Mathematical Analysis
& Convex Optimization

Vol. xxxxxx
https:\\ maco.lu.ac.ir
DOI: 10.22034/maco.6.1.1

Research Paper

A NEW METHOD FOR OBTAINING T3SSS IN THE T-WFE MODEL BASED ON P MATRICES

SAJJAD PIRADL*

Abstract. Type 3 sum of squares (T3SS) is used to test model comparisons that violate the marginality principle when testing main effects and are not dependent on the order of the model terms, and also, the sum of the individual effect SS is not equal to the total effect SS. This paper presents a new method to obtain T3SSs based on projection (P) matrices in two-way fixed effects (T-WFE) model. Previous research has shown that in analysis of variance (ANOVA) for unbalanced data, the total SS is not obtained by summing the T3SSs of the components considered as factors of variation. In fact, in such a case, there is a significant difference between the values of these two quantities. The method proposed in this paper uses P matrices to identify where differences occur and the magnitude of their differences, while the traditional ANOVA method cannot clearly explain these. This paper also discusses how to use the eigenvalues and eigenvectors of P matrices to obtain T3SSs. Then, a study was conducted on a real dataset based on an unbalanced dataset for the proposed model fitting method to calculate TESSs based on P matrices.

MSC(2010): 62J10.
Keywords: P matrices, T3SSs, Unbalance data, T-WFE model, ANOVA.

1. Introduction and motivation

Consider a case where the analysis of experimental data is performed assuming a T-WFE model. It is quite clear that in such a case, ANOVA is an analytical method that considers a linear model to test hypotheses regarding the effects of treatments that are to be compared in an experiment [?]. This method involves obtaining the total SS, which is the overall variation in the data, by summing the SS of the components considered as factors of variation. That is, in the ANOVA method, models are compared by decomposing the total SS into the specific SS of each factor. Recently, various studies have shown that, depending on the model fitting method, different techniques can be used to calculate factor-specific SS for ANOVA [?, ?, ?].

When calculating component SS through model comparison, it is classified into type 1 SS (T1SS), type 2 SS (T2SS), T3SS, etc., depending on the model fitting method [?, ?]. According to the model fitting method, SS gives the same value for each variation factor when the data is balanced [?]. However, in the case of unbalanced data where the number of data is not the same between treatments, the SS for each variation factor may vary depending on the model fitting method [?]. In the T1SS, regardless of whether the data are balanced or unbalanced, the total SS is decomposed into the SS of each factor, and the sum of these

Date: Received: , Accepted: .
*Corresponding author.

SSs is equal to the total variation of the data. The difference between the T1SS and T2SS can be identified through the analysis of the P matrices related to T2SS [?]. However, in the case of T3SS, if the data is unbalanced, the SS for each component is not provided by total variation decomposing of the data [?].

Although there are clear discussions in various studies about how to obtain SS according to the model fitting method, there is no clear discussion about where the difference in SS related to the fitting method comes from. When analyzing data in ANOVA, the concept of the P matrix allows for the interpretation of such differences. Some researchers discuss techniques for analyzing experimental data using P matrices for ANOVA [?, ?], and others discuss the use of P matrices for validating estimable functions in linear models with incomplete coefficients [?]. However, there is not much research on how to use P matrices in the general ANOVA method for data analysis.

This paper presents a method for estimating the model matrix derived from the T3SS model fitting method when the data are unbalanced, from the perspective of using the P matrix. When data analysis is done from this perspective, geometric interpretations can be simplified and descriptions of quantities that are not clearly specified in existing analytical methods can be concise and useful. Therefore, this method argues that the difference in data variation in the total SS decomposition using T3SSs can be explained specifically based on the P matrices. When SS is calculated based on the P matrix, various features of the matrix related to the vector space of observed vectors are extracted and used, which allows for the estimation and use of these features. Specifically, this paper discusses the calculation of SS using eigenvalues and eigenvectors of P matrices, as well as their applications. The content of this paper is organized as follows: Section **??** presents the definition and assumptions of the T-WFE model. Section 3 calculates the T3SSs in the T-WEF model using P matrices. In Section 4, a study on a real dataset is conducted based on an unbalanced dataset. Finally, Section 5 contains discussion and conclusions.

