Vignette to package 'samplingData'

Diana Trutschel

November 21, 2016

Contents

1 Introduction

1.1 Objective

To evaluate for example the match of different statistical model often simulation studies are used. In simulations the underlying data can be from a real study, but also simulated data given a distribution. For this purpose we provide a package to sampling data from normal distribution to mimic data of cluster randomised trials within different study designs, namely parallel, cross-over and stepped wedge design. Besides a traditional design collects sampling units within different groups, which should be compared, in the past years multilevel design, which collect additional units nested within the original sampling units, becomes very popular. Measurements of different patients nested within hospitals, also assigned as clusters, is one example of a two-level nested data. Another example of nested structures are the repeated measurements of patients. Furthermore, three-level nested data is obtained for example by the combination of both nesting examples. The complete data of such examples are then a multivariate distributed. The aim of this package is to provide a easy implementation of sampling multivariate normal distributed data for further investigations. Additionally, the requires power calculations for studies under these specific designs are given.

1.2 Different study types

Parallel, cross-over an stepped wedge designs. With this package we provide data sampling within three common used study design types: parallel, cross-over an stepped wedge designs (SWD). 1 shows examples of these kind of types, each with $C = 6$ cluster, followed over $T = 4$ time points. A parallel design is present, when two groups of treatments are given so that one group receives only the first while another group receives only second. In contrast to the parallel design in a crossover design trail each experimental unit (patient) receives different treatments during the different time points. Hence, it is a repeated measurements design. An alternative and more popular becoming design is the stepped wedge design. Here, the intervntion is rollout to different units sequential but random over different time points.

A	T_1	T_2	T_3	T_4
Center_1	0	0	0	0
Center_2	0	0	0	0
Center_3	0	0	0	0
Center_4	1	1	1	1
Center_5	1	1	1	1
Center_6	1	1	1	1

B	T_1	T_2	T_3	T_4
	0	0	1	1
	0	0	1	1
	0	0	1	1
	1	1	0	0
	1	1	0	0
	1	1	0	0

C	T_1	T_2	T_3	T_4
	0	1	1	1
	0	1	1	1
	0	0	1	1
	0	0	1	1
	0	0	0	1
	0	0	0	1

Table 1: Examples of different study design type of A) parallel, B) cross-over, and C) stepped wedge designs.

Cross-sectional versus longitudinal. In trials often subjects within clusters are followed over a period of time and measured to several measurment points. Two kinds of data collection is then possible: 1) cross-sectional data, if at each time point the measurment units (subjects) are different to the units at another timepoints, or 2) longitudinal data, the measurment units (subjects) are the same to all timepoints (known as repeated measurements). Hence, if it is a trail with C clusters and a clustersize of N each, which are follwed over T timepoints, then the total number of included subjects is $C \times N$ in a cross-sectional and $C \times T \times N$ in a longitudinal study.

2 Multivariate normal distributed data for multilevel data

For the situation of a cluster-randomized trail with T the number of time points, C the number of clusters and N the number of patients per cluster the complete dataset can be written as the vector of responses $\overrightarrow{Y} = \{Y_{ijk}\}$ of length $T \times C \times N$, which is sampled from a multidimensional normal distribution:

$$\overrightarrow{Y} \quad \sim \quad N\left(Zb, V\right),$$

where Zb is then the fixed effects full rank design matrix multiplied by the regression fixed effects coefficients and V the variance-covariance matrix. Y_{ijk} is then the observation in cluster i to time point k for the subject j in the cross-sectional case or the k-th measurement of the subject j in cluster i in the longitudinal case, respectively. The form of the random part (variance-covariance matrix) depends on the sampling of either cross-sectional or longitudinal study design, whereas the form of the fixed part depends on the study design type (parallel, cross-over an stepped wedge designs).

2.1 Fixed effects part

The regression fixed effects coefficients within the provided designs are defined as

$$b \quad = \quad (\mu, \beta_1, \cdots, \beta_I, \theta)\,,$$

where μ is the overall mean, θ is the intervention effect, β_k is the fixed time effect for time point $k, k = (1, \cdots, I)$.

The design matrix X of the model for such designs has the form

$$X \quad = \quad \begin{array}{c} \\ \text{cluster 1} \\ \vdots \\ \text{cluster i} \\ \vdots \\ \text{cluster C} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{time point 1} & \cdots & \text{time point k} & \cdots & \text{time point T} \end{array} \\ \begin{pmatrix} x_{11} & \cdots & \cdots & \cdots & x_{1T} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ x_{C1} & \cdots & \cdots & \cdots & x_{CT} \end{pmatrix} \end{array}$$

The fixed effects full rank design matrix Z is then a concatenation of all matrices Z_i of all clusters, which in turn are a concatenation of N replications of matrices Z_{ij} (hence for all j the Z_{ij} is the same) and which are created out of the design matrix X of the SWD model. Then is Z_{ij} for one subject in cluster i a column wise bind matrix of

1. a vector of ones (the same for all cluster)
2. a matrix A (the same for all cluster)
3. a vector, which is the corresponding row of the design matrix X to cluster i.

