MSOpt	*MSOpt*

Description

The `MSOpt` function allows the user to define the structure of the experiment, the set of optimization criteria and the a priori model to be considered. The output is a list containing all information about the settings of the experiment. According to the declared criteria, the list also contains the basic matrices for their implementation, such as information matrix, matrix of moments and matrix of weights. This function returns the `msopt` argument of the `Score` and `MSSearch` functions of the `multiDoE` package.

Usage

```
MSOpt(facts, units, levels, etas, criteria, model)
```

Arguments

`facts` — A list of vectors representing the distribution of factors across strata. Each item in the list represents a stratum and the first item is the highest stratum of the multi-stratum structure of the experiment. Within the vectors, experimental factors are indicated by progressive integer from 1 (the first factor of the highest stratum) to the total number of experimental factors (the last factor of the lowest stratum). Blocking factors are differently denoted by empty vectors.

`units` — A list whose i-th element, n_i, is the number of experimental units within each unit at the previous stratum ($i-1$). The first item in the list, n_1, represents the number of experimental units in the stratum 0. The latter is defined as the entire experiment, such that $n_0 = 1$.

`levels` — A vector containing the number of available levels for each experimental factor in `facts` (blocking factors are excluded). If all experimental factors share the number of levels one integer is sufficient.

`etas` — A list specifying the ratios of error variance between subsequent strata. It follows that `length(etas)` must be equal to `length(facts)-1`.

`criteria` — A list specifying the criteria to be optimized. It can contain any combination of:

- "I" : I-optimality
- "Id" : Id-optimality
- "D" : D-optimality
- "A" : Ds-optimality
- "Ds" : A-optimality
- "As" : As-optimality

More detailed information on the available criteria is given under **Details**.

`model` — A string which indicates the type of model, among "main", "interaction" and "quadratic".

Details

A little notation is introduced to show the criteria that can be used in the multi-objective approach of the `MultiDoE` package.

For an experiment with N runs and s strata, with stratum i having n_i units within each unit at previous stratum $(i-1)$ and stratum 0 being defined as the entire experiment $(n_0 = 1)$, the general form of the model can be written as:

$$y = X\beta + \sum_{i=1}^{s} Z_i \varepsilon_i$$

where y is a N-dimensional vector of responses $(N = \prod_{j=1}^{s} n_j)$, X is the $N by p$ model matrix,β is a p-dimensional vector containing the p fixed model parameters, Z_i is an $N by b_i$ indicator matrix of zero and ones for the units in stratum i (i.e. the (k,l)th element of Z_i is one if the kth run belongs to the lth block in stratum i and zero otherwise) and $b_i = \prod_{j=1}^{i} n_j$. Finally, the vector $\varepsilon_i \sim N(0, \sigma_i^2 I_{b_i})$ is a b_i-dimensional vector containing the random effects, which are all uncorrelated. The variance components $\sigma_i^2 (i = 1, \ldots, s)$ have to be estimated and this is usually done by using the REML method.

The best linear unbiased estimator for the parameter vector β is the generalized least square estimator:

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

This estimator has variance-covariance matrix:

$$Var(\hat{\beta}_{GLS}) = \sigma^2 (X'V^{-1}X)^{-1}$$

where $V = \sum_{i=1}^{s} \eta_i Z_i' Zi$, $\eta_i = \frac{\sigma_i^2}{\sigma^2}$ and $\sigma^2 = \sigma_s^2$. The variance components $\sigma_i^2 (i = 1, \ldots, s)$ have to be estimated. Finally, let $M = X'V^{-1}X$ be the information matrix of $\hat{\beta}$ when the GLS estimator is used to estimate model parameters in a multi-stratum experiment.

- **D-optimality.** The D-optimality criterion is based on minimizing the generalized variance of the parameter estimates. This can be done either by minimizing the determinant of the variance-covariance matrix of $\hat{\beta}$ or by maximizing the determinant of M. The objective function to be minimized is:

$$f_D(d; \eta) = \left(\frac{1}{\det(M)} \right)^{1/p}$$

  where d is the design with information matrix M and p is the number of model parameters.

- **A-optimality.** This criterion is based on minimizing the average variance of the estimates of the regression coefficients. The sum of the variances of the parameter estimates (elements of $\hat{\beta}$) is taken as a measure, which is equivalent to the trace of M^{-1}.
  The objective function to be minimized is:

$$f_A(d; \eta) = \mathtt{tr}(M^{-1})$$

  where d is the design with information matrix M.

- **I-optimality.** The I-optimality criterion seeks to minimize the average prediction variance. The objective function to be minimized is:

$$f_I(d;\eta) = \frac{\int_\chi f'(x)(X'V^{-1}X)^{-1}f(x)\,dx}{\int_\chi dx}$$

  where χ represents the design region.

  When there are k treatment factors and the experimental region is $[-1,+1]^k$, the objective function can also be written as:

$$f_I(d;\eta) = \texttt{tr}\left[(X'V^{-1}X)^{-1}B\right]$$

  where $B = 2^{-k}\int_\chi f'(x)f(x)\,dx$ is the moment matrix. The matrix B has a very specific structure for a full quadratic model, as shown in Hardin and Sloane (1991).

- **Id-optimality.** This criterion seeks to minimize the average prediction variance excluding the intercept from the set of parameters of interest. The objective function to be minimized is the same as the I-optimality criterion, where the first row and columns of the B matrix are deleted.

- **Ds-optimality.** The Ds-optimality criterion, as the D-optimality criterion, seeks to minimize the generalized variance of the parameter estimates excluding the intercept from the set of parameters of interest. The objective function to be minimized is:

$$f_{D_s}(d;\eta) = |(M_i^{-1})_{22}|$$

- **As-optimality.** This criterion, as the A-optimality criterion, is based on minimizing the average variance of the estimates of the regression coefficients excluding the intercept from the set of parameters of interest. The objective function to be minimized is:

$$f_{A_s}(d;\eta) = \texttt{tr}(W_i(M_i^{-1})_{22})$$

  where W_i is a diagonal matrix of weights, with the weights scaled so that the trace of W_i is equal to 1.

Value

`MSOpt` returns a list containing the following components:

- `facts`: The argument `facts`.
- `nfacts`: An integer. The number of experimental factors (blocking factors are excluded from the count).
- `nstrat`: An integer. The number of strata.
- `units`: The argument `units`.
- `runs`: An integer. The number of runs.
- `etas`: The argument `etas`.
- `avlev`: A list showing the available levels for each experimental factor.
- `levs`: A vector showing the number of available levels for each experimental factor.
- `Vinv`: The inverse of the variance-covariance matrix of the responses.
- `model`: The argument `model`.
- `crit`: The argument `criteria`.

- `ncrit`: An integer. The number of criteria.
- `M`: The moment matrix. Only with *I-optimality* criteria.
- `M0`: The moment matrix. Only with *Id-optimality* criteria.
- `W`: The diagonal matrix of weights. Only with *As-optimality* criteria.

More information on M, M0 and W can be found in the descriptions of the respective criteria in the **Details** section.

References

R. H. Hardin and N. J. A. Sloane. Computer generated minimal (and larger) response-surface designs: (II) The cube. Technical report, 1991.