Mathematical Statistics
Stockholm University

Comparison of within subject covariance matrices in 2×2 crossover trials with multivariate response

Niclas Sjögren

Research Report 2002:2

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.matematik.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2002:2**,
http://www.matematik.su.se/matstat

Comparison of within subject covariance matrices in 2×2 crossover trials with multivariate response

Niclas Sjögren*

January 2002

Abstract

Even though more than one response variable are measured after each treatment in crossover trials, they are usually analyzed separately using univariate methods. In a multivariate framework, it is shown how the two treatments in a 2×2 crossover trial with multivariate response can be compared with respect to both fixed treatment effects and within subject covariance matrices, marginally and simultaneously. The proposed exact statistical inferences are valid even with few subjects and without distributional assumption made about the between subject variability.

KEY WORDS: Wilks' lambda; U-distribution; semiparametric; simultaneous inference.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: Niclas.Sjogren@matematik.su.se.

1 Introduction

In a 2×2 crossover trial, each subject are randomly assigned to one of the two sequences, AB or BA. The subjects in the AB-sequence receive treatment A before treatment B, separated by a washout period to minimize an eventual carryover effect from period 1 to period 2. Each subject generates two responses, one after each treatment. Usually these responses are thought to be univariate and measured on a continuous scale. A lot of work have been done for such ordinary univariate 2×2 crossover trials and there are several books written on this topic, Jones and Kenward (1989), Ratkowsky *et al.* (1993) and Senn (1993), etc.

Even though the response often consist of more than one measured characteristic in practice, the usual way to analyze such a study is to analyze each characteristic one at a time using standard methods for univariate crossover trials. In this paper, we show how the two treatments in a 2×2 crossover trial with multivariate response can be compared with respect to both mean response levels and within-subject variability using multivariate methods. A multivariate response comes from either repeated measurements or a true multivariate design. In studies with repeated or longitudinal measurements, one single characteristic is measured on more than one occasion in each treatment period. Wallenstein and Fisher (1977) and Jones and Kenward (1989) discussed the split-plot, whereas Patel and Hearne (1980) presented a multivariate linear approach for this design. In a crossover trial with a true multivariate design, two or more characteristics are of interest and measured in each treatment period. Grender and Johnson (1993) used the general multivariate linear model to set up an unified approach that handles tests for treatment, period and carry-over effects as special cases. To test for direct treatment effects, Rodriguez-Carvajal and Freeman (1999) used Hotelling´s T^2 statistic together with a transformation similar to that of Jones and Kenward (1989) for the univariate case.

Guilbaud (1993) proposed a measure of the difference between the two within subject variances in a 2×2 crossover trial and showed how to make exact statistical inference about this measure without distributional assumption made about the between subject parameters. We extend this into the multivariate case. Thus, we propose a measure of the difference between the within subject covariance matrices and show how to make exact statistical inference of this measure without making distributional assumption about the between subject parameters. We also show how to construct a simultaneous confidence set for this measure and the difference in direct treatment effect vectors using the union of the two marginal confidence sets. The simultaneous confidence coefficient for such a set equals the product of the two marginal confidence coefficients.

The outline of this paper is as follows. The statistical model with assumptions and notations are presented in section 2. In section 3, we present the main ideas and the main results. In section 4, we use the results in section 3 to construct exact confidence regions for the measure of the difference between within subject covariance matrices. It is also described how to combine such a confidence region with a confidence region of the difference in fixed treatment effects. Finally, in section 5, we present computer simulations that has been a useful toy in the understanding of the results. The conclusion is that even if the distribution of the test statistic does not depend on the distribution of the between subject parameters, the decision drawn from such an inference depends on the actual values of the between subject variables.

2 Assumptions and notations

Let $\mathbf{Y}_{ij1}$ and $\mathbf{Y}_{ij2}$ denote the response column vectors from period 1 and period 2 for subject $i = 1, 2, \ldots, n_j$ in sequence group $j = 1, 2$, where sequence group 1 corresponds to the treatment sequence AB and group 2 corresponds to treatment sequence BA. $n_j \geq 2$ denotes the number of subjects randomized into sequence group j, $j = 1, 2$. Moreover, let

$$n = n_1 + n_2 \tag{1}$$

denote the total number of subjects in the study. Further, it is assumed that the random response $p \times 2$ matrix $(\mathbf{Y}_{ij1}, \mathbf{Y}_{ij2})$ can be represented as

$$(\mathbf{Y}_{ij1}, \mathbf{Y}_{ij2}) = \left(\boldsymbol{\mu}_{j1}, \boldsymbol{\mu}_{j2}\right) + \left(\boldsymbol{\xi}_{ij} + \boldsymbol{\varepsilon}_{ij1}, \boldsymbol{\xi}_{ij} + \boldsymbol{\varepsilon}_{ij2}\right), \tag{2}$$

where $\boldsymbol{\mu}_{j1}$ and $\boldsymbol{\mu}_{j2}$ are nonrandom vectors reflecting fixed effects such as treatment effects and period effects; $\boldsymbol{\xi}_{ij}$ is a random vector reflecting the between-subject variability; and $\boldsymbol{\varepsilon}_{ij1}$ and $\boldsymbol{\varepsilon}_{ij2}$ are random vectors reflecting the within-subject variability in each period.

The only distributional assumption made in this report concerns the n within vectors $\boldsymbol{\varepsilon}_{ij1}$ and $\boldsymbol{\varepsilon}_{ij2}$. They are assumed to be mutually independent and independent of the n "betwee" vectors $\boldsymbol{\xi}_{ij}$. With $=_d$ denoting equality in distribution, it is assumed that for $i = 1, 2, \ldots, n_j$,

$$(\boldsymbol{\varepsilon}_{ij1}, \boldsymbol{\varepsilon}_{ij2}) =_d \begin{cases} (\boldsymbol{\varepsilon}_A, \boldsymbol{\varepsilon}_B), & \text{if } j = 1 \\ (\boldsymbol{\varepsilon}_B, \boldsymbol{\varepsilon}_A), & \text{if } j = 2 \end{cases} \tag{3}$$

where $\boldsymbol{\varepsilon}_A$ and $\boldsymbol{\varepsilon}_B$ are independent multivariate normal distributed stochastic variables with zero mean and covariance matrix $\boldsymbol{\Lambda}_A$ respective $\boldsymbol{\Lambda}_B$, i.e. $\boldsymbol{\varepsilon}_A \sim N_p(\mathbf{0}, \boldsymbol{\Lambda}_A)$ and $\boldsymbol{\varepsilon}_B \sim N_p(\mathbf{0}, \boldsymbol{\Lambda}_B)$. The indexes $_A$ and $_B$ indicates treatments. In terms of these within-subject covariance matrices define the multivariate analogous, $\boldsymbol{\Gamma}$, to the univariate one defined in Guilbaud (1993) $\gamma = (\sigma_A^2 - \sigma_B^2)/(\sigma_A^2 + \sigma_B^2)$

$$\boldsymbol{\Gamma} = (\boldsymbol{\Lambda}_A - \boldsymbol{\Lambda}_B)(\boldsymbol{\Lambda}_A + \boldsymbol{\Lambda}_B)^{-1} \tag{4}$$

Thus, $\mathbf{\Gamma}$ is a measure of the difference between the covariance matrices Λ_A and Λ_B. If $\Lambda_A = \Lambda_B$, then $\mathbf{\Gamma}$ equals the $p \times p$ null matrix $\mathbf{0}$.

