
Appendix B. Statistical Methodology

Apéndice B. Metodología Estadística

INTRODUCTION

The 2007 Puerto Rico Census of Agriculture was primarily conducted by mail, with report forms sent to farm operators on the census mail list (CML). This mail list was supplemented by an area sample which accounted for farms not included in the mail list (NML). A single version of the report form, in Spanish, was used for the CML and NML contingents of the Census. Combining data from the CML and the area sample NML should provide complete coverage for the Puerto Rico Census of Agriculture.

CENSUS MAIL LIST COVERAGE AND WHOLE FARM NONRESPONSE

The initial CML consisted of 19,089 names and addresses stratified by size and type of farm. Nonresponse weighting was applied to account for farm operators who did not respond. For each municipio containing agricultural operations, nonresponse weights were computed by response homogeneity group (RHG); the RHGs were six size categories for historical total values of production reported for the 2007 farm records in that municipio. Each nonresponse weight was calculated as the total number of all agricultural operations listed on the CML for that RHG and municipio, divided by the total number of responding farms within that RHG and municipio. For each RHG found in a municipio, this weight, integerized, was applied to the data items of each responding operation of the RHG. This estimation procedure assumes that the distribution of farms in an RHG is the same for census nonrespondents as for census respondents. Records that were found to be undeliverable as addressed (“UAA”) by the U.S. Postal Service and which failed to respond to any follow-up attempts, were excluded from the

INTRODUCCIÓN

El Censo de Agricultura de 2007 en Puerto Rico se llevó a cabo en su mayoría a través del correo. Los formularios fueron enviados a los agricultores, usando el listado de correo del censo (CML, por sus siglas en inglés). Este listado fue complementado con una muestra de área que identificó a los agricultores que no se encontraban en la lista (NML, por sus siglas en inglés). Se utilizó un solo formulario, en español para ambos contingentes del censo (CML y NML). Combinando la información de la lista (CML) y de la muestra (NML) debemos obtener una cobertura completa para el Censo de Agricultura de Puerto Rico.

COBERTURA DE LA LISTA CENSAL Y LA NO RESPUESTA

La lista inicial (CML) consistió de 19,089 nombres y direcciones estratificados por tamaño y tipo de finca. Se aplicó la ponderación de no respuesta para incluir aquellos agricultores que no respondieron al censo. Para cada municipio donde hubo fincas, la ponderación fue calculada para respuestas agrupadas de forma homogénea (response homogeneity groups, or RHG). Los RHGs contenían seis categorías de valores de producción según informados en el 2007 en los récords de las fincas del municipio. Cada ponderación fue calculada como el total de todas las fincas listadas en CML, para todo RHG y su municipio, dividido por el número total de fincas que respondieron, dentro del RHG y el municipio. Por cada RHG encontrado en un municipio, esta ponderación, totalizada, fue aplicada a las partidas de cada finca dentro del RHG. Este procedimiento de estimación asume que la distribución de fincas en un RHG es la misma para la respuesta y la no respuesta. Récords que fueron devueltos por mala dirección postal (“UAA”) y que no respondieron a ningún procedimiento de seguimiento, fueron excluidos de las calculaciones de la no respuesta. Note que los récords de NML no

nonresponse calculations. Note that NML records were not subject to nonresponse weighting because they were subjected to complete nonresponse follow-up, eliminating the possibility of nonresponse for them.

Any CML operation designated as a “must” farm (the sixth RHG, based on very high value of production) was required to complete its census form; if a response could not be obtained directly from a “must” operation after multiple attempts to contact it, secondary or historical sources were used to provide the data. “Must” operations are considered unique; kept outside the normal nonresponse weighting process because they are too unusual to represent nonrespondents, they can only be self-representing; accordingly, all carry a nonresponse weight of one.

For the five non-“must,” non-self-representing RHGs, nonresponse follow-up was monitored to keep nonresponse weights from exceeding 2.0; if an RHG’s nonresponse weight exceeded this upper limit, a collapsing algorithm was applied to combine neighboring RHGs within a municipio, and recomputed the weights for the combined groups. For published tabulations of complete count items, the noninteger nonresponse weight was rounded to an integer weight of 1, 2, or 3, for each record, with seven records in low response municipios carrying a weight of 3. A random rounding algorithm based on cumulated fractional parts of the noninteger nonresponse weight values was used to produce an unbiased result at island level.

COVERAGE ADJUSTMENT

Due to the dynamic nature of mail lists, some farm operators may not be included in the CML. To account for this undercoverage, an area frame consisting of the entire island of Puerto Rico was sampled. For sampling, NASS stratified the Puerto Rico area frame on the basis of agricultural intensity, with strata consisting of: 1) land area with dense agriculture; 2) land area with sparse

fueron sometidos a la ponderación debido a que estos fueron sometidos al proceso de seguimiento eliminando así la posibilidad de no respuesta.

A toda finca incluida en CML y designada como “obligatoria” (el sexto RHG, basado en valores altísimos de producción) se le requirió completar el informe censal. Si no se pudo obtener la información directamente de una finca “obligatoria” después de tratar de contactarla varias veces, entonces usamos información de fuente administrativa, histórica o secundaria para completar el formulario para este tipo de finca. Las operaciones “obligatorias” se consideran únicas y se mantienen fuera de toda ponderación porque son poco usuales en representar los que no responden, solo pueden ser consideradas auto-representados o sea llevan una ponderación igual a uno (1).

