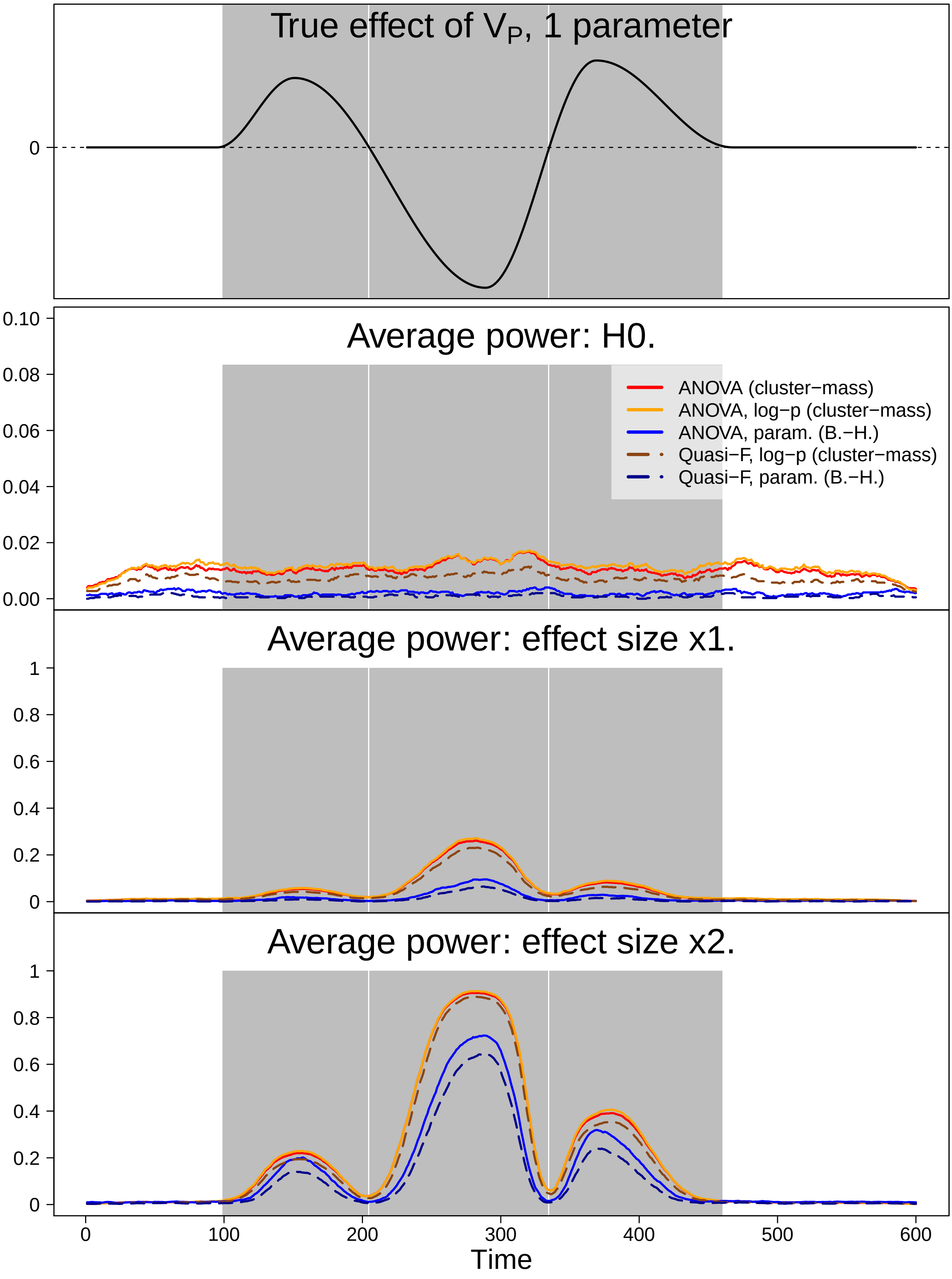

Figure F.4: Average power of the test of V_M for the model with 18 stimuli and 20 participants.