2. Definition and assumptions of the T-WFE model

In this section, a T-WFE model is assumed as a model for analyzing experimental data and its definition and assumptions are presented [?]. Suppose U and V are the two fixed factors considered in an experiment, where U represents the row factor consisting of levels in the row and V is considered as the column factor representing levels in the column. Also, it is assumed that the row factor U contains u levels and the column factor V contains v levels. A given treatment combination is represented by (k,l) which combines level k $(k = 1,2,3,\dots,u)$ of factor U and level l $(l = 1,2,3,\dots,v)$ of factor V. The experimental units used in the experiment are assumed to be homogeneous. It is also assumed that the treatment combination (k,l) is randomly assigned to the experimental units m_{kl}. Let $x_{klw},\ w = 1,2,3,\dots,m_{kl}$ be the observed response in the experimental unit in which a treatment combination (k,l) is administered. Since the two factors constituting the treatment combination are fixed factors, a T-WFE model is assumed as a model for the observed response of the experimental unit. If the level effect at level k of factor U is denoted by α_k, then k α_k's are fixed effects. Similarly, if the level effect at level l of factor V is represented by β_l, then the l β_l's also represent fixed effects. The fixed effects interaction in the treatment combination (k,l) of level k of factor U and level l of factor V is represented by γ_{kl}.

Now, based on all the above assumptions, the T-WFE model for data analysis can be presented as follows:

$$x_{klw} = \mu + \alpha_k + \beta_l + \gamma_{kl} + \delta_{klw}, \tag{2.1}$$

where μ is the overall mean of the data and δ_{klw}'s are the error terms, assuming that these terms are independent, identically distributed random variables and follow a normal probability distribution with a common mean of 0 and a common variance of a constant value σ^2_δ.

If the collected data of uv treatment combinations (k, l) represented by two-factor level combinations are expressed as a vector $\boldsymbol{x}$ of size $m \times 1$, the matrix model corresponding to equation (**??**) is expressed as follows:

$$\boldsymbol{x} = \mathbf{1}\mu + \boldsymbol{Y}_U\boldsymbol{\alpha} + \boldsymbol{Y}_V\boldsymbol{\beta} + \boldsymbol{Y}_{UV}\boldsymbol{\gamma} + \boldsymbol{\delta}, \tag{2.2}$$

where $m = \sum_{k=1}^{u}\sum_{l=1}^{v} m_{kl}$, the coefficient vector $\mathbf{1}$ of the population mean μ is a column vector with m elements all equal to 1, $\boldsymbol{Y}_U$ is a coefficient matrix of size $m \times u$ consisting of 0s and 1s, $\boldsymbol{\alpha}$ is a column vector representing u fixed effects of factor U with $u \times 1$ elements, $\boldsymbol{Y}_V$ is a coefficient matrix of size $m \times v$ consisting of 0s and 1s, $\boldsymbol{\beta}$ is a column vector representing v fixed effects of factor V with $v \times 1$ elements, $\boldsymbol{Y}_{UV}$ is a coefficient matrix of size $m \times uv$ consisting of 0s and 1s, $\boldsymbol{\gamma}$ is a column vector representing the fixed effects of the interaction uv in the treatment combination (k, l) of level k of factor U and level l of factor V with $uv \times 1$ elements, and $\boldsymbol{\delta}$ is a column random vector representing the error terms with $m \times 1$ elements.

The matrix representation in equation (**??**) shows that $\boldsymbol{x}$ consists of the sum of the multidimensional component vectors. These component vectors correspond to the P matrices on the subspaces corresponding to the factors that contribute to the total variation in the data. Therefore, SS of the distances to the P matrix of each subspace represents the SS based on the variation factor. The SS of distances by P matrix vary depending on the model fitting method, and there are differences between the calculated values.