Each row of Z_{ij} corresponds then to a identify entry of a fixed effects in regression fixed effects coefficients vector b.

$$Z_{ij} \quad = \quad \begin{array}{c} \\ \text{time point 1} \\ \vdots \\ \text{time point k} \\ \vdots \\ \text{time point I} \end{array} \begin{array}{c} \begin{array}{cccccc} \mu & \beta_1 & \beta_2 & \cdots & \beta_{I-1} & \theta \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & x_{i1} \\ \vdots & 0 & \ddots & & \vdots & \vdots \\ 1 & \vdots & \ddots & \ddots & \vdots & x_{ik} \\ \vdots & \vdots & & \ddots & 1 & \vdots \\ 1 & 0 & \cdots & \cdots & 0 & x_{iT} \end{pmatrix} \end{array}$$

Hence, Z_i is build by row wise bind N replicates of Z_{ij}

$$Z_i \quad = \quad \begin{array}{c} \text{subject 1} \\ \vdots \\ \text{subject N} \end{array} \begin{pmatrix} Z_{ij} \\ \vdots \\ Z_{ij} \end{pmatrix}$$

and Z by row wise bind C matrices Z_i

$$Z \quad = \quad \begin{array}{c} \text{cluster 1} \\ \vdots \\ \text{cluster C} \end{array} \begin{pmatrix} Z_i \\ \vdots \\ Z_i \end{pmatrix}$$

Hence, each row corresponds to one subject j of a cluster i to timepoint k and multiplied with the vector of regression fixed effects coefficients b result in the fixed effect part of the linear equation for this observation. Hence, it is the mean vector of the multivariate normal distribution and it is performed by the matrix multiplication Zb.

Parallel design For example with $I = 4$ cluster and $K = 3$ measurments, hence only two cluster for either control or treatment arm, the design matrix X is defined as

$$X = \begin{array}{c} \\ \text{cluster 1} \\ \text{cluster 2} \\ \text{cluster 3} \\ \text{cluster 4} \end{array} \begin{array}{c} \begin{array}{cccc} \text{time point 1} & \text{time point 2} & \text{time point 3} & \text{time point 4} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

and the matrix Z_i for cluster 1 and 2 is then

$$Z_1 = \begin{array}{c} \\ \text{time point 1} \\ \text{time point 2} \\ \text{time point 3} \\ \text{time point 4} \end{array} \begin{array}{c} \begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

and for cluster 3 and 4

$$Z_1 = \begin{array}{c} \\ \text{time point 1} \\ \text{time point 2} \\ \text{time point 3} \\ \text{time point 4} \end{array} \begin{array}{c} \begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

then, if $N = 2$ subjects are within each cluster the fixed effects full rank design matrix Z is

$$Z = \begin{array}{c} \\ \text{cluster 1} \\ \text{cluster 2} \\ \text{cluster 3} \\ \text{cluster 4} \end{array} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix}$$

$$= \begin{array}{c} \\ \text{cluster 1 subject 1} \\ \text{cluster 1 subject 2} \\ \text{cluster 2 subject 1} \\ \text{cluster 2 subject 2} \\ \text{cluster 3 subject 1} \\ \text{cluster 3 subject 2} \\ \text{cluster 4 subject 1} \\ \text{cluster 4 subject 2} \end{array} \begin{pmatrix} Z_{1j} \\ Z_{1j} \\ Z_{2j} \\ Z_{2j} \\ Z_{3j} \\ Z_{3j} \\ Z_{4j} \\ Z_{4j} \end{pmatrix}$$

$$
= \begin{array}{r}
\\
\text{cluster 1 subject 1 time point 1}\\
\text{cluster 1 subject 1 time point 2}\\
\text{cluster 1 subject 1 time point 3}\\
\text{cluster 1 subject 1 time point 4}\\
\text{cluster 1 subject 2 time point 1}\\
\text{cluster 1 subject 2 time point 2}\\
\text{cluster 1 subject 2 time point 3}\\
\text{cluster 1 subject 2 time point 4}\\
\text{cluster 2 subject 1 time point 1}\\
\text{cluster 2 subject 1 time point 2}\\
\text{cluster 2 subject 1 time point 3}\\
\text{cluster 2 subject 1 time point 4}\\
\text{cluster 2 subject 2 time point 1}\\
\text{cluster 2 subject 2 time point 2}\\
\text{cluster 2 subject 2 time point 3}\\
\text{cluster 2 subject 2 time point 4}\\
\text{cluster 3 subject 1 time point 1}\\
\text{cluster 3 subject 1 time point 2}\\
\text{cluster 3 subject 1 time point 3}\\
\text{cluster 3 subject 1 time point 4}\\
\text{cluster 3 subject 2 time point 1}\\
\text{cluster 3 subject 2 time point 2}\\
\text{cluster 3 subject 2 time point 3}\\
\text{cluster 3 subject 2 time point 4}\\
\text{cluster 4 subject 1 time point 1}\\
\text{cluster 4 subject 1 time point 2}\\
\text{cluster 4 subject 1 time point 3}\\
\text{cluster 4 subject 1 time point 4}\\
\text{cluster 4 subject 2 time point 1}\\
\text{cluster 4 subject 2 time point 2}\\
\text{cluster 4 subject 2 time point 3}\\
\text{cluster 4 subject 2 time point 4}
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array}\\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0\\
1 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 1 & 0\\
1 & 0 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 0\\
1 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 1 & 0\\
1 & 0 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 0\\
1 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 1 & 0\\
1 & 0 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 0\\
1 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 1 & 0\\
1 & 0 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 1\\
1 & 0 & 1 & 0 & 1\\
1 & 0 & 0 & 1 & 1\\
1 & 0 & 0 & 0 & 1\\
1 & 1 & 0 & 0 & 1\\
1 & 0 & 1 & 0 & 1\\
1 & 0 & 0 & 1 & 1\\
1 & 0 & 0 & 0 & 1\\
1 & 1 & 0 & 0 & 1\\
1 & 0 & 1 & 0 & 1\\
1 & 0 & 0 & 1 & 1\\
1 & 0 & 0 & 0 & 1\\
1 & 1 & 0 & 0 & 1\\
1 & 0 & 1 & 0 & 1\\
1 & 0 & 0 & 1 & 1\\
1 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
$$

