**Homework 9. Variance Estimation**

**MSDS 6370**

**Submitted by Hari Narayan Sanadhya**

**Objective:**

* For the student to learn more about replication methods for variance estimation.
* For the student to learn more about using SAS to implement the jackknife and BRR (Balanced Repeated Replication) methods of variance estimation for complex survey design.

**Introduction**

Asynchronous Lecture 9 focused on replication methods of variance estimation for complex survey designs. In particular, the discussion focused on the jackknife and BRR methods of variance estimation. This assignment builds on that discussion and extends it by exploring more options in SAS PROC SURVEYMEANS for implementing variance estimation.

**Exercise 1**

1. In the video this week, you are asked to compute the jackknife and BRR estimates of variance for the toy data we examined.

Submit the spreadsheet with the completed calculations for homework this week.

Fill in the last two rows of the table below:

|  |  |
| --- | --- |
| method of estimation for standard error | Standard error estimate |
| Taylor linearization without fpc | 0.009348 |
| Taylor linearization with fpc (as given in video) | 0.007637 |
| SAS computed jackknife estimate | 0.009425 |
| SAS computed BRR estimate | 0.009447 |
| jackknife estimate from spreadsheet calculation | 0.009424724 |
| BRR estimate from spreadsheet calculation | 0.009446672 |

****

**Exercise 2**

2. Go to the PROC SURVEYMEANS documentation. Look in the details section, under the topic “Replication Methods for Variance Estimation.” Read the intro, discussion of BRR, and the jackknife in this section. Follow the steps below:

**Step 1: Search Google for sas proc surveymeans and click on link**

**Step 2: Scroll down to Table 85.2 and click on Jackkife (BRR) Link to get explanation of methods**

**Step 3: Go back to main proc surveymeans sheet you found in Step 1, scroll down past Table 85.2 and find repweights link under Jackkife, Outweights to learn more about repweights (Look carefully since the link is not obvious).**

a. What is the purpose of the REPWEIGHTS (replicate weights) statement?

SOLUTION

Using REPWEIGHTS statement, the user can provide the replication weights to be used by SAS PROC SURVEYMEANS while using the BRR or JACKKNIFE variance estimation methods. If REPWEIGHTS are not provided, then the procedure constructs the replication weights for the analysis. REPWEIGHTS are nonnegative numerical numbers providing the weight for a single replicate. The number of replicates equals the number of REPWEIGHTS variables.

b. What does the documentation say about what to specify in your SAS code about the design if you use a REPWEIGHTS statement?

SOLUTION

REPWEIGHTS are nonnegative numerical numbers providing the weight for a single replicate. The number of replicates equals the number of REPWEIGHTS variables. If replicate weights are supplied to the PROC SURVEYMEANS with a REPWEIGHTS statement, there is no need to specify a CLUSTER or STRATA statement. If REPWEIGHTS statement is used and VARMETHOD is not supplied in the PROC SURVEYMEANS statement, then JACKKNIFE variance estimation method is used by default. If REPWEIGHTS statement is specified without WEIGHT statement, the procedure uses the average of each observation’s replicate weights as the observation’s weight.

Options available in the REPWEIGHTS statement are:

1. DF=df which specifies the degrees of freedom (positive number) for the analysis which by default equals the number of REPWEIGHTS variables
2. JKCOEFS=value option can be used when VARMETHOD=JACKKNIFE and it specifies a single value of the jackknife coefficient, which the procedure uses for all replicates.

NOTE: To specify different coefficients for different replicates, use the JKCOEFS=values or JKCOEFS=SAS-data-set option available in the PROC SURVEYMEANS directly instead of this option.

**Exercise 3: For steps a-d, provide the code and output of your answer**

**SAS Code**

libname xl XLSX '/home/harisanadhya0/sasuser.v94/MSDS 6370/Lab 9/HW9Dat.xlsx';

data jkDat; set xl.jk; run;

data brrDat; set xl.brr; run;

/\* Verify if data is imported correctly \*/

proc print data=brrDat(obs=10);

proc print data=jkDat(obs=10);

/\* Add a dataset with stratum size \*/

DATA stratumsize;

INPUT stratum \_total\_;

DATALINES;

1 7

2 6

3 5

4 6

;

/\* a. Mean and StdErr estimation using Taylor Method \*/

title "Taylor Method";

proc surveymeans data = jkDat total = stratumsize;

stratum stratum;

weight w;

cluster psu;

var y;

run;

/\* b. Mean and StdErr estimation using JACKKNIFE Method \*/

title "Jackknife Method";

proc surveymeans data = jkDat VARMETHOD = JK ;

stratum stratum;

weight w;

cluster psu;

var y;

run;

/\* c. Mean and StdErr estimation using JACKKNIFE Method with replication Weights \*/

title "Jackknife Method with REPWEIGHTS";

proc surveymeans data = jkDat VARMETHOD = JK;

repweights rw1 - rw8;

weight w;

var y;

run;

/\* d. Mean and StdErr estimation using BRR Method \*/

title "BRR Method";

proc surveymeans data = brrDat VARMETHOD = brr ;

stratum stratum;

weight w;

cluster psu;

var y;

run;

