SCIENCE RESOURCES STATISTICS

STATISTICAL GUIDELINES
For
SURVEYS and PUBLICATIONS

OUTLINE

Introduction

1. Standards for Surveys

2. Standards for Reports

Introduction

The purpose of statistical standards is not to develop a rigid set of rules, but to set a framework for assuring SRS data quality; that is, data and products, maintain an appropriate level of "fitness for use." The guidance found in this document is meant to help (1) increase the reliability and validity of data, (2) promote common understanding of desired methodology and processes, (3) avoid duplication and promote the efficient transfer of ideas across program areas, and (4) remove ambiguities and inconsistencies. The goal is to provide the clearest possible presentations of data. As statistical standards, this document focuses on technical issues involved in our work rather than issues of standards in contract management or publication formats.

In general, SRS aims to adhere to the ideals set forth in "Principles and Practices for a Federal Statistical Agency" (National Academy Press, 2nd ed., 2001). As a Federal statistical agency, SRS surveys must follow guidelines and policies as set forth in the Paperwork Reduction Act and other legislation related to surveys. For example, SRS surveys must follow the implementing guidance, survey clearance policies, response rate requirements, and related orders prepared by the Office of Management and Budget. The standards in this document are not meant to repeat materials found in such "rules," but focus on clarification or extensions of the principles outlined in the rules as they relate to SRS survey processes.

The following standards were developed from a thorough review of related standards developed at other Federal statistical agencies[1], a literature review, and group discussions with program staff from SRS. This documentation of standards communicates a core set of broad principles and methodological guidelines desired in SRS activities, trying to avoid overly prescriptive techniques. Thus, SRS personnel have wide latitude in selecting appropriate methodologies within the guidelines to meet the standards.

While it is important to define standards, the SRS management team expects ongoing staff discussion of these concepts with subsequent clarifications as necessary and, as a consensus develops for the need, additions, revisions, and/or more specific guidelines addressing best practices or policy decisions.

[1] This document draws from standards and guidelines found at other Federal Statistical Agencies, especially the National Center for Education Statistics. This was done intentionally to help promote a sense of uniformity in the quality of statistical products produced by the Federal statistical system and especially because other agencies produce data related to SRS products and SRS often uses their data in reports.

1. Standards for Surveys

This section addresses technical documentation needed for surveys and guidelines for methodology such as variance estimation and measurement of major nonsampling error sources.

1.1 Survey Documentation

PURPOSE OF DOCUMENTATION: Documentation of the survey must provide information for planning similar surveys and improving future replications of the same survey. The documentation must also enable the users of the survey data to understand the quality and limitations of the data.

Standard 1.1A Survey Documentation*: All surveys conducted by SRS must document the methodology used in the survey (both sample design and survey operations) in a methodology report or section and maintain such documentation on the web. To the extent feasible, documentation must be made of the kinds of nonsampling errors found in the survey and of the procedures (if any) used to correct for them.*

PROCEDURES: The documentation shall contain, at a minimum, the relevant detailed information used to develope the Supporting Statement of the OMB required Paperwork Reduction Act Submission and supporting data for the "survey quality" report kept on the SRS web site for each survey.

A methodology report/section will (1) describe the survey, (2) provide users with access to information about the quality of data, (3) serve as a guide in the development of an effective system to monitor and maintain data quality over time, (4) identify potential sources of nonsampling error, as available, and (5) provide references, when they exist, for more detailed documentation of the measurement of the magnitude of the various types of sampling and nonsampling error. Documentation should cover sample design, questionnaire definitions, coding, editing, imputation, etc.

Generally, this material will be prepared as part of the methodological report expected at the completion of each survey. Future survey contracts must include a requirement for a methodology report covering items addressed in this standard.

GUIDELINES FOR METHODOLOGY REPORTS/SECTIONS:

The methodology report/section should:

- Estimate sampling error for data items related to "purpose and use" issues described in the OMB clearance process and
- Discuss any current cycle and historic changes made to the survey and provide information related to the impact of the changes on the most important known nonsampling errors for the data items related to "purpose and use" issues. The user should get a clear idea of the impact of the change, especially when the data would be used in a time-series.
- Describe what is currently known about each source of error and its impact on the estimates (Sampling error is certainly a factor in measuring the quality of survey-based estimates, but nonsampling errors must be addressed since they can be equal, if not more important, in their impact); and

- Compare error types and measures for some key data items with those in other similar surveys (benchmarks). Where possible, nonsampling error estimates and bounds should make use of data from other surveys, administrative records, or censuses: any different source of information is useful as an indicator.