Let $\mathbf{Y}_{ij}^{+}$ and $\mathbf{Y}_{ij}^{-}$, $i = 1, 2, \ldots, n_j$ denote the column vectors containing within-subject sums and (A-B)-differences, i.e.

$$\mathbf{Y}_{ij}^{+} = \mathbf{Y}_{ij1} + \mathbf{Y}_{ij2} \tag{5}$$

$$\mathbf{Y}_{ij}^{-} = \begin{cases} \mathbf{Y}_{ij1} - \mathbf{Y}_{ij2},, & \text{if } j = 1 \\ \mathbf{Y}_{ij2} - \mathbf{Y}_{ij1},, & \text{if } j = 2 \end{cases} \tag{6}$$

Further, for any given $p \times p$ matrix $\mathbf{M}$, define $\mathbf{Y}_{ij}^{\mathbf{M}}$ as

$$\mathbf{Y}_{ij}^{\mathbf{M}} = \mathbf{Y}_{ij}^{+} - \mathbf{M}\mathbf{Y}_{ij}^{-} \tag{7}$$

The association between $\mathbf{Y}_{ij}^{\mathbf{M}}$ and $\mathbf{Y}_{ij}^{+}$ depends on the matrix $\mathbf{M}$, and the idea is to use this dependence to make statistical inference about $\mathbf{\Gamma}$.

The sum of squares and cross-product matrices corresponding to (5) and (6) are defined by

$$\mathbf{S}_{--} = \sum_{j=1}^{2}\sum_{i=1}^{n_j}(\mathbf{Y}_{ij}^{-} - \bar{\mathbf{Y}}_{.j}^{-})(\mathbf{Y}_{ij}^{-} - \bar{\mathbf{Y}}_{.j}^{-})',$$

$$\mathbf{S}_{-+} = \sum_{j=1}^{2}\sum_{i=1}^{n_j}(\mathbf{Y}_{ij}^{-} - \bar{\mathbf{Y}}_{.j}^{-})(\mathbf{Y}_{ij}^{+} - \bar{\mathbf{Y}}_{.j}^{+})',$$

with $\bar{\mathbf{Y}}_{.j}^{-} = \sum_{i=1}^{n_j} \mathbf{Y}_{ij}^{-}/n_j$ and $\bar{\mathbf{Y}}_{.j}^{+} = \sum_{i=1}^{n_j} \mathbf{Y}_{ij}^{+}/n_j$. $\mathbf{S}_{+-}$, $\mathbf{S}_{\mathbf{MM}}$, $\mathbf{S}_{-\mathbf{M}}$ and $\mathbf{S}_{\mathbf{M}-}$ are defined in the same way. Define $\mathbf{Y}_{j}^{-}$ to be the $p \times n_j$ matrix containing all of the $\mathbf{Y}_{ij}^{-}$'s, i.e.

$$\mathbf{Y}_{j}^{-} = \left(\mathbf{Y}_{1j}^{-}, \mathbf{Y}_{2j}^{-}, \ldots, \mathbf{Y}_{n_j j}^{-}\right)$$

In the similar way, define the $p \times n_j$ matrix $\bar{\mathbf{Y}}_{j}^{-}$ to be the "mean matrix", i.e.

$$\bar{\mathbf{Y}}_{j}^{-} = \left(\bar{\mathbf{Y}}_{j}^{-}, \bar{\mathbf{Y}}_{j}^{-}, \ldots, \bar{\mathbf{Y}}_{j}^{-}\right)$$

Further on, define the $p \times n$ matrices $\mathbf{Y}^{-}$ and $\bar{\mathbf{Y}}^{-}$ as

$$\mathbf{Y}^{-} = \left(\mathbf{Y}_{1}^{-}, \mathbf{Y}_{2}^{-}\right) \tag{8}$$

$$\bar{\mathbf{Y}}^{-} = \left(\bar{\mathbf{Y}}_{1}^{-}, \bar{\mathbf{Y}}_{2}^{-}\right). \tag{9}$$

With these definitions we can express $\mathbf{S}_{--}$ as

$$\mathbf{S}_{--} = (\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)'.$$

Defining $\mathbf{Y}^+$, $\mathbf{Y}^{\mathbf{M}}$, $\bar{\mathbf{Y}}^+$ and $\bar{\mathbf{Y}}^{\mathbf{M}}$ in the same way as in (8) and (9), we have

$$\begin{aligned}
\mathbf{S}_{+-} &= (\mathbf{Y}^+ - \bar{\mathbf{Y}}^+)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)', \\
\mathbf{S}_{\mathbf{MM}} &= (\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})', \quad (10) \\
\mathbf{S}_{-\mathbf{M}} &= (\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})'. \quad (11)
\end{aligned}$$

Note that the sum of cross-product matrices not are symmetric but that $\mathbf{S}_{+-} = \mathbf{S}'_{-+}$ and $\mathbf{S}_{\mathbf{M}-} = \mathbf{S}'_{-\mathbf{M}}$, where $'$ denotes the transpose of a matrix.

3 Ideas and basic results

We can write $\mathbf{Y}_{ij}^{\mathbf{M}}$ as

$$\mathbf{Y}_{ij}^{\mathbf{M}} = (\mathbf{\Gamma} - \mathbf{M})\mathbf{Y}_{ij}^- + \mathbf{Z}_{ij} + 2\boldsymbol{\xi}_{ij},$$

where

$$\mathbf{Z}_{ij} = \mathbf{Y}_{ij}^+ - \mathbf{\Gamma}\mathbf{Y}_{ij}^- - 2\boldsymbol{\xi}_{ij}. \quad (12)$$

Note that no randomness in $\mathbf{Z}_{ij}$ comes from the between subject random effects $\boldsymbol{\xi}_{ij}$, implying that $\mathbf{Z}_{ij}$ follows a multivariate normal distribution. Now, because $\mathrm{Cov}(\mathbf{Y}_{ij}^-, \mathbf{Z}_{ij}) = 0$, $\mathbf{Y}_{ij}^-$ and $\mathbf{Z}_{ij}$ are independent multivariate normal variables.

It is now evident that if $\mathbf{M} = \mathbf{\Gamma}$ then $\mathbf{Y}_{ij}^-$ and $\mathbf{Y}_{ij}^{\mathbf{M}}$ are independent. It is also evident that if $\mathbf{M} \neq \mathbf{\Gamma}$, then $\mathbf{Y}_{ij}^-$ and $\mathbf{Y}_{ij}^{\mathbf{M}}$ are not independent. Thus, the association between $\mathbf{Y}_{ij}^-$ and $\mathbf{Y}_{ij}^{\mathbf{M}}$ depends on the matrix, $\mathbf{M}$. The idea is to use this dependence on $\mathbf{M}$ to make statistical inference about $\mathbf{\Gamma} = (\mathbf{\Lambda}_A - \mathbf{\Lambda}_B)(\mathbf{\Lambda}_A + \mathbf{\Lambda}_B)^{-1}$.

For the moment, assume that the between subject parameters, $\boldsymbol{\xi}_{ij}$ are independent multivariate normal with zero mean vector and some proper covariance matrix. Now, under this assumption, we can derive the likelihood ratio test for testing independence between $\mathbf{Y}_{ij}^-$ and $\mathbf{Y}_{ij}^{\mathbf{M}}$. It will be shown in theorem 1 that the distribution of this test statistic does not depend on the distribution of the between subject parameters.