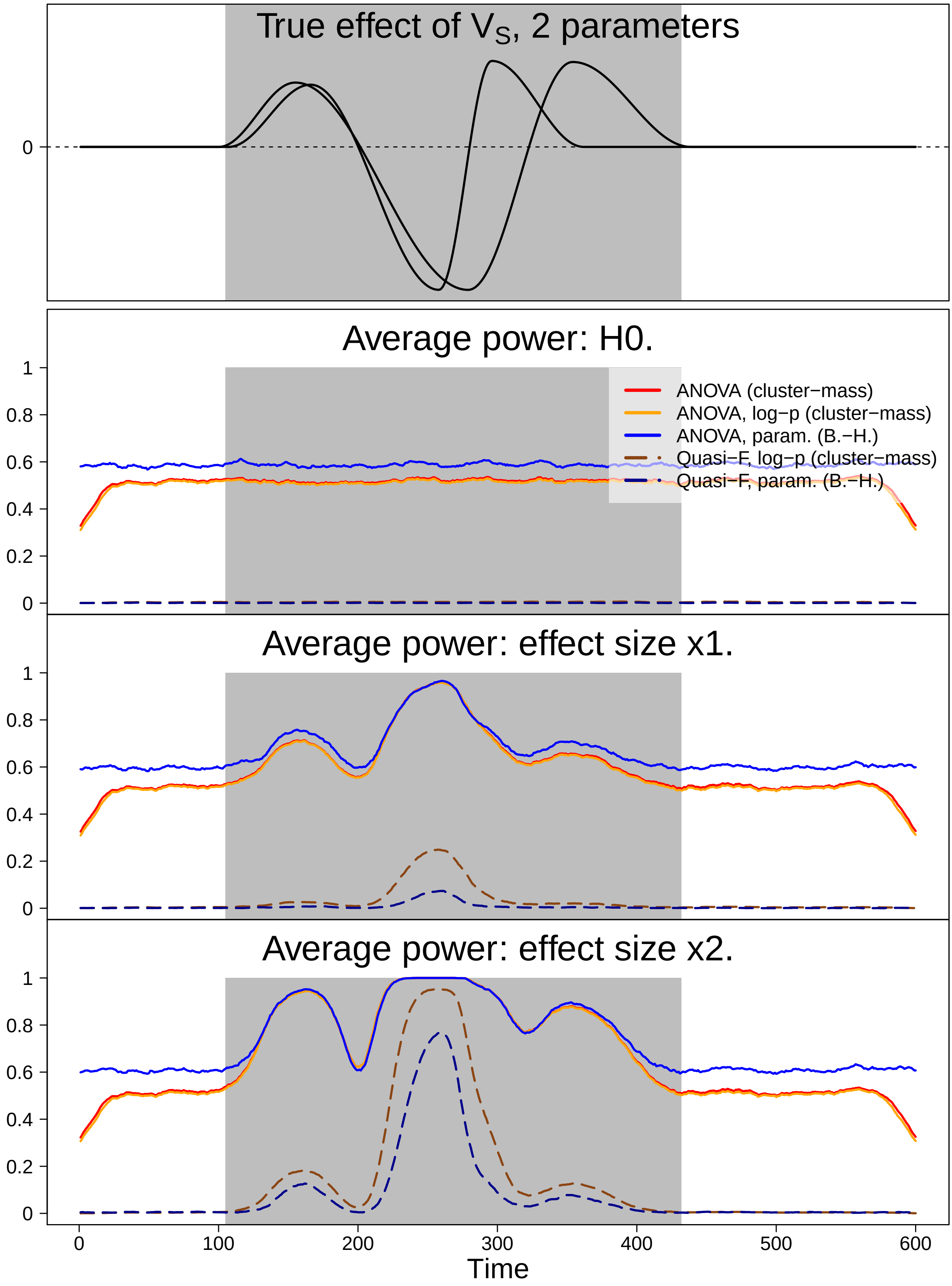

Figure F.5: Average power of the test of V_S for the model with 18 stimuli and 20 participants.

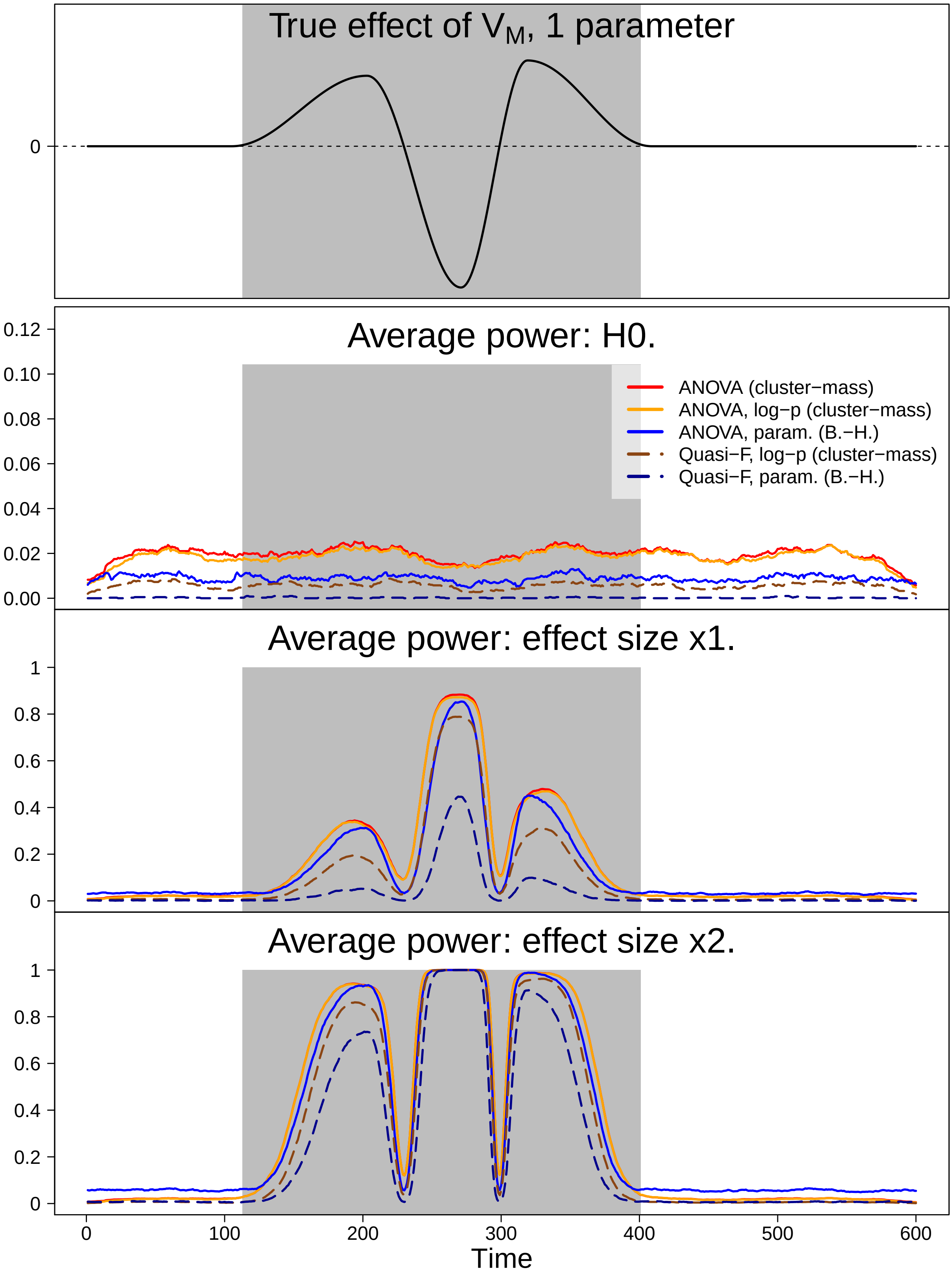

Figure F.6: Average power of the test of V_M for the model with 18 stimuli and 20 participants.

