

Supporting Document C

Technical Documentation Statements

Version 1.5

Issued: 09 Mar 06

Census Bureau Standard

Describing the Sources and Accuracy of Tabulations and Estimates

Authored by:

Jay Keller

Special Assistant to the Associate Director
for Methodology and Standards

USCENSUSBUREAU

Helping You Make Informed Decisions



Document Management & Control

Version	Issue Date	Approval	Description
1.0	23 Mar 01	Associate Directors	Initial Release
1.1	24 Sep 02	Associate Directors	Reissue
1.2	14 Jan 05	Configuration Mgr.	Reformatted to comply with Census Bureau Identity Standard and Quality Program Document Management Plan
1.3	04 Apr 05	M&S Council	Name change for standard
1.4	28 Dec 05	Configuration Mgr.	Updated e-mail address
1.5	09 Mar 06	Configuration Mgr.	Insert hyperlink for main standard.

The most current version of this document is maintained on the Census Bureau Intranet and may be accessed from the Quality Management Repository.

Technical Documentation Statements

Nonsampling error may occur during the development or execution of a survey. There are several sources of nonsampling error. These errors can occur because of circumstances created by the interviewer, the respondent, the survey instrument, or the way the data are collected and processed. For example, errors could occur because:

- the interviewer records the wrong answer, the respondent provides incorrect information, the respondent estimates the requested information, or an unclear survey question is misunderstood by the respondent (measurement error);
- some individuals or businesses which should have been included in the census or survey were omitted (coverage error);
- responses are not collected from all those in the sample (nonresponse error);
- forms may be lost, data may be incorrectly keyed, coded or recoded, etc. (processing error).

Information about nonsampling error for a specific program is provided or referenced with the data. The Census Bureau recommends that data users incorporate this information into their analyses, as nonsampling error could impact the conclusions drawn from the results.

Sampling error is the difference between an estimate based on a sample and the corresponding value that would be obtained if the estimate were based on the entire population (as from a census). Note that sample-based estimates will vary depending on the particular sample selected from the population. Measures of the magnitude of sampling error in direct survey tabular estimates (variances, standard deviations, or coefficients of variation) reflect the variation in the estimates over all possible samples that could have been selected from the population using the same sampling methodology.

Estimates of the magnitude of sampling errors for a specific estimates program are provided or referenced with the data. The Census Bureau recommends that data users incorporate this information into their analyses, as sampling error in survey estimates could impact the conclusions drawn from the results.

Error in model-based estimates arises from the effects of model error, sampling error, and nonsampling error. The relative contribution of these error components to the error in the model-based estimates depends on the model used and the properties of the data. Standard errors provided for the model-based estimates reflect, to the extent possible and subject to the model assumptions, the contributions of model error and sampling error, but do not reflect the contribution of nonsampling error.

Model error refers to error that would result in predictions from a statistical model even with no errors in the data (no sampling or nonsampling error). It can generally be broken down into three components: error that would occur in the predictions even if the true model were known; error

resulting from estimation of model parameters; and error resulting from differences between the form of the assumed model and that of the true, unknown, model.

Confidentiality

For data tabulations:

The Census Bureau has modified or suppressed some data on this site to protect confidentiality. Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's or business' data can be identified.

The Census Bureau's internal Disclosure Review Board sets the confidentiality rules for all data releases. A checklist approach is used to ensure that all potential risks to the confidentiality of the data are considered and addressed. For more information on how the Census Bureau protects the confidentiality of data, see the disclosure limitation topics listed below as appropriate.

For model-based estimates:

Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's or business' data can be identified. The Census Bureau's internal Disclosure Review Board sets the confidentiality rules for all data releases. For more information on how the Census Bureau protects the confidentiality of data used in producing model-based estimates, see the discussions of Disclosure Limitation and Release of Source Data.