In the T-WFE model of equation (**??**), x_{klw} is expressed as a sum of five components. Similarly, ANOVA decomposes the total SS of observed values into five component SS. These are SS based on μ, SS based on variation in the u level effects of the row factor U, SS based on variation in the v level effects of the column factor V, SS based on the variation in the interaction effects at the uv combined levels of the row factor U and the column factor V, and error SS based on variation in the error terms δs.

3. Obtaining T3SSs in the T-WEF model using P matrices

In this section, the model fitting method for T3SS based on P matrices is reviewed. That is, under the assumption of the T-WFE model defined in equation (**??**), T3SS is calculated as the difference between the variation resulting from fitting a model that excludes the effect of interest and the variation resulting from fitting a model that includes the effect of interest. Through this model comparison method, T3SS is used to test hypotheses about the estimable effects functions in a given model. Among the different types of model comparison methods, the model fitting method that calculates T3SS is preferred because0 it has been shown that type 3 (T3) hypotheses are of interest to many researchers.

Now, to examine the specific process based on a P matrix, suppose $\boldsymbol{Y} = (\mathbf{1}, \boldsymbol{Y}_U, \boldsymbol{Y}_V, \boldsymbol{Y}_{UV})$ is the model matrix in the matrix model equation (**??**). Also, let SSU be the variation due to the effects of level u of the row factor U. Where SSU is the difference between SS obtained from fitting the T-WFE model and SS obtained from fitting the reduced model without considering the main effects of the row factor U. When the model matrix $\boldsymbol{Y}$ is represented as $\boldsymbol{Y}_u = (\mathbf{1}, \boldsymbol{Y}_V, \boldsymbol{Y}_{UV})$, without considering the coefficient matrix associated with row effects, , the SS based on the P matrix will be as follows:

$$SSU = \boldsymbol{x}^T \left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T \right] \boldsymbol{x}, \tag{3.1}$$

where $(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}$ is the inverse of the matrix $\boldsymbol{Y}^T\boldsymbol{Y}$, $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T$ is the P matrix on $\boldsymbol{Y}$, $(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}$ is the inverse of the matrix $\boldsymbol{Y}_u^T\boldsymbol{Y}_u$, $\boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T$ is the P matrix on $\boldsymbol{Y}_u$, and T is the transpose symbol.

Also, let SSV be the variation due to the effects of level v on the column factor V. SSV is the difference between SS from fitting the T-WFE model and SS obtained from fitting the reduced model without considering the main effects of the column factor V. When the model matrix $\boldsymbol{Y}$, without considering the coefficient matrix related to column effects, is $\boldsymbol{Y}_v = (\mathbf{1}, \boldsymbol{Y}_U, \boldsymbol{Y}_{UV})$, the SS based on the P matrix is as follows:

$$SSV = \boldsymbol{x}^T \left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_v(\boldsymbol{Y}_v^T\boldsymbol{Y}_v)^{-1}\boldsymbol{Y}_v^T \right] \boldsymbol{x}, \tag{3.2}$$

where $(\boldsymbol{Y}_v^T\boldsymbol{Y}_v)^{-1}$ is the inverse of matrix $\boldsymbol{Y}_v^T\boldsymbol{Y}_v$ and $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T$ is the P matrix on $\boldsymbol{Y}_v$.

On the other hand, suppose $SSUV$ be the SS of the interaction vector $\boldsymbol{\gamma}$ between two factors U and V. $SSUV$ is the difference between the SS obtained from fitting the T-WFE model and the SS obtained from fitting the reduced model without considering the interaction effects of the two factors U and V. When the model matrix $\boldsymbol{Y}$, without considering the matrix of coefficients related to the interaction, is $\boldsymbol{Y}_{uv} = (\mathbf{1}, \boldsymbol{Y}_U, \boldsymbol{Y}_V)$, the SS based on the P matrix is as follows:

$$SSUV = \boldsymbol{x}^T \left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T \right] \boldsymbol{x}, \tag{3.3}$$

where $(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}$ is the inverse of matrix $\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv}$ and $\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T$ is the P matrix on $\boldsymbol{Y}_{uv}$.