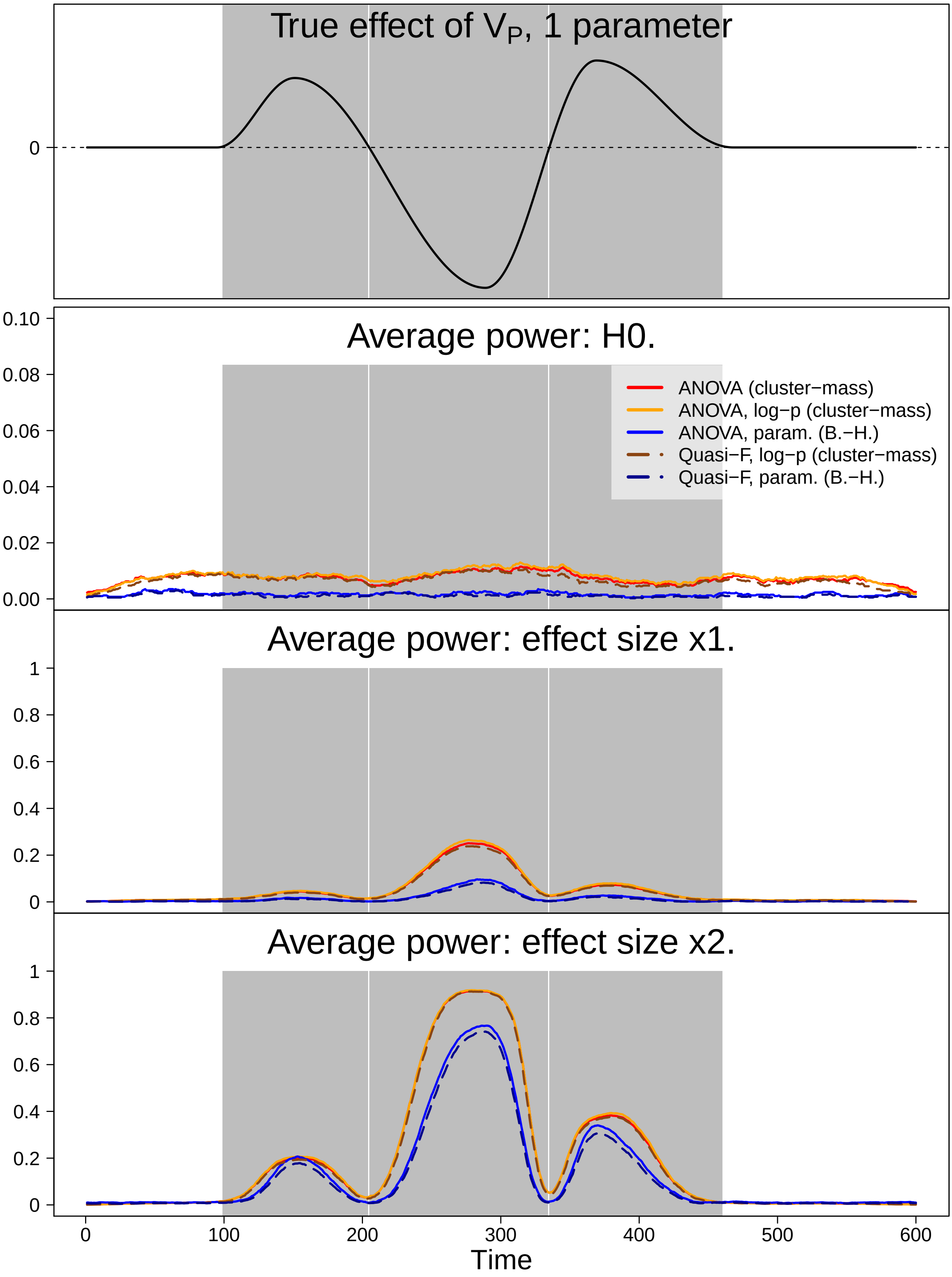

Figure F.7: Average power of the test of V_P for the model with 36 stimuli and 20 participants.

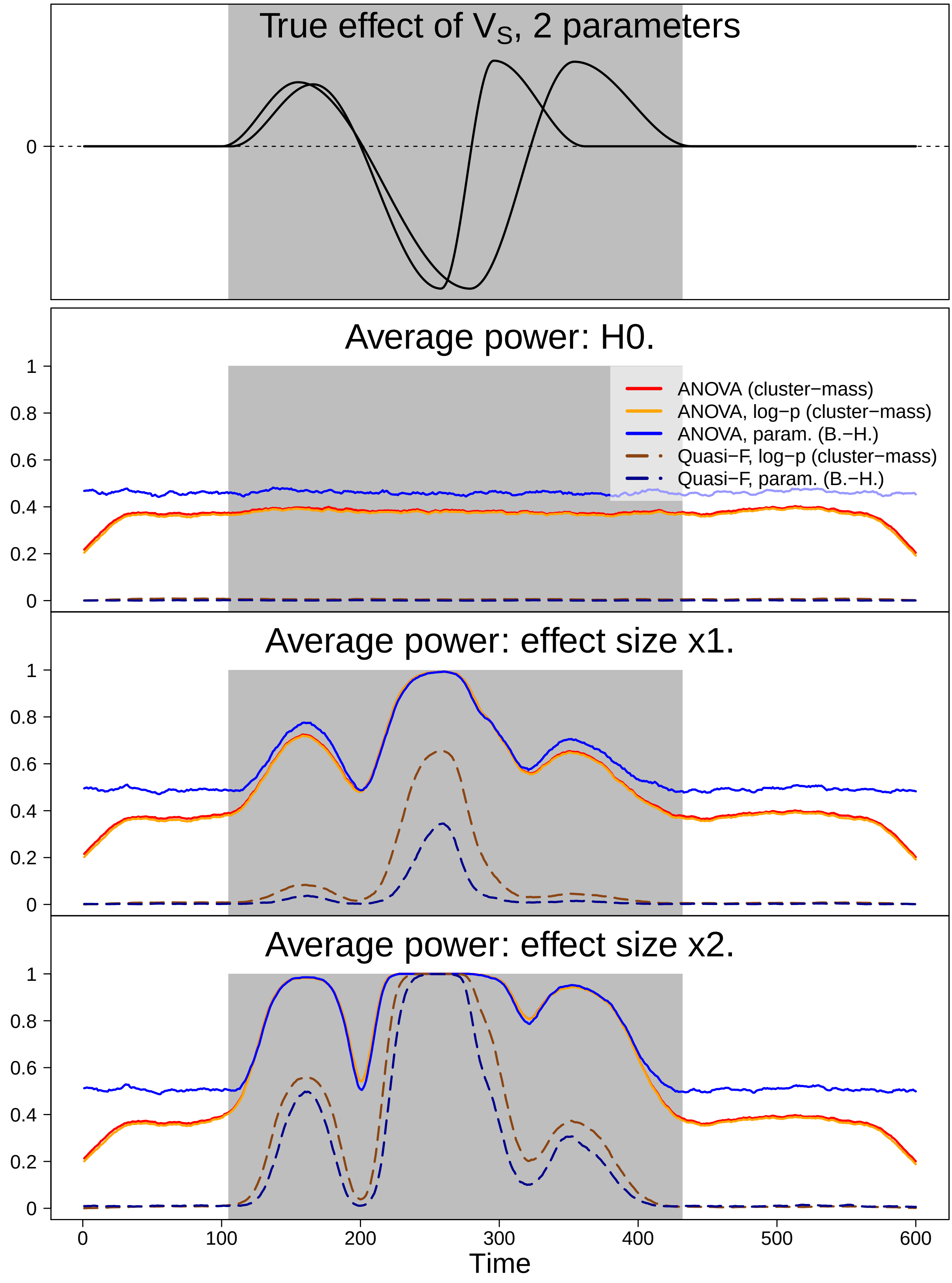

Figure F.8: Average power of the test of V_S for the model with 36 stimuli and 20 participants.

