

Problems creating MDM files

Most of the problems experienced when using HLM are with the creation of the Multivariate Data Matrix file (MDM file). In general, these problems are caused by invalid data or incorrect specifications when creating the MDM file. The most common of these problems, losing units of analysis when creating the MDM file, are illustrated and discussed below.

Data in the data files are not sorted correctly

Data in the level-1 data file have to be sorted by the ID field as used in the level-2 data file. If the data are not sorted, an incorrect number of units will be included in the MDM file. When creating a level-3 MDM file, level-2 IDs need to be sorted inside of the level-3 IDs. For example, suppose that you have only 3 observations (level-1) for 2 level-2 units. The correct order of these observations in the level-1 data file is:

1	(First	observation	for	first	level-2	unit)
1	(Second	observation	for	first	level-2	unit)
1	(Third	observation	for	first	level-2	unit)
2	(First	observation	for	second	level-2	unit)
2	(Second	observation	for	second	level-2	unit)
2	(Third observation for second level-2 unit)					

If, however, you accidentally specified it as

```
1
2
2
2
1
1 ,
```

HLM will not create the MDM file but warn you that the data in your level-1 file is incorrectly sorted. But, if you had this in your level-1 data file as the ID variable:

```
1
1
2
2
3
3
```

HLM will create an MDM file, with only 4 level-1 units and 2 level-2 units (remember that there is no unit 3 at level 2). The observations for which no corresponding ID could be found in the level-2 data file will be ignored. The *.sts file will list units for which no data could be found at the higher levels.

Presence of missing data

The number of units used in the creation of the MDM file will also differ from those in the data files if you have missing data. Missing data are only allowed on level-1 of the hierarchy, and you need to provide information on the missing data codes when creating the MDM file. Also note that the way in which HLM handles missing data differs from module to module. If the presence of missing data is not indicated, or missing data codes are not specified, the blanks and other codes, for example -99, will be treated as data. No missing data are allowed at the higher levels of the hierarchy.

The IDs may be incorrectly specified

In order for the program(s) to correctly read the data, the IDs need to conform to the following rules:

- The ID variables may be numeric or character (alphanumeric). If the data are in ASCII (plain text) format, you must use a character ID, as the programs look for the "A" in the format statements as indicators of which fields are the ID variables.
- The IDs must be sorted, and properly nested if there are more than two levels involved. For example, in three level data, the level-2 IDs must be sorted within the level-3 IDs. Level-1 cases must be grouped by their respective level-2 IDs, and the level-2 IDs within the level-3 IDs. The easiest way to assure this is to sort the data files as required prior to entering the data into HLM. Failure to do so will lead to the exclusion of data records from the MDM file, or the incorrect reading of your data.
- If the ID variable is numeric, it must be in the range $-(10^{13}+1)$ to $+(10^{13}-1)$ (i.e., 12 digits). Although the ID may be a number with a decimal point, only the integer part is used.
- If the ID variable is character, the length must not exceed 12 characters. Furthermore, the IDs at a given level must all be the same length. This is often a cause of problems. For example, imagine your data has IDs ranging from "1" to "100". You will need to recreate the IDs as "001" to "100". In other words, all spaces (blank characters) should be coded as zeros.

Although the programs are somewhat flexible in their input, it is strongly recommended that you use non-ASCII data with numeric IDs.