and the fixed part, hence the mean vector of the multivariate normal distribution, is then

$$\overrightarrow{\mu} \quad = \quad Z * \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \theta \end{pmatrix} = \begin{pmatrix} \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 + \theta \\ \mu + \theta \end{pmatrix}$$

Cross-over design For example with $I = 4$ cluster and $K = 4$ measurments, two cluster each switches treatment and control after time point 2, the design matrix X is defined as

$$X \quad = \quad \begin{array}{c} \\ \text{cluster 1} \\ \text{cluster 2} \\ \text{cluster 3} \\ \text{cluster 4} \end{array} \begin{array}{c} \begin{array}{cccc} \text{time point 1} & \text{time point 2} & \text{time point 3} & \text{time point 4} \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

and the matrix Z_i for cluster 1 and 2 is then

$$Z_1 = \begin{array}{c} \\ \text{time point 1} \\ \text{time point 2} \\ \text{time point 3} \\ \text{time point 4} \end{array} \begin{array}{c} \begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

and for cluster 3 and 4

$$Z_1 = \begin{array}{c} \\ \text{time point 1} \\ \text{time point 2} \\ \text{time point 3} \\ \text{time point 4} \end{array} \begin{array}{c} \begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

then, if $N = 2$ subjects are within each cluster the fixed effects full rank design matrix Z is

$$Z = \begin{array}{c} \text{cluster 1} \\ \text{cluster 2} \\ \text{cluster 3} \\ \text{cluster 4} \end{array} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix}$$

$$= \begin{array}{c} \text{cluster 1 subject 1} \\ \text{cluster 1 subject 2} \\ \text{cluster 2 subject 1} \\ \text{cluster 2 subject 2} \\ \text{cluster 3 subject 1} \\ \text{cluster 3 subject 2} \\ \text{cluster 4 subject 1} \\ \text{cluster 4 subject 2} \end{array} \begin{pmatrix} Z_{1j} \\ Z_{1j} \\ Z_{2j} \\ Z_{2j} \\ Z_{3j} \\ Z_{3j} \\ Z_{3j} \\ Z_{3j} \end{pmatrix}$$

$$
= \begin{array}{r}
\\
\text{cluster 1 subject 1 time point 1} \\
\text{cluster 1 subject 1 time point 2} \\
\text{cluster 1 subject 1 time point 3} \\
\text{cluster 1 subject 1 time point 4} \\
\text{cluster 1 subject 2 time point 1} \\
\text{cluster 1 subject 2 time point 2} \\
\text{cluster 1 subject 2 time point 3} \\
\text{cluster 1 subject 2 time point 4} \\
\text{cluster 2 subject 1 time point 1} \\
\text{cluster 2 subject 1 time point 2} \\
\text{cluster 2 subject 1 time point 3} \\
\text{cluster 2 subject 1 time point 4} \\
\text{cluster 2 subject 2 time point 1} \\
\text{cluster 2 subject 2 time point 2} \\
\text{cluster 2 subject 2 time point 3} \\
\text{cluster 2 subject 2 time point 4} \\
\text{cluster 3 subject 1 time point 1} \\
\text{cluster 3 subject 1 time point 2} \\
\text{cluster 3 subject 1 time point 3} \\
\text{cluster 3 subject 1 time point 4} \\
\text{cluster 3 subject 2 time point 1} \\
\text{cluster 3 subject 2 time point 2} \\
\text{cluster 3 subject 2 time point 3} \\
\text{cluster 3 subject 2 time point 4} \\
\text{cluster 4 subject 1 time point 1} \\
\text{cluster 4 subject 1 time point 2} \\
\text{cluster 4 subject 1 time point 3} \\
\text{cluster 4 subject 1 time point 4} \\
\text{cluster 4 subject 2 time point 1} \\
\text{cluster 4 subject 2 time point 2} \\
\text{cluster 4 subject 2 time point 3} \\
\text{cluster 4 subject 2 time point 4}
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

and the fixed part, hence the mean vector of the multivariate normal distribution, is then

$$\overrightarrow{\mu} \quad = \quad Z * \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \theta \end{pmatrix} = \begin{pmatrix} \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 \\ \mu \\ \mu + \beta_1 + \theta \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 \\ \mu \end{pmatrix}$$