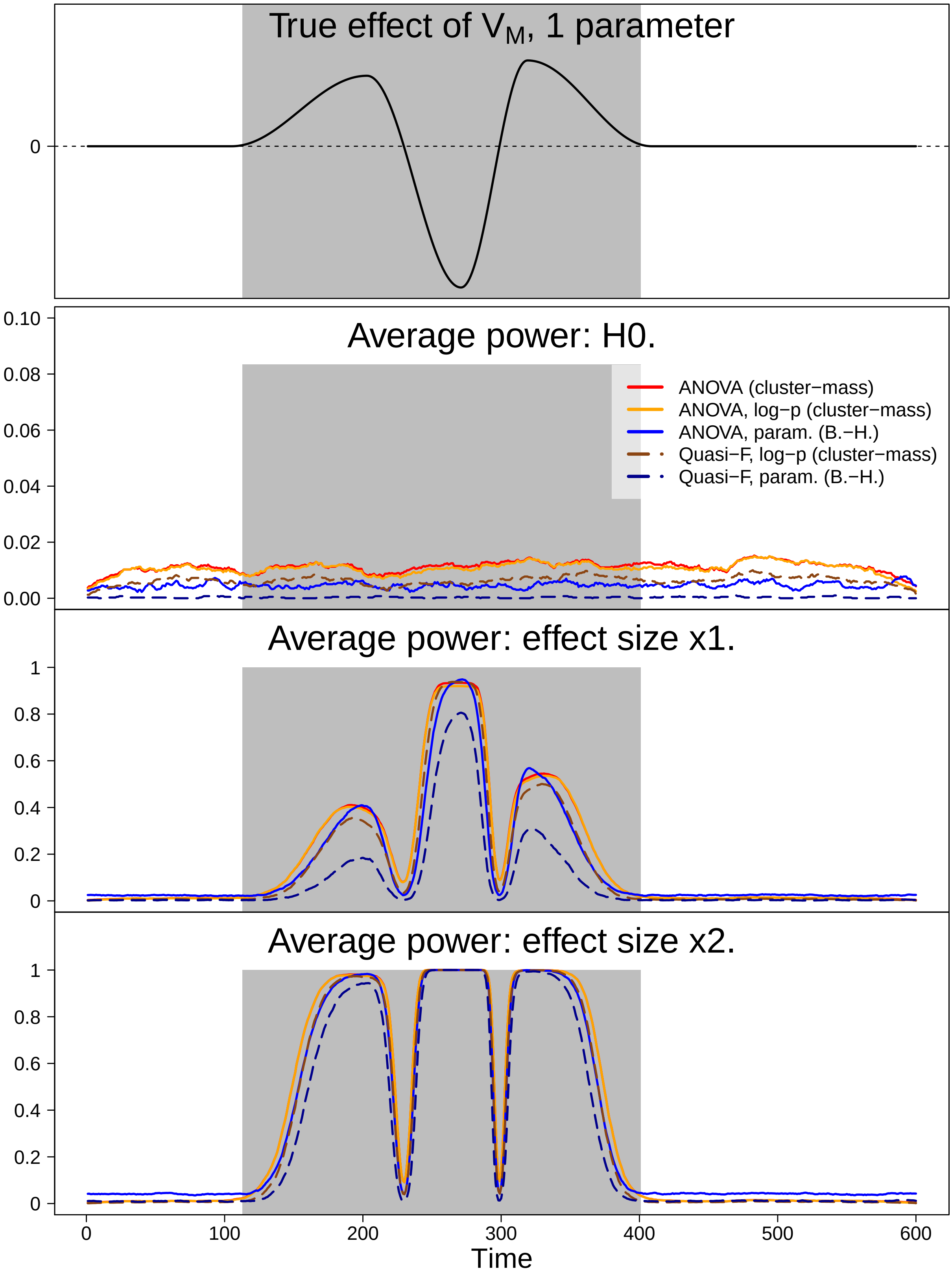

Figure F.9: Average power of the test of V_M for the model with 36 stimuli and 20 participants.

