How do I utilize the recode statement in SUDAAN?

Authored by

stats writer

July 1, 2024

The recode statement in SUDAAN is a useful tool for researchers to transform and manipulate data in order to fit their specific analysis needs. This statement allows users to recode existing variables or create new ones based on specific criteria, such as ranges, patterns, or missing values. By utilizing the recode statement, researchers can efficiently clean and organize their data to accurately represent their study population and conduct statistical analyses. This feature is particularly useful for working with complex survey data that may require recoding of categorical variables or adjusting for missing data. Overall, the recode statement in SUDAAN helps researchers to accurately and efficiently prepare their data for analysis, providing reliable and valid results for their research.

How can I use the recode statement in SUDAAN? | SUDAAN FAQ

You can use the recode statement with all procedures in SUDAAN (except proc records). This statement is especially useful when you need to create a categorical variable from a continuous variable. The original continuous variable is recoded "on the fly", and the recoded variable is not added to your data set; rather, it exists only for the duration of the procedure. In the first example below, a 0/1 variable is created from the continuous variable ab23. A cut-off value of 50 is given, so in the recoded variable, values less than 50 will be coded 0

and values equal to and
greater than 50 will be coded 1. Please see page 164 of the SUDAAN manual
for more information regarding the recode statement. On the var
statement, you need to specify the variable one time for each level of that
variable that appears on the catlevel statement. On the catlevel
statement, you need to specify the value of each level of the variable that you
want displayed in the output.

```
proc descript data=temp1 filetype=sas design = jackknife;
weight rakedw0;
jackwgts rakedw1--rakedw80 / adjjack=1;
recode ab23 = (50);
var ab23 ab23;
catlevel 0 1;
run;
```

Number of observations read : 55428 Weighted count : 23847415

Denominator degrees of freedom : 80

Variance Estimation Method: Replicate Weight Jackknife

by: Variable, One.

| | |

| Variable | | One

| | | 1 |

| | | |

| Age first told | Sample Size | 3709 |

| by doctor that | Weighted Size | 1380250.55 |

| have diabetes | Total | 709750.87 |

| or sugar | Percent | 51.42 |

| diabetes: 0 - | SE Percent | 1.03 |

| HIGH | | |

| | | |

| Age first told | Sample Size | 3709 |

| by doctor that | Weighted Size | 1380250.55 |

| have diabetes | Total | 670499.68 |

| or sugar | Percent | 48.58 |

| diabetes: 0 - | SE Percent | 1.03 |

| HIGH | | |

In the example below, the recode statement is used to create a
three-level variable from the continuous variable ab23. In the
recoded variable, values less than 20 will be coded as 0, values less than or
equal to 30 will be coded as 1, and values less than 70 will be coded as 2.

```
proc descript data=temp1 filetype=sas design = jackknife;
weight rakedw0;
jackwgts rakedw1--rakedw80 / adjjack=1;
recode ab23 = (20 30 70);
var ab23 ab23 ab23;
catlevel 0 1 2;
run;
```

| | |
| Variable | | One

| | | 1 |

| | | |

| Age first told | Sample Size | 3709 |

| by doctor that | Weighted Size | 1380250.55 |

| have diabetes | Total | 76494.78 |

| or sugar | Percent | 5.54 |

| diabetes: 0 - | SE Percent | 0.52 |

| HIGH | | |

| | | |

| Age first told | Sample Size | 3709 |

| by doctor that | Weighted Size | 1380250.55 |

| have diabetes | Total | 113516.38 |

| or sugar | Percent | 8.22 |

| diabetes: 0 - | SE Percent | 0.58 |

| HIGH | | |

| | | |

| Age first told | Sample Size | 3709 |

| by doctor that | Weighted Size | 1380250.55 |

| have diabetes | Total | 1055360.64 |

| or sugar | Percent | 76.46 |

| diabetes: 0 - | SE Percent | 0.88 |

| HIGH | | |

The example below shows how you can use the recode statement to
recode a 0/1 variable into a 1/2 variable. Although this is not needed
for proc descript, this is used because the recoding is so clearly
shown in the output.

```
data temp01;
set temp1;
newvar1 = 0;
if _n_ ge 4050 then newvar1 = 1;
run;

proc descript data=temp01 filetype=sas design =
jackknife;
weight rakedw0;
jackwgts rakedw1--rakedw80 / adjjack=1;
recode newvar1 = (0 1);
var newvar1 newvar1;
catlevel 1 2;
```

```
run;

Number of observations read : 11000 Weighted count : 4801259
Denominator degrees of freedom : 80
Variance Estimation Method: Replicate Weight Jackknife
by: Variable, One.
----------------------------------------------------
| | |
| Variable | | One
| | | 1 |
----------------------------------------------------
| | | |
| NEWVAR1: 1 | Sample Size | 11000 |
| | Weighted Size | 4801258.85 |
| | Total | 1763363.22 |
| | Percent | 36.73 |
| | SE Percent | 0.58 |
----------------------------------------------------
| | | |
| NEWVAR1: 2 | Sample Size | 11000 |
| | Weighted Size | 4801258.85 |
| | Total | 3037895.63 |
```