Stepped wedge design For example with $I = 3$ cluster and $K = 4$ measurments, hence only one cluster switches per timepoint, the design matrix X is defined as

$$X \quad = \quad \begin{array}{c} \\ \text{cluster 1} \\ \text{cluster 2} \\ \text{cluster 3} \end{array} \begin{array}{c} \begin{array}{cccc} \text{time point 1} & \text{time point 2} & \text{time point 3} & \text{time point 4} \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

and the matrix Z_i for cluster 1 is then

$$Z_1 \quad = \quad \begin{array}{c} \\ \text{time point 1} \\ \text{time point 2} \\ \text{time point 3} \\ \text{time point 4} \end{array} \begin{array}{c} \begin{array}{ccccc} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

then, if $N = 2$ subjects are within each cluster the fixed effects full rank design matrix Z is

$$
\begin{aligned}
Z &= \begin{matrix} \text{cluster 1} \\ \text{cluster 2} \\ \text{cluster 3} \end{matrix} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \\
&= \begin{matrix} \text{cluster 1 subject 1} \\ \text{cluster 1 subject 2} \\ \text{cluster 2 subject 1} \\ \text{cluster 2 subject 2} \\ \text{cluster 3 subject 1} \\ \text{cluster 3 subject 2} \end{matrix} \begin{pmatrix} Z_{1j} \\ Z_{1j} \\ Z_{2j} \\ Z_{2j} \\ Z_{3j} \\ Z_{3j} \end{pmatrix}
\end{aligned}
$$

$$
= \begin{matrix} \\ \text{cluster 1 subject 1 time point 1} \\ \text{cluster 1 subject 1 time point 2} \\ \text{cluster 1 subject 1 time point 3} \\ \text{cluster 1 subject 1 time point 4} \\ \text{cluster 1 subject 2 time point 1} \\ \text{cluster 1 subject 2 time point 2} \\ \text{cluster 1 subject 2 time point 3} \\ \text{cluster 1 subject 2 time point 4} \\ \text{cluster 2 subject 1 time point 1} \\ \text{cluster 2 subject 1 time point 2} \\ \text{cluster 2 subject 1 time point 3} \\ \text{cluster 2 subject 1 time point 4} \\ \text{cluster 2 subject 2 time point 1} \\ \text{cluster 2 subject 2 time point 2} \\ \text{cluster 2 subject 2 time point 3} \\ \text{cluster 2 subject 2 time point 4} \\ \text{cluster 3 subject 1 time point 1} \\ \text{cluster 3 subject 1 time point 2} \\ \text{cluster 3 subject 1 time point 3} \\ \text{cluster 3 subject 1 time point 4} \\ \text{cluster 3 subject 2 time point 1} \\ \text{cluster 3 subject 2 time point 2} \\ \text{cluster 3 subject 2 time point 3} \\ \text{cluster 3 subject 2 time point 4} \end{matrix}
\begin{matrix} \begin{matrix} \mu & \beta_1 & \beta_2 & \beta_3 & \theta \end{matrix} \\ \begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1
\end{pmatrix} \end{matrix}
$$

and the fixed part, hence the mean vector of the multivariate normal distribution, is then

$$\overrightarrow{\mu} \quad = \quad Z * \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \theta \end{pmatrix} = \begin{pmatrix} \mu + \beta_1 \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 + \theta \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 + \theta \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu + \theta \\ \mu + \beta_1 \\ \mu + \beta_2 \\ \mu + \beta_3 \\ \mu + \theta \end{pmatrix}$$

2.2 Random part

All clusters are independent from each other, hence the variance-covariance matrix V is a block-diagonal matrix of the matrices V_i of all clusters (and all others are zeros), where for all i the V_i are the same for all cluster.

$$V \quad = \quad \begin{array}{c} \\ \text{cluster 1} \\ \\ \\ \text{cluster C} \end{array} \begin{array}{c} \begin{array}{cccc} \text{cluster 1} & & & \text{cluster C} \end{array} \\ \begin{pmatrix} V_i & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & V_i \end{pmatrix} \end{array}$$

and

$$V_i \quad = \quad \begin{array}{c} \\ \text{subject 1} \\ \\ \\ \text{subject N} \end{array} \begin{array}{c} \begin{array}{cccc} \text{subject 1} & & & \text{subject N} \end{array} \\ \begin{pmatrix} V_{i1,i1} & V_{i1,i2} & \cdots & V_{i1,iN} \\ V_{i1,i2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & V_{iN-1,iN} \\ V_{i1,iN} & \cdots & V_{iN-1,iN} & V_{iN,N} \end{pmatrix} \end{array}$$

Therefor we define $V_{ij,i\tilde{j}}$ as a submatrix of V_i for the entities corresponding to the measurements of subject i and the measurement of subject $\tilde{j}$.

