

The General Method for Variable Recoding and Its Applications in SAS

Randi Chen Pacific Health Research Institute, Honolulu, Hawaii
Ka-On Fong Pacific Health Research Institute, Honolulu, Hawaii

ABSTRACT

Variable recoding is a time consuming but important part in SAS programming. However, there are few methods available. Using the IF-THEN statements is a common practice but it is tedious and error-prone when the variable to be recoded, say X, has many levels. Although the user-defined format is an alternative, the drawback is that this method is inflexible, e.g. it cannot directly recode the combination of two variables, say X and Y, into Z. By extracting the essential components of the IF-THEN statements, we constructed a general method or a template for variable recoding: (1) to create a transformation function $F(X)$, where X can be multiple variables; (2) to assign values to Y using the function, $Y=F(X)$. In this paper, we will illustrate several examples to recode discrete variables and hope, by doing so, to inspire SAS users to develop new applications using the general method.

INTRODUCTION

Survey or questionnaire data, which are mostly discrete variables, usually cannot be used directly for data analysis or report without recoding. The IF-THEN statements are often used to recode variables in SAS. However, IF-THEN is very tedious when the variable Y and the variable X to be recoded have too many levels.

The need for a simple and more comprehensive method is apparent. We will discuss several ways of variable recoding in addition to IF-THEN approach.

SAMPLE DATA SET

The data set Report contains data from a survey of sports activity. There are twenty sports activities reported in the survey, and each subject can only report a maximum of ten sports he or she participated in last week. The following SAS program will create data set Report.

```
Data Report;
Input ID 5. (ACT1-ACT10) (5.);
Cards;
  1   3   9
  2   6
  3  12   4   9  11  15
  4   1   3   4  13  19   3   4   8  11  15
  5   8   8   1  16
;
Run;
```

Each value for ACT1 to ACT10 represents a particular sports activity, such as 1 is for bicycling; 2 is for swimming; 3 is for walking, etc. Every sport has a MET value which is the ratio of work metabolic rate to resting metabolic rate.

The MET values for the sports (SPT) activities (1-20) in this data set are:

SPT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MET	7	8	4	3	5	5.5	6	7	4	8	3	6.5	7	4	7	2	7	4	6	7.5

In calculating total physical activity for the subjects, it involves the conversion of the type of sport to the corresponding MET value. We will present several methods to show how to convert the sports activity into the MET value.

RECODING METHODS

USING IF-THEN STATEMENTS

```

IF ACT1 = 1 THEN METofACT1=7;   IF ACT1 = 2 THEN METofACT1=8;
IF ACT1 = 3 THEN METofACT1=4;   IF ACT1 = 4 THEN METofACT1=3;
IF ACT1 = 5 THEN METofACT1=5;   IF ACT1 = 6 THEN METofACT1=5.5;
IF ACT1 = 7 THEN METofACT1=6;   IF ACT1 = 8 THEN METofACT1=7;
IF ACT1 = 9 THEN METofACT1=4;   IF ACT1 =10 THEN METofACT1=8;
IF ACT1 =11 THEN METofACT1=3;   IF ACT1 =12 THEN METofACT1=6.5;
IF ACT1 =13 THEN METofACT1=7;   IF ACT1 =14 THEN METofACT1=4;
IF ACT1 =15 THEN METofACT1=7;   IF ACT1 =16 THEN METofACT1=2;
IF ACT1 =17 THEN METofACT1=7;   IF ACT1 =18 THEN METofACT1=4;
IF ACT1 =19 THEN METofACT1=6;   IF ACT1 =20 THEN METofACT1=7.5;

```

SAS OUTPUT

Obs	ID	ACT1	METofACT1
1	1	3	4.0
2	2	6	5.5
3	3	12	6.5
4	4	1	7.0
5	5	8	7.0

USING AN ARRAY

```

ARRAY MET(20) _temporary_
(7 8 4 3 5 5.5 6 7 4 8
3 6.5 7 4 7 2 7 4 6 7.5);

METofACT1 = MET(ACT1);  ** recode the first reported activity *;

```

We can use a temporary array to recode the same variable. In the assignment statement, METofACT1 = MET(ACT1), the value of the reported ACT1 is used as the index of the array MET to reference the MET value. Take the first observation as an example, ACT1 is 3, the value of MET(ACT1) will be the third element in the array MET which is 4. These codes generate the same results as the IF-THEN statements do. We can easily use a do-loop to recode all ten activity variables with the array MET. The array method can be easily extended to recode two discrete variables into one variable using a two-dimensional array. The sample codes are omitted.