Para los otros cinco no “obligatoria”, no auto-representados del RHG, la ponderación aplicada fue monitoreada someramente para asegurarnos que no excediera un valor mayor de 2.0; si una ponderación de RHG excedía el límite superior, se le aplicó un algoritmo para combinar RHG en la vecindad dentro del municipio, y se computó la ponderación para el grupo combinado. Para propósitos de la tabulación, en cada record, la ponderación de no respuesta para fracciones de finca se redondeó a una ponderación de 1, 2 ó 3, con siete récords en municipios de baja respuesta, llevando una ponderación de 3. Para garantizar un resultado sin sesgo, a nivel de Isla, utilizamos un algoritmo de redondeo aleatorio para calcular los remanentes acumulativos.

AJUSTE DE COBERTURA

Se presume que una lista de nombres y direcciones nunca está completa. Por lo tanto; y para ampliar la cobertura, se utilizó un marco muestral de área representativo de toda la Isla de Puerto Rico. El marco muestral se estratificó por el NASS de acuerdo con la actividad agrícola. Las estratas se clasificaron como: 1) áreas con mucha agricultura, 2) agricultura dispersa y pocas viviendas, 3) agricultura dispersa y muchas

agriculture and few houses; 3) land area with sparse agriculture and many houses; 4) cities with no apparent agricultural activity; and 5) ostensibly nonagricultural land such as parks and military reservations. Primary Sampling Units (PSUs) were created based on specific size requirements and permanent boundaries. An additional sampling enhancement involved the grouping of municipios with similar agriculture into nine clusters. Within each stratum and cluster, a random sample of PSUs was selected and then further subdivided into target sampling units called segments. Of approximately 7,500 segments available for sampling, 300 were selected into sample. Aerial photography and maps for the 300 segments were provided to support field data collection. All NML farms discovered within the 300 sampled segments were included in the area sample.

PRE-CENSUS AREA SCREENING

Prior to the actual census data collection, enumerators using aerial photos and municipio maps identified all farm operators within each assigned area segment. Enumerators recorded the farm operator's name, address, and cuerdas operated within the segment. Farm operators from the selected segments were then checked against the Census Mail List. If no match was found, a census report form was sent to the NML operator, with enumerators making follow-up visits to collect the data. For those farm operators discovered in an area segment but included in the Census Mail List, only the CML report was used, to avoid duplication. During the pre-screening process, 453 NML farm operators were found in the 300 sampled area segments, but only 334 of these original NML records were determined to be actual farm operators.

NML DISTRIBUTION AT THE MUNICIPIO LEVEL

Although the area sample size that determined the NML component ensured acceptable precision at the island level, the sample was not designed for reliable estimation of municipio-level data. To redistribute the island-level NML component in a

viviendas, 4) ciudades con cero agricultura y 5) áreas con cero agricultura tales como parques e instalaciones militares. Las unidades muestrales primarias, se crearon utilizando criterios de tamaño y límites permanentes. Además, se crearon nueve grupos de municipios con características agrícolas similares. Dentro de cada estrata y cada grupo, se seleccionó una muestra aleatoria de unidades primarias que luego se subdividió en unidades de muestreo de enfoque llamadas segmentos. Se seleccionaron 300 de alrededor de 7,500 segmentos disponibles para muestra. A los enumeradores se les proveyó una serie de fotografías aéreas y mapas de los 300 segmentos para el trabajo de campo. Toda operación NML que se encontró dentro de los 300 segmentos y que no estaba en la lista fue designada como operación de la muestra.

SELECCION DE LA MUESTRA PRE-CENSO

Antes de la toma del censo, los enumeradores usando fotografías aéreas y mapas de los municipios, identificaron todas aquellas fincas que estaban dentro de cada segmento. Anotaron el nombre del agricultor, la dirección postal, y el total de cuerdas dentro del segmento. Los nombres de los agricultores dentro del segmento especificado fueron cotejados contra la lista postal del censo. Si el nombre no se encontraba, este operador NML se designó como para recibir un formulario y un enumerador le visitó para obtener la información. Para evitar duplicación, aquellos agricultores encontrados en el segmento, pero que estaban en la lista, solo el reporte CML fue utilizado. Durante esta selección de la muestra se encontraron un total de 453 fincas potenciales (NML) de los 300 segmentos del área de muestreo, pero solo 334 de estos récords originales NML resultaron ser operaciones agrícolas.

DISTRIBUCION DE LA MUESTRA (NML) AL NIVEL DE MUNICIPIO

Aunque el tamaño de la muestra que determinó el componente de NML asegura una precisión aceptable a nivel de Isla, la muestra no está designada para hacer estimados adecuados a nivel de municipio. Para redistribuir el componente de NML a nivel de Isla de

more reasonable manner, a statistical calibration model was developed using information from similar CML and NML records. The 2007 model, based on a prototype used for the 1997 Puerto Rico Census, but extended to include small, marginal farms which may be omitted from the Census list sources, applies the area sample design “cluster” definition cited above, as well as the census categorical variable, farm type. For each municipio and farm type, the expanded estimate of total farms of that farm type was calculated for CML farms and combined with the count of unweighted NML farms of that farm type within the municipio. An analogous total, combining the CML estimate of total farms and the unweighted NML count for that farm type, was also computed at cluster level for the municipio’s cluster: the municipio’s share of each NML farm found in its cluster was calculated as the ratio of the municipio-level total farms of that type, described above, to the analogous total farms of that type at cluster level. The municipio-level expansion weight for an NML farm in a cluster was developed as the product of its conventional area frame weight and the municipio’s share of the farms in that farm type. Within an original NML record’s cluster, a replicate record was created to carry the fractionalized weight and the data allocated to any other municipio in the cluster with a share in NML farms of that farm type. The integerization algorithm applied to the municipio-level NML weight was designed to prevent the sum of the redistributed municipio level farm count to deviate by more than 2 from the island level NML farm count.