For all two different subjects j and $\tilde{j}$ ($j \neq \tilde{j}$) this submatrix is defined by

$$V_{ij,i\tilde{j}} = \begin{array}{c} \\ \text{timpoint 1} \\ \\ \text{timepoint T} \end{array} \begin{array}{c} \begin{array}{ccc} \text{timpoint 1} & & \text{timepoint T} \end{array} \\ \begin{pmatrix} \sigma_e^2 & \cdots & \sigma_e^2 \\ \vdots & & \vdots \\ \sigma_e^2 & \cdots & \sigma_e^2 \end{pmatrix} \end{array}$$

The difference in the distributions of the observations within a cross-sectional and longitudinal SWD is in the random part of the model. Thus the variance-covariance matrix V of the normal distribution $N(Zb, V)$ and hence the form of the V_i or $V_{ij,ij}$ respectively differ.

Variance-Covariance matrix within a cross-sectional design.

$$V_{ij,ij} = \begin{array}{c} \\ \text{timpoint 1} \\ \\ \\ \text{timepoint T} \end{array} \begin{array}{c} \begin{array}{cccc} \text{timpoint 1} & & & \text{timepoint T} \end{array} \\ \begin{pmatrix} \sigma_\alpha^2 + \sigma_e^2 & \sigma_e^2 & \cdots & \sigma_e^2 \\ \sigma_e^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_e^2 \\ \sigma_e^2 & \cdots & \sigma_e^2 & \sigma_\alpha^2 + \sigma_e^2 \end{pmatrix} \end{array}$$

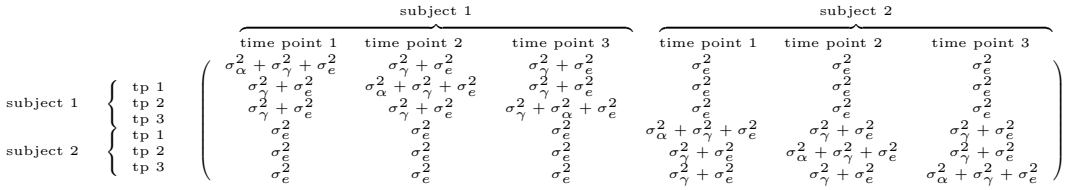
For our example of ($I = 2$ cluster,) $K = 3$ timepoints and $N = 2$ subjects each cluster the Variance-Covariance matrix V_i for each cluster is then

$$\begin{array}{cc} & \begin{array}{cccccc} \overbrace{}^{\text{subject 1}} & & & \overbrace{}^{\text{subject 2}} & & \end{array} \\ & \begin{array}{cccccc} \text{tp 1} & \text{tp 2} & \text{tp 3} & \text{tp 1} & \text{tp 2} & \text{tp 3} \end{array} \\ \begin{array}{c} \text{subject 1} \\ \\ \\ \text{subject 2} \\ \\ \end{array} \begin{array}{c} \left\{ \begin{array}{l} \text{tp 1} \\ \text{tp 2} \\ \text{tp T} \end{array} \right. \\ \left\{ \begin{array}{l} \text{tp 1} \\ \text{tp 2} \\ \text{tp T} \end{array} \right. \end{array} & \begin{pmatrix} \sigma_\alpha^2 + \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 \\ \sigma_e^2 & \sigma_\alpha^2 + \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 \\ \sigma_e^2 & \sigma_e^2 & \sigma_\alpha^2 + \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 \\ \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_\alpha^2 + \sigma_e^2 & \sigma_e^2 & \sigma_e^2 \\ \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_\alpha^2 + \sigma_e^2 & \sigma_e^2 \\ \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \sigma_\alpha^2 + \sigma_e^2 \end{pmatrix} \end{array}$$

Variance-Covariance matrix within a longitudinal design.

$$V_{ij,ij} = \begin{array}{c} \\ \text{timpoint 1} \\ \\ \\ \text{timepoint T} \end{array} \begin{array}{c} \begin{array}{cccc} \text{timpoint 1} & & & \text{timepoint T} \end{array} \\ \begin{pmatrix} \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_e^2 & \sigma_\gamma^2 + \sigma_e^2 & \cdots & \sigma_\gamma^2 + \sigma_e^2 \\ \sigma_\gamma^2 + \sigma_e^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_\gamma^2 + \sigma_e^2 \\ \sigma_\gamma^2 + \sigma_e^2 & \cdots & \sigma_\gamma^2 + \sigma_e^2 & \sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_e^2 \end{pmatrix} \end{array}$$

For our example of ($I = 2$ cluster,) $K = 3$ timepoints and $N = 2$ subjects each cluster the Variance-Covariance matrix V_i for each cluster is then



3 Manual

Use the following Package under GPL and load to the library:

```
#load the package
library(samplingData)
```

3.1 Design matrices

In each study a design matrix of values of explanatory variables can be used to describe the study type. Here, each row represents an study unit (cluster) and the cell entities are the encoding of recieving the treatment or not (zeros and ones). Table1 shows such design matrices or different study types.

In contrast each row of the design matrix of the complete data of the trail represents a measurement with the successive columns corresponding to the variables (effects) and their specific values for that. The design matrix of the complete data corresponds to the fixed part of the multivariate normal distribution.