USING A TABLE

We can also create a table (SPTMET) to set up the relation for the variable SPT to the new variable MET, and then use SAS/SQL procedure to recode ACT1 to METofACT1. The table SPTMET can be made within the SAS System or other software, such as MS EXCEL which needs to be imported into SAS.

SPT	MET
1	7
2	8
3	4
4	3
5	5
6	5.5
7	6
8	7
9	4
10	8
11	3
...	...

```

Proc SQL;
Create table final as
Select m.*, a.MET as METofACT1
From report as m left join SPTMET as a
On M.ACT1 = a.SPT
Order by ID;

```

The SAS/SQL procedure is used to merge the definition of the MET by SPT variable. It has the same functionality as the assignment statement in the DATA step.

USE A USER-DEFINED SAS FORMAT

```

Proc SQL;
Create table fmtTable as
Select SPT as Start, MET as Label, "fmtX" as fmtName
From SPTMET;

Proc format CNTLIN = fmtTable;

Data final; Set Report;
METofACT1 = PUT (ACT1, fmtX.) + 0;
Run;

```

Some users have proposed to use SAS/FORMAT procedure to recode the variables. This method is similar to the table method since it requires a table to define the relationship between the original variable to the new variable. In addition, a SAS function PUT with the ACT1 and fmtX as the parameters is used to assign values for METofACT1.

DISCUSSION

Variable recoding is, in fact, a mathematical transformation from variable X to Y through a relation F, i.e. $Y = F(X)$. If we look into the IF-THEN statement (IF ACT1=1 THEN METofACT1=7) we may find that it carries out two actions, to set up the relation for the value of ACT1 with the value of METofACT1 ($F: 1 \rightarrow 7$), and to assign the value 7 for METofACT1 given ACT1 = 1, i.e. METofACT1 = F (ACT1).

Since data transformation is an abstract form, we propose a general method as a working model or a template for variable recoding:

1. to define a transformation function F(X), where X can be multiple variables.
2. to assign values to Y using the function F, $Y=F(X)$.

There are many ways to define transformation function and to perform assignment. We can use an array to define the relation of X to Y, where X is a sequence of continuous integers; use a two-dimensional array to construct a multiple function F(X1, X2) to recode two discrete variables; use a table to define a transformation function that may involve even more discrete variables, such as F(X1, X2, X3); and use SAS/FORMAT procedure to create a user-defined format as a transformation function. If using SAS functions or user-defined SAS functions, direct assignment can be applied in DATA step. Otherwise, data merging and SAS/SQL procedure can be used for assignment.

CONCLUSIONS

The general method provides a systematic technique for variable recoding. The advantage of this method is to help the users understand and solve complex recoding problems.

REFERENCES

- Ainsworth BE, et al (1993), "Compendium of Physical activities: classification of energy costs of human physical activities," Medicine and Science in Sports and Exercise. The American College of Sports Medicine.
- SAS Institute Inc. (1990), SAS Language: Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENT

The authors would like to express their appreciation to the following people for their critical review for this paper:
Mrs. Eva Ardo, Pacific Health Research Institute
Mrs. Jennifer Smith, Pacific Health Research Institute

CONTACT INFORMATION

Randi Chen
Pacific Health Research Institute
846 South Hotel Street, Suite 205A
Honolulu, HI 96815
Phone: (808) 524-4411
Fax: (808) 524-5559
Email: rchen@phrihawaii.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.