ESTIMATION

After weighting adjustment of the CML farm records, for nonresponse, and of the NML farm records, for sample expansion and municipio allocation, the CML and NML components were combined to provide a single estimate. Since the CML and NML contingents are mutually exclusive, our combined estimate should reflect complete and unduplicated coverage, provided that there is no significant nonresponse bias, and other nonsampling error is not operative.

B-4 APPENDIX B/APÉNDICE B

manera más razonable, se desarrollo un modelo de calibración usando información de réconds similares en la lista y en la muestra réconds CML y NML. El modelo de 2007, esta basado en el modelo usado en el censo de Puerto Rico de 1997, pero se expandió para incluir fincas pequeñas que fueran excluidas de las fuentes del listado; aplica la definición del diseño de grupos mencionados anteriormente, así como la variable censal categórica, tipo de finca. Para cada municipio y cada tipo de finca, el estimado total de fincas expandidas por tipo de finca, se calculó para fincas CML en la lista y se combinó con el total no ponderado de fincas NML en la muestra para ese tipo de finca dentro del municipio. Un total análogo, combinando el estimado CML del total de fincas y el conteo no ponderado NML para ese tipo de finca, fue también calculado a nivel de grupo para los municipios agrupados-la parte del municipio en cada finca de muestra NML, encontrado dentro de cada grupo, fue calculada como la razón del total de fincas a nivel de municipio de ese tipo de finca, al total análogo de ese tipo de finca a nivel de grupo. La ponderación a nivel de municipio, para una finca de la muestra dentro de un grupo, se crea como el producto de toda ponderación convencional y la participación del municipio en ese tipo de finca. Dentro de los réconds originales NML en los grupos de la muestra, se replicó un record para que llevara la ponderación fragmentada y la información recreada a cualquier otro municipio dentro del grupo con participación en ese tipo de finca en la muestra. Este algoritmo de números enteros aplicado a la muestra a nivel de municipio, se designó para prevenir que la suma de las fincas redistribuidas a nivel de municipio rebasara el nivel de 2 previamente establecido para la muestra en el conteo de fincas NML a nivel de isla.

ESTIMACION

Después del ajuste ponderado de los réconds de finca CML, por no respuesta, y de los réconds de finca NML, por expansión de la muestra y distribución de municipio, ambos componentes (CML y NML) fueron combinados para proveer un solo estimado. Como ambos componentes CML y NML son mutuamente exclusivos, nuestro estimado total debe reflejar una cobertura completa e irreproducible, siempre y cuando no exista sesgo de no respuesta y los errores de muestreo y otros sesgos sean mínimos.

EDITING DATA AND IMPUTING FOR ITEM NONRESPONSE

For the 2007 Puerto Rico Census of Agriculture, all reported data were keyed and then edited by computer. The edits were used to determine whether the reports met the criterion for satisfying the census farm definition: at least 500 dollars of reported sales. The complex edit and imputation system provided the basis for deciding to accept, impute, delete, or alter the reported value for each data record item.

Whenever possible, edit imputations, deletions, and revisions were based on component data, for a sum, or else on other related data provided in the respondent's report form. For certain missing items, such as operator characteristics, a record's 2002 census data were used when available. Values for other data items, either missing or reported and rejected by the edit, were calculated based on reported quantities and known fixed price parameters. When these and similar methods were not available and values had to be supplied, the imputation process used other local information available on the particular item in question. For example, a farm operation that reported cuerdas of pigeon peas harvested but failed to report quantity harvested was assigned a quantity based on an average pound per cuerda of pigeon peas harvested. This average came from the Puerto Rico Department of Agriculture and previous census data.

After the initial computer edit, all keyed reports not meeting the census farm definition were reviewed to ensure that the data had been keyed correctly. Edit referrals were generated for about 60% of the reports whose operations were validated as farms; these cases were also reviewed for keying accuracy and to ensure that computer edit actions were correct. If the results of the computer edit were not acceptable, corrections were made and the record reedited.

CORRIGIENDO LA INFORMACION Y ATRIBUYENDO POR LA NO RESPUESTA

Para el Censo de Agricultura de Puerto Rico del 2007, la información recopilada se entró en las computadoras y se sometió a una revisión detallada. La revisión se usó para determinar si la información recopilada llenaba los requisitos necesarios para la definición censal de finca: como mínimo 500 dólares de ventas reportadas. El sistema de revisión e imputación proveyó la base para decidir si aceptar, atribuir, eliminar o alterar la información recopilada para cada partida del formulario.

Siempre que fue posible, los cambios e imputaciones se realizaron a base de componentes de información reportada por el agricultor en el formulario censal. Para ciertas partidas, como en el caso de características del agricultor, se usó información del censo del 2002 si estaba disponible. Los valores para las partidas en blanco o inaceptables, fueron calculados basándose en información previamente recopilada o en parámetros ya establecidos. Cuando estos medios no estaban disponibles, los valores fueron asignados usando información local disponible para la partida específica. Por ejemplo, si una finca reportó cuerdas de gandules cosechadas, pero no informó la cantidad cosechada, se le asignó una cantidad basándose en el rendimiento promedio (lbs.) por cuerda de gandules cosechada según parámetros existentes en el Departamento de Agricultura y en el censo anterior.