All matrices could also be implemented manually using the function `matrix()`, but, instead of ordering an amount of zeros and ones, the provided functions in this package make it easy to recieve this complex matrices for simple study designs using only some parameters (balanced, equal number of clusters per switch).

designMatrix The design matrix for the study type of the three types a) parallel, b) crossover, and c) SWD can be performed by using the function `designMatrix()`, which require four parameters: the number of clusters within the trail, the number of measurement time points, the number of cluster, which switch over from control to intervention at each time point and the study type ("SWD" as default).

```
I<-6 #number of cluster
K<-4  #number of  timecpoints

#Design matrix for parallel study, see Table 1
sw<-3 #number of cluster switches
designMatrix(nC=I, nT=K, nSw=sw, design="parallel")
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    0    0    0    0
## [2,]    0    0    0    0
## [3,]    0    0    0    0
## [4,]    1    1    1    1
## [5,]    1    1    1    1
## [6,]    1    1    1    1

#Design matrix for cross-over study, see Table 1
designMatrix(nC=I, nT=K, nSw=sw, design="cross-over")

##      [,1] [,2] [,3] [,4]
## [1,]    0    0    1    1
## [2,]    0    0    1    1
## [3,]    0    0    1    1
## [4,]    1    1    0    0
## [5,]    1    1    0    0
## [6,]    1    1    0    0

#if swP is set, then the timepoint of switch is setted manually
designMatrix(nC=I, nT=K, nSw=sw, swP=1, design="cross-over")

##      [,1] [,2] [,3] [,4]
## [1,]    0    1    1    1
## [2,]    0    1    1    1
## [3,]    0    1    1    1
## [4,]    1    0    0    0
## [5,]    1    0    0    0
## [6,]    1    0    0    0

#Design matrix for SWD study, see Table 1
sw<-2 #number of cluster switches
designMatrix(nC=I, nT=K, nSw=sw)

##      [,1] [,2] [,3] [,4]
## [1,]    0    1    1    1
## [2,]    0    1    1    1
## [3,]    0    0    1    1
## [4,]    0    0    1    1
## [5,]    0    0    0    1
## [6,]    0    0    0    1
```

completeDataDesignMatrix The function `completeDataDesignMatrix()` performes the design matrix for complete data within given study design. It requires a design matrix of a study and the number of subject within each 'cell'.

```r
K<-4  #number of  time points
J<-2 #number of subjects, each cluster and timepoint

##### for parallel study #####

I<-4 #number of cluster
sw<-2 #number of cluster switches
# create a design matrix
(X<-designMatrix(nC=I, nT=K, nSw=sw, design="parallel"))
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    0    0    0    0
## [2,]    0    0    0    0
## [3,]    1    1    1    1
## [4,]    1    1    1    1
```

```r
# create the corresponding complete data design matrix
completeDataDesignMatrix(J, X)
```

```
##       [,1] [,2] [,3] [,4] [,5]
##  [1,]    1    0    0    0    0
##  [2,]    1    1    0    0    0
##  [3,]    1    0    1    0    0
##  [4,]    1    0    0    1    0
##  [5,]    1    0    0    0    0
##  [6,]    1    1    0    0    0
##  [7,]    1    0    1    0    0
##  [8,]    1    0    0    1    0
##  [9,]    1    0    0    0    0
## [10,]    1    1    0    0    0
## [11,]    1    0    1    0    0
## [12,]    1    0    0    1    0
## [13,]    1    0    0    0    0
## [14,]    1    1    0    0    0
## [15,]    1    0    1    0    0
## [16,]    1    0    0    1    0
## [17,]    1    0    0    0    1
## [18,]    1    1    0    0    1
## [19,]    1    0    1    0    1
## [20,]    1    0    0    1    1
## [21,]    1    0    0    0    1
## [22,]    1    1    0    0    1
## [23,]    1    0    1    0    1
## [24,]    1    0    0    1    1
## [25,]    1    0    0    0    1
## [26,]    1    1    0    0    1
## [27,]    1    0    1    0    1
## [28,]    1    0    0    1    1
```

```
## [29,]    1    0    0    0    1
## [30,]    1    1    0    0    1
## [31,]    1    0    1    0    1
## [32,]    1    0    0    1    1

##### for cross-over study #####

# create a design matrix
(X<-designMatrix(nC=I, nT=K, nSw=sw, design="cross-over"))

##      [,1] [,2] [,3] [,4]
## [1,]    0    0    1    1
## [2,]    0    0    1    1
## [3,]    1    1    0    0
## [4,]    1    1    0    0

# create the corresponding complete data design matrix
completeDataDesignMatrix(J, X)

##       [,1] [,2] [,3] [,4] [,5]
##  [1,]    1    0    0    0    0
##  [2,]    1    1    0    0    0
##  [3,]    1    0    1    0    1
##  [4,]    1    0    0    1    1
##  [5,]    1    0    0    0    0
##  [6,]    1    1    0    0    0
##  [7,]    1    0    1    0    1
##  [8,]    1    0    0    1    1
##  [9,]    1    0    0    0    0
## [10,]    1    1    0    0    0
## [11,]    1    0    1    0    1
## [12,]    1    0    0    1    1
## [13,]    1    0    0    0    0
## [14,]    1    1    0    0    0
## [15,]    1    0    1    0    1
## [16,]    1    0    0    1    1
## [17,]    1    0    0    0    1
## [18,]    1    1    0    0    1
## [19,]    1    0    1    0    0
## [20,]    1    0    0    1    0
## [21,]    1    0    0    0    1
## [22,]    1    1    0    0    1
## [23,]    1    0    1    0    0
## [24,]    1    0    0    1    0
## [25,]    1    0    0    0    1
## [26,]    1    1    0    0    1
## [27,]    1    0    1    0    0
## [28,]    1    0    0    1    0
```

```
## [29,]    1    0    0    0    1
## [30,]    1    1    0    0    1
## [31,]    1    0    1    0    0
## [32,]    1    0    0    1    0

##### for SWD study #####

I<-3 #number of cluster
# create a design matrix
(X<-designMatrix(nC=I, nT=K, nSw=1))

##      [,1] [,2] [,3] [,4]
## [1,]    0    1    1    1
## [2,]    0    0    1    1
## [3,]    0    0    0    1

# create the corresponding complete data design matrix
completeDataDesignMatrix(J, X)

##       [,1] [,2] [,3] [,4] [,5]
##  [1,]    1    0    0    0    0
##  [2,]    1    1    0    0    1
##  [3,]    1    0    1    0    1
##  [4,]    1    0    0    1    1
##  [5,]    1    0    0    0    0
##  [6,]    1    1    0    0    1
##  [7,]    1    0    1    0    1
##  [8,]    1    0    0    1    1
##  [9,]    1    0    0    0    0
## [10,]    1    1    0    0    0
## [11,]    1    0    1    0    1
## [12,]    1    0    0    1    1
## [13,]    1    0    0    0    0
## [14,]    1    1    0    0    0
## [15,]    1    0    1    0    1
## [16,]    1    0    0    1    1
## [17,]    1    0    0    0    0
## [18,]    1    1    0    0    0
## [19,]    1    0    1    0    0
## [20,]    1    0    0    1    1
## [21,]    1    0    0    0    0
## [22,]    1    1    0    0    0
## [23,]    1    0    1    0    0
## [24,]    1    0    0    1    1
```