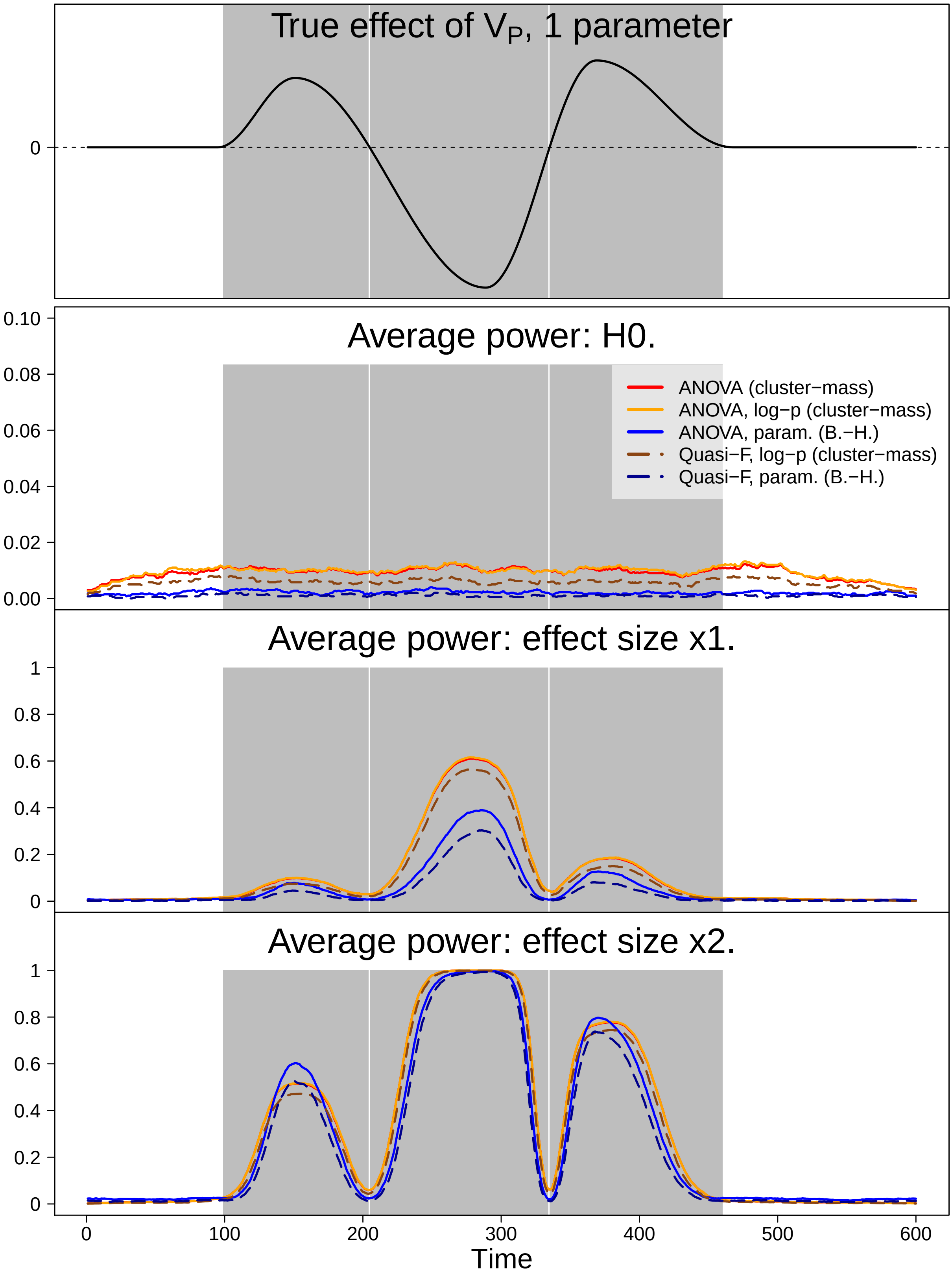

Figure F.10: Average power of the test of V_P for the model with 36 stimuli and 40 participants.

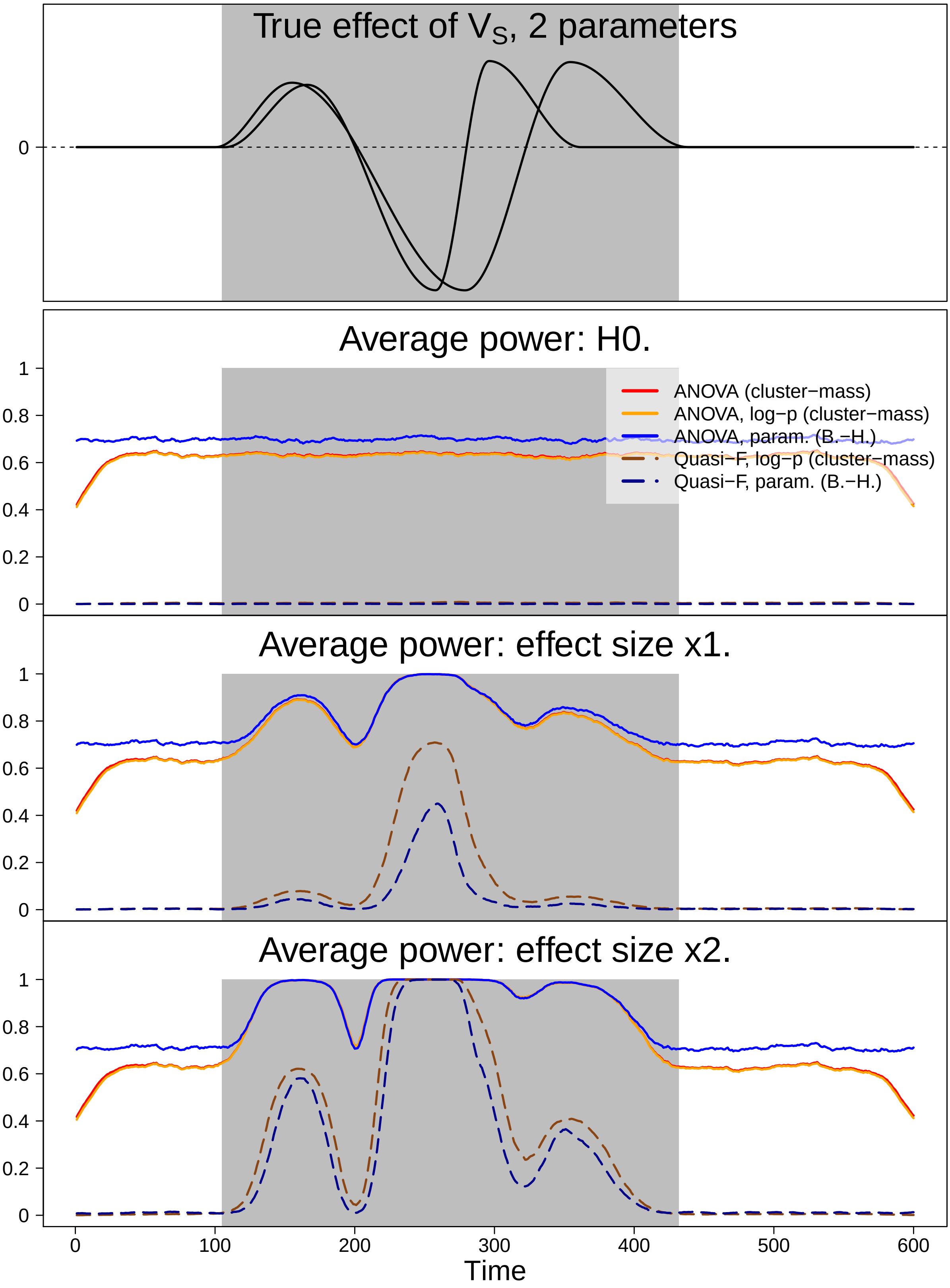

Figure F.11: Average power of the test of V_S for the model with 36 stimuli and 40 participants.

