

**Main function:** DCM (M1, M2, ...)

Takes two data matrices, with variables in rows and samples in columns. Searches for DC sets, or sets of variables with higher pairwise correlation in M1 than in M2.

**Options:** *option (default value)*

- *Max.groups (5)*: maximum number of DC sets to find before terminating search
- *echo (FALSE)*: function reports progress
- *resid.full (FALSE)*: should the data matrices be residualized before running DCM. Choose TRUE if strong global correlation structure is suspected.
- *max.iter (100)*: maximum number of iterations algorithm should undergo before forced termination
- *max.time (160 min)*: time until forced termination of algorithm
- *est.size (nrow(M1)/10)*: size of initial set to find
- *QN (FALSE)*: quantile normalize data before running DCM
- *start (NULL)*: optionally define a specific initial set instead of using built-in initializing function
- *strict ('low')*: choose 'high' to be more stringent in throwing out unreliable data – see *sanitize* below
- *alpha (0.05)*: level at which to control FDR

**Alternate function:** DCM\_Classic (M1, M2, ... )

Runs DCM algorithm, with a classic test of equality in place of the new method. Usage exactly as *DCM()*, with analog *run\_DCM\_Classic*. Dependency *classicTest()* implements correlation testing method of Larntz and Perlman (1985).

## Dependencies

- *run\_DCM*: performs a single runthrough of the DCM algorithm until convergence
  - *bhy*: applies method of Benjamini and Yekutieli to p-values for multiple testing
  - *makeTau\_in*: calculates test statistic variance estimator for variables in current DC set
  - *makeTau\_out*: calculates test statistic variance estimator for variables outside current DC set
- *prepData\_DCM*: prepares two data matrices for analysis via DCM
  - *sanitize*: finds and removes rows or columns with too many values at threshold and/or with missing data
  - *stdize*: row-standardizes matrices, for simple correlation calculation
  - *quantNorm*: quantile normalize matrix by row
- *init\_DCM*: uses greedy algorithm on difference of fisher transformed correlation matrices to select initial DC set
  - *fisher*: fisher transforms a sample correlation value, truncated at boundaries
- *resid\_DCM*: residualizes matrices, removing effect of discovered DC set
  - *emfa\_DCM*: implementation of Bishop (2006) of EM algorithm for factor analysis