With the normality assumption described above we have that $(\mathbf{Y}_{ij}^-, \mathbf{Y}_{ij}^{\mathbf{M}})'$ follows a multivariate normal distribution, i.e.

$$\begin{pmatrix} \mathbf{Y}_{ij}^- \\ \mathbf{Y}_{ij}^{\mathbf{M}} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_j^- \\ \boldsymbol{\mu}_j^{\mathbf{M}} \end{pmatrix}, \mathbf{\Sigma} \right), \quad i = 1, 2, \ldots, n_j \quad j = 1, 2 \quad (13)$$

Where $\boldsymbol{\mu}_j^-$ and $\boldsymbol{\mu}_j^{\mathbf{M}}$ are the proper mean vectors and where the covariance matrix, $\boldsymbol{\Sigma}$ can be partitioned as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{--} & \boldsymbol{\Sigma}_{-\mathbf{M}} \\ \boldsymbol{\Sigma}_{\mathbf{M}-} & \boldsymbol{\Sigma}_{\mathbf{MM}} \end{pmatrix}$$

Note that $\mathbf{Y}_{ij}^-$ and $\mathbf{Y}_{ij}^{\mathbf{M}}$ are independent if and only if $\mathbf{M} = \boldsymbol{\Gamma}$ $(\boldsymbol{\Sigma}_{-\mathbf{M}} = \mathbf{0})$. We want to test

$$\begin{array}{c} H_0\text{: } \mathbf{Y}_{ij}^- \text{ and } \mathbf{Y}_{ij}^{\mathbf{M}} \text{ are independent } (\boldsymbol{\Sigma}_{-\mathbf{M}} = \mathbf{0}) \\ \text{versus} \\ H_A\text{: } \mathbf{Y}_{ij}^- \text{ and } \mathbf{Y}_{ij}^{\mathbf{M}} \text{ are not independent } (\boldsymbol{\Sigma}_{-\mathbf{M}} \neq \mathbf{0}). \end{array} \tag{14}$$

Adopting standard techniques used for determining likelihood ratio tests in multivariate normal models (see for example Seber (1984)) in model (13), it follows that the likelihood ratio test for testing (14) is

$$\widetilde{\Lambda}(\mathbf{M}) = \frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_{--}||\widehat{\boldsymbol{\Sigma}}_{\mathbf{MM}}|},$$

where $\widehat{\boldsymbol{\Sigma}}$, $\widehat{\boldsymbol{\Sigma}}_{--}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{MM}}$ are the likelihood estimates for the covariance matrices, $\widehat{\boldsymbol{\Sigma}}_{--} = \mathbf{S}_{--}/n$, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{MM}} = \mathbf{S}_{\mathbf{MM}}/n$ and $\widehat{\boldsymbol{\Sigma}} = (\mathbf{S}_1 + \mathbf{S}_2)/n$, where $\mathbf{S}_j = \sum_{i=1}^{n_j}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)'$. Note that $(n_1 - 1)\mathbf{S}_1 \sim W_p(n_1 - 1, \boldsymbol{\Sigma})$ and $(n_2 - 1)\mathbf{S}_2 \sim W_p(n_2 - 1, \boldsymbol{\Sigma})$ are independent. It follows that $(n-2)(\mathbf{S}_1 + \mathbf{S}_2) \sim W_p(n-2, \boldsymbol{\Sigma})$, where $W_p(m, \boldsymbol{\Sigma})$ denotes the Wishart distribution with m degrees of freedom. Instead of $\widetilde{\Lambda}(\mathbf{M})$, we can express the likelihood ratio test as

$$\Lambda(\mathbf{M}) = \frac{|\mathbf{S}|}{|\mathbf{S}_{--}||\mathbf{S}_{\mathbf{MM}}|}. \tag{15}$$

Using

$$|\mathbf{S}| = |\mathbf{S}_{\mathbf{MM}}||\mathbf{S}_{--} - \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-}| \tag{16}$$

and setting $\mathbf{E} = \mathbf{S}_{--} - \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-}$ and $\mathbf{H} = \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-}$ we have

$$\Lambda(\mathbf{M}) = \frac{|\mathbf{S}_{--} - \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-}|}{|\mathbf{S}_{--}|} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}. \tag{17}$$

The statistic of the form (17) first showed up as the likelihood ratio statistic for testing linear hypotheses by Wilks (1932). Therefore, it is sometimes named "Wilks' likelihood ratio test" or "Wilks' Λ" in the literature. There are other tests than the likelihood ratio test available, as for example "the maximum root test" and "Pillai´s trace statistic". More can be read about these tests in books in multivariate analysis, see for example Seber (1984). The distribution of (17) when $\mathbf{E}$ and $\mathbf{H}$ are independently Wishart variables is by some authors called "the U-distribution". Applying Lemma 2.10 and its corollary in

Seber (1984) [page 50-51] to $\mathbf{S}$, we have that, when $H_0 : \mathbf{\Sigma}_{-\mathbf{M}} = \mathbf{0}$ is true, the random matrices $\mathbf{E}$ and $\mathbf{H}$ are independently distributed as $W_p\left(n-p-2, \mathbf{\Sigma}_{--}\right)$ and $W_p\left(p, \mathbf{\Sigma}_{--}\right)$, respectively. We adopt the name "U-distribution" and use the commonly used notation

$$\Lambda(\mathbf{\Gamma}) \sim U_{p,p,n-p-2}.$$

Upper quantiles for the U-distribution can for example be found in Seber (1984). However, various approximations have been obtained. Overviews of those can be found in books in multivariate statistics such as for example Seber (1984). Bartlett (1938) showed that $-(n-p-1.5)\log\Lambda(\mathbf{\Gamma})$ is approximately χ^2 distributed with p^2 degrees of freedom for large n. By Seber (1984)[page 41], "This approximation is surprisingly accurate for the usual critical values". Thus, this approximation should be adequate for most circumstances in practice.

The following theorem assures that we can use the test statistic (17) to make exact statistical inference about $\mathbf{\Gamma}$, even with no distributional assumption made about the between subject parameters, $\boldsymbol{\xi}_{ij}$.

Theorem 1:
The test statistic,

$$\Lambda(\mathbf{M}) = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \frac{|\mathbf{S}_{--} - \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-}|}{|\mathbf{S}_{--}|} \tag{18}$$

for testing

$$H_0 : \mathbf{\Gamma} = \mathbf{M} \quad \textit{versus} \quad H_A : \mathbf{\Gamma} \neq \mathbf{M}$$

has the following properties.

(a) $\Lambda(\mathbf{\Gamma}) \sim U_{p,p,n-p-2}$*, where* U *stand for the U-distribution described above. This holds with no distributional assumption made about the between subject parameters,* $\boldsymbol{\xi}_{ij}$*. In particular, the* $\boldsymbol{\xi}_{ij}$*'s are not assumed to be independent or identically distributed.*

(b) $\Lambda(\mathbf{\Gamma})$ *and* $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$ *are independent.*

(c) $\Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}) = 1$.

(d) $\Lambda(.)$ *is symmetric around* $\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$*, i.e.* $\Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1} + \mathbf{A}) = \Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1} - \mathbf{A})$ *for all* $p \times p$ *matrices* $\mathbf{A}$.

Proof: See the Appendix.

□

We can use property (a) to construct statistical tests as well as confidence regions for $\mathbf{\Gamma}$. In section 4.3 we use (b) to construct simultaneous confidence sets for the difference in direct treatment effects and $\mathbf{\Gamma}$ at a certain significance level. The properties (c) and (d) motivates that $\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$ can be seen as an estimator of $\mathbf{\Gamma}$.

$\Lambda(\mathbf{M})$ takes values between 0 and 1, where the probability observing a value close to 0 is small if H_0 is true. Thus, H_0 is rejected in favor of H_A when $\Lambda(\mathbf{M})$ is too small. $\Lambda(\mathbf{M})$ takes the value 1 for $\mathbf{M} = \mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$, which can be seen as an estimator of $\mathbf{\Gamma}$. Compare this with the univariate case, where $\gamma^* = S_{-+}/S_{--}$ is an exactly median unbiased estimator of $\gamma = (\sigma_A^2 - \sigma_B^2)/(\sigma_A^2 + \sigma_B^2)$.

We now discuss some theoretical properties of the U-distribution, more can for example be found in Anderson (1984)[Chapter 8].