3.2 Covariance-Variance-Matrices

Covariance-Variance matrix are needed besides the mean vector to specify the kind of multivariate normal distribution. The form depends on the kind of multilevel structure. In our examples of cluster randomized studies with measurements over time there are two possibilities: 1) two-level data within cross-sectional studies and 2) three-level data within longitudinal studies.

CovMat_Design The corresponding covariance-Variance matrices can be performed with the provided `CovMat_Design()`. The function required the design parameter K number of timepoints, I number of clusters, J number of subjects within each cluster to each timepoint, and also the variances corresponding to each level. If 'sigma.2.q' is not given, then it a cross-sectional, otherwise a longitudinal design is performed.

```
#study design parameter
K<-3  #number of measurement (or timepoints)
I<-2 #number of cluster
J<-2 #number of subjects

### for cross-sectional data
sigma.1<-0.1
sigma.3<-0.9
CovMat_Design(K, J, I,
              sigma.1.q=sigma.1, sigma.3.q=sigma.3)
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
##  [1,]  1.0  0.9  0.9  0.9  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [2,]  0.9  1.0  0.9  0.9  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [3,]  0.9  0.9  1.0  0.9  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [4,]  0.9  0.9  0.9  1.0  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [5,]  0.9  0.9  0.9  0.9  1.0  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [6,]  0.9  0.9  0.9  0.9  0.9  1.0  0.0  0.0  0.0   0.0   0.0   0.0
##  [7,]  0.0  0.0  0.0  0.0  0.0  0.0  1.0  0.9  0.9   0.9   0.9   0.9
##  [8,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  1.0  0.9   0.9   0.9   0.9
##  [9,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  1.0   0.9   0.9   0.9
## [10,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  0.9   1.0   0.9   0.9
## [11,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  0.9   0.9   1.0   0.9
## [12,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  0.9   0.9   0.9   1.0
```

```
### for longitudinal data
sigma.1<-0.1
sigma.2<-0.4
sigma.3<-0.9
CovMat_Design(K, J, I,
              sigma.1.q=sigma.1, sigma.2.q=sigma.2, sigma.3.q=sigma.3)
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
##  [1,]  1.4  1.3  1.3  0.9  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [2,]  1.3  1.4  1.3  0.9  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
##  [3,]  1.3  1.3  1.4  0.9  0.9  0.9  0.0  0.0  0.0   0.0   0.0   0.0
```

```
##  [4,]  0.9  0.9  0.9  1.4  1.3  1.3  0.0  0.0  0.0   0.0   0.0   0.0
##  [5,]  0.9  0.9  0.9  1.3  1.4  1.3  0.0  0.0  0.0   0.0   0.0   0.0
##  [6,]  0.9  0.9  0.9  1.3  1.3  1.4  0.0  0.0  0.0   0.0   0.0   0.0
##  [7,]  0.0  0.0  0.0  0.0  0.0  0.0  1.4  1.3  1.3   0.9   0.9   0.9
##  [8,]  0.0  0.0  0.0  0.0  0.0  0.0  1.3  1.4  1.3   0.9   0.9   0.9
##  [9,]  0.0  0.0  0.0  0.0  0.0  0.0  1.3  1.3  1.4   0.9   0.9   0.9
## [10,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  0.9   1.4   1.3   1.3
## [11,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  0.9   1.3   1.4   1.3
## [12,]  0.0  0.0  0.0  0.0  0.0  0.0  0.9  0.9  0.9   1.3   1.3   1.4
```

