InfoDecompuTE: an R package for **info**rmation **decomp**osition of the **t**wo phase **e**xperiments.

Here, we presented an R package that allows the users to enter any design with the relationship of the random and fixed factors using the Wilkinson and Rogers’ syntax. (Structural formula)

**Ideas for the paper**

**When can we use this package?**

The researchers can use this package to generate the structure of the analysis of variance table with the coefficients of the variance components for the expected mean square from a complex one/two-phase experimental design.

It is important to study the coefficients of the EMS, because it allows us to check for any valid statistical test.

This package allows the statistical researcher to study how the information is decomposed or split to different strata and different sources of variation.

Currently, there is no any freely available statistical software that can compute the coefficients of the variances components of the expected mean square from a given design.

Genstat and aov function of R program can produce the structure of ANOVA table with the degrees of freedom for each source of variation, but they do not give the coefficients of the variance components of the EMS. SAS and Minitab have the ability to compute the coefficients by matching the suffices of each terms in the linear model, but neither programs can cope the two phase experiments.

**How useful is this package?**

The package is written in R program; hence most statistical researchers should be familiar with this programming language.

The package is open sourced, so the user can study and adjust any part of these functions.

This package can also analysis the balanced incomplete block design and produce the efficiency factors for each source of variation.

The user also can use the treatment contrasts for each treatment effects. The users can also break down the treatment effects into multiple orthogonal contrasts, since the confounding may only occur on the specific contrasts of the treatment effects.

The function also aware the orders of the factors are fitted.

**What does not this package do?**

This package can only analysis the experiments up to two phases. If another phase is added, it will increase the computation time from n^2 to n^3, because an additional for loop is require to define the block structure of the additional phase.

Users still need to have some understandings on how to build the model using the structural formula for each tier.

**The parameters that the function takes and what is the output likes.**

The parameters that the function takes can be divided into two main parts. The first part is the design of the experiment; the second part is the relationships between random and fixed factors of the experiment.

The first parameter consists of the experimental design in a data frame format. The values of the data frame should be factors. For example…..

The relationship between random and fixed factors can be shown using the Wilkinson and Rogers’ syntax and written as the structural formula. (Maybe a section discussing the crossing and nesting??)

The one-phase experiment is represented by two parameters, one for random factors and the other for fixed factors. These parameters are random.terms and fixed.terms.

The two-phase experiment is represented by three parameters: these are the random factors of phase 2 experiments, the random factors of the phase 1 experiment and the fixed factors of the overall experiment. The parameters are random.terms2, random.term1 and fixed.terms.

The var.comp parameter allows the researcher to decide which variance components to show on the analysis of variance table and their order.

The statistical researchers can also specify the contrasts for the fixed factors using the parameter trt.contr. The parameter trt.contr is a list of matrices for each of these contrasts. If the researchers use this parameter, then they need to set the contrasts for every fixed factor. The order of fixed factors for these contrasts needs to be the exact same order of the fixed factors in the parameter fixed.term.

The researchers can break down the fixed factors into multiple orthogonal contrasts; hence the fixed factors that are represented by multiple orthogonal contrasts become a list containing matrices for each of these contrasts. Then, the parameter trt.contr is still a list, but now it contains either a fixed factor that corresponds by a list of orthogonal contrasts or a single contrast matrix.

For a complex experimental design, it can be constituted by many random factors. This can become difficult to read on the analysis of variance table. The parameter table.legend allow the researcher to use the legend, which are the lower case alphabets, for the variance components of the analysis of variance table.

**How and where to install the package??**

Under <http://www.bioinformatics.org.nz/> or in cran….

**Theory: Information decomposition and how does this software works?**

Eigenvalue decomposition (Nelder ) or generalized inverse (John and Williams) to invert the singular matrices.

Using Projection and Pythagorean Theorem to define the block structure of the second phase.

We applied their formula by treating the random factors of the first phase experiment as the fixed factors

This overcomes the confounding that may occur between the random factors of second phase and first phase as shown in the design by Brien and Payne, 1999.

**Example**

**Brien and Payne, 1999**

Describe their design…. This is a nice example showing the confounding between the phase 1 and phase 2 random factors.

Specify the structure formula for each tier

Discuss the output, i.e. the ANOVA table

**May need another design that uses the contrasts**

**Conclusion**

How can be we improve this package?

Re-implement the matrix calculation in C to speed up the computation time.