$\Lambda(\mathbf{M})$ can be seen as a function of the eigenvalues of $\mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-}\mathbf{S}_{--}^{-1}$. The eigenvalues are distinct with probability 1 and they can therefore be ordered as $\theta_1 > \theta_2 > \ldots > \theta_p$.

$$\Lambda(\mathbf{M}) = \prod_{k=1}^{p}(1 - \theta_k).$$

It is known that $0 \leq \theta_i < 1$ and if we write $\theta_k = r_k^2$, then the positive square root r_k is called the kth sample canonical correlation between $\mathbf{Y}_{ij}^-$ and $\mathbf{Y}_{ij}^{\mathbf{M}}$. In the one dimensional case (p=1) we have that $\theta_1 = S_{-M}^2/(S_{--}S_{MM})$ where M is a scalar. Thus, r_1 is the sample correlation between the one dimensional stochastic variables Y_{ij}^- and Y_{ij}^M. $\Lambda(M) = (1 - r_1^2)$.

$\mathbf{S}_{--}$ and $\mathbf{S}_{\mathbf{MM}}$ can be interchanged in the expression of $|\mathbf{S}|$ in (16). Thus, $|\mathbf{S}|$ can be written as

$$|\mathbf{S}| = |\mathbf{S}_{--}||\mathbf{S}_{\mathbf{MM}} - \mathbf{S}_{\mathbf{M}-}\mathbf{S}_{--}^{-1}\mathbf{S}_{-\mathbf{M}}|. \tag{19}$$

Property (c) together with the symmetric property (d) imply that the test statistic $\Lambda(\mathbf{M})$ can be expressed as a function of the difference between the matrix $\mathbf{A}$ and the central point $\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$. Using (19) in (15) together with evaluating $\mathbf{S}_{\mathbf{MM}}$, $\mathbf{S}_{-\mathbf{M}}$ and $\mathbf{S}_{\mathbf{M}-}$ for $\mathbf{M} = \mathbf{S}_{+-}\mathbf{S}_{--}^{-1} + \mathbf{A}$ using (7), (10) and (11) yield

$$\Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1} + \mathbf{A}) = \frac{|\mathbf{S}_{++} - \mathbf{S}_{+-}\mathbf{S}_{--}^{-1}\mathbf{S}_{-+}|}{|\mathbf{S}_{++} - \mathbf{S}_{+-}\mathbf{S}_{--}^{-1}\mathbf{S}_{-+} + \mathbf{A}\mathbf{S}_{--}\mathbf{A}'|}, \tag{20}$$

which is easier to work with than (17). Inverting the right hand side of (20) and using the fact that the determinant of a matrix equals the product of its eigenvalues we have that (20) can be written as

$$\Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1} + \mathbf{A}) = 1/\prod_{i=1}^{p}(1 + \lambda_i),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{A}\mathbf{S}_{--}^{-1}\mathbf{A}'(\mathbf{S}_{++} - \mathbf{S}_{+-}\mathbf{S}_{--}^{-1}\mathbf{S}_{-+})^{-1}$.

3.1 Connection to the univariate case

For the one dimensional case ($p = 1$), the U-distribution has the following property (see for example page 40 in Seber (1984)),

$$\frac{(n-3)(1-\Lambda(\gamma))}{\Lambda(\gamma)} \sim F_{1,n-3}.$$

Here we write γ instead of Γ to indicate that we consider the one dimensional case. The F distribution with 1 degrees of freedom in the nominator and $n-3$ in the denominator is the square of the t-distribution with $n-3$ degrees of freedom. Thus,

$$\left(\frac{(n-3)(1-\Lambda(\gamma))}{\Lambda(\gamma)}\right)^{1/2} \sim t_{n-3}.$$

Further on, Guilbaud (1993) showed how to make exact inference about γ by using that the test statistic

$$T(c) = (\gamma^* - c)/s_* \tag{21}$$

is t-distributed with $n-3$ degrees of freedom when $c = \gamma$. Here $\gamma^* = S_{-+}/S_{--}$ and $s_*^2 = (S_{++}/S_{--} - (\gamma^*)^2)/(n-3)$.

The test statistic $\Lambda(.)$ is a function of the test statistic (21) in the one dimensional case,

$$T(c) = \left(\frac{(n-3)(1-\Lambda(\gamma))}{\Lambda(\gamma)}\right)^{1/2}.$$

Thus, in the one dimensional case, making exact statistical inference about γ using theorem 1 is equivalent to the exact statistical inference proposed by Guilbaud (1993) using (21).

4 Confidence regions

In this section we first describe how to construct confidence regions for the difference in fixed treatment effects. This is done by adopting the same technique as in the univariate case with the distinction that it ends up in the Hotelling´s T^2 instead of the Student´s t distribution. Then we show how to use the distributional result (a) in theorem 1 to construct a confidence set for $\mathbf{\Gamma}$. Finally, we show how to combine those marginal confidence sets using the independence result (b) in theorem 1.

4.1 Confidence region for the (A-B) difference in fixed treatment effects

Inference about fixed treatment effects is made by examining the $\mathbf{Y}_{ij}^{-}$'s defined in (6). We have

$$\mathbf{Y}_{ij}^{-} = \boldsymbol{\mu}_j^{-} + \boldsymbol{\varepsilon}_{ij}^{A} - \boldsymbol{\varepsilon}_{ij}^{B} \tag{22}$$

where, $\boldsymbol{\mu}_1^{-}$ and $\boldsymbol{\mu}_2^{-}$ are fixed effects and $\boldsymbol{\varepsilon}_{ij}^{A}$ and $\boldsymbol{\varepsilon}_{ij}^{B}$ are independent multivariate normal distributed with zero means and covariance matrices $\boldsymbol{\Lambda}_A$ and $\boldsymbol{\Lambda}_B$ respectively. Assume that the direct treatment effects and the period effects are fixed and additive with no other disturbing fixed effects being present, that is no carry-over effects. This assumption implies that the fixed effects in (22) can be written as

$$\boldsymbol{\mu}_1^{-} = \boldsymbol{\tau} + \boldsymbol{\pi}$$
$$\boldsymbol{\mu}_2^{-} = \boldsymbol{\tau} - \boldsymbol{\pi}$$

where, the vector of constants $\boldsymbol{\tau}$ equals the (A-B) difference of the fixed direct treatment effects and the vector of constants $\boldsymbol{\pi}$ equals the (period 1-period 2) difference of the fixed period effects. Exact multivariate statistical inference about $\boldsymbol{\tau}$ can be made using the statistic $\mathbf{D}$ defined by

$$\mathbf{D} = (\bar{\mathbf{Y}}_1^{-} + \bar{\mathbf{Y}}_2^{-})/2,$$

where $\bar{\mathbf{Y}}_1^{-} = (1/n_1)\sum_{i=1}^{n_1} \mathbf{Y}_{i1}^{-}$ and $\bar{\mathbf{Y}}_2^{-} = (1/n_2)\sum_{i=1}^{n_2} \mathbf{Y}_{i2}^{-}$ are the mean differences of (treatment A- treatment B) effects in sequence group 1 respective sequence group 2. Now we have that

$$\mathbf{D} \sim N_p(\boldsymbol{\tau}, \boldsymbol{\Sigma}_{\mathbf{D}}), \tag{23}$$

where the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}}$ is

$$\boldsymbol{\Sigma}_{\mathbf{D}} = (1/n_1 + 1/n_2)\boldsymbol{\Sigma}_{--}/4,$$

and is estimated by

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{D}} = (1/n_1 + 1/n_2)\mathbf{S}_{--}/4(n-2).$$

Consider the Hotelling´s T^2, which in this case is

$$T^2 = (\mathbf{D} - \boldsymbol{\tau})' \, \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} \, (\mathbf{D} - \boldsymbol{\tau}) \, .$$

Then, we can write a $(1-\alpha)$ confidence region for $\boldsymbol{\tau}$, $\mathbf{C}_{\boldsymbol{\tau}}$ as a function of $\mathbf{D}$ and $\mathbf{S}_{--}$ as

$$\mathbf{C}_{\boldsymbol{\tau}}(\mathbf{D}, \mathbf{S}_{--}) = \left\{ \boldsymbol{\tau} : T^2 \le \frac{p(n-2)}{n-p-1} F_{p,n-p-1}^2(\alpha) \right\}, \tag{24}$$

where $F_{p,n-p-1}(\alpha)$ denotes the upper α quantile in the F distribution with p and $n-p-1$ degrees of freedom. This confidence set has coverage probability $1-\alpha$, this holds of course even without distributional assumption on the between-subject parameters $\boldsymbol{\xi}_{ij}$'s.

