

### **Example 3: Linear Regression**

In the following sample code, Body Mass Index (*bmi*) is examined in relation to race (*racehpr2*), sex (*srsex*), and age (*srage*) while controlling for each other. Note that *racehpr2* and *srsex* are categorical variables, and White (*racehpr2=6*) and Male (*srsex=1*) are used as their reference categories.

#### **SAS:**

```
PROC SURVEYREG DATA = data NOMCAR VARMETHOD=TAYLOR; a
STRATA tsvarstr;
CLUSTER tsvrunit; b
WEIGHT rakedw0;
FORMAT racehprf. srsex srsex.;
CLASS racehpr2 srsex; c
MODEL bmi = srsex racehpr2 srage/SOLUTION; d
RUN;
```

<sup>a</sup> In SAS, the *NOMCAR* option presents the assumption that missing values are not completely at random.

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

<sup>c</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 alphabetically last as a reference group.

<sup>d</sup> *SOLUTION* option provides the parameter estimates when using a *CLASS* statement.

#### **SUDAAN:**

```
PROC SORT DATA = data;
BY tsvarstr tsvrunit;
RUN;

PROC REGRESS DATA = data FILETYPE=SAS DESIGN=WR;
NEST tsvarstr tsvrunit; e
WEIGHT rakedw0;
SUBGROUP racehpr2 srsex;
LEVELS 7 2;
REFLEVEL racehpr2=6 srsex=1;
MODEL bmi = racehpr2 srsex srage;
RUN;
```

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

#### **Stata:**

##### **\*Sample design specification step\* <sup>f</sup>**

```
use "DATASET LOCATION"
svyset TSVRUNIT [pw=rakedw0], strata (TSVARSTR) g
```

##### **\*Analysis\***

```
Recode racehpr2 (6=1) (1=2) (2=3) (3=4) (4=5) (5=6) (7=7), gen(race) h
xi: svy: regress bmi i.srsex i.race srage
```

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

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

<sup>h</sup> Reordering is done in order to choose "White" (*racehpr2=6*) as the reference group.

**SPSS:**

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

\* 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<sup>b</sup>

/ESTIMATOR TYPE=WR.

**\*Analysis\***

RECODE<sup>c</sup>

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 General Linear Model.

CSGLM bmi 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 SE CINTERVAL

/PRINT SUMMARY VARIABLEINFO SAMPLEINFO

/TEST TYPE=F PADJUST=LSD

/MISSING CLASSMISSING=EXCLUDE

/CRITERIA CILEVEL=95.

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

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

<sup>c</sup> SPSS CSGLM automatically chooses the highest value of the categorical independent variables as the reference groups. Therefore, re-coding categorical variables is necessary to select the desired reference categories if they are different than the categories with the highest values.