3.3 Sample data under a given study design

We provide a function to sample a complete data set from multivariate normal distribution to mimic data of cluster randomised trials within different study designs, namely parallel, cross-over and stepped wedge design and different type of longitudinal or cross-sectional data.

sampleData Therefore, we provide the `sampleData()`, where the mean vector and the covariance-variance matrix of the distribution under such studies has to be given.

```r
#desing parameter
K<-4  #number of  time points
J<-25 #number of subjects, each cluster and timepoint

#variances of each level
sigma.1<-0.1
sigma.2<-0.4
sigma.3<-0.9

#regression paramters
mu.0<-0
theta<-1
betas<-rep(0, K-1)
parameters<-c(mu.0, betas, theta)

##### for parallel study #####

I<-4 #number of cluster
sw<-2 #number of cluster switches
# create a design matrix
X<-designMatrix(nC=I, nT=K, nSw=sw, design="parallel")
# create the corresponding complete data design matrix
D<-completeDataDesignMatrix(J, X)
#performe covariance-Variance matrix for longitudinal design
V<-CovMat_Design(K, J, I, sigma.1.q=sigma.1, sigma.2.q=sigma.2, sigma.3.q=sigma.3)
#sample data within the design
sample.data<-sampleData(type = "long", K=K,J=J,I=I, D=D, V=V, parameters=parameters)
```

```r
#need the lme4 package for analysis
library(lme4)
lmer(val~intervention+measurement + (1|cluster)+(1|subject), data=sample.data)

## Linear mixed model fit by REML ['lmerMod']
## Formula: val ~ intervention + measurement + (1 | cluster) + (1 | subject)
##    Data: sample.data
## REML criterion at convergence: 508.2947
## Random effects:
##  Groups   Name        Std.Dev.
##  subject  (Intercept) 0.5864
##  cluster  (Intercept) 0.0000
##  Residual             0.3215
## Number of obs: 400, groups:  subject, 100; cluster, 4
## Fixed Effects:
##  (Intercept)  intervention  measurement2  measurement3  measurement4
##     -1.17630       1.47996       0.01892      -0.01455       0.06178
```

```r
## [1] 0.8332087
## [1] 0.6272227
## [1] 0.3155692
## [1] 0.8922119
```

```r
# ##### for cross-over study #####

# create a design matrix
X<-designMatrix(nC=I, nT=K, nSw=sw, design="cross-over")
# create the corresponding complete data design matrix
D<-completeDataDesignMatrix(J, X)
#performe covariance-Variance matrix for longitudinal design
V<-CovMat_Design(K, J, I, sigma.1.q=sigma.1, sigma.2.q=sigma.2, sigma.3.q=sigma.3)
#sample data within the design
sample.data<-sampleData(type = "long", K=K,J=J,I=I, D=D, V=V, parameters=parameters)

#analysis of the three-level data
lmer(val~intervention+measurement + (1|cluster)+(1|subject), data=sample.data)

## Linear mixed model fit by REML ['lmerMod']
## Formula: val ~ intervention + measurement + (1 | cluster) + (1 | subject)
##    Data: sample.data
## REML criterion at convergence: 512.7811
## Random effects:
##  Groups   Name        Std.Dev.
##  subject  (Intercept) 0.6617
##  cluster  (Intercept) 1.3903
##  Residual             0.3036
```

```
## Number of obs: 400, groups:  subject, 100; cluster, 4
## Fixed Effects:
##  (Intercept)  intervention  measurement2  measurement3  measurement4
##     -1.15953       0.97857       0.05056       0.04240       0.01554
```

```
## [1] 0.8334145
## [1] 0.618346
## [1] 0.3172744
## [1] 0.9959059
```

```
##### for SWD study #####

I<-3 #number of cluster
# create a design matrix
X<-designMatrix(nC=I, nT=K, nSw=1)
# create the corresponding complete data design matrix
D<-completeDataDesignMatrix(J, X)
#performe covariance-Variance matrix for cross-sectional design
V<-CovMat_Design(K, J, I, sigma.1=sigma.1, sigma.3=sigma.3)
#sample data within the design
sample.data<-sampleData(type = "cross-sec", K=K,J=J,I=I, D=D, V=V, parameters=parameters)

#analysis of the two-leveldata
lmer(val~intervention+measurement + (1|cluster), data=sample.data)

## Linear mixed model fit by REML ['lmerMod']
## Formula: val ~ intervention + measurement + (1 | cluster)
##    Data: sample.data
## REML criterion at convergence: 146.9664
## Random effects:
##  Groups   Name        Std.Dev.
##  cluster  (Intercept) 0.8272
##  Residual             0.2932
## Number of obs: 300, groups:  cluster, 3
## Fixed Effects:
##  (Intercept)  intervention  measurement2  measurement3  measurement4
##     -0.35075       0.99072      -0.09816      -0.05106      -0.02214
```

```
## [1] 0.7792388
## [1] 0.3157349
## [1] 0.997281
```

3.4 Power calculations

calcPower.SWD

4 Summary