4.2 Exact confidence region for $\mathbf{\Gamma}$

We now use the results in theorem 1 to construct an exact $1-\alpha$ confidence region, $\mathbf{C}_{\mathbf{\Gamma}}$, for $\mathbf{\Gamma} = (\mathbf{\Lambda}_A - \mathbf{\Lambda}_B)(\mathbf{\Lambda}_A + \mathbf{\Lambda}_B)^{-1}$. This is done by including matrices $\mathbf{M}$ that are not statistically significant on the α-level, i.e.

$$\mathbf{C}_{\mathbf{\Gamma}} = \{\mathbf{M} : \Lambda(\mathbf{M}) \geq U_{p,p,n-p-2}(\alpha)\}. \tag{25}$$

Thus, the confidence region consists of all $p \times p$ matrices $\mathbf{M}$ that are not significant different from $\mathbf{\Gamma}$ on level α. Using the expression (20) for $\Lambda(\mathbf{M})$ we can construct the confidence region by first including the center point $\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$ and then expand the region by also including non significant surroundings.

4.3 Simultaneous confidence regions

As in the univariate case (see Guilbaud (1993)), the exact confidence coefficient for a confidence region for $(\boldsymbol{\tau}, \mathbf{\Gamma})$ defined by the union of the two marginal confidence regions for $\boldsymbol{\tau}$ (24) and $\mathbf{\Gamma}$ (25) equals the product of the two associated marginal confidence coefficients. That is, as if the two random regions were independent.

In fact, there is as in the univariate case a certain dependence between the random regions $\mathbf{C}_{\boldsymbol{\tau}}$ and $\mathbf{C}_{\mathbf{\Gamma}}$ through $\mathbf{S}_{--}$ in (24) and (25). The confidence coefficient can nevertheless be factorized into the two marginal confidence coefficients because of the same reason as in the univariate case. That is that the two coverage events $\{\boldsymbol{\tau} \in \mathbf{C}_{\boldsymbol{\tau}}\}$ and $\left\{\mathbf{\Gamma} \in \mathbf{C}_{\mathbf{\Gamma}}\right\}$ are independent. Here, this independence holds because the event $\left\{\mathbf{\Gamma} \in \mathbf{C}_{\mathbf{\Gamma}}\right\}$ can be expressed in terms of $\Lambda(\mathbf{\Gamma})$ and the event $\{\boldsymbol{\tau} \in \mathbf{C}_{\boldsymbol{\tau}}\}$ in terms of $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$, which by theorem 1 are independent.

We can construct a simultaneous confidence region at a desired exact coverage probability by choosing the two marginal confidence coefficient appropriately. We may choose different marginal confidence coefficient even though the most appropriate way may would be to choose the same.

5 Simulations

In this section, we show computer simulations that illustrate the behavior of the test statistic $\Lambda(\mathbf{M})$ for different types of distributions on the between subject variables, $\boldsymbol{\xi}_{ij}$. It is sufficient to show the behavior in the one dimensional case because the same pattern also is seen in the multivariate case. As mentioned in section 3.1, the test based on $T(.)$ defined in (21) is equivalent with the test based on $\Lambda(.)$ in the one dimensional case. We choose to present the results of the simulations in terms of $T(.)$ because the t-distribution is available in standard statistical software whereas the U-distribution is not.

We simulated 5000 studies, each included 20 subjects whose responses were simulated from the univariate case of model (2). It was assumed that there were no fixed period and no fixed treatment effects present, i.e. $\pi = 0$ and $\tau = 0$. Further, the two within subject random variables ε_A and ε_B were assumed to be normal distributed with expectation zero and variance one, i.e. $\varepsilon_A \sim N(0,1)$ as well as $\varepsilon_B \sim N(0,1)$. Each subject was given a simulated value of $(\varepsilon_A, \varepsilon_B)$. They were also given four simulated values of the between subject parameter ξ_{ij}. Thus, each subject gets four simulated sets of (Y_{ij1}, Y_{ij2}), where the values on the within subject variables are fixed in the four sets. Thus, the only thing differing between the sets are the values on ξ_{ij}. The following cases are considered:

- Normal 1: $\xi_{ij} \sim N(0,10)$. That is when ξ_{ij} is simulated from the normal distribution with zero expectation and variance 10.
- Normal 2: Same as "Normal 1" but new random numbers, $\xi_{ij} \sim N(0,10)$.
- Constant: $\xi_{ij} = 0$, $i = 1, 2, \ldots, n_j$, $j = 1, 2$.
- Exponential: $\xi_{ij} = X_{ij} - 1$, where X_{ij} is an exponential distributed random variable with expectation 1.

The first two cases, "Normal 1" and "Normal 2", contains the same normal distribution where the variance is large compared to the variance of the within subject variables. The "Constant"-case can be seen as if is no between subject variation present, i.e the between subject variables are constant equal to zero, $\xi_{ij} = 0$, for $i = 1, 2, \ldots, n_j$, $j = 1, 2$. Finally, the "Exponential"-case represents a shewed distribution with the same variance as the within subject variables.

Each study gives rise to four test statistics, one for each case of ξ_{ij}. Thus, the simulation gives us 5000 quadruples of test statistics. In Figure 1 we show scatter plots, when these are plotted pairwise against each other. Figure 2 shows histograms of the 5000 simulated test statistics for each of the four cases. The probability density function of Student's t distribution with $n - 3$ degrees of freedom is included as a reference curve in each of the four histograms.

We know from theorem 1 that the test statistic follow the same distribution whatever the underlying distribution is on the between subject variables ξ_{ij}. Though, we see in Figure 1 that the values of the test statistics depend on the values of the between subject variable.