Luego de la revisión inicial, los informes que no llenaban los requisitos de finca definida por el censo fueron nuevamente cotejados, para asegurarnos de que la información fue entrada correctamente. Se refirieron alrededor de 60% de los informes clasificados como fincas; estos casos también fueron revisados para cerciorarnos de que la entrada de datos y la redacción fuesen correctas. Si los resultados corregidos en la computadora no eran aceptables, se hicieron cambios y los récords se enviaron nuevamente para revisión.

CENSUS SURVEY ERROR

Because the NML component of the census data is an estimate, and surveys in general are prone to human error, the possibility of both sampling and nonsampling errors is always present. Sampling errors occur when only a portion of a population is selected. For samples with known probabilities of selection, the precision for a sample can be determined and confidence intervals calculated. In addition, the precision for list nonresponse can be determined and confidence intervals calculated as well. Nonsampling errors are associated with mistakes in reporting or keying the data as well as imputing for nonresponse.

CENSUS SAMPLING ERROR

The 2007 Puerto Rico Census of Agriculture, like the two Censuses preceding it, used an area frame to estimate the NML component from a sample: thus there is sampling error associated with this estimator. The 300 segments sampled represented only a small proportion of the approximately 7,500 segments potentially available. If a different 300 had been selected, the results would have been different. The error resulting from the difference between the sample actually selected, and the universe it was designed to represent, is called sampling error. If the sample of 300 is selected with known probabilities, then the sampling errors can be estimated from just one sample. The sampling error of an estimate is generally expressed in terms of the variance of the estimate, or its square root, the standard error, a measure of precision allowing the user to gauge the accuracy of the estimate. These measures are based on the average sum of the squared differences between each value of the reported data item and the mean value for that item.

Assuming that for each municipio, nonresponding CML farms in a given CML nonresponse stratum resemble the responding farms in that stratum, we can also apply the results of sampling theory to quantify the variability generated by nonresponse in the portion of any estimate coming from the CML

ERROR DE LA ENCUESTA

Dado que el componente NML de los datos de censo es un estimado, y las encuestas están usualmente expuestas a errores humanos, la posibilidad de errores de muestreo siempre existe. Los errores de muestreo ocurren cuando solo se selecciona una porción de la población total. En muestras con una probabilidad conocida de selección, la precisión de la muestra puede determinarse y los intervalos de confiabilidad pueden calcularse. Adicionalmente, la precisión a la no respuesta puede determinarse y los intervalos de confiabilidad pueden calcularse también. Los errores ajenos al muestreo se asocian con errores cometidos al anotar o entrar los datos y con imputaciones a la no respuesta.

ERROR DE MUESTREO

El Censo Agrícola de Puerto Rico del 2007, al igual que los dos censos anteriores, usó un marco muestral de área para estimar el componente de NML, por lo tanto existe error de muestreo asociado con este estimador. En Puerto Rico, los 300 segmentos seleccionados representan solamente una pequeña porción de los casi 7,500 segmentos disponibles. Si se hubiesen seleccionado otros 300 segmentos diferentes, el resultado hubiese sido diferente al obtenido. A la magnitud de estas diferencias se le llama error de muestreo. Si la muestra de 300 segmentos es seleccionada con una probabilidad conocida, entonces el error de muestreo podría estimarse con una sola muestra. El error de muestreo en un estimador es generalmente expresado en términos de la varianza del estimado o su raíz cuadrada, el error estándar, que es una medida que provee al usuario una mejor comprensión de la medida de la precisión del estimador. Estas medidas se basan en la suma promedio de los cuadrados de las diferencias entre cada valor de las partidas reportadas y el valor promedio para esa partida.

Asumiendo que para cada municipio, las fincas CML en la lista que no respondieron y que estaban dentro de una misma estrata CML, tenían características similares a las fincas que completaron el formulario, también podemos aplicarles los resultados de teoría de muestreo para demostrar la varianza generada por la

contingent. For that estimate, we compute CML-based nonresponse variance for a stratified sample, assuming that when there is nonresponse within a nonresponse stratum, the respondents represent a random sample of all eligible farms, responding or nonresponding, within that stratum, and that, within municipio and stratum, response is independent of the response occurring within any other municipio and stratum combination. In our Census, nonresponse can occur only for the non-self-representing CML cases. Under the assumptions of random and independent nonresponse, we sum the individual CML variances for an estimate, from each combination of municipio and stratum, to obtain the total nonresponse variance of that estimate at Island level. For each estimate, the CML-based nonresponse variance is added to the sampling variance for that estimate from the NML contingent, to obtain the total variance of the estimate for the 2007 Puerto Rico Census of Agriculture.

The figures presented in Table B-1 are estimates of the percent relative standard error of selected quantities for Puerto Rico. The percent relative standard error is equal to the standard error of the estimate divided by the estimate itself, multiplied by 100. The estimates of relative standard error displayed in table B-1 were derived to reflect the total sampling error, which comes from Area Sampling, and also, under our assumptions, from the nonresponse process operating on CML records.

The particular sample selected for this census is one of many different samples, which could have been chosen. Had many different samples been taken - that is, had sampling been performed repeatedly under the same general conditions with the same sample design - many different estimates of each of the quantities being estimated would have been obtained. It is in this scenario that the following general rules about the standard error apply.