The SSs given in the quadratic form $\boldsymbol{x}$ are the best possible partitions of the total variation, but their sum does not equal the total SS. The reason for this is that the P matrix space generated by the model matrix $\boldsymbol{Y}$ must contain subspaces representing the four components, including the mean, but the model fitting method for finding the T3SS is simply divided into two orthogonal subspaces.

The values computed in the subspaces representing the components of the variation factors are not given by the SS of the P matrices in each subspace, due to the redundancy of these spaces. In this case, one way to calculate the redundancy associated with each variation component is to use the P matrix of the model matrix $\boldsymbol{Y}$ on a subspace consisting of non-overlapping component vectors. The space of the P matrix corresponding to the model matrix $\boldsymbol{Y}$ can be divided into two spaces: one with the $\boldsymbol{Y}_{uv}$ matrix and one without it. Furthermore, $\boldsymbol{Y}_{uv}$ does not include the matrix $\boldsymbol{Y}_{UV}$, which is usually included in the P matrices representing the subspace when computing the variation due to the main effects of the factors U and V. Therefore, the variation due to the main effects of the row factor U and the column factor V can be calculated by applying the P matrix to the space generated by $\boldsymbol{Y}_{uv}$ that does not include $\boldsymbol{Y}_{UV}$. Also, since $\boldsymbol{Y}_{uv}$ consists of three component vectors, SS is calculated corresponding to the sequential fitting of the model.

Now, suppose that SST is the product of the distance of one P matrix to another on $\boldsymbol{Y}_{uv}$ as follows:

$$SST = \boldsymbol{x}^T\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T\boldsymbol{x} + \boldsymbol{x}^T\left[\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T\right]\boldsymbol{x}, \tag{3.4}$$

where $(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}$ is the inverse of matrix $\mathbf{1}^T\mathbf{1}$ and $\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T$ is the P matrix on $\mathbf{1}$.

Assuming that the second term on the right side of equation (**??**) is equal to SSt, it can be expressed as the sum of two components $SStu$ and $SStv$. $SStu$ can be written as follows:

$$SStu = \boldsymbol{x}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{x}, \tag{3.5}$$

where $\boldsymbol{A} = \left[\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T\right]\boldsymbol{Y}_U$.

We can see that $SStu$ is the quadratic form of $\boldsymbol{x}$. The P matrix corresponding to this quadratic form can be written as follows:

$$\boldsymbol{Y}_{tu}(\boldsymbol{Y}_{tu}^T\boldsymbol{Y}_{tu})^{-1}\boldsymbol{Y}_{tu}^T = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T. \tag{3.6}$$

Also, $SStv$ is obtained as follows:

$$SStv = \boldsymbol{x}^T\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{x}, \tag{3.7}$$

where $\boldsymbol{B} = \left[\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T - \boldsymbol{Y}_{tu}(\boldsymbol{Y}_{tu}^T\boldsymbol{Y}_{tu})^{-1}\boldsymbol{Y}_{tu}^T\right]\boldsymbol{Y}_V$.

We can see that $SStv$ is also a quadratic form of $\boldsymbol{x}$. The P matrix corresponding to this quadratic form can be written as follows:

$$\boldsymbol{Y}_{tv}(\boldsymbol{Y}_{tv}^T\boldsymbol{Y}_{tv})^{-1}\boldsymbol{Y}_{tv}^T = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T. \tag{3.8}$$

Therefore, it follows that the P matrix $\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T$ on $\boldsymbol{Y}_{uv}$ can be expressed as the sum of three orthogonal P matrices as follows:

$$\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T + \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T + \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T. \tag{3.9}$$

Then, using equations (**??**) to (**??**), the SST formula in equation (**??**) can be rewritten as follows:

$$SST = \boldsymbol{x}^T\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-\mathbf{1}}\mathbf{1}^T\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{x}. \tag{3.10}$$

Equation (**??**) describes the differences in T3SS for each factor, where the variation in the overall data does not correspond to the variation in the entire data set. This shows that the P matrix space corresponding to the model fitting method for calculating factor variation is expressed as a sum of mutually orthogonal P matrices, when they are partitioned orthogonally. Using these matrices, it is possible to identify where the difference between the sum of the component variation and the total variation occurs and determine the extent of that difference.

