**Lab 10. Practicing ratio estimation with PROC SURVEYMEANS**

**MSDS 6370**

**Submitted by Hari Narayan Sanadhya**

**Objective:**

* For the student to learn more about ratio estimation.
* For the student to practice using Proc Surveymeans for ratio estimation.

**Introduction**

The topic of Asynchronous week 10 was a discussion of ratio estimation and aspects of using models in forming estimates with survey data. In this lab, we continue to study forming estimates with a ratio estimator using SAS.

**Estimating the mean with data from a cluster sample using SAS**

Today you will learn how to use SAS PROC SURVEYMEANS to produce a ratio estimator.

You will learn one new feature of PROC SURVEYMEANS.

Note the option below:

Ratio *varname1/varname2*

This option is used to specify the estimate of the ratio where *varname1* is the variable in the numerator and *varname2* is variable in the denominator.

You will use the data collected by a wildlife biologist in the Excel file lab10Dat.xlsx.

Follow the directions for Exercises 1, 2, and 3 on the next page.

**Lab 10. Results and Exercises**

The biologists selected 12 “pockets” of bush at random from 248 in the area, and wants to estimate the number of grouse in each pocket. She also knows the area of each pocket in hectares, as well as the total number of hectares in the study area, which is 3015.

**Exercise 1**

1. There are two ways you can estimate total number of grouse in the study area.

(a) The first way is to calculate

where N is the total number of pockets and is the mean number of grouse per pocket.

Calculate in the spreadsheet. (Spreadsheet attached at the end of the document)

(b) The other is to use the area of the pockets to help in the estimation, by calculating a ratio estimator of total number of grouse. First examine the graph in spreadsheet labeled Prelim Plot. Does it look like using the size of the pocket would be helpful for predicting number of grouse? \_\_\_\_ Why? \_\_\_\_\_

**Solution**:

This is because it appears that the Grouse number is given based of the Area of Pocket. There appears to be a relation between the two. As the area of pocket increases, the higher is the number of grouse inside the bush. So, the number of grouse appears to be some function of Area of Pocket. This makes sense as well because the more is the area of the pocket, its more likely to find greater number of grouse inside of it.

(c) The ratio estimator is

.

where tx is the total number of hectares in the study area, and is the mean number of hectares per pocket in the sample.

Calculate in the spreadsheet. (Spreadsheet attached at the end of the document)

**Exercise 2**

2. Now you will calculate these two estimates using PROC SURVEYMEANS. Make sure you get the estimates that you calculated by hand. (Note: SAS will look for a weight in the input data set. Since you are not using proc survey select which automatically calculates a weight you will need to set the samplingWeight = 1 in the input dataset. In PROC SURVEYMEANS be sure to include the statement, WEIGHT SamplingWeight;).

(a) First calculate Before writing the SAS code, consider the following questions: What are the weights? What design features do you need to tell SAS? What is the population size? Submit your SAS code to complete this part of the lab.

**Solution**:

SAS Code

**data** wildlife;

input area grouse samplingWeight ;

datalines;

8.9 24 1

2.7 3 1

6.6 10 1

20.6 36 1

3.7 8 1

4.1 8 1

25.8 60 1

1.8 5 1

20.1 35 1

14.0 34 1

10.1 18 1

8.0 22 1

;

**run**;

/\* Actual weight = N/n; \*/

**data** wildlife;

set wildlife;

samplingWeight = 248/12;

**run**;

**proc print** data=wildlife;

**run**;

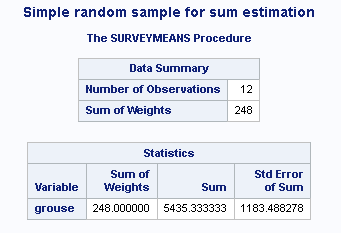
title "Simple random sample for sum estimation";

**proc surveymeans** data = wildlife sum total=248 sumwgt;

weight samplingWeight;

var grouse;

**run**;



(b) Next calculate . Remember that you must use the ratio statement, where the numerator is a variable you create that is tx\*y. Submit your SAS code to complete this part of the lab.

**Solution**:

SAS Code

**data** wildlife;

set wildlife;

txGrouse = 3015\*grouse;

**run**;

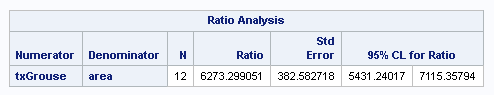
title " Simple random sample with ratio estimation";

**proc surveymeans** data = wildlife total=248;

weight samplingWeight;

ratio txGrouse/area;

**run**;



The ratio above is the estimate of the sum of the number of Grouse computed using ratio estimation.

(c) Recall that the default method of calculating variance (Taylor) tends to underestimate the variance when the sample size is small. In this example, the sample size is only 12, which could be considered small. Recalculate the ratio estimator, using the jackknife estimator of variance. Does it make an appreciable difference? \_\_\_\_ Which standard error estimate do you think is better to use and why? Submit your SAS code to complete this part of the lab.

**Solution**:

SAS Code

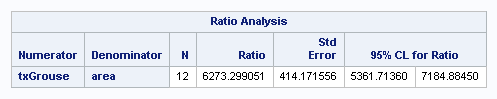
title " Simple random sample with ratio estimation - JackKnife variance estimate";

**proc surveymeans** data = wildlife varmethod=jackknife;

weight samplingWeight;

ratio txGrouse/area;

**run**;



From Taylor variance extimation, we get the standard error of the ratio estimate (estimate of number of Grouse) comes out to be 382.58 whereas using JackKnife, the standard error comes out to be 414.71. Though the ratio estimate is the same in both the cases, JackKnife gives a higher standard error. I would prefer the results from the Taylor estimates in the given scenario mainly due to the reason that this design is a non-complex design – no strata or clustering involved so JackKnife is not presenting any advantage as its beneficial when generating nonlinear estimates. Since here we have a simple survey design, its an example of linear estimation.

(d) Fill in the table below:

|  |  |  |
| --- | --- | --- |
| Estimators | Estimate of total | Standard error of estimate |
|  | 5435.333333 | 1183.488278 |
|  | 6273.299051 | 382.582718 |
|  | 6273.299051 | 414.171556 |

(e) Is collecting the area of the pockets worthwhile for estimating total number of grouse? Why?

**Solution:**

Yes, because it appears that the Grouse number is given based of the Area of Pocket. There appears to be a relation between the two. As the area of pocket increases, the higher is the number of grouse inside the bush. Due to this relation of the Pocket size with the number of Grouse, its worthwhile to collect its area which will lead to more accurate estimates.

**Exercise 3**

3. Now suppose the wildlife biologist would like to estimate the number of grouse per hectare. Make an estimate along with its standard error.

Solution:

Grouse per hectare using simple inflation estimator = 1.803

Grouse per hectare using ratio estimator = 2.081