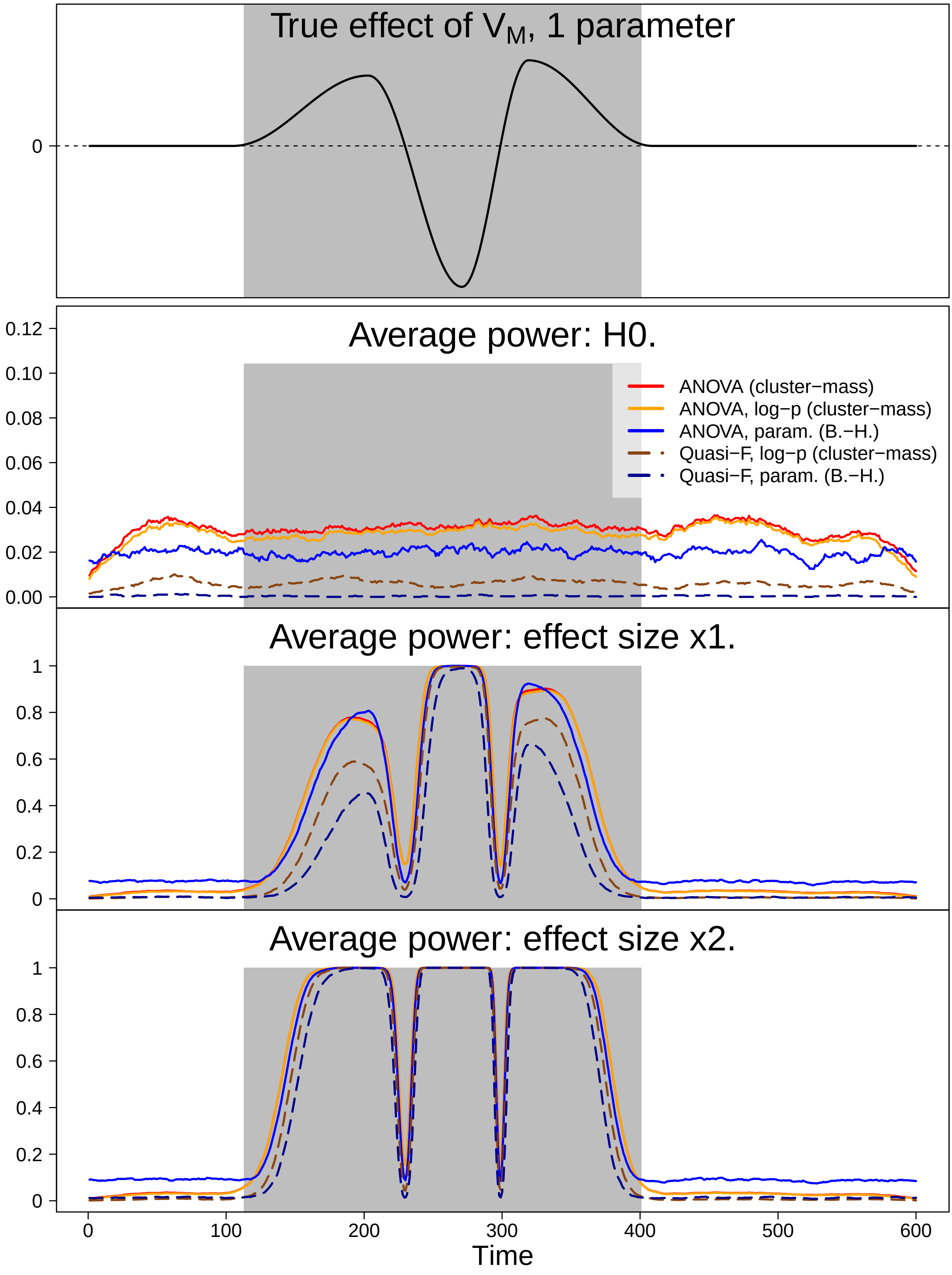

Figure F.12: Average power of the test of V_M for the model with 36 stimuli and 40 participants.

Bibliography

Abrahamsen, P. (1997). *A Review of Gaussian Random Fields and Correlation Functions.* Norsk Regnesentral/Norwegian Computing Center.

Agresti, A. and B. A. Coull (1998). Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician 52*(2), 119–126.

Allen, S. L., R. Bonduriansky, and S. F. Chenoweth (2018). Genetic constraints on microevolutionary divergence of sex-biased gene expression. *Philosophical Transactions of the Royal Society B: Biological Sciences 373*(1757), 20170427.

Anderson, M. and C. T. Braak (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation 73*(2), 85–113.

Anderson, M. J. and P. Legendre (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of statistical computation and simulation 62*(3), 271–303.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language 68*(3), 255–278.

Basso, D. and L. Finos (2012). Exact Multivariate Permutation Tests for Fixed Effects in Mixed-Models. *Communications in Statistics - Theory and Methods 41*(16-17), 2991–3001.

Bates, D. and S. DebRoy (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis 91*(1), 1–17.

Bates, D., R. Kliegl, S. Vasishth, and H. Baayen (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software 67*(1), 1–48.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B 57*(1), 289–300.

Boisgontier, M. P. and B. Cheval (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews 68*, 1004–1005.

Box, G. E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *The Annals of Mathematical Statistics 25*(2), 290–302.

Box, G. E. P. and N. R. Draper (1987). *Empirical Model-Building and Response Surfaces.* John Wiley & Sons.

Brown, M. B. and A. B. Forsythe (1974). 372: The Anova and Multiple Comparisons for Data with Heterogeneous Variances. *Biometrics 30*(4), 719–724.

Bullmore, E. T., J. Suckling, S. Overmeyer, S. Rabe-Hesketh, E. Taylor, and M. J. Brammer (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE transactions on medical imaging 18*(1), 32–42.

Bürki, A., J. Frossard, and O. Renaud (2018). Accounting for stimulus and participant effects in event-related potential analyses to increase the replicability of studies. *Journal of Neuroscience Methods 309*, 218–227.

Cabrera, J. L. O. (2018). Locpol: Kernel Local Polynomial Regression. R-Packages.

Cardinal, R. N. and M. R. Aitken (2013). *ANOVA for the Behavioral Sciences Researcher.* Psychology Press.

Carpenter, J. R., H. Goldstein, and J. Rasbash (2003). A Novel Bootstrap Procedure for Assessing the Relationship between Class Size and Achievement. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 52*(4), 431–443.

Chambers, J. M. (2009). *Software for Data Analysis: Programming with R.* Springer-Verlag.

Cheval, B., P. Sarrazin, L. Pelletier, and M. Friese (2016). Effect of Retraining Approach-Avoidance Tendencies on an Exercise Task: A Randomized Controlled Trial. *Journal of Physical Activity and Health 13*(12), 1396–1403.

Cheval, B., E. Tipura, N. Burra, J. Frossard, J. Chanal, D. Orsholits, R. Radel, and M. P. Boisgontier (2018). Avoiding sedentary behaviors requires more cortical resources than avoiding physical activity: An EEG study. *Neuropsychologia 119*, 68–80.

Chung, E. and J. P. Romano (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics 41*(2), 484–507.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior 12*(4), 335–359.

Cornfield, J. and J. Tukey (1956). Average Values of Mean Squares in Factorials. *The Annals of Mathematical Statistics 27*(4), 42.

Cox, D. R. and D. V. Hinkley (1979). *Theoretical Statistics.* Chapman and Hall/CRC.

Csardi, G. and T. Nepusz (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

David, H. A. (2008). The Beginnings of Randomization Tests. *The American Statistician 62*(1), 70–72.

Dekker, D., D. Krackhardt, and T. A. B. Snijders (2007). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika 72*(4), 563–581.

Draper, N. R. and D. M. Stoneman (1966). Testing for the Inclusion of Variables in Linear Regression by a Randomisation Technique. *Technometrics 8*(4), 695.