4. **Study on a real dataset**

In this section, an unbalanced dataset is used as an example of a study on a real dataset to calculate T3SSs. Data were obtained from a small two-way (T-W) treatment structure experiment conducted in a completely randomized design structure. For more details, see [**?**]. Table **??** presents these experimental data from the combined levels of two treatment factors, U and V, at levels U_1 and U_2 and V_1, V_2 and V_3, respectively:

U \ V	$\boldsymbol{V_1}$	$\boldsymbol{V_2}$	$\boldsymbol{V_3}$
$\boldsymbol{U_1}$	19, 20, 21	24, 26	22, 25, 25
$\boldsymbol{U_2}$	25, 27	21, 24, 24	31, 32, 33

TABLE 1. An unbalanced T-W experimental data.

Here, according to the matrix model of equation (**??**), an attempt is made to calculate T3YY based on the variation in factors using the P matrices. Using equation (3.1), SSU is obtained as 61.71. SSU can also be calculated using the eigenvectors of the P matrix associated with the quadratic form of equation (**??**). Specifically, the eigenvectors that have non-zero eigenvalues of the P matrix $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T$ of equation (3.1) are considered as follows. This is because $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T$ is an idempotent matrix, meaning that its eigenvalues are either 0 or 1. On the other hand, since the row factor U has two levels, its degree of freedom (df) is 1. Therefore, the P matrix $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T$ represents a matrix in a one-dimensional subspace and has only one non-zero eigenvalue, equal to 1. The eigenvector corresponding to the eigenvalue 1 is given by $\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 \\ J_2 \end{pmatrix}$, where the vector $\boldsymbol{J}_1^T$ is equal to $(-0.22, -0.22, -0.22, -0.33, -0.33, -0.22, -0.22, -0.22)$ and the vector $\boldsymbol{J}_2^T$ is equal to $(0.33, 0.33, 0.22, 0.22, 0.22, 0.22, 0.22, 0.22)$.

Therefore, the SSU calculation is performed using the eigenvector $\boldsymbol{J}$, which is given by $\boldsymbol{x}^T\boldsymbol{J}(\boldsymbol{J}^T\boldsymbol{J})^{\boldsymbol{-1}}\boldsymbol{J}^T\boldsymbol{x}$ and yields a value of 61.70, which is almost the same as equation (**??**). It should be noted that the difference of 0.01 is due to a rounding error in the notation of the eigenvector component. Furthermore, eigenvalues and eigenvectors not only facilitate the calculation of variation due to factors, but also help in understanding the number of df and the dependence of vector spaces. Then, the amount of variation due to the level of factor V is obtained by equation (**??**) as $SSV = 77.17$. Also, the SS resulting from the $SSUV$ interaction is obtained using equation (3.3) as 71.63.

The P matrix on the model matrix $\boldsymbol{Y}$ is equal to $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{x}$, so the SS of the distances to the P matrix is equal to $\boldsymbol{x}^T\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{x}$, and this value is equal to 10.19. The SS resulting from the overall mean μ is $\boldsymbol{x}^T\boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{\boldsymbol{-1}}\boldsymbol{1}^T\boldsymbol{x}$ which is equal to 9950.06. Therefore, the variation due to the main effects and the interaction of the factors U and V is equal to $\boldsymbol{x}^T\left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{\boldsymbol{-1}}\boldsymbol{1}^T\right]$ or 238.94. This quantity is not presented as the sum of SS due to the main effect of factor U (= 61.71), SS due to the main effect of factor V (= 77.17), and SS due to the interaction (= 71.63), because the sum of these three quantities is 210.51. The difference between these variations is because the partitioning of the P matrix space according to the model fitting method does not include mutually orthogonal subspaces.

