

Lab4

Part 1) Ratio and Regression estimation under SRS

Introduction

In this material we will calculate ratio estimators using data **from a simple random sample** in order to estimate population parameters. We will construct a ratio estimator using summary statistics from the sample and using SURVEYMEANS. We will also use SURVEYREG to obtain a regression estimate. Ratio and regression estimators will be compared with the mean estimator of the population mean.

Goals:

- 1) Learn how to use SAS to obtain sample correlations.
- 2) Learn how to estimate a ratio using SURVEYMEANS.
- 3) Understand how to construct a ratio estimator for the population total and mean.
- 4) Understand how to estimate standard errors for ratio estimators.
- 5) Be able to compare ratio and sample mean-based estimators.
- 6) Learn how to estimate regression coefficients using SURVEYREG.
- 7) Learn how to estimate a regression estimator for a population mean or total using SURVEYREG.

Examining Ratio Model Assumptions

1. We will be working with the SRS sample from the Agricultural Census used in examples in Chapter 2 of Lohr. We will be using the number of acres of farms in 1987 and the number of acres of farms in 1992 in our ratio and regression estimates.

Download the agsrs.csv dataset from the course webpage to your desktop.

open the file **lab4.sas** in SAS.

Submit the DATA step at the beginning of the program to read in the data.

Note that the SRS sampling weight is calculated as $w_i = N / n$ with the values entered for the problem.

2. Ratio estimators and regression estimators for means and totals do best when there is a positive linear relationship. Ratio estimators are better than regression estimators when the relationship between the response and other variables passes through the origin. With this in mind, it is a good idea to check on these assumptions via plots and sample correlations.

Submit the plot of acres of farms in 1992 versus acres of farms in 1987. Notice that the data shows a linear trend through the origin.

Submit the PROC CORR statement. Note that the COV option is needed to calculate sample variances for y (1992 farm acres in county) and x (1987 farm acres in county).

- a. PROC CORR provides a few descriptive statistics for both the numerator (y) and denominator (x) variables. MEAN is the sample mean, STD DEV is s , SUM is the sum of the observed values.
- b. PROC CORR also produces a sample correlation matrix. Notice that the correlation between the two years is near 1.

Estimating a Ratio Using PROC SURVEYMEANS

3. We can estimate the ratio using PROC SURVEYMEANS, which will also provide an estimate of the standard error of the ratio estimator.

NOTE: The SAS standard error of the ratio estimator uses the sample mean rather than the known population mean for the x (or denominator) variable.

Submit the PROC SURVEYMEANS statement. If we specify VAR and STDERR, SAS will generate the estimated variance and standard error of the ratio.

The output from this SURVEYMEANS contains an estimate of the ratio of 1992 farm acres relative to 1987 farm acres and its standard error.

Estimating a Population Mean and Total Using a Ratio Estimator

4. Now that we have the ratio estimate for acres of farms in 1992 over acres of farms in 1987 and the population mean and total for 1987 (from lecture), we can estimate the mean acres of farms per county in 1992 and the total acreage devoted to farms in 1992 using ratio estimation. We will use a DATA step as a calculator to do this for us. SAS will not do this directly in the PROC SURVEYMEANS.

Submit the DATA step to calculate our estimates.

Submit the PROC PRINT statement to view the estimates.

Comparing Ratio and Sample Mean-Based Estimators

5. We now have two ways to estimate population means and totals. Submit the next SURVEYMEANS statement in the program to calculate the sample mean-based estimates for the population mean and total for acres of farms in 1992. Notice the differences in estimated values between our two approaches. Also note the differences in the estimated standard errors for our two types of estimators.

Regression Estimators

6. If we believed that the relationship between acres devoted to farms in 1992 and 1987 did not pass through the origin, we might consider using an intercept in a regression model. Submit the SURVEYREG statement to estimate the regression coefficients.

The CLPARM option in the MODEL statement provides confidence intervals for your estimate and the estimated regression coefficients.