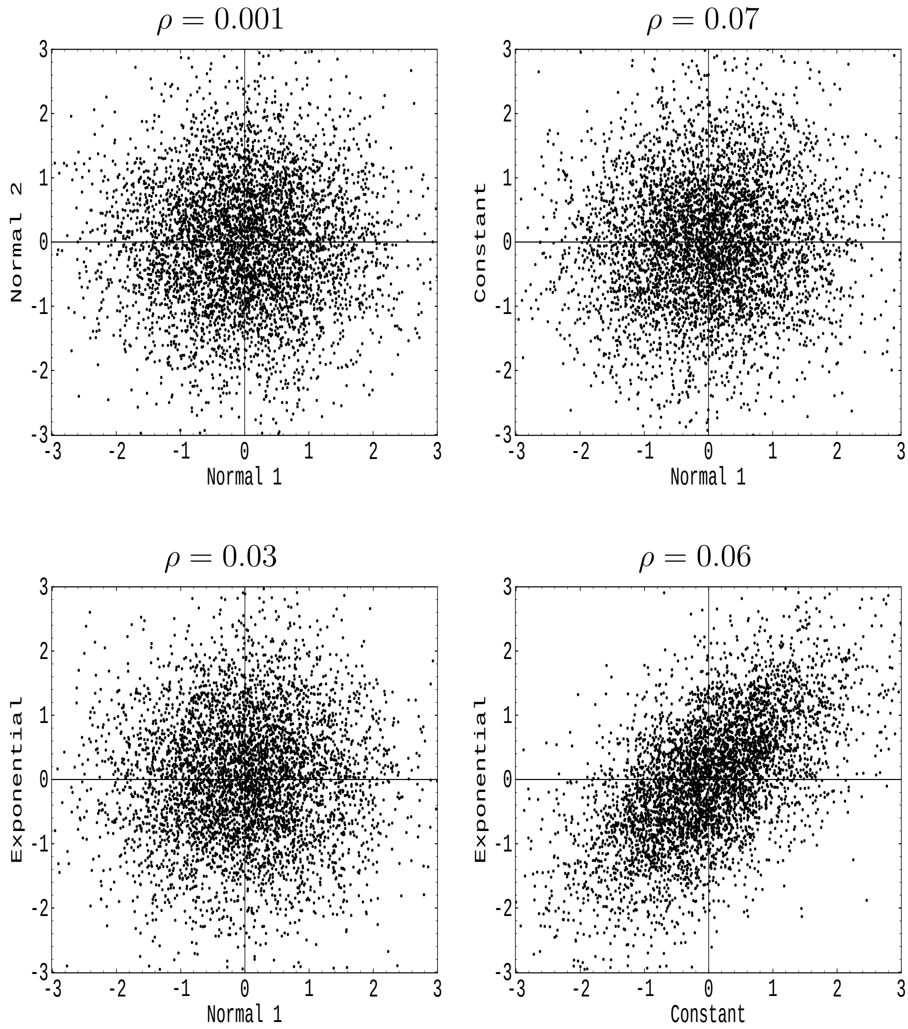

Figure 1: *Scatter plots of the simulated test statistics. "Normal 1" and "Normal 2" indicate that the between subject variables, ξ_{ij}, are simulated from a normal distribution; "Constant" indicates that they are constant equally to zero; and "Exponential" that they are simulated from an exponential distribution, i.e. a shewed distribution.*

If there were no such dependence, the four simulated test statistics would have generated the same value on the test statistic for all 5000 repetitions. Remember that the only thing differing between the four test statistics is the values on the between subject variables, ξ_{ij}. Even though we not necessary reach the same decision from the four test statistics, they follow the same distribution. This is seen in the histograms in Figure 2, where we see that the distribution of the four test statistics follows the t-distribution with $n - 3$ degrees of freedom.

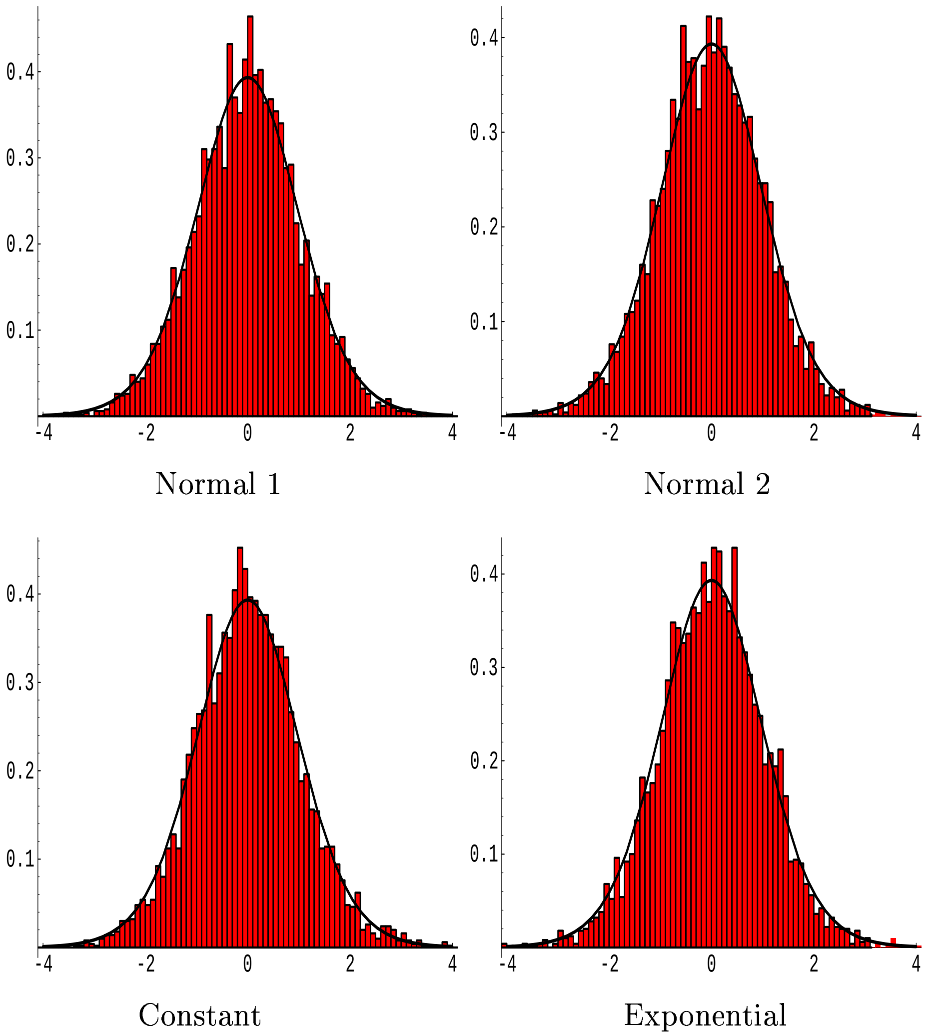

Figure 2: *Histograms of the simulated test statistics. In each histogram, we have also plotted the probability density function of Student's t distribution with* $n-3$ *degrees of freedom as a reference curve.*

The decision drawn from a test statistic depend on the values on the between subject variables, but the distribution on the test statistic does not depend on the distribution on the between subject parameters. The dependence between test statistics depend on the relation between the within subject variability and the between subject variability. Small within subject variability compared to between subject variability generates almost independent statistics (see Figure 1).

Appendix: Proof of theorem 1

Proof:

(a)

If we can show that $\mathbf{E} \sim W_p(n-p-2, \mathbf{\Sigma})$, $\mathbf{H} \sim W_p(p, \mathbf{\Sigma}_{--})$ and that $\mathbf{E}$ and $\mathbf{H}$ are statistically independent, then it follows that $\Lambda(\mathbf{\Gamma}) = |\mathbf{E}|/|\mathbf{E}+\mathbf{H}| \sim U_{p,p,n-p-2}$. Using the expressions (10) and (11) we can express $\mathbf{H}$ in terms of $\mathbf{Y}^-$ as follows,

$$\begin{aligned}
\mathbf{H} &= \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-} = \\
&= \mathbf{Y}^-(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})' \left[(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})'\right]^{-1} (\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})(\mathbf{Y}^-)' = \\
&= (\mathbf{Y}^-)\mathbf{A}(\mathbf{Y}^-)'.
\end{aligned}$$

Here we have defined the $p \times p$ matrix $\mathbf{A}$ as

$$\mathbf{A} = (\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})' \left[(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})'\right]^{-1} (\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}}).$$

Thus, $\mathbf{A}$ contains only $\mathbf{Y}^{\mathbf{M}}$. Let for the moment $\mathbf{Y}^{\mathbf{M}}$ and thus also $\mathbf{A}$ be fixed. When H_0 is true, fixing $\mathbf{Y}^{\mathbf{M}}$ does not influence the distribution of $\mathbf{Y}^-$ because $\mathbf{Y}^-$ and $\mathbf{Y}^{\mathbf{M}}$ are independent. A theorem in multivariate statistics says that if $\mathbf{A}$ is idempotent, then

$$(\mathbf{Y}^-)\mathbf{A}(\mathbf{Y}^-)' \sim W_p\left(k, \mathbf{\Sigma}_{--}, \boldsymbol{\mu}^-\mathbf{A}(\boldsymbol{\mu}^-)'\right) \tag{26}$$