Dunn, O. J. (1958). Estimation of the Means of Dependent Variables. *The Annals of Mathematical Statistics 29*(4), 1095–1111.

Edgington, E. and P. Onghena (2007). *Randomization Tests, Fourth Edition*. Chapman and Hall/CRC.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics 7*(1), 1–26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Number 38 in CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

Efron, B. and C. Morris (1977). Stein's paradox in statistics. *Scientific American 236*(5), 119–127.

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Eklund, A., T. E. Nichols, and H. Knutsson (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences 113*(28), 7900–7905.

Erickson, K. I., M. W. Voss, R. S. Prakash, C. Basak, A. Szabo, L. Chaddock, J. S. Kim, S. Heo, H. Alves, S. M. White, T. R. Wojcicki, E. Mailey, V. J. Vieira, S. A. Martin, B. D. Pence, J. A. Woods, E. McAuley, and A. F. Kramer (2011). Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences 108*(7), 3017–3022.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Chapman and Hall/CRC.

Fay, M. P. and P. A. Shaw (2010). Exact and asymptotic weighted logrank tests for interval censored data: The interval R package. *Journal of statistical software 36*(2), 1–34.

Finos, L. and D. Basso (2014). Permutation tests for between-unit fixed effects in multivariate generalized linear mixed models. *Statistics and Computing 24*(6), 941–952.

Finos, L., w. c. b. D. Basso, A. Solari, J. Goeman, and M. Rinaldo (2014). Flip: Multivariate Permutation Tests. R-Packages.

Fisher, R. A. (1948). Combining Independent Tests of Significance. *The American Statistician 2*(5), 30–31.

Fisher, S. R. A. (1935). *The Design of Experiments.* Oliver and Boyd.

Freedman, D. and D. Lane (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business & Economic Statistics 1*(4), 292.

Friedrich, S., E. Brunner, and M. Pauly (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis 153*, 255–265.

Friedrich, Sarah, Konietschke, Frank, and Pauly, Markus (2017). GFD: An R Package for the Analysis of General Factorial Designs. *Journal of Statistical Software 79*(1), 1–18.

Frossard, J. and O. Renaud (2018). Permuco: Permutation Tests for Regression, (Repeated Measures) ANOVA/ANCOVA and Comparison of Signals. R-Packages.

Frossard, J. and O. Renaud (2019). The correlation structure of mixed effects models with crossed random effects in controlled experiments. *arXiv:1903.10766 [stat].*

Gelman, A. (2005). Analysis of variance–why it is more important than ever. *The Annals of Statistics 33*(1), 1–53.

Gentle, J. (2007). *Matrix Algebra : Theory, Computations, and Applications in Statistics.* Springer-Verlag.

Godfrey, M., S. Hepburn, D. J. Fidler, T. Tapera, F. Zhang, C. R. Rosenberg, and N. Raitano Lee (2019). Autism spectrum disorder (ASD) symptom profiles of children with comorbid Down syndrome (DS) and ASD: A comparison with children with DS-only and ASD-only. *Research in Developmental Disabilities 89*, 83–93.

Green, P. J. and B. W. Silverman (1993). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall/CRC.

Greene, W. (2011). *Econometric Analysis.* Prentice Hall.

Hartmann, M., N. R. Sommer, L. Diana, R. M. Müri, and A. K. Eberhard-Moscicka (2019). Further to the right: Viewing distance modulates attentional asymmetries (‘pseudoneglect’) during visual exploration. *Brain and Cognition 129*, 40–48.

Heritier, S., E. Cantoni, S. Copt, and M.-P. Victoria-Feser (2009). *Robust Methods in Biostatistics.* John Wiley & Sons.

Hoaglin, D. C. and R. E. Welsch (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician 32*(1), 17–22.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics 6*(2), 65–70.

Hothorn, T., K. Hornik, M. A. Van De Wiel, A. Zeileis, et al. (2008). Implementing a class of permutation pests: The coin package. *Journal of Statistical Software 28*(8), 1–23.

Howell, D. C. (2012). *Statistical Methods for Psychology.* Cengage Learning.

Huber, P. (1964). Robust Estimation of location Parameter. *The Annals of Mathematical Statistics 35*(1), 73–101.

Huh, M.-H. and M. Jhun (2001). Random Permutation Testing in Multiple Linear Regression. *Communications in Statistics - Theory and Methods 30*(10), 2023–2032.

Huynh, H. and L. S. Feldt (1970). Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions. *Journal of the American Statistical Association 65*(332), 1582–1589.

Huynh, H. and L. S. Feldt (1976). Estimation of the Box Correction for Degrees of Freedom from Sample Data in Randomized Block and Split-Plot Designs. *Journal of Educational Statistics 1*(1), 69–82.

Janssen, A. (2005). Resampling student's t-type statistics. *Annals of the Institute of Statistical Mathematics 57*(3), 507–529.

Janssen, A. and T. Pauls (2003). How Do Bootstrap and Permutation Tests Work? *The Annals of Statistics 31*(3), 768–806.

Jayakumar, G. D. S. and A. Sulthan (2014). Exact Distribution of Hat Values and Identification of Leverage Points. *Journal of Reliability and Statistical Studies 7*(1), 61–78.

Judd, C. M., J. Westfall, and D. A. Kenny (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology 103*(1), 54–69.

Karniski, W., R. C. Blair, and A. D. Snider (1994). An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. *Brain Topography 6*(3), 203–210.

Kennedy, P. E. (1995). Randomization Tests in Econometrics. *Journal of Business & Economic Statistics 13*(1), 85.

Kenward, M. G. and J. H. Roger (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics 53*(3), 983.

Kern, E. M. A. and R. B. Langerhans (2019). Urbanization Alters Swimming Performance of a Stream Fish. *Frontiers in Ecology and Evolution 6*, 229.

Khatri, C. G. and C. R. Rao (1968). Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, 167–180.

Kherad Pajouh, S. and O. Renaud (2010). An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA. *Computational Statistics & Data Analysis 54*, 1881–1893.

Kherad-Pajouh, S. and O. Renaud (2015). A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs. *Statistical Papers 56*(4), 947–967.

Koenker, R. and G. Bassett (1978). Regression Quantiles. *Econometrica 46*(1), 33–50.

Konietschke, F., A. C. Bathke, S. W. Harrar, and M. Pauly (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis 140*, 291–301.

Krieglmeyer, R. and R. Deutsch (2010). Comparing measures of approach–avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition and Emotion 24*(5), 810–828.

Kruskal, W. H. and W. A. Wallis (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association 47*(260), 583–621.

Kuznetsova, A., P. B. Brockhoff, and R. H. Christensen (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software 82*(13), 1–26.

Lachaud, C. M. and O. Renaud (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics 32*(2), 389–416.

