



Academic

Survey Sampling and Analysis

by Katrien Declercq

Researchers often use sample survey methodology to obtain information about a large aggregate or population by selecting and measuring a sample from the population. Due to the variability of characteristics among items in the population, researchers apply scientific sample designs in the sample selection process to reduce the risk of a distorted view of the population, and they make inferences about the population based on the information from the sample survey data. In order to make statistically valid inferences for the population, they must incorporate the sample design in the data analysis.

Traditional SAS procedures, such as the MEANS procedure and the GLM procedure, compute statistics under the assumption that the sample is drawn from an infinite population by simple random sampling. These procedures generally do not correctly estimate the variance of an estimate if they are applied to a sample drawn by a complex sample design. SAS/STAT software (version 8) now provides the SURVEYSELECT, SURVEYMEANS, and SURVEYREG procedures for survey sampling and analysis.

The SURVEYSELECT procedure provides a variety of methods for selecting probability-based random samples. The procedure can select a simple random sample or a sample according to a complex multistage sample design that includes stratification, clustering, and unequal probabilities of selection. With probability sampling, each unit in the survey population has a known, positive probability of selection.

PROC SURVEYSELECT provides methods for both equal probability sampling and sampling with probability proportional to size (PPS). In PPS sampling, a unit's selection probability is proportional to its size measure. Available PPS methods include without replacement, with replacement, systematic, and sequential with minimum replacement. The procedure can apply these methods for stratified and replicated sample designs.

The SURVEYMEANS and SURVEYREG procedures perform statistical analysis for survey data. These analytical procedures take into account the design used to select the sample. The sample design can be a complex sample design with stratification, clustering, and unequal weighting. You can use the SURVEYMEANS procedure to compute the following statistics:

- population total estimate, its standard deviation, and a corresponding t test
- population mean estimate, its standard error, and a corresponding t test
- proportion estimate for a categorical variable and corresponding t test
- confidence limits for estimates of the population total, population mean, and proportions
- data summary information

PROC SURVEYREG fits linear models for survey data and computes regression coefficients and their variance-covariance matrix. The procedure also provides significance tests for the model effects and for any specified estimable linear functions of the model parameters.

More detailed information can be found in the SAS V8 Online Doc.