

## Example 4: Logistic Regression

In the following sample code, current asthma status (*astcur*) is examined, controlling for race (*racehpr2*), sex (*srsex*), and age (*srage\_p*). SUDAAN and Stata require the dependent variables to be coded as 0 and 1 for logistic regression, so 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 VARMETHOD=JACKKNIFE;
FORMAT astcur astcurf. racehpr2 racehprf. srsex srsex.;a
WEIGHT rakedw0;
REPWEIGHT rakedw1-rakedw80/JKCOEFS=1;b
CLASS astcur (REF="NO CURRENT ASTHMA") racehpr2 (REF="WHITE")
srsex (REF="MALE")/PARAM=REF;c
MODEL astcur = racehpr2 srsex srage_p;
RUN;
```

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

<sup>b</sup> Jackknife coefficients are necessary for accurate variance calculations, and jackknife coefficients of 1 in SAS will produce equal variance calculations as those produced in SUDAAN. However, for SAS V.9.2(TS1M0) and earlier, a value of 1 will not be accepted; as a substitute, 0.9999 can be entered. Without this specification, the default value of the jackknife coefficients will be  $[(\# \text{ replicate weights} - 1)/\# \text{ replicate weights}]$ ; for CHIS, this would be  $[(80 - 1)/80] = 0.9875$ .

<sup>c</sup> 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=JACKKNIFE;
WEIGHT rakedw0;
JACKWGTS rakedw1-rakedw80/ADJJACK=1;
SUBGROUP racehpr2 srsex;
LEVELS 7 2;
REFLEVEL racehpr2 = 6 srsex = 1;
MODEL ast = racehpr2 srsex srage_p;
RUN;
```

Example 4: Logistic Regression continued.

**Stata:**

**\*Sample design specification step\*<sup>a</sup>**

```
use "DATASET LOCATION"  
svyset [pw=rakedw0], jkrw(rakedw1-rakedw80, multiplier(1)) vce(jack)  
mse
```

**\*Analysis\***

```
recode astcur (2=0) (1=1) (-9=.), gen (ast)  
  
xi: svy: logit ast srage_p i.race i.srsexb  
xi: svy: logistic ast srage_p i.race i.srsexc
```

<sup>a</sup> In Stata, the sample design specification step should be included before conducting any analysis.

<sup>b</sup> This statement produces parameter estimates.

<sup>c</sup> 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.