

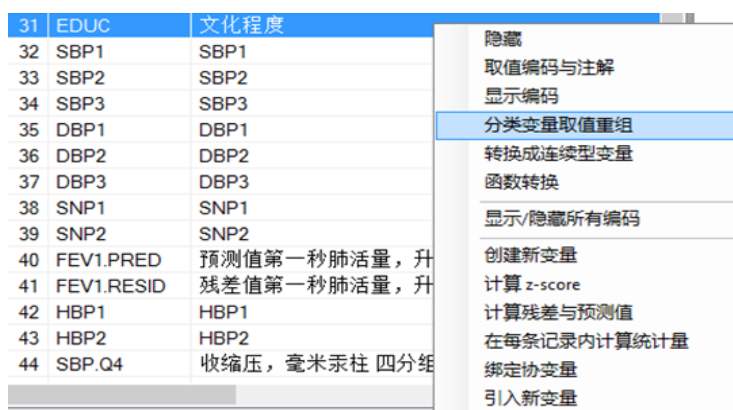
分类型变量取值编码重组

对分类型变量进行重新组合分类，合并频数较少的类或差异较小的类，是数据分析中最常见的变量操作。

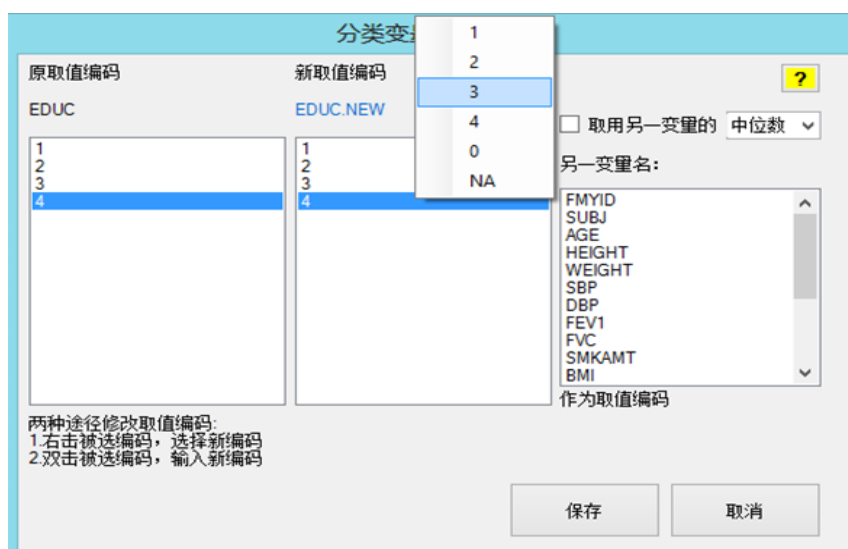
另外一种情况是用于趋势检验，如要对 AGE 对 SBP 的作用进行趋势检验，通常首先把 AGE 分成 4（或 5）等分组变成分类型变量，编码为 0、1、2、3，下一步如果直接把这编码为 0、1、2、3 的分类变量按连续性变量放入回归模型中做趋势检验，因为这 4 个等分组间间距不等，实为不妥。需要把编码 0、1、2、3 分别替换为对应的等分组内 AGE 的中位数，然后按连续性变量放入回归模型中。

例 1：“demo.xls” 中文化程度变量 EDUC：1=小学、2=初中、3=高中/中专、4=大学及以上。因为只有极少数人落在第 4 组（大学及以上），分析时有必要将第 3 组与第 4 组合并成“高中及以上”。

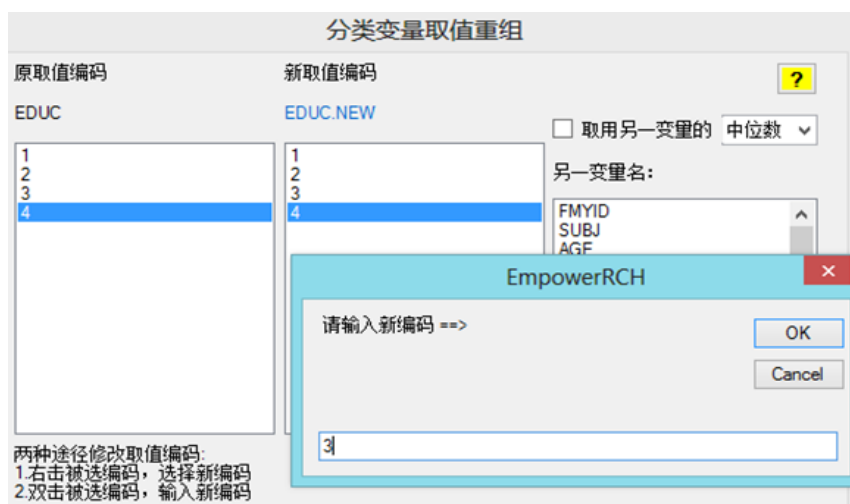
右击变量名“EDUC”，选择“分类变量取值重组”。



在弹出窗口内，左边是原变量名与取值编码，中间是新变量名与取值编码，在此处右击“4”，在弹出窗口中选择“3”即将原编码“4”改成与第 3 组相同的编码“3”。



或双击取值编码，在弹出窗口中输入新编码（如下）。



例 2. “demo.xls” 中年龄变量 AGE 生成四等分组变量后名为 AGE.Q4，其编码为 0、1、2、3，将其重新编码为各对应的四等分组内年龄的中位数。

输入界面如下：



使用本模块将 AGE.Q4 的编码 0、1、2、3 替换成其对应的各等分组内 AGE 的中位数。由于系统每次运行时自动计算各组内 AGE 中位数，如有数据更新，各组内中位数将自动更新。用户不必先计算各组中位数然后手动输入，更不必担心数据更新带来的问题。