

APPENDIX B

Accuracy of Sample Data

CONTENTS

Confidentiality of the Data.....	B-1
Editing of Unacceptable Data	B-1
Sources of Error.....	B-2

The data contained in this report are based on the VICS 2001 sample interviewed in the second quarter of 2001. The VICS is designed to provide accurate estimates for the population and housing units for the United States Virgin Islands on annual basis. The VICS, like any other survey, is subject to error. The purpose of appendix B and C is to provide data users with a basic understanding of the accuracy of the data, estimation methodology and the sample design.

CONFIDENTIALITY OF THE DATA

Every effort has been made to protect the confidentiality of the 2001 U.S. Virgin Islands Community Survey data, and to make sure that published data do not disclose information about specific individuals, households, or housing units. All full-time or part-time employees of the University who handle the questionnaires are required to swear to the same oath of confidentiality that is administered by the Bureau of the Census in the collection of census data.

EDITING OF UNACCEPTABLE DATA

The objective of the processing operation was to produce a set of data that describes the population as clearly and accurately as possible. The questionnaire contained two sections: population and housing.

Questionnaires were reviewed and edited during field data collection operations by

field supervisors for consistency, completeness, and acceptability. Questionnaires were also reviewed by edit clerks in the Survey office for omissions, certain inconsistencies, and improper sample selection. For example, write-in entries such as “Don’t know” or “NA” was considered unacceptable in certain quantities and/or in conjunction with other data omissions.

As a result of this review operation, a follow-up telephone or personal visit was made to obtain missing information. Potential coverage errors were included in the follow-up, as well as questionnaires with omissions or inconsistencies beyond the completeness and quality tolerances specified in the review procedures.

Subsequent to field operations, remaining incomplete or inconsistent information on the questionnaires was assigned using imputation procedures during the final automated edit of the collected data. Allocations—computer assignments of acceptable data in place of unacceptable entries or blanks—were needed most often when an entry for a given item was lacking, or when the information reported for person, housing unit or household expenditure was inconsistent with other information for that same person or housing unit. As in previous surveys, the general procedure for changing unacceptable entries was to assign an entry for a person or housing unit that was consistent with entries for persons or housing units with similar characteristics. The assignment of acceptable data in place of blanks or unacceptable entries enhanced the usefulness of the data.

Another way to make corrections during the computer editing process is substitution. Substitution is the assignment of a full set of characteristics for a person or housing unit.

SOURCES OF ERRORS

The results of a survey are used to make quantitative inferences about the population surveyed. A survey error occurs when there is discrepancy between the inferences and the reality. Sources of errors are generally divided into two major types: sampling errors and nonsampling errors. Sampling errors are result of studying a subset of the whole population. Sampling errors are controlled by designing a sample using sound sampling techniques and random selection method, so that the final sampling units have the same probability of selection. The extent of sampling error in survey estimators depends on the survey's sample design. The standard error formulae can be used to calculate the variance of the sample mean.

In any large-scale statistical operation, such as the 2001 U.S. Virgin Islands Community Survey, human- and machine-related errors do occur. These errors are commonly referred to as non-sampling errors. Such errors include not enumerating every sampled household or every person in the selected sample, not obtaining all required information from the respondents, obtaining incorrect or inconsistent information, and recording information incorrectly. In addition, errors can occur during the field review of the enumerators' work, during clerical handling of the questionnaires, or during the electronic processing of the questionnaires.

To reduce various types of non-sampling errors, a number of techniques were implemented during the planning, data

collection, and data processing activities. Quality assurance methods were used throughout the data collection and processing phases of the survey to improve the quality of the data.

Nonresponse

The failure to collect the survey data from some sampled units (non-response), is a serious survey problem that seems to have grown as respondents have become less willing to participate in surveys. If non-respondents differ from respondents with regard to survey variables, the survey estimates based on the respondents alone will be biased estimates of the overall population parameters.

The nonresponse in surveys could occur at two levels: *total nonresponse* occurs when no information is collected from a sampled unit, and *item nonresponse* occurs when some but not all the information is collected. In interview surveys, the mode used for VICS, total non response could be classified into: refusals, not at home, and language problem. The interviewers were carefully trained in approaches to use to avoid refusals. They were instructed to return to conduct an interview at a time more convenient to the respondent if necessary. The assurance of anonymity and confidentiality were provided to eliminate any fears. 'Not at homes' were treated by at least five callbacks. Most of the interviewing was done during the evenings and weekends. The callbacks were made on different days and different times of day.

Item non-response, may occur for a variety of reasons. The respondents may not know the answers to certain questions or they may refuse to answer some questions. They may consider them sensitive or irrelevant. The interviewer may skip over a question under the pressure of the interviewing. A recorded

answer on the questionnaire may be rejected during editing because it is inconsistent with other answers.

The imputation methods have been devised to compensate for item nonresponse. These methods operate by assigning values for the missing values, using the responses to other questions on the questionnaire. One method known as class mean method, divides the sample into classes on the basis of the responses to other relevant items, and then assigns the respondent class mean for all the item non-responses in that class. A variant of this method is to assign for each item non-response one of the item responses in the same class. The U.S. Census Bureau uses a similar method, known as 'hot deck method'.

Undercoverage

It is possible for some sample housing units or persons to be missed by the survey. A major way to avoid undercoverage in a

survey is to ensure that its sampling frame, an address list in this case, is as accurate as possible. In absence of the Master Address File (MAF) for the United States Virgin Islands, a complete 'House Listing' was done of the selected BGs. The house listing was used as the sampling frame, which was up to date.

Processing Error

The many phases involved in the processing of the survey data represent potential sources for the introduction of non-sampling error. The processing of the questionnaire includes the keying of data from completed questionnaires, automated clerical review, and follow-up by telephone, the manual coding of some variables, and the electronic data processing.