The methodology report/section should include information on the following aspects of the survey:

- Survey design and operational considerations
  a. Target population and sample frame
  b. Sample design, including sample size and target coefficient of variation (CV) for key items
  c. Data collection techniques (modes, number of contacts, copies of contact materials: instruments, letters, notices)
  d. Estimation techniques
  e. Methods of mixed modes
  f. Methods used to develope the questionnaire
  g. Definitions of items in the questionnaire

- Possible sources of error (GPRA web reporting supporting data)
  a. Sampling
  b. Coverage
  c. Unit nonresponse
  d. Major item nonresponse
  e. Measurement

- Changes affecting the time-series of survey estimates

The methodology report should also contain information about the following aspects of the survey:

- Any formal pretesting of new questionnaires or new or substantially modified questionnaire items;
- Documentation of imputation specifications;
- Debriefing or focus group sessions with small groups of respondents to learn how they respond to the questionnaire and how they interpreted content items;
- Any review of relevant cognitive research from outside surveys;
- Research and decisions on methods to effectively make nonresponse weighting adjustments;
- Development and documentation of imputation methods for handling item nonresponse;
- Documentation for the editing of the survey, including an appropriate set of cross-wave edits; and;
- Standards for the methodology to be used to control error sources during the conduct of the survey. This effort should be aimed at achieving targets for sufficiently small total survey error in order to achieve reliable data relative to the intended uses of the data (e.g., testing between groups or over time).

The checklist presented in Appendix A gives ideas on potential methods for evaluation of sources of survey error.

1.2 Survey Methodology

PURPOSE: To ensure that there is adequate planning for estimation, variance estimation, and indicators of major nonsampling error sources in surveys, and that appropriate procedures are used.

Standard 1.2 Survey Methodology: *Each recurring SRS survey shall develop and use methodology for estimation, variance estimation, and measurements/control of nonsampling error that is consistent with generally accepted professional standards for surveys and, as appropriate, consistent across similar SRS surveys.*

GUIDELINES FOR VARIANCE ESTIMATION:

- Variance estimates shall be obtained for all key variables, those so identified in the OMB clearance document, for which estimates will be presented.
- Variance estimators will be based on theoretical and empirical considerations related to the sample design.

    a) Estimation should, where possible, make use of other data from the survey, from prior surveys, or from administrative records or censuses.
    b) Ratio or regression estimators should be used if they provide some reduction in variance without substantially complicating subsequent analyses.
- Variance estimators need to be designed to minimize the mean squared error of the resulting estimates.

    a) Taylor Series, Jackknife estimators and balanced repeated replication should generally be considered superior to random group estimators because of their greater stability in most cases.

- When replication methods are used to produce variance estimates, the number of replicates should be justified if that number is less than thirty because of the concern for the stability of the variance estimate.
- Subject to confidentiality constraints, the data file resulting from full-scale data collection will contain all the information needed for estimation and variance estimation, e.g., weights, PSU codes, replicates.
- If a large number of statistics are being estimated and presented, consideration should be given to modeling variances and presenting the summary form of the model.

GUIDELINES FOR COMPUTATION OF RESPONSE RATES:

- Response rate reporting for random digit dial and in-person household surveys should follow the "Statistical Definitions: Final Disposition of Case Codes and Outcome Rates for RDD Telephone Surveys and In-Person Household Surveys" as prepared by the American Association for Public Opinion Research.
- Response rates for sample surveys are to be calculated on weighted data (using the sample base weights); and response rates for censuses or universes are based on unweighted data (or, equivalently, where the weights are equal one). When the sampling unit is not the unit of analysis, it is appropriate to multiply the sampling weight of the sampling unit by the sampling weight of the unit of analysis. For example, the sample may be institutions, but interest is in estimates that are functions of the number of students enrolled in science or

engineering. In such a case, multiply the sample weight of the institution by the sampling weight of the students enrolled. This concept could also apply to a census. Comparisons to the unweighted response rates are encouraged.

