\section[Introduction]{Introduction}

A primary objective of comparative experiments is to contrast measurements made on experimental units of material (e.g.\ humans, animals, plants, tissues, cells, etc.) in response to the interventions, or \emph{treatments}, to which they are subjected. Many practical situations arise in which the response variable of interest cannot be measured directly from the experimental units in a single experiment (Phase~1). Instead, the experimental units must be further processed in a subsequent experiment (Phase~2) before the measurements can be made. Such \emph{two-phase experiments} were introduced by \cite{McIntyre1955} in the context of a study of the effects of four light treatments on the synthesis of tobacco mosaic virus in tobacco leaves. Healthy tobacco plants were inoculated with the virus and subjected to different light treatments (Phase~1 experiment). To measure the disease severity, sap was first expressed from the experimental tobacco plants and then injected into the leaves of specific assay plants (Phase~2 experiment) on which lesions subsequently appeared and were counted.

Efforts have been made to develop a general theory for the design of two-phase experiments \citep{Brien1983, Wood1988, Brien1999, Jarrett2008}. In this development, \cite{Brien1983} presented a set of rules for deriving the analysis of variance (AVOVA) table with degrees of freedom (DF) based on the experimental structures. Within a same experimental structure, a set of factors sharing the same randomisation scheme is defined as a \emph{tier}. In general, two-phase experiments involve three tiers of factors: two tiers of block factors and one tier of treatment factors. Consequently, two-phase experiments are also known as \emph{multi-tiered experiments}. Tiers 1 and 2 comprise block factors from the Phase~2 and 1 experiments, respectively, while Tier 3 contains the treatment factors of the phase 2 experiments. The design procedure for the two-phase experiments comprises a two-step process: (1) the allocation of treatments to experimental units in the Phase~1 experiment, and (2) the allocation of experimental units from the Phase~1 experiment to the experimental units in the Phase~2 experiment. Hence, the randomisation procedure generally must be performed twice for both these two allocations. \cite{Brien2006b} thus named the randomisation procedure for the multi-tiered experiments, including the two-phase experiments, \emph{multiple randomisations}. \cite{Brien2009, Brien2010} discussed the decomposition of the space span by the vector of observations, or \emph{data vector}, for different types of multiple randomisation. The decomposition of the data space is a method to separate the total variability into different known sources of variation, more simply referred to as \emph{information decomposition}. If the subspaces, that are partitioned from the data space, are orthogonal to one another, then the method of information decomposition is also known as the ANOVA. \cite{Brien2011} summarised the previous publications by \citeauthor{Brien2004}, together with some fundamentals of the design of two-phase experiments. Additionally, \cite{Brien2011} provided a set of basic rules for deriving the \emph{expected mean square} (EMS), but these rules were applied manually and only for the generally balanced designs \citep{Payne2003}.

Two-phase experiments are common in studies that use high-throughput biotechnologies to identify and quantify different intracellular molecular species, such as gene transcripts, proteins, metabolites, etc. \cite{Jarrett2008} conducted a detailed comparative study of the properties of two competing designs -- multiple dye-swap and alternating loop \citep{Churchill2002}-- of the same size (i.e. same number of replicates of each treatment) for a two-colour microarray experiment at Phase~2, where the Phase~1 experiment was arranged in a completely randomised design (CRD). By constructing theoretical ANOVA tables for these two designs, they demonstrated that the distribution of the treatment information across the strata of the ANOVA table was dependent on the Phase~2 design. More specifically, while the multiple dye-swap design could be analysed using a simple ANOVA, the alternating loop design required a more involved analysis to test for treatment effects. This is a consequence of the sources of variation introduced in the Phase~2 experiment interacting with those introduced in Phase~1. Thus, \cite{Jarrett2008} illustrated the importance of considering the sources of variation introduced at each phase when designing two-phase experiments, and showed that constructing the relevant ANOVA tables can achieve this end.

Construction of ANOVA tables with complete information decomposition, including all known sources variation together with their associated DF and expected mean squares (EMSs), is a laborious manual task, even for small two-phase experiments. The commercial statistical software \proglang{Genstat} can be used to decompose known sources of variation from an experiment into their associated DF in the ANOVA, but only for generally balanced designs that require the use of pseudofactors \citep{Monod1992}. While the \code{AMTIER} procedure available within \proglang{Genstat} eliminates the need for pseuodfactors by enabling the fitting of three model formulae, two corresponding to the block structures of the Phase~1 and 2 experiments and one for the treatment structure, it also generates ANOVA tables with only the DF decomposition \citep{Brien2006a}. The limitation of these approaches is that they do not generate the treatment and residual EMSs needed to assess the preferred design(s) from a competing design. The \code{GLM} procedures in \proglang{JMP} and \proglang{SAS} and the \code{ANOVA} command in \proglang{Minitab} can compute the EMSs, but neither program can do this for two-phase experiments. Meanwhile, the \proglang{R} package \pkg{dae}, developed by \cite{Brien2011a}, can perform information decomposition of a given design. This package can only generate a projection matrix for each factor and perform a decomposition within a single stratum. Therefore, no existing statistical software package can perform all these tasks simultaneously and straightforwardly. Thus, we introduce an \proglang{R} package called \pkg{infoDecompuTE} for the {\bf info}rmation {\bf decomp}osition of {\bf T}wo-phase {\bf E}xperiments.

Our \pkg{infoDecompuTE} quickly generates complete theoretical ANOVA tables for the designs of any given single- or two-phase experiment, showing the distribution of the known sources of variation in the experiment, and thereby enabling researchers to compare the properties of competing designs. These tables also make it possible to quickly assess whether a particular design yields a valid F-tests for the treatment comparison of interest.

This article demonstrates the theoretical concepts and methods underlying \pkg{infoDecompuTE} as well as its use. Section~\ref{sec:infoDecomp} explains the information decomposition for a single-phase experiment. Section~\ref{sec:infoiDecompTwoPase} then generalises this to two-phase experiments, and shows how it differs to from that a single-phase experiment. Next, Section~\ref{sec:exampleTwoPase} shows an example of a two-phase experiment with the theoretical ANOVA tables. Section~\ref{sec:package} then demonstrates the use of \pkg{infoDecompuTE}. Finally, Section~\ref{sec:example} illustrates the application of infoDecompuTE to a two-phase viticulture-sensory evaluation experiment from a seminal paper on two-phase experiments \citep{Brien1999}.