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

**MSDS 6370**

**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 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.

The option

Ratio *varname1/varname2*

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 MSDS\_6370\_Lab10\_data.xlsx.

Excel file MSDS\_6370\_Lab10\_data.xlsx also shows the calculation for a ratio estimator and a mean-per-unit (or simple inflation estimator) of total number of grouse on a study area.

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

***Lab 10. Results and Exercises***

*Excel file MSDS\_6370\_Lab10\_data.xlsx shows the calculation for a ratio estimator and a mean-per-unit (or simple inflation estimator) of total number of grouse on a study area.*

*The biologists selected 12 “pockets” of bush at random from the 248 in the area, and attempted to count the number of grouse in each. 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.

*(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?*

Yes, there appears to be a positive correlation between the two variables.

*(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.

**Exercise 2**

*2. Now you will calculate these two estimates using PROC SURVEYMEANS. Make sure you get the estimates that you calculated by hand.*

*(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 in an appendix.*

*(b) Next calculate . Remember that you must use the ratio statement, where the numerator is a variable that you create that is tx\*y. Submit your SAS code in an appendix.*

*(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?*

It makes a difference but appears to be rather minimal (Taylor: Var = 392; JK: Var = 414).

*Which standard error estimate do you think is better to use and why?*

The sample size is low, so the conservative approach is to use the jackknife method, it will prevent underestimating variance as Taylor may be vulnerable to.

*Submit your SAS code in an appendix.*

*(c) Fill in the table below:*

|  |  |  |
| --- | --- | --- |
| Estimators | Estimate of total | Standard error of estimate |
|  | 5435 | 1213.2 |
|  | 6273 | 392.2 |

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

Yes, there appears to be a clear correlation and when area is removed from PROC SURVEYMEANS, the estimated variance triples.

**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.*

|  |  |  |
| --- | --- | --- |
| Estimators | Estimate of ratio | Standard error of estimate |
|  | 2.080696 | 0.137370 |