- Overall effective response rates (R_o) are to be calculated as the ratio of the number of completed interviews to the number of sample respondents drawn minus respondents considered to be out-of-scope. As an example, in a household interview, this would be the number of units sampled minus the number of vacant units, condemned units, or units converted from residential to business use. A subsample of "unable-to-contact" sample units could be drawn for further study, if it is unknown whether the units are in-scope or out-of-scope and if they represent a significant proportion of the initial sample. If the unable-to-contact sample unit is known to be in-scope but doesn't respond, the unit is classified as a noninterview. For unable-to-contact units for which it is unknown whether they are in-scope or out-of-scope, a portion of them can be considered out-of-scope (the ratio of the # of the weighted known out-of-scope to the weighted # of units sampled less the weighted # of unable-to-contacts) for the purpose of calculating Ro.

$$R_o = \frac{\text{Weighted \# of completed interviews}}{\text{Weighted \# of units sampled} - \text{weighted \# out-of-scope units}}$$

- Item response rates (R_i) are to be calculated as the ratio of the number of respondents for which an in-scope response was obtained to the number of completed interviews for which the question (or questions if a composite variable) was intended to be asked.

$$R_i = \frac{\text{weighted \# of respondents with in-scope response}}{\text{weighted \# of completed interviews for which question intended}}$$

GUIDELINES FOR IMPUTATION OF ITEM NONRESPONSE:

PURPOSE: To ensure that there is adequate planning for SRS imputation in sample surveys and in censuses, that appropriate imputation procedures are used, and that these procedures are tested and found to be valid and to yield consistent estimates.

- For all new surveys and new data collections of ongoing surveys, the set of items to be imputed should be identified during the planning phase, and all appropriate missing data imputed for these items. For example, items to consider are those for which important univariate or cross tabulated estimates are to be published by SRS.

    a) Missing data for key items should be imputed in all surveys, ongoing as well as new (except for items for which the response rate is too low to permit estimation and analysis).

- Imputation procedures should be consistent across cycles of the survey, must be based on theoretical and empirical considerations, and must make use of the most relevant data available.

    a) There should be empirical evidence that imputation procedures have produced desirable results, e.g., unbiased estimates for simple means and totals under the assumption that nonrespondents are similar to respondents within imputation classes (cells), maintenance of covariance structures, etc.
    b) When changes are made to improve imputation, some information on the impact of comparisons with past data should be provided.

- Published estimates based on incomplete data for which there has been no imputation must appropriately note this fact. Data used in text, where the missing data could impact the proper use in the text, should be accompanied by a statement indicating the proportion of missing data. For tabular data, use footnotes, technical appendices or a rederence to methodological work on the web, as appropriate.

    a) If item nonresponse is above 20 percent, publication of estimates using the item needs program management approval, with consultation with the Chief Statistician.

- All imputed values on a data file must be clearly identified as such.

2. Standards for Reports

This section addresses issues of data quality, statistical testing and graphical and tabular presentation of data from surveys.

PURPOSE: To ensure that the substantive, statistical, graphical, and tabular content of SRS reports is valid methodologically and used appropriately for any conclusions drawn or implied.

2.1 Standards for Analysis and Statistical Comparisons

PURPOSE: To ensure that comparisons, conclusions and inferences cited in reports are based on appropriate statistics.

STANDARD 2.1A Data Validity*: The analysis of an item (used in the text of a report or as the highlight of a table or graphic) shall adequately consider or be contingent on an acceptable level of "nonsampling error" in the survey estimate for the item. The consideration of "nonsampling error" must include the possible impact on the analysis from known or highly suspected influences including survey nonresponse, item nonresponse, response validity, and coverage error.*

GUIDELINES:

a) A reasonable measure of "total item nonresponse: is:
   total item nonresponse rate= (1 – survey response rate * item response rate).
b) This measure should not be excessive for any analyzed variable and when large (e.g. .4), the representativeness of the sample should be evaluated with a non-response bias analysis.
c) If the measure is unacceptable for the whole survey but is acceptable for a given stratum or analytic sub-universe, analysis of the variable at these levels may be done.

