Statistical Sample Design, Estimation, and Analysis

Many of the nation's most important and widely used surveys in the fields of education, health, and the economy have benefited from Westat's careful approach, in which our expertise and innovative thinking are applied to solving unique research challenges.

Westat statisticians work closely with clients in developing and implementing sample survey designs that meet the requirements, challenges, and resource allocations for their specific research goals. These designs include plans for identifying potential survey subjects, selecting an unbiased sample of these subjects, imputation of survey responses after data collection, sample weighting, variance estimation, and producing analysis-ready data that are representative of the targeted population and maintain the confidentiality of all survey respondents. Westat statisticians have extensive background and special expertise in statistical model development applied to complex survey data that can address many demanding analysis tasks.

Sample Design and Selection

A sound sample design begins with carefully selecting an efficient and unbiased sample of the targeted population. Westat has extensive experience with area surveys of the population in households and group quarters, surveys of and within schools and other establishments, as well as surveys of unique and hard-to-reach populations.

Westat has decades of experience in selecting samples based on standard and custom sampling frames. High-quality general population surveys use address-based sampling frames consisting of locatable addresses from the U.S. Postal Service (USPS), and when appropriate, additional addresses not locatable by their USPS address found by on-the-ground enumeration. Other frames are obtained from school districts, healthcare providers, or academic institutions, to name a few.

Selection strategies, including multistage probability samples and methods that control overlap across samples or oversample some subgroups to achieve higher quality estimates for those groups, are customized for each project.

Imputation

Imputation is a process of assigning values for missing responses to produce a complete set of survey data.

The main reason for conducting imputation is to facilitate the analysis of the data, using subject-matter knowledge and information gained during the data collection process.

Westat has developed and implemented efficient procedures for hot deck and similar random imputation techniques, regression and other model-based imputation, and multiple and fractional imputation. Additionally, Westat statisticians have developed automated procedures for imputing whole questionnaires, which combine several of these techniques to address the challenges of imputing data from complex surveys while preserving the covariance structures and distributions of the underlying survey data.

Weighting

Results from probability-based sample surveys need to be weighted to reflect probabilities of selection under the complex sample design employed to select the samples. In addition, statistical adjustments to compensate for nonresponse

and noncoverage are often applied to improve representativeness and precision of the survey-based estimates.

Westat has extensive experience with a variety of weight adjustment and calibration techniques. Westat also has experience with weighting strategies needed to combine independent samples together so that estimates may be made using data from the combined surveys.

Variance Estimation

Providing measures of precision is essential to the interpretation of survey results. When data are collected as part of a complex sample survey, care is needed to produce design-consistent estimates of variance. In a complex sample survey setting, variance estimates will be incorrect if computed using standard statistical software packages that assume simple random sampling.

Westat has considerable experience with variance estimation techniques that can be used to estimate sampling errors, including direct variance estimation, replication methods (balanced repeated replication, jackknife replication, bootstrap), and Taylor series linearization methods. When creating weighted datasets, we typically provide replicate weights for variance estimation, which ensure that data users can correctly calculate measures of precision in a wide range of software programs. In addition to the weighted datasets, Westat has developed software that can generate customized and high-volume tabulations and supports online analytic processing on the web.

Statistical Disclosure Control

The ability to protect the confidentiality of survey respondents has become increasingly important in recent years. In order to uphold the privacy pledge extended to survey participants, many Federal and private agencies require both an assessment of disclosure risk and application of measures to minimize disclosure prior to release of tabular data or microdata.

Statistical disclosure control techniques reduce the risk of disclosing identifiable information that may reveal the identity of an individual, business, or other entity that has contributed to the dataset. Judicious application of these techniques maximizes the amount of information that may be provided from a dataset for analytic purposes while protecting data confidentiality.

Westat statisticians have extensive experience in assessing data disclosure risk and applying statistical disclosure control techniques. We have developed software that analyzes the disclosure risk elements in the dataset, performs coarsening treatments and perturbation methods, and assesses the impact of the perturbation and coarsening methods on the explanatory power of the data.

Small Area Estimation

The essence of small area estimation is the use of regression techniques and mixed models to predict estimates for small domains, often defined geographically. The resultant estimators are known as "indirect" estimators since they are distinct from standard survey or "direct" estimators that do not depend on the validity of a particular statistical model.

Westat staff includes experts who have made significant contributions to the theory and application of small area estimation and related methods, including data integration.

Analysis of Survey Data

In analyzing data from complex surveys, the sample design needs to be incorporated into the analysis to obtain statistically valid estimates. These analyses often require the use or development of specialized statistical software. Westat statisticians have extensive experience performing such analyses, as well as providing advice to and collaborating with client researchers in their analyses.