In other words, the P matrix space for obtaining SS according to the main effect of factor U is generated by the P matrix $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T$ in equation (**??**), but for calculating SS it is not orthogonal to the P matrix space because the main effect of factor V is generated by the P matrix $\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_v(\boldsymbol{Y}_v^T\boldsymbol{Y}_v)^{-1}\boldsymbol{Y}_v^T$ in equation (**??**). The orthogonality of these two subspaces can be verified using the P matrix. The one-dimensional P matrix space for calculating SS due to the main effect of factor U is a subspace that is not orthogonal to the two-dimensional P matrix space for calculating SS due to the main effect of factor V.

The T3SS calculation, which does not consider the overlap of the P matrix space, does not correspond to the amount of variation in the P matrix space produced by the model matrix. Therefore, to understand the difference in the values of variation, we can see that SSU in equation (**??**) and SSV in equation (**??**) are computational methods that allow for the overlap of P matrix spaces. Also, since the P matrix space defined by $\boldsymbol{Y}_{uv}$ is the only space that can be decomposed into three orthogonal subspaces, the difference is calculated using the orthogonal decomposition of the P matrix $\boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T$.

The value of $SStu$ in equation (**??**) is equal to 76.56. The difference between $SStu$ and SSU is 14.85. From equation (**??**)(3.7), the value of $SStv$ is obtained as 90.74. Also, the difference between $SStu$ and SSV is 13.58. The sum of $SStu$, $SStv$, and $SSUV$ is $\boldsymbol{x}^T\left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{-\boldsymbol{1}}\boldsymbol{1}^T\right]\boldsymbol{x}$, which is 238.94.

All the results of the T3SS calculation based on the factors of variation are shown in the form of related quadratic equations in Table **??**:

Source of variation	**df**	**SS**	**MS**
$\boldsymbol{U}$ **factor**	1	$\boldsymbol{x}^T\left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_u(\boldsymbol{Y}_u^T\boldsymbol{Y}_u)^{-1}\boldsymbol{Y}_u^T\right]\boldsymbol{x} = 61.71$	61.71
$\boldsymbol{V}$ **factor**	2	$\boldsymbol{x}^T\left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_v(\boldsymbol{Y}_v^T\boldsymbol{Y}_v)^{-1}\boldsymbol{Y}_v^T\right]\boldsymbol{x} = 77.17$	38.59
UV factor	2	$\boldsymbol{x}^T\left[\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T - \boldsymbol{Y}_{uv}(\boldsymbol{Y}_{uv}^T\boldsymbol{Y}_{uv})^{-1}\boldsymbol{Y}_{uv}^T\right]\boldsymbol{x} = 71.63$	35.82
Error factor	10	$\boldsymbol{x}^T\left[\boldsymbol{I} - \boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\right]\boldsymbol{x} = 20.00$	2.00
Total	15	$\boldsymbol{x}^T\left[\boldsymbol{I} - \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{-\boldsymbol{1}}\boldsymbol{1}^T\right]\boldsymbol{x} = 258.94$	17.26

TABLE 2. T3SS values and their quadratic forms for the data in Table **??**.

where $MS = \frac{SS}{df}$ is the mean square and $\boldsymbol{I}$ is an identity matrix with $m \times m$ elements.

5. Discussion and conclusion

This paper discusses a P matrix-based method for calculating T3SS under the assumption of the T-WFE model. T3SS is obtained using the P matrix derived from the model fitting method. That is, the calculation of the variation due to the parameter vector of each component is performed using the P matrix in the corresponding subspace as the distance SS to the P matrix. Here, the distance SS to the P matrix is the quadratic form of $\boldsymbol{Y}$. In the quadratic form of $\boldsymbol{Y}$, the matrix is given by the P matrix in the subspace and is idempotent. When the P matrix space corresponding to the model matrix $\boldsymbol{Y}$ is decomposed into subspaces according to the model fitting method to calculate T3SS, the P matrix space corresponding to $\boldsymbol{Y}$ includes non-orthogonal and overlapping subspaces.

