

Lab 2: Simple random sample selection and analysis

Introduction

This material deals with the topics of SRS design, selection, and analysis using SAS.

Goals:

1. Learn how to use PROC SURVEYSELECT to draw an SRS.
2. Be able to use SURVEYMEANS to obtain estimates from an SRS.
3. Understand the principles behind simulations.

Using PROC SURVEYSELECT

1. Download the SAS program file lab2.sas and data file agpop.csv from the course website. NOTE: the majority of information about this material is included within the lab2.sas program comments.
2. Highlight and submit the first two DATA steps to read in the data. The first DATA step reads the whole data set ($N = 3,078$) and second one removes the observations with missing data in them. ($N = 3,041$)
3. Submit the PROC MEANS and PROC FREQ section of the program to view the population means and totals from MEANS and proportions from FREQ.
4. Submit the PROC SURVEYSELECT section to draw a SRS of size 30 from the data frame.
 - (1) The METHOD is the sampling design such as SRS for simple random sampling and SYS for systematic sampling.
 - (2) The n=30 is the number of sampling units we want to select from our frame.
 - (3) The SEED part is the random number SAS uses for selecting the sample.
 - If this is missing, SAS will use the system time to obtain a SEED number.
 - The SEED number must be some positive integer.
 - (4) The OUT part is where we provide a name for the outputted data set containing the sample.
5. Look at the output from running PROC SURVEYSELECT. Most of the information we already know, but note that SAS tells us the selection probability for a sampling unit.
6. Submit the PROC PRINT statement to display the 30 selected records from our frame.
7. Change the SEED number to 8104 in the PROC SURVEYSELECT statement and submit the PROC SURVEYSELECT and PROC PRINT statements again.

Using PROC SURVEYMEANS to Analyze Data from a SRS

To analyze data from an SRS we will use PROC SURVEYMEANS. Keep in mind though that if the variable we are interested in is categorical, then we should include a CLASS statement within PROC SURVEYMEANS.