Langsrud, Ø. (2005). Rotation tests. *Statistics and computing 15*(1), 53–60.

Lehmann, E. L. and J. P. Romano (2008). *Testing Statistical Hypotheses*. Springer-Verlag.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods 49*(4), 1494–1502.

Manly, B. F. J. (1991). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall/CRC.

Mann, H. B. and A. Wald (1943). On Stochastic Limit and Order Relationships. *The Annals of Mathematical Statistics 14*(3), 217–226.

Maris, E. and R. Oostenveld (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods 164*(1), 177–190.

Modugno, L. and S. Giannerini (2015). The Wild Bootstrap for Multilevel Models. *Communications in Statistics - Theory and Methods 44*(22), 4812–4825.

Mogg, K., B. P. Bradley, M. Field, and J. De Houwer (2003). Eye movements to smoking-related pictures in smokers: Relationship between attentional biases and implicit and explicit measures of stimulus valence. *Addiction (Abingdon, England) 98*(6), 825–836.

Montgomery, D. C. (2017). *Design and Analysis of Experiments*. John Wiley & Sons.

Morris, J. S. (2002). The BLUPs are not “best” when it comes to bootstrapping. *Statistics & Probability Letters 56*(4), 425–430.

Musariri, T., N. Pegg, J. Muvengwi, and F. Muzama (2018). Differing patterns of plant spinescence affect blue duiker (Bovidae: Philantomba monticola) browsing behavior and intake rates. *Ecology and Evolution 8*(23), 11754–11762.

Nash, J. C. and R. Varadhan (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software 43*(9), 1–14.

Nelder, J. A. and R. Mead (1965). A Simplex Method for Function Minimization. *The Computer Journal 7*(4), 308–313.

Nichols, T., G. Ridgway, M. Webster, and S. Smith (2008). GLM permutation-nonparametric inference for arbitrary general linear models. *NeuroImage 41*(S1), S72.

O'Gorman, T. W. (2005). The Performance of Randomization Tests that Use Permutations of Independent Variables. *Communications in Statistics - Simulation and Computation 34*(4), 895–908.

Pauly, M. (2011). Discussion about the quality of F-ratio resampling tests for comparing variances. *Test 20*(1), 163–179.

Pauly, M., E. Brunner, and F. Konietschke (2015). Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society B 77*(2), 461–473.

Pernet, C., M. Latinus, T. Nichols, and G. Rousselet (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of Neuroscience Methods 250*, 85–93.

Pernet, C. R., N. Chauveau, C. Gaspar, and G. A. Rousselet (2011). LIMO EEG: A toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational intelligence and neuroscience 2011*, 3.

Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*, Volume 240. Wiley Chichester.

Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, 26–46.

Pursell, L. and S. Y. Trimble (1991). Gram-Schmidt Orthogonalization by Gauss Elimination. *The American Mathematical Monthly 98*(6), 544–549.

Raaijmakers, J. G. W., J. M. C. Schrijnemakers, and F. Gremmen (1999). How to Deal with "The Language-as-Fixed-Effect Fallacy": Common Misconceptions and Alternative Solutions. *Journal of Memory and Language 41*(3), 416–426.

Romano, J. P. (1990). On the Behavior of Randomization Tests Without a Group Invariance Assumption. *Journal of the American Statistical Association 85*(411), 686.

Rouanet, H. and D. Lepine (1970). Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology 23*(2), 147–163.

Salibian-Barrera, M. (2005). Estimating the p-values of robust tests for the linear model. *Journal of Statistical Planning and Inference 128*(1), 241–257.

Salibian-Barrera, M. and R. H. Zamar (2002). Bootstrapping robust estimates of regression. *Annals of Statistics 30*(2), 556–582.

Sassenhagen, J. and D. Draschkow (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology 56*(6), 1–8.

Schaalje, G. B., J. B. McBride, and G. W. Fellingham (2002). Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models. *Journal of Agricultural, Biological, and Environmental Statistics 7*(4), 512–524.

Searle, S. R. (2006). *Linear Models for Unbalanced Data.* John Wiley & Sons.

Seber, G. A. F. and A. J. Lee (2012). *Linear Regression Analysis.* John Wiley & Sons.

Shmueli, G. (2010). To explain or to predict? *Statistical science 25*(3), 289–310.

Smith, S. and T. Nichols (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage 44*(1), 83–98.

Soetaert, K. (2017). plot3D: Plotting Multi-Dimensional Data. R-Package.

Soler, J., B. Arias, J. Moya, M. I. Ibáñez, G. Ortet, L. Fañanás, and M. Fatjó-Vilas (2019). The interaction between the ZNF804A gene and cannabis use on the risk of psychosis in a non-clinical sample. *Progress in Neuro-Psychopharmacology and Biological Psychiatry 89*, 174–180.

ter Braak, C. J. F. (1992). Permutation Versus Bootstrap Significance Tests in Multiple Regression and Anova. In K.-H. Jöckel, G. Rothe, and W. Sendler (Eds.), *Bootstrapping and Related Techniques*, pp. 79–85. Springer-Verlag.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society B 58*(1), 267–288.

Tipura, E., O. Renaud, and A. Pegna (2017). Attention shifting and subliminal cueing: An EEG study using emotional faces (Submitted).

Troendle, J. F. (1995). A Stepwise Resampling Method of Multiple Hypothesis Testing. *Journal of the American Statistical Association 90*(429), 370–378.

Weiss, N. A. (2015). wPerm: Permutation Tests. R-Packages.

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika 34*(1-2), 28–35.

Welch, B. L. (1951). On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika 38*(3/4), 330–336.

Welch, W. and L. Gutierrez (1988). Robust Permutation Tests for Matched-Pairs Designs. *Journal of the American Statistical Association 83*(402), 450–455.

Welch, W. J. (1987). Rerandomizing the median in matched-pairs designs. *Biometrika 74*(3), 609–614.

Wheeler, B. and M. Torchiano (2016). lmPerm: Permutation Tests for Linear Models. R-Package.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin 1*(6), 80–83.

Wilson, J. P., K. Hugenberg, and N. O. Rule (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology 113*(1), 59–80.

Winer, B. J. (1962). *Statistical Principles in Experimental Design*. McGraw-Hill Book Company.

Winkler, A. M., G. R. Ridgway, G. Douaud, T. E. Nichols, and S. M. Smith (2016). Faster permutation inference in brain imaging. *NeuroImage 141*, 502–516.

Winkler, A. M., G. R. Ridgway, M. A. Webster, S. M. Smith, and T. E. Nichols (2014). Permutation inference for the general linear model. *NeuroImage 92*, 381–397.