/\* e. Mean and StdErr estimation using BRR Method with replication Weights \*/

title "BRR Method with REPWEIGHTS";

proc surveymeans data = brrDat VARMETHOD = brr;

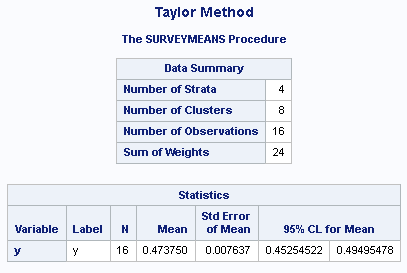
repweights rw1\_1111 rw2\_2121 rw3\_1221 rw4\_2211 rw5\_1112 rw6\_2122 rw7\_1222 rw8\_2212;

weight w;

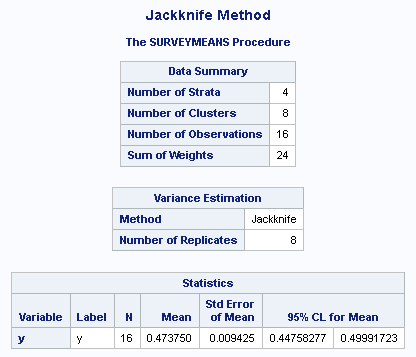
var y;

run;

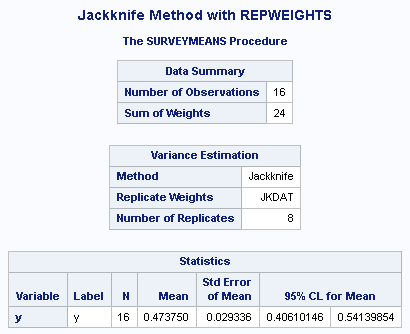
1. Using the data set HW9Dat.xlsx, use proc surveymeans and the Taylor Series Method to form an estimate of the population mean and its standard error.



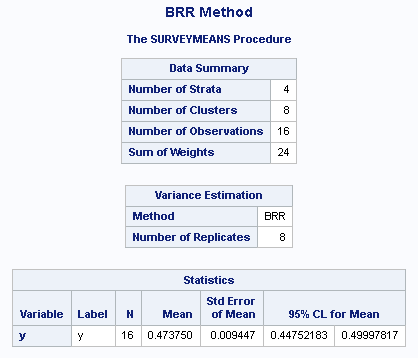
1. With the same input data use proc surveymeans and the Jackknife Method to form an estimate of the population mean and its standard error (see Section 9.6 of the Asynchronous lectures. Note, the total option is not used here even though it appears in the code from the lectures. You should get the same answer as in the lecture.).



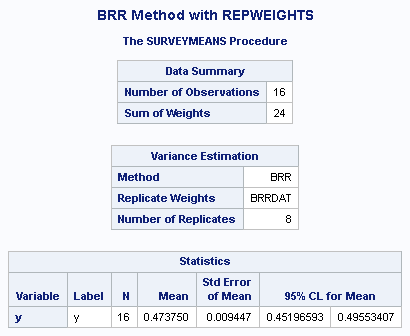
1. Look through the data and identify the variables in the sheet “jk” that are the repweights. Use the repweights statement and the jackknife method in proc surveymeans to find an estimate of the population mean and its standard error (Note: You can use the code in part b above without the cluster and stratum statements. Just add the repweights statement and list the repweight variables from the data set.)



1. With the input data use proc surveymeans and the BRR Method to form an estimate of the population mean and its standard error (see Section 9.6 of the Asynchronous lectures. Note, the total option is not used here even though it appears in the code from the lectures. You should get the same answer as in the lecture.).



1. Look through the data and identify the variables in the sheet “brr” that are the repweights. Use the repweights statement and the brr method in proc surveymeans to find an estimate of the population mean and its standard error (Note: You can use the code in part c above without the cluster and stratum statements. Just add the repweights statement and list the repweight variables from the data set.)



1. Compare the replication methods output. Why do you think the output from the Jackknife with the repweights statement is so different from the other results using replication methods (Do not agonize over this answer too much. If you are not sure of the answer, look at the formulas given by the sheets in the Workbook used in Exercise 1, and take a guess.)

SOLUTION

The replication weights provided in the dataset does not match with the replication weights that were computed by SAS PROC SURVEYMEANS when the REPWEIGHTS statement was not used. When looked at the weights provided in the sample dataset, it appears that the replication weights in the columns rw7 and rw8 are not correct. For stratum 4, the jackknife weights should be zero from first PSU and doubled for second PSU (weight adjustment) but instead both the PSU's in stratum 4 have weights as zero for replication weight specified in rw7 column. In rw8 column, the replication weight for first PSU should be doubled (weight adjustment) and zero for second PSU but instead weights of first PUS's were quadrupled (4 times) though the second PSU had weight of ZERO which was correct. If the weights have been correctly provided, the answers for JACKKNIFE estimates using repweights statement would not be so different then other cases.