1. In repeated sampling, approximately 90% of all estimates obtained from the different samples would fall within 1.65 standard errors of the actual value

no respuesta en la parte del estimado proveniente del contingente de CML. Para este estimado, calculamos la varianza para una muestra estratificada basándonos en el CML, asumiendo que cuando existe la no respuesta dentro de una estrata de no respuesta, el entrelistado representa una muestra aleatoria de todas las fincas elegibles, respuesta o no respuesta, dentro de la estrata, y esta respuesta, dentro del municipio y la estrata, es independiente de la respuesta que ocurre dentro de la combinación de cualquier otro municipio y la estrata. En este Censo, la no respuesta solo puede darse para los casos de CML que no son representativos de por si. Bajo esta premise, de no respuesta independiente y aleatoria, sumamos la varianza individual de CML para un estimado, de cada combinación de municipio y estrata, para obtener el total de varianza de no respuesta del estimado a nivel de Isla. Para cada estimado, se sumo la varianza de no respuesta basada en la lista, a la varianza de muestra para este estimado en el contingente de la muestra NML, para obtener la varianza total del estimado para el Censo de Agricultura de 2007 en Puerto Rico.

Las cifras presentadas en la Tabla B-1 son estimados del por ciento relativo de error estándar para ciertas cantidades en Puerto Rico. El por ciento relativo de error estándar es igual al error estándar del estimador, dividido por el estimado y multiplicado por 100. Los estimados del error estándar relativo presentados en la tabla B-1 fueron derivados para reflejar el error de muestreo, que resulta del muestreo por área, y también, bajo lo asumido, del proceso de no respuesta que opera en los récords CML.

La muestra de segmentos seleccionada para este Censo de Agricultura es una de varias muestras que pudieron haberse usado. Si se hubiesen usado varias muestras diferentes- esto es, si el muestreo se hubiese ejecutado repetidamente bajo las mismas condiciones con el mismo diseño de muestra - pudieron haberse obtenido muchos estimados diferentes de cada una de las cantidades estimadas. Es bajo estas condiciones que aplicamos las siguientes reglas generales sobre el error estándar.

1. En el muestreo repetitivo, aproximadamente un 90% de todos los estimados obtenidos de las diferentes muestras caerán dentro de un error estándar de 1.65 del

which would have been obtained had a complete canvass or enumeration been performed, with no error present in the estimates.

2. In repeated sampling, approximately 95% of all estimates obtained from the different samples would fall within 1.96 standard errors of the actual value which would have been obtained had a complete canvass or enumeration been performed, with no error present in estimates.

Use of these rules can be illustrated with the following example from Table B-1. Our estimate of total Island-level production expenditures for the purchase of gasoline and other fuel and oil products was \$17,568,756; the relative standard error obtained for this estimate was 3.5% (or .035). The standard error of the estimate is then $.035 * \$17,568,756$, or \$614,906.46 (rounded to three decimal places). We may construct an interval that ranges from 1.65 standard errors below the estimate to 1.65 standard errors above the estimate: from \$16,554,160, the interval's lower limit, or $(\$17,568,756 - 1.65 * \$614,906.46)$, to its upper limit of \$18,583,352, computed as $(\$17,568,756 + 1.65 * \$614,906.46)$, rounding to the nearest unit.

This interval is called a 90% confidence interval. The importance of this interval is given in the following statement: in repeated sampling, under the same general conditions, 90% of all such intervals constructed using the estimate and the relative standard error obtained for each sample selected would contain the actual value of agricultural chemicals purchased.

Similarly, a 95% confidence interval may be constructed: from a lower limit of \$16,363,539, or $(\$17,568,756 - 1.96 * \$614,906.46)$ to an upper limit of \$ 18,773,973, computed as $(\$17,568,756 + 1.96 * \$614,906.46)$, with rounding to the nearest unit.

The estimates of relative standard error displayed in table B-1 are for Island-level estimated totals and farm counts and are a measure only of the sampling error of the estimates; they do not account for error in the estimates due to other sources of error, called

valor actual obtenible mediante una enumeración completa, sin error alguno en los estimados.

2. En el muestreo repetitivo, aproximadamente un 95 % de todos los estimados obtenidos de las diferentes muestras caerán dentro de un error estándar de 1.96 del valor actual obtenible mediante una enumeración completa, sin error alguno en los estimados.

El uso de estas reglas puede verse en la Tabla B-1. Asumiendo que el estimado total de gastos por compra de gasolina, aceite y otros combustibles para uso de las fincas en Puerto Rico es \$17,568,756; asumiendo también que el error estándar relativo de este estimado es 3.5 por ciento (.035). El error estándar del estimado es entonces $(.035 * \$17,568,756)$, o \$614,906.46. Podríamos construir un intervalo que se extendiera desde un error estándar de 1.65 bajo el estimado hasta 1.65 sobre el estimado: \$16,554,160 $(\$17,568,756 - 1.65 * \$614,906.46)$ hasta \$18,583,352 $(\$17,568,756 + 1.65 * \$614,906.46)$ redondeados a la unidad más cercana.

A este intervalo se le llama intervalo de confianza de 90 por ciento. La importancia de este intervalo se explica en la siguiente aseveración: en muestreo repetitivo, bajo las mismas condiciones, un 90 por ciento de todos los intervalos construidos usando los estimados y el error estándar relativo obtenido por cada muestra seleccionada contendrá el valor actual de los productos químicos agrícolas comprados.

De la misma forma, puede construirse un intervalo de confianza de 95 por ciento: \$16,363,539 $(\$17,568,756 - 1.96 * \$614,906.46)$ a \$18,773,973 $(\$17,568,756 + 1.96 * \$614,906.46)$ redondeados a la unidad más cercana.