We don't do this in the lab4.sas program, but if we were to use SURVEYREG without an intercept term (using / NOINT option in the MODEL statement), we would not get the same value as if we used RATIO in SURVEYMEANS.

The regression parameter estimates do NOT give you the estimated mean. You will need to use an ESTIMATE statement to obtain a linear estimate based on the estimated coefficients. The syntax for estimating the mean is:

```
ESTIMATE '<label>' INTERCEPT 1 <SAS name for x variable> <value of xbarU>;
```

This essentially calculates the formula we learned in lecture by inserting the b-hats and population mean for x (\bar{x}_U) into the equation. In other words,

```
ESTIMATE 'Reg est for mean 1992 farm acres' intercept 1 acres87 313016.4;
```

is equivalent to:

$$\hat{y}_{reg} = \hat{B}_0 + \hat{B}_1 \bar{x}_U = -2548.12(1) + 0.9950(313016.4) = 308,903.2$$

Part 2) Domain estimation under SRS

Introduction

Domain estimation is used to estimate parameters for subpopulations that are not defined by strata. It is a special case of ratio estimation. Now we will learn two ways to use SAS to calculate estimates for the population mean.

Goals:

- 1) Learn how to obtain domain estimates from SAS using SURVEYMEANS.
- 2) Explore relationship between domain and ratio estimators by calculating a ratio estimate for a domain using SURVEYMEANS.

We still work on the SRS sample from the Agricultural Census used in examples in Chapter 2 of Lohr and we are interested in the acres92 of west.

Note that "west" in the DATA step defines the subpopulation domains.

Submit the SURVEYMEANS statements. The output should contain the overall estimate for the population mean using all of the data. In addition, there should be a section that shows domain estimates for the mean for each level of west.

how to use "SAS help and documentation"?

1) click the "help" and choose " SAS help and documentation"

The screenshot shows the SAS software interface. The 'Help' menu is open, displaying options like 'SAS Help and Documentation', 'Getting Started with SAS Software', 'Learning SAS Programming', 'SAS on the Web', and 'About SAS 9'. The main window displays the output of a PROC SURVEYREG procedure. The output includes a title, a data summary, and a table of statistics.

The SURVEYMEANS Procedure

Data Summary

Number of Observations	300
Sum of Weights	3078

Statistics

Variable	Mean	Std Error of Mean	Var of Mean	Sum	Std Dev	Var of Sum
acres92	297897	18898	357150824	916927110	58169381	3.3836769E15

The code editor shows the following SAS code:

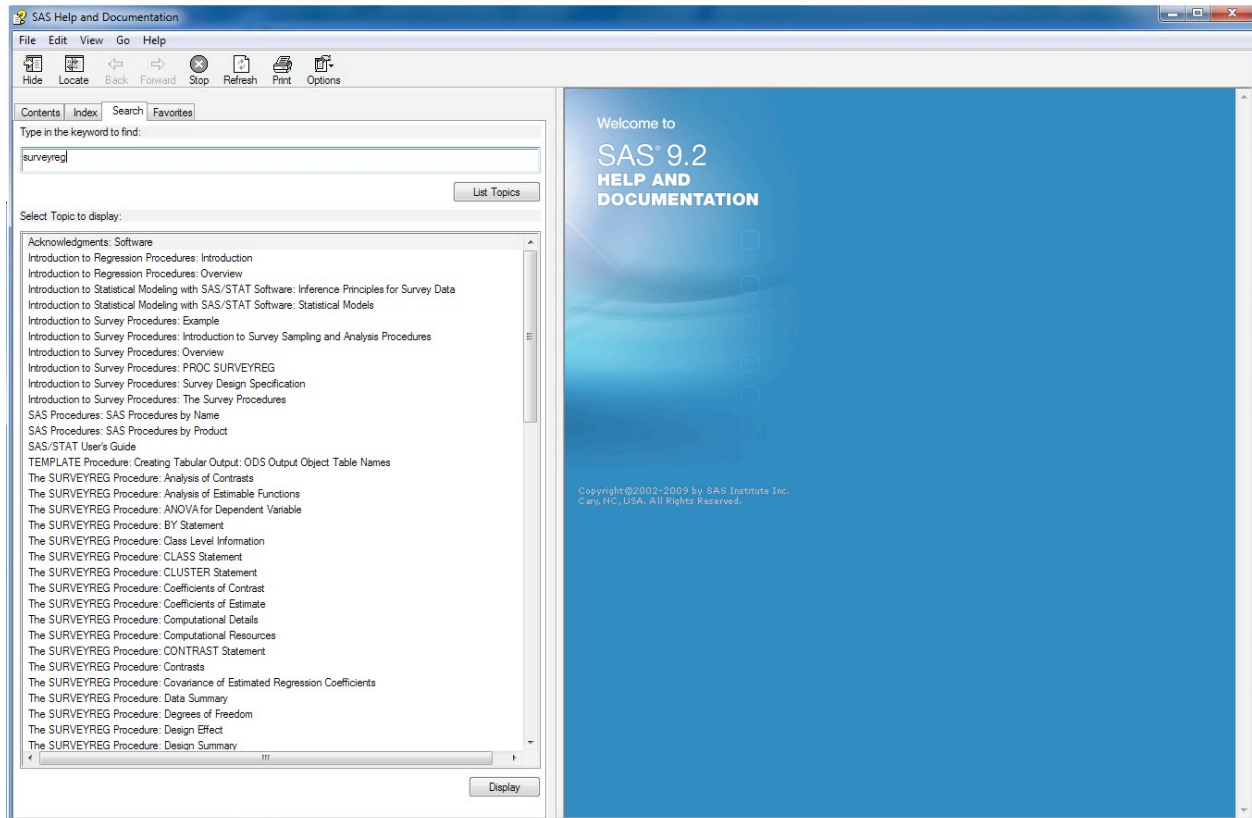
```

/* If we did not believe that the relationship between
x and y can be explained using a straight line we could
fit a line with an intercept and a slope. PROC
SURVEYREG allows us to do this and estimate a mean or total
at the same time, without extra calculations.

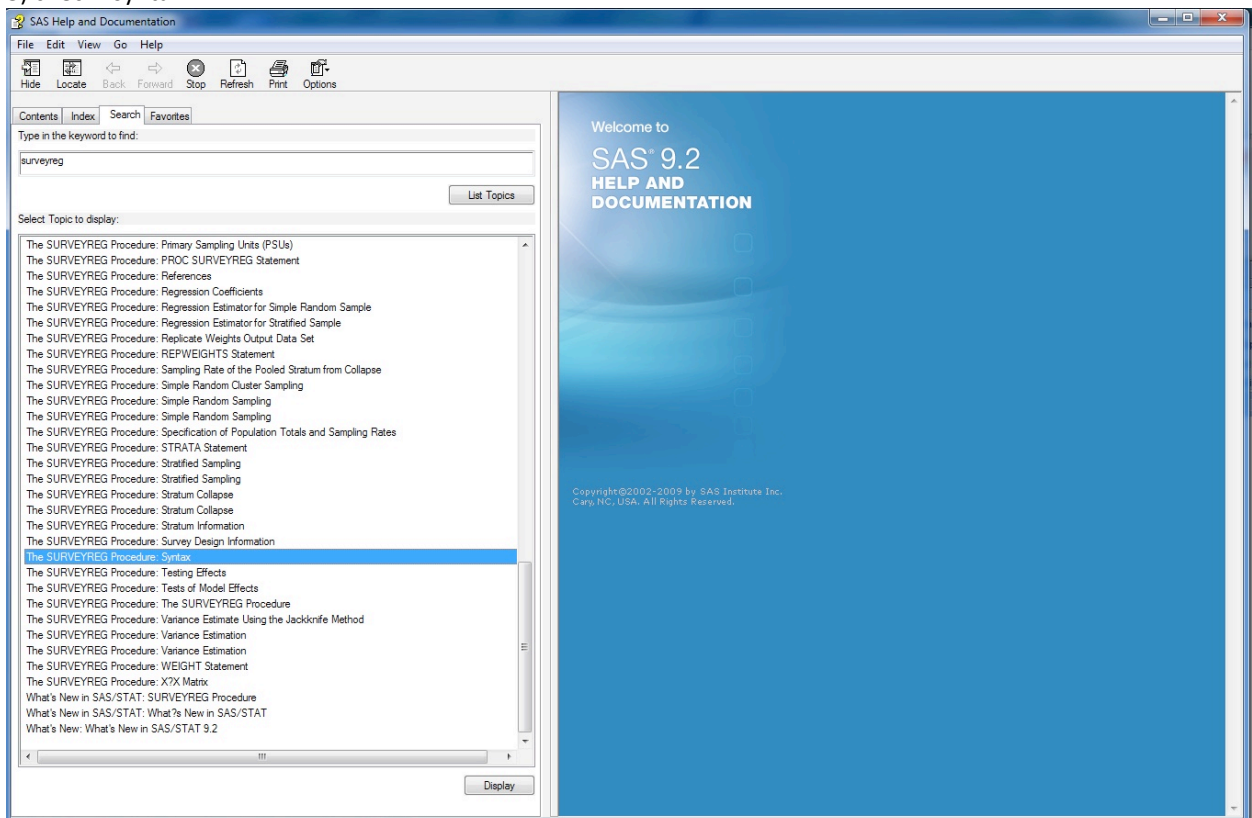
The estimate statement gives the estimated value and approximate
estimated standard deviation for the estimate for coefficients
you provide.
*/
PROC SURVEYREG DATA=agsamp TOTAL=3078 ;
MODEL acres92 = acres87 / CLFARM;
ESTIMATE 'Regression estimate for mean 1992 farm acres' intercept 1 acres87 313343.283;
WEIGHT wgt;
TITLE 'Regression estimator for mean 1992 farm acres per county';
RUN;

