Example 3: Linear Regression

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

**SAS:**

```sas
PROC SURVEYREG DATA = data VARMETHOD=JACKKNIFE;
WEIGHT rakedw0;
REPWIGHT rakedw1-rakedw80/JKCOEFS=1 or 0.5;  
FORMAT racehpr2 racehprf. srsex srsex.; CLASS racehpr2 srsex;  
MODEL bmi_p = srsex racehpr2 srage_p/SOLUTION;
RUN;
```

a. 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 [(# replicate weights - 1)/# replicate weights]; for CHIS, this would be [(80 - 1)/80] = 0.9875. Following external statistical consultation and research on two-PSU-per-stratum paired jackknife replicate designs, CHIS recommends that data users use a jackknife coefficient of 0.5 instead of the historical 0.9999 beginning in 2019, in conjunction with the methodological redesign. The use of a jackknife coefficient algebraically equivalent to 1 for paired jackknife replicates has produced more conservative estimates of standard errors in recent years. This change will be reflected in all CHIS data products including AskCHIS and Data Access Center (DAC) projects using CHIS 2019 data as well as all subsequent years moving forward.

b. 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.

c. SOLUTION option provides the parameter estimates when using a CLASS statement.

**SUDAAN:**

```sas
PROC REGRESS DATA = data FILETYPE=SAS DESIGN=JACKKNIFE;
WEIGHT rakedw0;
JACKWGTS rakedw1-rakedw80/ADJJACK=1 or 0.5;  
SUBGROUP racehpr2 srsex;
LEVELS 7 2;
REFLEVEL racehpr2=6 srsex=1;
MODEL bmi_p = srsex racehpr2 srage_p;
RUN;
```

**Stata:**

```stata
*Sample design specification step*  
use "DATASET LOCATION"
svyset [pw=rakedw0], jkrw(rakedw1-rakedw80, multiplier(1 or 0.5)) vce(jack) mse

*Analysis*  
recode racehpr2 (6=1) (1=2) (2=3) (3=4) (4=5) (5=6) (7=7), gen(race)
xi: svy: regress bmi_p i.srsex i.race srage_p
```

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

e. Recoding is done in order to choose “White” (racehpr2=6) as the reference group.