CDD user guide

2013-05-28 PsN 3.6.2

Introduction

The Case Deletions Diagnostics (CDD) algorithm is a tool primarily used to identify influential components of the dataset, usually individuals. The CDD works by identifying groups in the data set and creating one new data set for each member of the group, where that member has been removed. The model is run once with each new data set. The PsN implementation of the CDD can take any column as base for the grouping and all rows with the same value in that column will be considered a group as long as no individual contain multiple values in that column.

One should take care that the grouping creates sensible individual records. PsN will renumber the ID column so that two individuals with the same ID will not end up next to each other.

Examples

cdd moxonidine.mod -case_column=1

cdd pheno.mod -case_column=AGE

Input and options

Required input

A model file is required on the command-line.

-case_column=<name | number >	The column on which the case-deletion is done. You can either use the name of the column as specified in the $INPUT record in the model file or you can use the column number. Numbering starts with 1.

Optional input

-bins=N	Sets the number of databins, or cdd datasets, to use. If the number of unique values, or factors, in the based_on column is higher than N then one or more factors will be deleted in each cdd dataset. Specifying N as higher than the number of factors will have no effect. N is then reset to the number of factors. Default value = Number of unique values in the based_on column.
-xv	Default true. Run the cross-validation step (-xv) or not (-no-xv).
-selection_method=<random | consecutive>	Default consecutive. Specifies whether the factors selected for exclusion should be drawn randomly or consecutively from the datafile.
-outside_n_sd_check=X	Default 2. Mark the runs with CS-CR outside X standard deviations of the PCA.

Some common PsN-options

For a complete list of common options see common_options_defaults_versions.pdf, or psn_options -h on the commandline.

-directory=cdd_dirN	The directory in which the script will run NONMEM can be named. The default name is "cdd_dirN" where N is increased by 1 each time you run the script. If the run is aborted or crashes, setting the directory to the one from which the script was running earlier can be done. PsN will then not run the model files that had finished, saving time. Note that same set of options must be given as when the run was started the first time.
-threads=N	The number of parallel processes to start on a parallel computer.
-help	With -help a longer help message will be printed.
-last_est_complete	is optional and only applies with NONMEM7 and cdd option -xv. See common_options_defaults_versions.pdf for details.

Known bugs/issues

If NONMEM6 is used with the cdd and the S matrix is algorithmically singular (message in lst-file, checked also by sumo script) the Cook scores cannot be trusted. The cdd does not check for this error.