Los estimados del error estándar relativo presentados en la tabla B-1 son solo para totales estimados a nivel de Isla y totales de fincas y son una medida del error de muestreo en el estimado; no toman en cuenta por errores en el estimado debido a otras fuentes, llamados

nonsampling error. The following sections discuss different types of nonsampling errors that may be present in the estimates.

CENSUS NONSAMPLING ERROR

Many other types of error are not caused by sampling and are not so easily quantified: *nonsampling error* may be introduced into the census through incorrect responses by reporters, or through mistakes made by programmers or data entry staffers, as in inappropriate adjustment of missing data. Nonsampling error due to mail list incompleteness and duplication or misclassification of records on the mail list is called coverage error, discussed in greater detail below. The purpose of the area frame is to eliminate nonsampling error associated with CML incompleteness, coverage error, but not nonresponse.

RESPONDENT AND ENUMERATOR ERROR

Incorrect or incomplete responses to the census report form or to the questions posed by an enumerator can introduce error into the census data. To reduce reporting error, detailed instructions for completing the report form were provided to each respondent. Questions were phrased as clearly as possible, based on previous tests of the report form. In addition, each respondent's answers were checked for completeness and consistency by the complex edit program.

ITEM NONRESPONSE

Nonresponse to particular questions on the census report form, which we would logically or statistically expect to be present, may result in a type of nonsampling error. When information reported for another farm with ostensibly similar characteristics is used to impute for item nonresponse, the final estimates may be biased, if the unobserved characteristics of the nonrespondents differ systematically from those reported by respondents at either the record level (for an individual farm operation) or collectively, if respondent records are averaged or totaled.

errores no debidos a la muestra. Las secciones siguientes discuten otros tipos de errores no de muestreo y que podrían estar presentes en el estimado.

ERRORES DE CENSO AJENOS AL MUESTREO

El otro tipo de error en un censo ocurre como resultado de obtener respuestas incorrectas, de errores cometidos por las personas entrando los datos o de ajustes inapropiados cuando las partidas están en blanco. A estos se les conoce como errores ajenos al muestreo. También pueden surgir de la insuficiencia y duplicación de la lista, así como de clasificación errónea de los récords, y se les conoce como errores de cobertura, se discute con mayor detalle más adelante. Más aún, el marco muestral de área se usó para eliminar errores ajenos al muestreo asociados con la insuficiencia de la lista (CML) errores de cobertura, pero no por la no respuesta.

ERROR DEL ENUMERADOR O DEL ENTREVISTADO

Respuestas incorrectas o incompletas en el formulario censal o las preguntas formuladas por el enumerador podrían introducir error en la información censal. Para reducir este error, se dieron instrucciones detalladas de como completar el formulario a cada uno de los agricultores. Las preguntas fueron fraseadas lo más claramente posible basándonos en evaluaciones anteriores del formulario. Además, cada respuesta fue cotejada para verificar su integridad y consistencia.

PARTIDAS EN BLANCO

Partidas en blanco en ciertas preguntas del cuestionario, para las cuales se esperaba una respuesta, pueden causar errores ajenos al muestreo. Cuando imputamos los datos para la partida en blanco usando información recopilada para otra finca con características similares, el estimado final puede estar sesgado. Esto resulta porque las características omitidas por los casos que no han sido reportados, pueden ser diferentes de las características reportadas por los operadores, lo mismo al nivel de elemento (operación de finca individual) o en el promedio o total.

PROCESSING ERROR

All phases of processing a census report form may introduce nonsampling error into the estimates. The processing of census report forms includes clerical screening for farm activity, computerized check-in of report forms and follow-up of nonrespondents, keying and transmittal of completed report forms, computerized editing of inconsistent and missing data, review and correction of individual records referred from the computer edit, review and correction of tabulated data, and electronic data processing. These operations undergo a number of quality control checks to ensure as accurate an application as possible, yet some errors may ultimately escape detection.

COVERAGE ERROR

Farms incorrectly excluded or included in the census result in coverage errors. Possible coverage errors include the following:

Overcount due to farms duplicated or enumerated more than once.

Undercount due to farms misclassified as nonfarms.

Overcount due to nonfarms misclassified as farms.

These coverage errors were minimized with computerized duplication identification programs and careful enumeration of all land area within a sampled area segment. In addition, field enumerators were required to follow up CML cases identified as major agricultural operations, if they did not respond by mail. If all follow-up attempts failed, either the NASS survey database, the census historic database, or other more current sources were used to impute data for the record.

ERROR DE PROCESAMIENTO

Todas las fases del procesamiento para cada formulario del censo son fuentes para la introducción de errores ajenos al muestreo. El procesamiento de las formas censales incluye una revisión clerical de las actividades agrícolas, el registro computadorizado del recibo de los formularios y el seguimiento de los casos que no han respondido, la entrada y transmisión electrónica de los datos en archivos computadorizados; la depuración computadorizada de la información inconsistente u omitida, el repaso y corrección de casos con errores rechazados por el proceso de depuración de datos, un análisis y corrección de los datos tabulados; y el procesamiento electrónico de los datos. Estas operaciones conllevan un número de cotejo de control de calidad para asegurar que la aplicación sea lo más precisa posible. Sin embargo, varios de estos errores pueden escaparse sin ser detectados.

ERROR DE COBERTURA

Fincas incorrectamente excluidas o incluidas en el censo resultan como error de la cobertura. Los siguientes son posibles errores de cobertura:

Cifras exageradas debido a duplicidad o enumeración de fincas por más de una vez.