STANDARD 2.1B Statistical testing*: When estimates from sample surveys are compared to one another in the text of reports, comparisons must be subjected to statistical testing for significance. The level of significance (alpha) should be stated (in general .05 is used).*

GUIDELINES:

a) Appropriate statistical tests are required whenever terms such as increased/decreased, was different, trended upward/downward, and similar comparative terms are used.
b) Exclusive reliance on t-tests should be avoided when other simple procedures, such as chi square analysis or analysis of variance, are proper statistical practice.
c) When a chi square analysis or an analysis of variance does not show a statistically significant overall difference among categories, it is inappropriate to t-test individual differences for pairs of categories.
d) When a test is implied, the test must be documented. Informal write-ups of statistical tests and sources of variance estimates, or a statement of where such documentation may be found when there are many tests, should be included in the report review jacket.
e) Occasionally, finding no change is of practical importance. At such times, the author must determine that the statement is not simply the result of an insufficient sample size to measure the change, that is, that the test had enough statistical power to measure an important or practical amount of change had the change occurred.
f) The report review file (red jacket) must indicate where the documentation of testing is kept. The documentation must be available upon request for two years after publication of the report.

STANDARD 2.1C Design Effects*: Hypothesis tests should incorporate the design effects (cluster, stratification, etc.) of the survey associated with the estimates being tested.*

GUIDELINE: A permissible (although not necessarily preferred) way to incorporate design effects into chi square analysis and analysis of variance is to reduce effective sample sizes.

STANDARD 2.1D Multiple Comparisons*: When t-tests are used to compare means or proportions, multiple tests need to be combined into sets based on the nature of the conclusions to be drawn, and a Bonferroni (or other reasonable) correction for multiple testing should to be incorporated.*

GUIDELINES:

a) In general, the data items to be combined for joint testing would be the items used within major sections of short reports or those used in concept chapters of longer publications.
b) Multiple comparison procedures include Bonferroni, Scheffe, and Tukey tests. Multiple comparison procedures should be used to control the level of type I error for simultaneous inferences.
c) In analyses involving multiple variables, factors, or levels, an overall F or chi-square test of significance should be performed to determine which variables or factors are significant before conducting multiple comparisons. The overall test of significance could be from an analysis of variance, a two-way contingency table, or the test of a log-linear model in the case of higher dimensional tables of categorical data. When multiple comparisons are used, the number of comparisons and the overall type I error rate should be given for each simultaneous inference.
d) Consideration should be given to use of multivariate techniques such as regression and log-linear models rather than a sequence of cross tabulations, especially in analyses involving complex covariate structure.

STANDARD 2.1E Relevance*: Findings should be "practically and substantively" as well as "statistically" significant.*

GUIDELINE:
Authors must assure that small differences are "substantively" as well as "statistically" significant, because data sets with large samples often show small differences to be statistically significant.

Standard 2.1F Use of Outside Research
The author and SRS reviewers should pay particular attention to assessing the scientific integrity of cited research, especially when it is from sources outside SRS. Reports must clearly indicate sources.

GUIDELINES:

a) In general, material from a scientifically acceptable peer reviewed (including many agency review processes) source is desired.
b) When using non-peer reviewed or unpublished material, the author must have assessed the methodology used in the research and noted in the review folder their assessment of the appropriateness of the methodology and the adequacy of tests used in the work. In this case, the analyst should use the source data to the extent practical.
c) If the author wants to cite descriptive terms that should have depended on statistical tests, the author should make notes to the review jacket when data are from a successfully adjudicated outside report. But if the author is citing data from tables or charts, and making "new" comparisons that weren't in the original report, then tests are needed.
d) When insufficient information from the outside source is available to determine the quality of the data or whether statements have been appropriately tested, use of the data should be avoided.

2.2 Standards for Tabular and Graphical Presentations

PURPOSE: To ensure that tables and graphs used in SRS publications display data in such a way that readers can conveniently evaluate the accuracy of the results presented.

Standard 2.3A Use of Tables
Tables must be able to "stand alone," having adequate statistical information to assess basic data quality issues without reading text in the report.