```

2) type "surveyreg" in the search dialog



3) check "syntax"



4)

The screenshot shows the SAS Help and Documentation window. The search bar contains 'surveyreg'. The left pane lists various topics, with 'The SURVEYREG Procedure: Syntax' selected. The right pane displays the following content:

Previous Page | Next Page

The [SURVEYREG](#) Procedure

Syntax: [SURVEYREG](#) Procedure

The following statements are available in PROC [SURVEYREG](#):

```
PROC SURVEYREG <options> ;
  BY variables ;
  CLASS variables ;
  CLUSTER variables ;
  CONTRAST 'label' effect values < ... effect values > </ options > ;
  DOMAIN variables <variable=variable variable=variable=variable ... > ;
  ESTIMATE 'label' effect values < ... effect values > </ options > ;
  MODEL dependent = <effects > </ options > ;
  OUTPUT <keyword <=variable-name> ... keyword <=variable-name>> </ option > ;
  REPWEIGHTS variables < / options > ;
  STRATA variables </ options > ;
  WEIGHT variable ;
```

The PROC [SURVEYREG](#) and MODEL statements are required. If your model contains classification effects, you must list the classification variables in a CLASS statement, and the CLASS statement must precede the MODEL statement. If you use a CONTRAST statement or an ESTIMATE statement, the MODEL statement must precede the CONTRAST or ESTIMATE statement.

The CLASS, CLUSTER, STRATA, CONTRAST, and ESTIMATE statements can appear multiple times. You should use only one MODEL statement and one WEIGHT statement.

- [PROC \[SURVEYREG\]\(#\) Statement](#)
- [BY Statement](#)
- [CLASS Statement](#)
- [CLUSTER Statement](#)