(see for example Arnold (1981)). Here, $\boldsymbol{\mu}^-\mathbf{A}(\boldsymbol{\mu}^-)'$ is the noncentrality matrix; $\boldsymbol{\mu}^-$ denotes the expected mean matrix corresponding to $\mathbf{Y}^-$, i.e.

$$\boldsymbol{\mu}^- = (\boldsymbol{\mu}_1^-, \boldsymbol{\mu}_1^-, \ldots, \boldsymbol{\mu}_1^-, \boldsymbol{\mu}_2^-, \boldsymbol{\mu}_2^-, \ldots, \boldsymbol{\mu}_2^-);$$

and k is the rank of the $\mathbf{A}$ matrix. $\mathbf{A}$ is idempotent because $\mathbf{A}^2 = \mathbf{A}$. The rang of $\mathbf{A}$ equals the rang of $\mathbf{Y}^{\mathbf{M}}$, which has the rang p (with probability 1). The noncentrality matrix, $\boldsymbol{\mu}^-\mathbf{A}(\boldsymbol{\mu}^-)'$, equals the null matrix because $\boldsymbol{\mu}^-(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})' = \mathbf{0}$. So for fixed $\mathbf{A}$, $\mathbf{H}$ follows the central Wishart distribution with p degrees of freedom, i.e. $\mathbf{H} \sim W_p(p, \mathbf{\Sigma}_{--})$. Since this distribution does not depend on $\mathbf{Y}^{\mathbf{M}}$, $\mathbf{H}$ is unconditionally $W_p(p, \mathbf{\Sigma}_{--})$ and independent of $\mathbf{Y}^{\mathbf{M}}$. When $H_0 : \mathbf{\Sigma}_{-\mathbf{M}} = \mathbf{0}$ is true, $\mathbf{Y}^{\mathbf{M}}$ and $\mathbf{Y}^-$ are independent even if we only assume that the between subject parameters, $\boldsymbol{\xi}_i$ are independent of the within subject parameters, ε_{ij1} and ε_{ij2}. This mean that $\mathbf{H} \sim W_p(p, \mathbf{\Sigma}_{--})$ holds with no distributional assumption made on $\boldsymbol{\xi}_{ij}$. In particular, they are not assumed to be independent or identically distributed.

The same argument can be used to show that $\mathbf{E} \sim W_T(n-2-T, \mathbf{\Sigma}_{--})$ holds with no distributional assumption made about $\boldsymbol{\xi}$. Let $\mathbf{b}_1$ be the $n \times 1$ vector ,

$$\mathbf{b}_1 = 1/\sqrt{n_1}(1, 1, \ldots, 1, 0, 0, \ldots, 0)',$$

in which there are n_1 nonzero elements. Define $\mathbf{b}_2$ in the similar way, i.e.

$$\mathbf{b}_2 = 1/\sqrt{n_2}(0, 0, \ldots, 0, 1, 1, \ldots, 1)',$$

but here with n_2 nonzero elements. Now, we can write $\bar{\mathbf{Y}}^-$ as

$$\bar{\mathbf{Y}}^- = \mathbf{Y}^-(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')$$

Using that the $n \times n$ matrix $\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2'$ is symmetric, we can express $\mathbf{S}_{--}$ as

$$\begin{aligned}\mathbf{S}_{--} &= \left(\mathbf{Y}^- - \bar{\mathbf{Y}}^-\right)\left(\mathbf{Y}^- - \bar{\mathbf{Y}}^-\right)' = \\ &= \mathbf{Y}^-\left(\mathbf{I}_{(n\times n)} - (\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')\right)(\mathbf{Y}^-)',\end{aligned}$$

where $\mathbf{I}_{(n\times n)}$ is the $n \times n$ identity matrix. With this notation, $\mathbf{E}$ can be expressed as

$$\begin{aligned}\mathbf{E} &= \mathbf{S}_{--} - \mathbf{S}_{-\mathbf{M}}\mathbf{S}_{\mathbf{MM}}^{-1}\mathbf{S}_{\mathbf{M}-} = \\ &= \mathbf{S}_{--} - \mathbf{H} = \\ &= \mathbf{Y}^-\left(\mathbf{I}_{(n\times n)} - (\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')\right)(\mathbf{Y}^-)' - \mathbf{Y}^-\mathbf{A}(\mathbf{Y}^-)' = \\ &= \mathbf{Y}^-\left(\mathbf{I}_{(n\times n)} - (\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2') - \mathbf{A}\right)(\mathbf{Y}^-)' = \\ &= \mathbf{Y}^-\mathbf{B}(\mathbf{Y}^-)',\end{aligned}$$

where we have defined the $p \times p$ matrix $\mathbf{B}$ as

$$\mathbf{B} = \left(\mathbf{I}_{(n\times n)} - (\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2') - \mathbf{A}\right).$$

The rank of $\mathbf{B}$ is $n - 2 - p$ and to show that $\mathbf{B}$ is idempotent we use the following facts

$$\begin{aligned}&\mathbf{A}^2 = \mathbf{A} \\ &(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2') = (\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2') \\ &(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}}) = \mathbf{0} \\ &(\mathbf{Y}^{\mathbf{M}} - \bar{\mathbf{Y}}^{\mathbf{M}})'(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2') = \mathbf{0}.\end{aligned}$$

The noncentrality matrix, $\boldsymbol{\mu}^-\mathbf{B}(\boldsymbol{\mu}^-)' = \mathbf{0}$ because

$$\boldsymbol{\mu}^-\mathbf{I}_{(n\times n)}(\boldsymbol{\mu}^-)' = \boldsymbol{\mu}^-(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')(\boldsymbol{\mu}^-)'$$

and $\boldsymbol{\mu}^-\mathbf{A}(\boldsymbol{\mu}^-)' = \mathbf{0}$. Arguing as before shows that $\mathbf{E} \sim W_p\left(n - 2 - p, \boldsymbol{\Sigma}_{--}\right)$ holds with no distributional assumption made about the $\boldsymbol{\xi}$'s when H_0 is true.

It remains to show that $\mathbf{E}$ and $\mathbf{H}$ are statistically independent when H_0 is true. We use the result from multivariate statistics that says: "the Wishart variables $\mathbf{YAY}'$ and $\mathbf{YBY}'$ are independent if and only if $\mathbf{AB} = \mathbf{0}$" (see for example Arnold (1981)). So, $\mathbf{E}$ and $\mathbf{H}$ are independent if and only if $\mathbf{BA} = \mathbf{0}$. Thus, $\mathbf{E}$ and $\mathbf{H}$ are independent because

$$(\mathbf{b}_1\mathbf{b}_1' + \mathbf{b}_2\mathbf{b}_2')\mathbf{A} = \mathbf{0}$$

and $\mathbf{A}^2 = \mathbf{A}$.