On the other hand, this paper examines how SS in a subspace is not represented by all components, and this variation is due to the overlapping nature of the subspaces of the P matrix and how the nestedness of the subspaces can be verified using the eigenvectors of each space. In computing T3SS using P matrices, it is discussed how the non-orthogonality of the P matrix subspaces relates to the model fitting method. It is shown that the nestedness of the subspaces can be verified using the eigenvectors of the corresponding P matrices.

Furthermore, the form of the P matrix in each subspace for calculating the T3SS obtained from the model fitting method has been identified, and it is shown that these P matrices are quadratic form related matrices that represent the SS of each component.

Also, this paper discusses how to use the corresponding P matrix subspace for the T3SS corresponding to each variation factor as a quadratic form and shows how the eigenvalues and eigenvectors of this matrix can be used to calculate the SS. Finally, this paper describes the overlap of subspaces in which the T3SS caused by the variation factor does not correspond to the total variation in the data and discusses how to use the eigenvectors of the P matrix to specifically identify this overlap.

Acknowledgments

This research received no external funding.

References

[1] N. Huntington-Klein, *The effect: An introduction to research design and causality*, Routledge / Taylor & Francic, UK, 2022.

[2] L. K. Skipper, P. J. A. Skipper, Experimental design and scientific data analysis, *Routledge / Taylor & Francic*, UK, 2024.

[3] G. P. Quinn, M. J. Keough, Experimental design and data analysis for biologists, 2th edition , *Cambridge University Press*, UK, 2023.

[4] R. A. Johnson, R. G. K. Bhattacharyya, Statistics: Principles and methods, 8th edition, *John Wiley & Sons*, USA, 2019.

[5] C. Hirotsu, Advanced analysis of variance, *John Wiley & Sons*, USA, 2017.

[6] G. A. Milliken, D. E. Johnson, Analysis of Messy data volume 1: Designed experiments, 2th edition, *Chapman and Hall / CRC*, USA, 2009.

[7] L. Durell, N. Mastin, L. M. Lovin, W. Baylor Steele IV, B. W. Brooks, A. S. Hering, Balanced two-way functional ANOVA: A case study for toxicological photolocomotor response studies, *Journal of Agricultural, Biological and Environmental Statistics*, 1–30, 2025.

[8] L. R. LaMotte, Testing ANOVA effects: A resolution for unbalanced models, *Communications in Statistics – Theory and Methods*, **53**(22):7860-7870, 2024.

[9] X. Cao, J. Fang, Z. Yao, Strong sum of projections in type II factors, *Journal of Functional Analysis*, **281**(5):1–11, 2021.

[10] S. Goldstein, S. A. Paszkiewicz, Linear Combinations of Projections in Type III Factors, *Integral Equations and Operator Theory*, **94**(4):1–32, 2022.

[11] W. Yu, Q. Zhang, W. Li, High-dimensional projection-based ANOVA test, *Journal of Multivariate Analysis*, **207**:1–15, 2025.

[12] J. Tyssedal, S. Hussain, Factor screening in nonregular two-level designs based on projection-based variable selection, *Journal of Applied Statistics*, **43**(3):490–508, 2016.

[13] K. Aishima, Consistent estimation with the use of orthogonal projections for a linear regression model with errors in the variables, *Linear Algebra and its Applications*, **681**(1):101–126, 2024.

[14] J. Aguilar-Loyo, A comparative analysis of two-way fixed effects estimators in staggered treatment designs, *Journal of Econometrics*, **251**:1–16, 2025.

(Sajjad Piradl) Department of Statistics, Payame Noor University, Tehran, Iran
Email address: `sajjadpiradl@pnu.ac.ir`