Cifras subestimadas debido a fincas erróneamente clasificadas como no fincas.

Cifras exageradas debido a no fincas erróneamente clasificadas como fincas.

Estos errores de cobertura se minimizaron a través de una revisión ejecutada por las computadoras para identificar y eliminar récords duplicados y una enumeración somera de todo terreno agrícola dentro de cada segmento en la muestra. Además, los enumeradores visitaron agricultores de CML identificados como de una gran producción agrícola y que no respondieron a través del correo. Si después del seguimiento no hubo respuesta, se utilizó o la encuesta del NASS o la fuente de datos del censo anterior o otras fuentes más recientes para llenar los datos del record.

BIAS AND PRECISION OF THE ESTIMATES

The nonsampling errors discussed above can be sources of bias leading to underestimates or overestimates, based on the actual processed data. Since the formulas used to calculate estimates of relative standard error do not account for the nonsampling error sources, the figures published in the following tables may also be somewhat biased. It is important to note that the largest portion of the variability in the estimates probably comes from sampling error; it is the relative standard error of the estimates due to this source of error that is reported in the following table.

SESGO Y PRECISION DE LOS ESTIMADOS

Los errores ajenos al muestreo discutidos anteriormente, pueden ser fuentes de sesgo que podrían causar que los datos publicados contengan estimados por debajo o sobre lo normal. Como las fórmulas usadas para calcular estimados del error estándar relativo no son responsables por los errores ajenos al muestreo, las cifras publicadas podrían estar un tanto sesgadas. Es importante tomar en cuenta que la porción mayor de variabilidad en los estimados proviene de errores de muestreo. Las tablas siguientes contienen el error estándar relativo de los estimados procedente de estas fuentes.

Table B-1. Estimates of Precision for Puerto Rico Totals: 2007

[For meaning of abbreviations and symbols, see introductory text.]

Item	All Farms	Relative Standard Error of Estimate
FARMS AND LAND IN FARMS		
Farms	15,745	1.7
Land in farms	557,530	3.4
Average size of farm	35.4	3.8
Farms by size:		
Less than 10 cuerdas	7,502	4.1
farms		
cuerdas	28,405	5.5
10 to 19 cuerdas	3,545	6.8
farms		
cuerdas	49,322	7.0
20 to 49 cuerdas	2,680	7.2
farms		
cuerdas	78,496	6.7
50 to 99 cuerdas	865	10.0
farms		
cuerdas	58,741	9.0
100 to 174 cuerdas	524	9.6
farms		
cuerdas	70,444	10.2
175 to 259 cuerdas	207	5.6
farms		
cuerdas	43,195	5.5
260 cuerdas or more	422	9.3
farms		
cuerdas	228,926	8.0
LAND IN CROPLAND		
Cropland	14,074	2.1
farms		
cuerdas	392,726	3.9
Farms by cuerdas in cropland:		
Less than 10 cuerdas	9,847	3.1
10 to 19 cuerdas	2,714	8.0
20 to 49 cuerdas	1,668	8.0
50 to 99 cuerdas	661	9.7
100 to 174 cuerdas	408	10.6
175 to 259 cuerdas	178	14.8
260 cuerdas or more	262	11.4
TENURE OF OPERATOR		
Full owners	11,402	2.7
farms		
cuerdas	287,231	5.4
Part owners	1,918	6.3
farms		
cuerdas	159,251	7.8
Tenants	2,425	9.6
farms		
cuerdas	111,048	6.2
PRINCIPAL OCCUPATION OF OPERATOR		
Agricultural	5,456	4.1
Nonagricultural	8,947	3.4
Not reported	1,342	7.5
FARMS BY TYPE OF ORGANIZATION		
Individual or family	13,958	2.1
farms		
cuerdas	399,977	3.6
Partnership	49	17.6
farms		
cuerdas	7,699	12.6
Corporation	575	11.1
farms		
cuerdas	108,615	12.0
Other	1,163	13.7
farms		
cuerdas	41,240	9.0
FARMS BY MARKET VALUE OF AGRICULTURAL PRODUCTS SOLD		
Less than \$1,000	4,442	6.6
\$1,000 to \$2,499	2,771	8.0
\$2,500 to \$4,999	2,428	8.8
\$5,000 to \$7,499	1,206	11.5
\$7,500 to \$9,999	882	14.0
\$10,000 to \$19,999	1,497	9.1
\$20,000 to \$39,999	1,030	7.9
\$40,000 to \$59,999	281	8.4
\$60,000 or more	1,208	5.6
MARKET VALUE OF AGRICULTURAL PRODUCTS SOLD		
Total sales	15,745	1.7
farms		
dollars	515,685,532	2.3
Average per farm	32,752	2.8
dollars		
Crops including nursery and greenhouse, and other agricultural products		
farms	10,206	2.8
dollars	218,835,477	3.0
Coffee	5,678	2.8
farms		
dollars	41,823,565	8.8
Pineapples	53	28.0
farms		
dollars	526,908	12.3
Plantains	3,756	6.3
farms		
dollars	44,874,655	5.8
Bananas	1,980	7.8
farms		
dollars	10,082,177	5.5
Grains	969	16.1
farms		
dollars	1,553,126	13.0
Root crops or tubers	1,691	11.7
farms		
dollars	6,682,651	11.8
Fruits and coconuts	2,350	7.0
farms		
dollars	19,903,773	2.7
Vegetables or melons including hydroponic crops	1,007	12.7
farms		
dollars	35,440,431	8.5

--continued

Table B-1. Estimates of Precision for Puerto Rico Totals: 2007 - Con.

