检查重复编码与记录

此模块可以用来：

（1） 检查数据错误；

（2） 对重复测量数据进行统计报告。

基本定义：

（1） 重复记录：两条或多条记录所有的变量（字段）都相同。

（2） 重复 ID：两条或多条记录的 ID 相同。

重复的产生：

如果某记录错误地重新输入了一次或多次，如果多次输入无误，这样所有字段都相同，数据中将会出现两条完全相同的记录，即有重复的记录。有时因为你的数据文件中可能有一个自动递增的指示变量，这个变量不会相同。因此在检查重复记录时就不能考虑该变量。

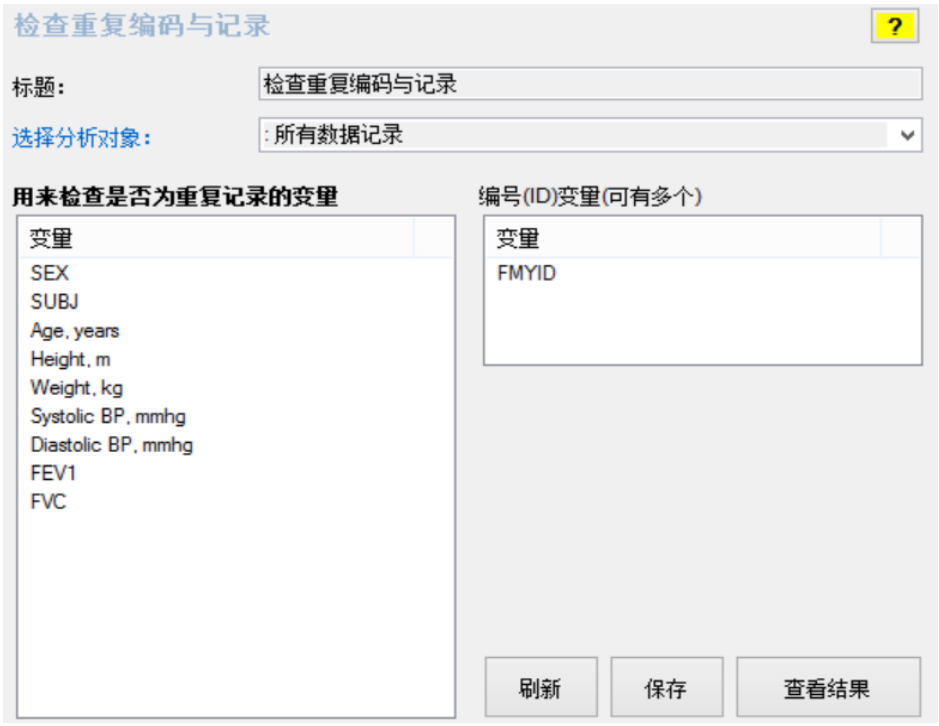
要检查重复记录，本模块要求给出一个变量清单，用于判断两记录是否完全相同。如所列变量完全相同，则认为两记录重复，不考虑其它变量。

如果重复录入过程中有输入错误，这样就不是所有字段都相同，如果研究对象编号输入相同，则数据中将有重复的编号，而不是两条重复的记录。重复编号的产生也可能是由于同一个编号错误地分配给了两个或多个研究对象。

使用本模块可以检出这些错误，并保存一个清洁的数据文件。

重复测量数据：即一个研究对象有多次观测，每次观测一条记录，则数据就有很多个重复的研究对象编号。

例如，某数据含 N 个核心家系，每个家系有 1-10 人，每人有一个记录，共享一个核心家系编号。使用此模块可以报告按核心家系大小（人数）统计的核心家系数。



输出结果：

检查重复编码与记录

总记录数: 832

检测重复记录

记录数	* 出现数	合计删除的重复数
832	1	0

检查重复编码 FMYID

Number of FMYID	* 出现数	合计删除的重复数
5	1	0
49	2	98
15	3	45
37	4	148
38	5	190
30	6	180
14	7	98
4	8	32
4	9	36

636 编码(FMYID) 有重复, 总有重复编码的记录数: 827
删除 827 条有重复编码的记录后
输出文件: demo_25_tb11_nodupid.x1s 有 5 条记录

Notes:
1. 重复的编码: 编码(FMYID)相同
2. 删除重复编码: 如果 3 条记录编码ID相同, 3 条记录均删除

结果解释：

首先根据所列出的变量检查是否有重复记录，结果所列变量都不同的有832条，每条出现次数都是1，即没有重复记录。没有重复记录被删除。

如果有N条记录完全相同，删除时将删除N-1条，保留一条。

接着检查重复FMYID，发现：
有5 个FMYID，每个只有一条记录；
有49个FMYID ，每个有2条记录；
有15个FMYID ，每个有3条记录；
……

总共636条件记录有重复的FMYID：
49 +15*2+37*3+38*4+30*5+14*6+4*7+4*8=636
可以计算出独特的FMYID数 = 832-636 = 196独特的FMYID。

如果删除重复的FMYID的记录，则只剩下5条记录，因为如果有几条记录的FMYID相同，这些记录都将被删除。