

Example 4: Logistic Regression

In the following sample code, current asthma status (*astcur*) is examined, controlling for race (*racehpr2*), sex (*srsex*), and age (*srage*). As SUDAAN and Stata require the dependent variables coded as 0 and 1 for logistic regression, a new dependent variable *ast* is created and assigned 1 where *astcur*=1 (“Current asthma”) and 0 where *astcur*=2 (“No current asthma”). The category “No current asthma” is used as the reference in the analysis.

SAS:

```
PROC SURVEYLOGISTIC DATA = data NOMCAR VARMETHOD=TAYLOR; a
FORMAT astcur astcurf. racehpr2 racehprf. srsex srsex.; b
STRATA tsvarstr;
CLUSTER tsvrunit; c
WEIGHT rakedw0;
CLASS astcur (REF="NO CURRENT ASTHMA") racehpr2 (REF="WHITE")
srsex (REF="MALE")/PARAM=REF; d
MODEL astcur = racehpr2 srsex srage;
RUN;
```

^a In SAS, the NOMCAR option presents the assumption that missing values are not completely at random. This, along with the DOMAIN statement, is the appropriate approach for domain analyses, which uses the entire sample for variance estimation.

^b When the values are formatted either in the data setup or in the procedure, SAS automatically picks the category of the categorical variables whose label is in the last alphabetical order as the reference group. In PROC SURVEYLOGISTIC, the reference category of the independent and dependent variables may be specified in a CLASS statement.

^c When using concatenated data across adults, adolescents, and/or children, use *tsvrunit*; when using separate data files, delete the commands associated with *tsvrunit*.

^d PARAM=REF is specified to ensure dummy coding of the categorical independent variables.

SUDAAN:

```
DATA newdata;
SET data;
IF astcur=1 THEN ast=1;
ELSE IF astcur=2 THEN ast=0;
RUN;

PROC RLOGIST data = newdata FILETYPE=SAS DESIGN=WR;
WEIGHT rakedw0;
NEST tsvarstr tsvrunit;
SUBGROUP racehpr2 srsex;
LEVELS 7 2;
REFLEVEL racehpr2 = 6 srsex = 1;
MODEL ast = racehpr2 srsex srage;
RUN;
```

Example 4: Logistic Regression continued.

Stata:

***Sample design specification step*^a**

```
use "DATASET LOCATION"
```

```
svyset TSVRUNIT [pw=rakedw0], strata (TSVARSTR)b
```

Analysis

```
recode astcur (2=0) (1=1) (-9=.), gen (ast)
```

```
xi: svy: logit ast srage i.race i.srsexc
```

```
xi: svy: logistic ast srage i.race i.srsexd
```

Stata, the sample design specification step should be included before conducting any analysis.

^b When using concatenated data across adults, adolescents, and/or children, use `tsvrunit`; when using separate data files, delete the commands associated with `tsvrunit`.

^c This statement produces parameter estimates.

^d This statement produces odds ratios. Stata automatically chooses the lowest value of the categorical variable as the reference group for the independent and dependent variables.

In

Example 4: Logistic Regression continued.

SPSS:

***Sample design specification step*^a**

* Analysis Preparation Wizard.

CSPLAN ANALYSIS

/PLAN FILE='\\PATH FOR COMPLEX SURVEY PLAN FILE\FILENAME.csaplan'

/PLANVARS ANALYSISWEIGHT=RAKEDW0

/PRINT PLAN

/DESIGN STRATA= TSVARSTR CLUSTER=TSVRUNIT^b

/ESTIMATOR TYPE=WR.

Analysis

RECODE

srsex

(1=2) (2=1) INTO newsex.

VARIABLE LABELS newsex 'NEWSEX'.

EXECUTE.

RECODE

racehpr2

(1=1) (2=2) (3=3) (4=4) (5=5) (6=7) (7=6) INTO newrace.

VARIABLE LABELS newrace 'NEWRACEHPR2'.

EXECUTE.

* Complex Samples Logistic Regression.

CSLOGISTIC astcur BY newsex newrace WITH srage

/PLAN FILE = '\\PATH FOR COMPLEX SURVEY PLAN FILE\FILENAME.csaplan'

/MODEL newsex newrace srage

/INTERCEPT INCLUDE=YES SHOW=YES

/STATISTICS PARAMETER EXP SE CINTERVAL

/TEST TYPE=F PADJUST=LSD

/ODDSRATIOS FACTOR=[newsex]

/ODDSRATIOS FACTOR=[newrace]

/ODDSRATIOS COVARIATE=[srage]

/MISSING CLASSMISSING=EXCLUDE

/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE]

LCONVERGE=[0] CHKSEP=20 CILEVEL=95

/PRINT SUMMARY VARIABLEINFO SAMPLEINFO.

^a In SPSS, the sample design specification step should be included before conducting any analysis.

^b When using concatenated data across adults, adolescents, and/or children, use `tsvrunit`; when using separate data files, delete the commands associated with `tsvrunit`.

^c SPSS CSLOGISTIC automatically chooses the highest value of the categorical variable as the reference group for the independent variables as well as the dependent variable. Therefore, recoding categorical variables is necessary to select the desired reference categories if they are different than the categories with highest values.