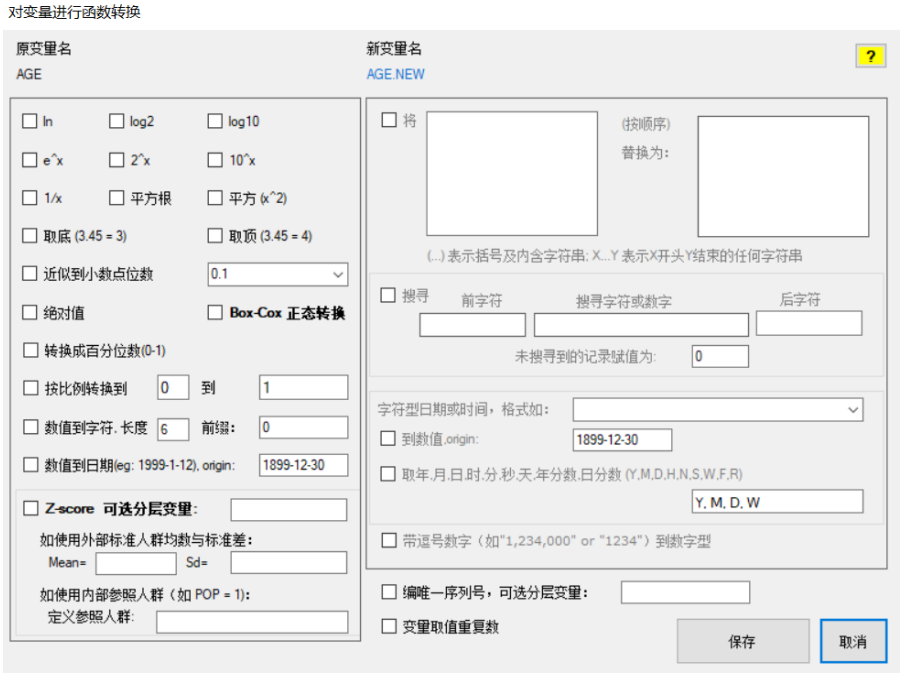
# 变量函数转换

对变量进行函数转换：在变量名清单窗口，右击变量名，选择“函数转换”。

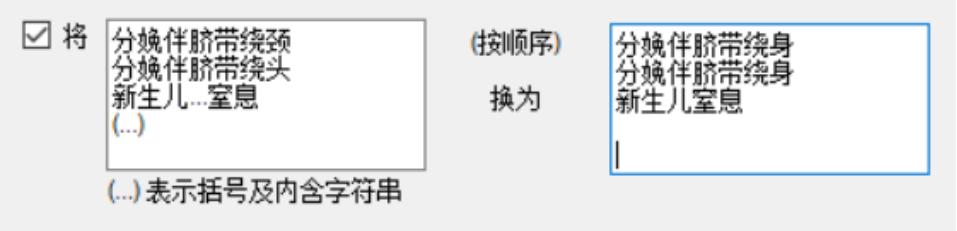



除常见数学函数外，《易侕统计》增设了几种常见的字符型数据转换方法：

1）统一字符串长度。如研究对象编号变量，原为数字，有的长，有的短，在短的前面增加一个字符如“0”或“空格”。

2）取代原字符串中某字符。如日前变量原格式为“02-15-2012”，现要用“/”取代“-”，变成"02/15/2012"。操作如下：


可以同时做多种替换，原字符串依次列在左边，新字符串依次列右边。如新字符串为空，置空行并排在最后即可。原字符串“(…)”表示括号及括号内的任何内容；“STR1...STR2”表示 STR1 开头 STR2 结束的任何字符串。如下图表示依次将（1）分娩伴脐带绕颈、分娩伴脐带绕头均替换成分娩伴脐带绕身，（2）新生儿...窒息：如新生儿轻度窒息、新生儿中度窒息等替换成新生儿窒息；（3）括号及括号内的内容除掉。




3）搜寻字符（串）：

可给定前字符与后字符，可搜寻前后字符之间的数字或/和字符，要搜寻的信息填在搜寻字符或数字处，匹配方法可以是完全匹配或内含待搜寻信息。默认匹配方法是内含待搜寻信息，如要定义完全匹配，在待搜寻信息前加~号，即表示前字符与后字符之间的内容与待搜寻信息完全相同。

i）搜寻并提取数字:

搜寻字符中用#表示 0-9 的数字，一个#表示一个数字，#号中间可以有一个小数号，如 ##.#，#.###。

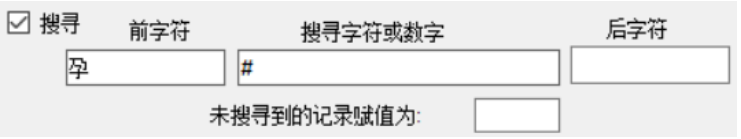
前字符与后字符中用[0-9]表示 1 位数字，同理，[0-9] [0-9]表示 2 位数字。

前字符、后字符、搜寻字符中用||号表示或。

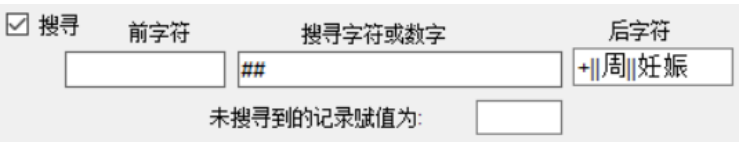
如有以下三种原始记录：

1. 1. IVF-ET 术后；2. 甲状腺功能减退；3. 乙肝携带者(1,4,5,+)；4. 孕 4 产 1 40+2 周妊娠 LOA 待产
2. 1. 前次剖宫产；2. 孕 4 产 138+5 妊娠（LOA）待产
3. 妊娠期糖尿病；孕 2 产 0 40 周妊娠 LOA 待产，亚临床甲减

要提取孕字后面的 1 位数孕次信息，前字符：孕，后字符置空；也可以后字符：产，前字符置空；搜寻字符为：#。

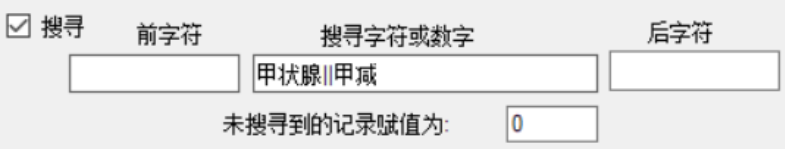



要提取产次后面的 2 位数孕周信息，前字符：产[0-9]；后字符置空。也可以定义后字符：+||周||妊娠；前字符置空。搜寻字符为：##。




ii）搜寻并提取字段:

如上例，要提取“甲状腺功能减退、亚临床甲减”，原始记录中同样的术语可能有多种写法，可以定义前字符置空，后字符置空，搜寻字符：甲，即找出带甲字的字段。也可以定义搜寻字符为：甲状腺||甲减，即找出带甲状腺或带甲减的字段。




4）将字符型日期转换成数字，定义一个起点日期，计算要转换的日期与起点日期所差天数。

5）编唯一序列号：如将字符型变量（如姓名)转换成数字型的编号，如果姓名有重复，则其编号也相同。

6）将原字符型数字变量转换成数值型。

a. 如原数据含逗号，除去逗号。如将 1，234，000 转换成 1234000。
b. 如原数据是带引号的数字，除去引号。如将“1234”转换成 1234。

其它几个增设的函数，简介如下：

1）把数值型数据按比例放大或缩小到（如“0 至 1”）某两个值之间的范围内。如果原数据从最小到最大为 20-80，转换到 0-1 之间，即将 20 换成 0，80 换成 1，其它中间数据按比例缩小，如 50 将换成：（50-20）/（80-20）=0.5。

2）把数值型数据转换成百分位数。即将原数据排序，计算各点对应的百分位数。

3）变量取值重复数：如果某变量重复，如家系测量数据中的家系编号，计算该变量的重复数即得出每个家系的记录数（调查人数），然后可以按该变量挑选出符合要求的家系（如调查人数>=3）。

4）Box-Cox 正态性转换：自动寻找最佳转换函数对原变量进行函数转换，尽可能达到正态或接近正态的对称分布。通过查看“新变量生成记录”可得到转换函数公式。点击项目名如“demo”,得“项目信息”页面，然后点击“查看新变量生成记录: xxx_datastep.lst”


可知,如下结果（例）：

```
[1] "Creating new variable: SBP.NEW"
 boxCox transform formula:
    "(X^(-2) -1)/(-2)"
 After boxCox transform, calculate z-score and then
shift to positive
```

```
     (x - 0.499968364026935)/ 8.88114819496485e-06 +
8.70786996390631

   [1] "Creating new variable: DBP.NEW"
boxCox transform formula:
"(X^(-1) -1)/(-1)"

   [1] "Creating new variable: AGE.Q5"
    Min   20%   40%   60%   80%   Max
   15.60 25.68 31.80 39.00 50.52 77.00
```

5）Z-score 转换：计算 Z-score 时可以定义均数与标准差，也可以定义内部参照人群，从参照人群计算均数与标准差，然后进行转换。