GUIDELINES:

To achieve this purpose the following should be met:

- Adequately small standard errors (se's) or confidence intervals (CI's) on statistics in tables that are referenced in text of the report, or would very likely be used by typical readers, are implied in short reports.
- Standard errors or relative variances should be presented in long technical reports and detailed statistical tables. For long reports, reporting can be handled in two ways. For publications that are targeted to general audiences and for detailed statistical table publications, a separate table of standard errors on key statistics may be presented in a technical appendix. The second method, often preferred for more technical

publications, is to include standard errors, relative errors or confidence intervals in the table being presented, either in a separate subsection of the table or in columns accompanying the statistics being presented.

- "Low quality" cells in a table, those having a relative error greater than 50 percent, are best suppressed. Publication of cells based on an n less than 5 need the approval of program management with consultation with the Chief Statistician.

Standard 2.3B Rounding in Tables

Rounding should be appropriate for the level of standard error of the estimate.

GUIDELINES:

a) Rounding should neither be so little as to imply a level of precision that is not present in the data, nor should it be much larger than the standard error of the estimate for key variables in a table (those typically used in comparisons).
b) Percentages should generally be rounded to no more than three significant digits. (e.g., to one decimal place if any percentage is 10 or more, to two decimal places if all percentages are under ten, andto three decimal places if all percentages are one or less.)
c) Rounding should be small enough to avoid notable differences in a calculated value from rounded versus unrounded data when the calculated value is used in the text. For example, a problem occurs when percent changes or ratios of key comparisons are computed with unrounded data when used in the text, but the data are highly rounded in tables.
d) When possible, calculations to produce summary data and computations performed for purposes of estimating standard errors should use unrounded data.
    1. When adding, multiplying, or dividing figures that have been rounded to different significant digits, the product can only be stated in terms of the number with the fewest significant digits.
    2. Sums of column (row) figures in a table must be derived using unrounded figures, with appropriate rounding of the total after its derivation. To handle the problem of column (row) figures not summing to a rounded total, an explanatory footnote must be used. For example, "Because of rounding, details may not add to totals."
    3. The final rounded value must be obtained from the original values, if available, not from a series of roundings (e.g. 7.1748 can be 7.175 or 7.17 or 7.2 or 7 but not 7.18). This situation typically arises when researchers round percentages from already rounded tables.

Appendix A

A CHECKLIST DESIGNED TO EVALUATE DIFFERENT TYPES OF NONSAMPLING ERROR

(1) Nonresponse

Unit Nonresponse:

- Description of characteristics of units not responding
- Examination of changes in response rates over time
- Intensive follow-up of a sample of nonrespondents to ascertain nonresponse bias
- Assessment of method of handling nonresponse: imputation or weighting procedures

Item Nonresponse:

(2) Coverage – adequacy of frame

- Matching studies to earlier versions of the same data source or to other data sources
- Comparison of a very reliable external source for some subset of the current population to the current frame
- Analysis of survey returns for deaths, duplicates, changes in classification, and out of scope unites
- Field work – such as area listings
- Review of frames by appropriate interested/knowledgeable parties
- Comparison of estimated counts with estimated counts from another source

(3) Validity – accuracy of responses

- Re-interview Study with adjudication of disagreements
- Site visits to examine administrative records and compare to subject responses
- Comparison with outside data sources

(4) Reliability – consistency of responses

- Re-interview Study
- Examination of changes in response over repeated questioning
- Agreement among multiple respondents in a sampled unit (e.g. parent/student)
- Agreement of statistics derived from different sections of the questionnaire or different questionnaires.

(5) Survey Design Issues

- Pretests to determine the efficacy of devices to improve response (e.g. offering incentives)
- Pretests to compare alternative ways of collecting or processing data
- Comparison of final design effects with estimated design effects used in survey planning
- Ina longitudinal survey, establishment of a small independent replicate of the sample for use as a test group while the survey continues to allow an evaluation of the effect of changes

(6) Estimation

- Examination of the choice of estimator/design
- Possibility of fitting survey distributions to known distributions from other sources to reduce variance and bias
- Re-estimation using alternative techniques, such as alternative outlier treatments, alternative imputation procedures, and alternative variance estimation techniques
- Analysis of the imputation process – frequency of imputation, initial and final distribution of the variables
- Use of generalized variance curves
- Effect of changes in data processing procedures on survey estimates