| | Percent | 63.27 |
| | SE Percent | 0.58 |

The example below shows how you can use the recode statement to
recode a 1/2 variable into a 0/1 variable. According to the SUDAAN
website, you cannot use the recode statement to recode a value of 2 to
0 (2 = 0).

```
data temp01a;
set temp01;
if newvar1 = 0 then newvar2 = 1;
if newvar1 = 1 then newvar2 = 2;
run;

proc descript data=temp01a filetype=sas design = jackknife;
weight rakedw0;
jackwgts rakedw1--rakedw80 / adjjack=1;
recode newvar2 = (2 3);
var newvar2 newvar2;
```

catlevel 0 1;

run;

| | |

| Variable | | One

| | | 1 |

| | | |

| NEWVAR2: 0 | Sample Size | 55428 |

| | Weighted Size | 23847415.32 |

| | Total | 1763363.22 |

| | Percent | 7.39 |

| | SE Percent | 0.15 |

| | | |

| NEWVAR2: 1 | Sample Size | 55428 |

| | Weighted Size | 23847415.32 |

| | Total | 22084052.10 |

| | Percent | 92.61 |

| | SE Percent | 0.15 |

The values on the recode statement temporarily recode

newvar2
to be a 0/1 variable for use as a dependent variable in a logistic regression.

```
proc rlogist data=temp01a filetype=sas design = jackknife;
weight rakedw0;
jackwgts rakedw1--rakedw80 / adjjack=1;
recode newvar2 = (2 3);
model newvar2 = racehpra ae21a ab23;
subgroup racehpra;
levels 4;
run;
```

Number of zero responses : 37
Number of non-zero responses : 404

Independence parameters have converged in 8 iterations

Number of observations read : 55428 Weighted count: 23847415
Observations used in the analysis : 441 Weighted count: 221626

Denominator degrees of freedom : 80

Maximum number of estimable parameters for the model is 6

Sample and Population Counts for Response Variable NEWVAR2
0: Sample Count 37 Population Count 27513
1: Sample Count 404 Population Count 194113

R-Square for dependent variable NEWVAR2 (Cox & Snell, 1989): 0.022326

-2 * Normalized Log-Likelihood with Intercepts Only : 330.83
-2 * Normalized Log-Likelihood Full Model : 320.87
Approximate Chi-Square (-2 * Log-L Ratio) : 9.96
Degrees of Freedom : 4

Note: The approximate Chi-Square is not adjusted for clustering.
Refer to hypothesis test table for adjusted test.
Variance Estimation Method: Replicate Weight Jackknife
Working Correlations: Independent

Link Function: Logit

Response variable NEWVAR2: NEWVAR2

Independent P-value

Variables and Beta T-Test

Effects Coeff. SE Beta T-Test B=0 B=0

Intercept 0.45 0.96 0.47 0.6399

Race - UCLA CHPR

Definition

LATINO 0.50 0.44 1.12 0.2642

PACIFIC ISLANDER inf . . .

AIAN 2.54 0.81 3.13 0.0024

ASIAN 0.00 0.00 . .

Minutes doing

walking or

bicycling 0.01 0.01 1.52 0.1329

Age first told by

doctor that have

diabetes or sugar

diabetes 0.02 0.02 0.94 0.3491

Contrast Degrees
of P-value
Freedom Wald F Wald F

OVERALL MODEL 5 33.64 0.0000
MODEL MINUS
INTERCEPT 4 3.16 0.0182
INTERCEPT . . .
RACEHPRA 2 4.91 0.0097
AE21A 1 2.30 0.1329
AB23 1 0.89 0.3491

Independent
Variables and Lower 95% Upper 95%
Effects Odds Ratio Limit OR Limit OR

Intercept 1.57 0.23 10.63
Race - UCLA CHPR
Definition
LATINO 1.64 0.68 3.94
PACIFIC ISLANDER . . .
AIAN 12.64 2.52 63.27
ASIAN 1.00 1.00 1.00

Minutes doing walking or bicycling 1.01 1.00 1.03

Age first told by doctor that have diabetes or sugar diabetes 1.02 0.98 1.05