Disclosure Limitation Procedures

Suppression (for data tabulations)

Data Swapping (for data tabulations)

Protection of Microdata Files (for data tabulations)

Release of Source Data (for model-based estimates)

Questions about confidentiality may be addressed to: [POL.Policy.Office@census.gov]

Title 13, United States Code: Title 13 of the United States Code authorizes the Census Bureau to conduct censuses and surveys. Section 9 of the same Title requires that any information collected from the public under the authority of Title 13 be maintained as confidential. Section 214 of Title 13 and Sections 3559 and 3571 of Title 18 of the United States Code provide for the imposition of penalties of up to five years in prison and up to \$250,000 in fines for wrongful disclosure of confidential census information.

Disclosure Limitation: Disclosure limitation is the process for protecting the confidentiality of data. A disclosure of data occurs when someone can use published statistical information to identify either an individual or business that has provided information under a pledge of

confidentiality. For data tabulations the Census Bureau uses disclosure limitation procedures to modify or remove the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual or business, the Census Bureau has taken steps to disguise or suppress the original data while making sure the results are still useful. The techniques used by the Census Bureau to protect confidentiality in tabulations vary, depending on the type of data. For model-based estimates the Census Bureau uses other procedures to assure that the estimates and related information that are released cannot be used to disclose individual data.

Suppression: Suppression is a method of disclosure limitation used to protect individuals' confidentiality by not showing (suppressing) the cell values in tables of aggregate data for cases where only a few individuals or businesses are represented. The cells that are not shown are called *primary suppressions*. To make sure the primary suppressions cannot be closely estimated by subtracting the other cells in the table from the marginal totals, additional cells are also suppressed. These additional suppressed cells are called *complementary* or *secondary suppressions*. The process of suppression does not change the marginal totals, so the integrity of the data is not adversely affected. Before the Census Bureau releases data, computer programs check published tables for both primary and complementary disclosures. Suppression was used for the 1980 Census of Population and Housing and is now used for economic surveys and censuses.

Example -- With Disclosure

Value of Shipments by County and Industry

County	Industry				Total
	W	X	Y	Z	
Alpha	15	1*	3*	1*	20
Beta	20	10	10	15	55
Gamma	3*	10	10	2*	25
Delta	12	14	7	2*	35
Total	50	35	30	20	135

NOTE: * Indicates cells in which data may be identifiable due to the low number in the cell.

Example -- Without Disclosure, Protected by Suppression**Value of Shipments
by County and Industry**

County	Industry				Total
	W	X	Y	Z	
Alpha	15	D	D	D	20
Beta	20	10	10	15	55
Gamma	D	D	10	D	25
Delta	D	14	D	D	35
Total	50	35	30	20	135

NOTE: D indicates data withheld to limit disclosure.

Data Swapping: Data swapping is a method of disclosure limitation designed to protect confidentiality in tables of frequency data (the number or percent of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases when creating a table. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas that have similar characteristics (such as the same number of adults and same number of children). Because the swap often occurs within a neighboring area, there is no effect on the marginal totals for the area or for totals that include data from multiple areas. Because of data swapping, users should not assume that tables with cells having a value of one or two reveal information about specific individuals. Data swapping procedures were first used in the 1990 Census, and were used for Census 2000 and for the Census Bureau's American Community Survey. For a description of the disclosure limitation procedures used in the economic census and surveys, see the discussion on [suppression](#).

Protection of Microdata Files: The Census Bureau sometimes releases microdata files which contain data from the censuses of the United States population and household surveys, which it conducts. These files contain individuals' responses that represent only samples of the population and have had all individual identifiers (such as name and address) removed from the records. In addition, to protect confidentiality, the Census Bureau may modify distinguishing characteristics (such as high levels of income), and restrict geographic identifiers (such as the name of a city) so that populations are composed of at least 100,000 people. This is done to protect the identity of individuals.

Release of Source Data: Source data used in the production of model-based estimates are released to the public only when such release would not disclose individual data or violate other confidentiality restrictions applicable to the source data. This includes both source data obtained from Census Bureau surveys and censuses and also data obtained from other sources such as other government agencies.