(b)

Define $\mathbf{C}_{ij}$ as

$$\mathbf{C}_{ij} = (\mathbf{Z}_{ij} - \bar{\mathbf{Z}}_j) + 2(\boldsymbol{\xi}_{ij} - \bar{\boldsymbol{\xi}}_j),$$

where $\mathbf{Z}_{ij}$ is defined in (12) and is independent of $\mathbf{Y}_{ij}^-$. Thus, the $\mathbf{Y}_{ij}^-$'s and the $\mathbf{C}_{ij}$'s are independent. Let $\mathbf{C}_j$ be the $p \times n_j$ matrix containing the $\mathbf{C}_{ij}$'s from treatment sequence j, i.e.

$$\mathbf{C}_j = (\mathbf{C}_{1j}, \mathbf{C}_{2j}, \ldots, \mathbf{C}_{n_j j}), \qquad j = 1, 2.$$

Further, let

$$\mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2).$$

With this notation, we can express the sum of squares and cross-product matrices $\mathbf{S}_{--}, \mathbf{S}_{-\mathbf{M}}, \mathbf{S}_{\mathbf{M}-}$ and $\mathbf{S}_{\mathbf{MM}}$ in terms of $\mathbf{Y}^-$, $\bar{\mathbf{Y}}^-$ and $\mathbf{C}$ as

$$\begin{aligned}
\mathbf{S}_{--} &= (\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)' \\
\mathbf{S}_{-\mathbf{M}} &= (\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)'(\boldsymbol{\Gamma} - \mathbf{M}) + (\mathbf{Y}^- - \bar{\mathbf{Y}}^-)\mathbf{C}' \\
\mathbf{S}_{\mathbf{M}-} &= (\boldsymbol{\Gamma} - \mathbf{M})(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)' + \mathbf{C}(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)' \\
\mathbf{S}_{\mathbf{MM}} &= (\boldsymbol{\Gamma} - \mathbf{M})(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)'(\boldsymbol{\Gamma} - \mathbf{M}) + (\boldsymbol{\Gamma} - \mathbf{M})(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)\mathbf{C}' + \\
&\quad +\mathbf{C}(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)'(\boldsymbol{\Gamma} - \mathbf{M}) + \mathbf{C}\mathbf{C}'.
\end{aligned}$$

The test statistic under H_0, is with these notations

$$\Lambda(\boldsymbol{\Gamma}) \quad = \quad \frac{|(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)'|}{|(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)(\mathbf{Y}^- - \bar{\mathbf{Y}}^-)'|}.$$

Thus, the test statistic is a function of the random variables $\mathbf{Y}_{ij}^-$ and $\mathbf{C}_{ij}$. According to the assumptions made earlier, the $\mathbf{Y}_{ij}^-$'s follow a multivariate normal distribution and are independent of the $\mathbf{C}_{ij}$'s, whereas the distribution of the $\mathbf{C}_{ij}$'s is unknown. However, we know from the proof of (a) that the conditional distribution of $\Lambda(\boldsymbol{\Gamma})$ given $\mathbf{C}$ has the same distribution as the unconditional distribution, i.e. $\Lambda(\boldsymbol{\Gamma})$ as well as $\Lambda(\boldsymbol{\Gamma})$ given $\mathbf{C}$ follows an U-distribution, $\Lambda(\boldsymbol{\Gamma}) \sim U_{p,p,n-p-2}$. Because the distribution does not depend on the parameters, $(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}_{--})$, specifying the distribution of the $\mathbf{Y}_{ij}^-$'s, it follows from Basus' theorem (see for example Lehmann (1991)) that $\Lambda(\boldsymbol{\Gamma})$ is conditionally independent of the sufficient statistic for $(\boldsymbol{\mu}^-, \boldsymbol{\Sigma}_{--})$ given $\mathbf{C}$. A sufficient statistic for the mean vector and the covariance matrix in the multivariate normal distribution is of course $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$, which is a function of the $\mathbf{Y}_{ij}^-$'s. Thus $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$ and $\mathbf{C}$ are independent.

The fact that $\Lambda(\boldsymbol{\Gamma})$ is conditional independent of $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$ given $\mathbf{C}$ and that $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$ and $\mathbf{C}$ are independent, imply that $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$ and $(\Lambda(\boldsymbol{\Gamma}), \mathbf{C})$ are independent. Recall that the conditional distribution of $\Lambda(\mathbf{\Gamma})$ given $\mathbf{C}$ has the same distribution as the unconditional one. This mean that $\Lambda(\mathbf{\Gamma})$ and $\mathbf{C}$ are independent, i.e. $(\bar{\mathbf{Y}}_1^-, \bar{\mathbf{Y}}_2^-, \mathbf{S}_{--})$ and $\Lambda(\mathbf{\Gamma})$ are independent.

(c)

Using (7) together with (11) we see that $\mathbf{S}_{-\mathbf{M}}$ can be written as

$$\mathbf{S}_{-\mathbf{M}} = \mathbf{S}_{-+} - \mathbf{S}_{--}\mathbf{M}' \tag{27}$$

Thus, using that $\mathbf{S}'_{+-} = \mathbf{S}_{-+}$ we have for $\mathbf{M}^* = \mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$

$$\begin{aligned} \mathbf{S}_{-\mathbf{M}^*} &= \mathbf{S}_{-+} - \mathbf{S}_{--}(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1})' = \\ &= \mathbf{S}_{-+} - \mathbf{S}_{--}\mathbf{S}_{--}^{-1}\mathbf{S}_{-+} = \mathbf{0} \end{aligned}$$

and it follows that $\Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}) = 1$.

(d)

Set $\mathbf{M}^- = \mathbf{S}_{+-}\mathbf{S}_{--}^{-1} - \mathbf{A}$ and $\mathbf{M}^+ = \mathbf{S}_{+-}\mathbf{S}_{--}^{-1} + \mathbf{A}$. Then, because of (27) we have

$$\mathbf{S}_{-\mathbf{M}^-} = \mathbf{S}_{--}\mathbf{A} = -\mathbf{S}_{-\mathbf{M}^+}. \tag{28}$$

Moreover, using (7) together with (10) we have

$$\mathbf{S}_{\mathbf{M}^+\mathbf{M}^+} = \mathbf{S}_{\mathbf{M}^-\mathbf{M}^-} = \mathbf{S}_{++} + \mathbf{A}\mathbf{S}_{--}\mathbf{A}'. \tag{29}$$

Thus, using (28) and (29) in the expression (18) of the test statistic $\Lambda(\mathbf{M})$ we see that

$$\Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1} - \mathbf{A}) = \Lambda(\mathbf{S}_{+-}\mathbf{S}_{--}^{-1} + \mathbf{A}),$$

i.e. $\Lambda(\mathbf{M})$ is symmetric around $\mathbf{S}_{+-}\mathbf{S}_{--}^{-1}$.

□

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis.* John Wiley & Sons, Inc.

Arnold, S. F. (1981). *The theory of linear models and multivariate analyses.* John Wiley & Sons, Inc.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Cambridge Philis. Soc.*, **34**, 33–40.

Grender, J. M. and Johnson, W. D. (1993). Analysis of crossover designs with multivariate response. *Statistics in medicine*, **12**, 69–89.

Guilbaud, O. (1993). Exact inference about the within-subject variability in 2×2 crossover trials. *Journal of the American Statistical Association*, **88**(423), 939–946.

Jones, B. and Kenward, M. (1989). *Design and analysis of cross-over trials.* Chapman and Hall.

Lehmann, E. L. (1991). *Testing statistical hypothesis.* John Wiley & Sons, Inc.

Patel, H. I. and Hearne, E. M. (1980). Multivariate analysis for the two-period repeated measures crossover design with application to clinical trials. *Communications in statistics -theory and methods*, **A9**(18), 1919–1929.

Ratkowsky, D., Evans, M., and Alldredge, J. (1993). *Crossover experiments: design, analysis and application.* Marcel Dekker.

Rodriguez-Carvajal, L. A. and Freeman, G. H. (1999). Multivariate AB-BA crossover trial. *Journal of applied statistics*, **26**(3), 393–403.

Seber, G. A. F. (1984). *Multivariate obeservations.* John Wiley & Sons, Inc.

Senn, S. (1993). *Cross-over trials in clinical research.* John Wiley & Sons, Inc.

Wallenstein, S. and Fisher, A. C. (1977). The analysis of the two-period repeated measures crossover design with application to clinical trials. *Biometrics*, **33**, 261–269.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471–494.