[For meaning of abbreviations and symbols, see introductory text.]

Item	All Farms	Relative Standard Error of Estimate
MARKET VALUE OF AGRICULTURAL PRODUCTS SOLD - Con.		
Crops including nursery and greenhouse, and other agricultural products - Con.		
Nursery, greenhouse, floriculture, and sod	farms 524	13.3
	dollars 44,575,883	1.5
Grasses and other crops	farms 256	16.1
	dollars 13,372,308	8.1
Livestock, poultry, and their products	farms 5,662	4.9
	dollars 296,850,055	3.2
Cattle and calves	farms 3,568	6.3
	dollars 33,005,235	4.6
Poultry and poultry products	farms 1,079	14.7
	dollars 63,573,676	1.7
Dairy products	farms 354	5.3
	dollars 184,542,585	2.2
Hogs and pigs	farms 1,075	14.2
	dollars 6,238,744	6.8
Aquaculture	farms 40	14.8
	dollars 832,725	4.5
Other livestock and livestock products	farms 787	12.5
	dollars 8,657,090	5.1
OTHER FARM-RELATED INCOME		
Custom farm work done for others	farms 441	20.6
	dollars 4,575,759	3.9
Agritourism and recreational services	farms 30	47.0
	dollars 524,203	41.5
Rent from farmland	farms 292	16.8
	dollars 2,157,351	5.5
Sale of farm by-products	farms 23	17.3
	dollars 361,125	27.9
Participation in government farm programs	farms 2,295	5.0
	dollars 30,171,304	11.6
Other farm-related income	farms 261	33.5
	dollars 4,464,755	15.2
PRODUCTION EXPENSES		
Livestock, poultry, and fish purchased	farms 4,627	5.6
	dollars 41,604,668	4.4
Feed purchased for livestock, poultry, and fish	farms 7,409	4.1
	dollars 115,243,997	3.4
Medicines and drugs purchased for livestock and poultry	farms 6,742	4.5
	dollars 8,877,191	3.6
Veterinarian services	farms 789	12.2
	dollars 2,292,929	1.8
Professional services	farms 2,257	5.7
	dollars 4,509,397	2.6
Seeds, bulbs, plants, and trees purchased	farms 8,696	3.3
	dollars 22,330,647	4.0
Commercial fertilizer purchased	farms 6,436	4.2
	dollars 10,059,841	5.1
Gasoline and other fuel and oil products purchased	farms 6,481	3.8
	dollars 17,568,756	3.5
Wages and salaries of hired farm workers and family members	farms 7,705	3.5
	dollars 107,786,397	3.5
Contract labor	farms 725	12.9
	dollars 3,730,959	30.2
Machine hire and customwork	farms 2,459	8.0
	dollars 6,311,472	3.9
Agricultural chemicals purchased	farms 5,013	4.7
	dollars 6,723,989	3.5
Machinery repair and maintenance expenses	farms 2,487	4.2
	dollars 9,299,979	2.5
Building repair and maintenance expenses	farms 1,006	7.5
	dollars 5,944,376	4.2
Expenses for water	farms 1,072	9.2
	dollars 3,439,116	33.9
All other expenses	farms 7,758	3.8
	dollars 72,049,313	3.8
HIREF FARM WORKERS, AGREGADOS, AND SHARECROPPERS		
Workers who worked		
5 months or more	farms 5,416	4.5
	number 18,392	4.5
Less than 5 months	farms 3,485	5.1
	number 11,249	3.8
Agregado and sharecropper families living on place		
December 31, 2007	farms 1,024	8.9
	number 2,067	6.0
INVENTORY OF LIVESTOCK, POULTRY, AND THEIR PRODUCTS		
Cattle and calves	farms 4,671	5.5
	number 267,542	3.2
Hogs and pigs	farms 1,513	12.5
	number 69,892	5.5
Horses	farms 2,883	8.8
	number 15,737	8.4
Sheep	farms 568	16.2
	number 11,137	11.0
Goats	farms 664	18.5
	number 7,359	6.4

--continued

Table B-1. **Estimates of Precision for Puerto Rico Totals: 2007 - Con.**

[For meaning of abbreviations and symbols, see introductory text.]

Item	All Farms	Relative Standard Error of Estimate
INVENTORY OF LIVESTOCK, POULTRY, AND THEIR PRODUCTS - Con.		
Rabbits	480	23.5
..... farms		
..... number	29,145	17.4
Pullets for replacement	71	11.0
..... farms		
..... number	308,110	0.1
Layers	516	20.5
..... farms		
..... number	642,279	0.2
Poultry for meat	456	19.5
..... farms		
..... number	5,176,402	2.1
Roosters	1,802	10.4
..... farms		
..... number	67,260	11.4
CROPS HARVESTED		
Coffee grown in the shade	2,627	4.3
..... farms		
..... cuerdas	11,807	3.0
..... cwt	42,684	3.5
Coffee grown without shade	3,808	3.8
..... farms		
..... cuerdas	26,728	2.5
..... cwt	137,910	2.8
Pineapples	124	36.6
..... farms		
..... cuerdas	330	7.0
..... tons	2,145	9.1
Plantains	4,569	5.7
..... farms		
..... cuerdas	18,034	6.4
..... thousands	249,948	6.2
Bananas	2,196	7.8
..... farms		
..... cuerdas	7,015	3.7
..... thousands	373,667	5.0