

DATA-PROCESSING PROCEDURES

THE PROCESSING PLAN

It had been estimated originally that 2,300,000 persons and approximately 530,000 housing units would be enumerated in Puerto Rico, 25 percent of which would fall in the sample.

The availability of these and other more detailed estimates of the population and housing were necessary for the development of the census processing plan.

Once the framework of the census enumeration was known, as well as the type of questionnaires to be used (since they were the source documents for the data processing) and the specifications for the tables to be published (the end product of the data processing), it was then possible to develop a set of procedures, forms, and other pertinent material for processing the data according to a plan designed to produce the required statistics in the most efficient and economical manner. Instructions and forms were prepared in both English and Spanish.

The development of the overall processing plan called for the preparation of procedures and instructions for use in the following steps in the data processing:

- Receipt, check-in, distribution, and control of the questionnaires.
- Geographic identification coding.
- Matching and transcribing information from supplemental forms.
- Block check (inspection of the enumeration and numbering of blocks).
- "General" coding.
- Industry and occupation coding.
- Key punching (preparing punchcards).
- Quality control:
 - Verification of general coding.
 - Verification of industry and occupation coding.
 - Verification of card punching.
- Mechanical recoding.
- Ratio estimating.
- Tabulation.
- Editing, adding, balancing, and inflation of sample data, as well as posting figures from the machine tabulations to table worksheets.
- Verification, calculation, and review of tables.

DISTRIBUTION AND CONTROL OF WORK

The flow of work from one operation to another was subjected to a systematic control designed to maintain uninterrupted continuity in the work and enable the supervisors to ascertain the status of the operations at any given moment. A reporting system which included the use of individual daily production records was developed so that progress of the operations could be reviewed periodically and estimates of expected production as well as of staffing or time schedules could be made.

In addition, the cover page of the enumeration book was designed in such a way that space was provided for the clerks engaged in the different phases of the data processing to enter the kind of work they had just completed, their name, and the date of completion. A similar label was pasted on the portfolios used for handling the sample questionnaires.

RECEIPT AND CHECK-IN OF QUESTIONNAIRES

The first data-processing operation was receiving and inspecting the questionnaires sent to the Operations Office by the field offices. Each enumerator had carried the books of questionnaires in a specially designed plastic portfolio. These portfolios containing the questionnaires and other material used by the enumerator were sent to the Operations Office where they were checked to make sure that the contents were complete.

The Procedure

A control was established to make certain either that a portfolio with the corresponding completed questionnaires was received for every ED, or that an explanation had previously been submitted to the Operations Office to account for exceptions. Examples of such exceptions were ED's which did not have any population and ED's which had been combined with others.

The procedure for the receipt and check-in was based on the use of form F-281PR, the Minor Civil Division (MCD) Sheet (see example in appendix) which provided a listing of the ED's in each of the 76 municipios in Puerto Rico. This form identified the minor civil divisions or smaller geographic areas, in alphabetical order, and within each area the ED's into which it had been divided were listed.

Discrepancies between the materials received from the enumerators and the MCD Sheet listings were resolved either by reference to maps or by referring the problem, together with an explanation of the error, back to the field office for review and correction.

Results of the Operation

The receipt and check-in of the questionnaires for the 2,966 ED's in Puerto Rico required 612 man-days and cost \$6,527. Of an approximate total of 150,000 sample questionnaires, only 25, or less than 0.2 percent, were not accounted for.

By May 13, 1960, the Operations Office had received from the field offices the portfolios containing the questionnaires for all the 76 municipios. The receipt and check-in process was then well underway, and the next phase in the operations could be started with the material that was ready.

GEOGRAPHIC IDENTIFICATION/CODING

Since the census publications were to provide the data pertaining to the characteristics of the population and housing within the many specific geographic areas of Puerto Rico, the geographic identification of the data was a carefully planned phase of the processing operations designed to fit into the overall plan of publication.

The basic approach to the geographic identification was the development of a code scheme for identifying each individual ED by geographic area so as to differentiate the various areas for which data were to be published. In Puerto Rico these areas were: (1) the political subdivisions which are the municipios and cities; (2) the towns,

villages, barrios, subbarrios; and (3) the standard metropolitan statistical areas (SMSA's), urbanized areas, and census tracts. It was also necessary to distinguish urban areas from rural areas.

Geographic identification coding was based on the MCD Sheet on which each ED had been listed in geographic sequence. A 14-digit code was assigned to each ED, as follows: The 1st digit was used to identify Puerto Rico as one of the outlying areas. The 2nd and 3rd digits were used to identify the municipio. The 4th, 5th, and 6th digits were used to identify the barrios or subbarrios within each one of the municipios. The 7th through the 10th digits were used to identify the ED's located in an SMSA and census tract or outside the SMSA. The 11th through the 14th digits were used to identify each ED and to indicate a split ED.

After completion of the identification coding, the codes were transferred to a master set of punchcards which were used to provide the geographic identification for each of the data cards punched for the census.

The geographic identification operations took only four man-days for coding, verification of coding, and the key-punching of the identification information.

BLOCK CHECK

Since this was the first time a Puerto Rico census program called for preparation of statistics for city blocks, the work done by the enumerators with respect to the assignment of the block numbers to the census questionnaires required an unusually detailed and thorough examination.

The procedures for checking the correctness of the block enumeration required that the enumerators' block number entries on the questionnaires and listing books be checked against the ED maps to determine whether there were any missing block numbers, extra block numbers, or erroneous block number assignments. Where discrepancies were noted which could not be resolved, a field check was made so that the block numbers could be correctly assigned.

Because the ED maps for some of the areas requiring block enumeration had not been delivered to the field in time to instruct the crew leaders in all of the technical requirements of this phase of the enumeration, there had been some improvisation. For example, it was found during the block check operation that in some areas enumerators had ignored the block numbers on the ED maps and assigned their own numbering system. This complicated the review operation and resulted in a considerable number of changes. A total of 795 ED's were included in the block check programs, or about 23 percent of the total ED's for Puerto Rico.

The block check was a slow and time-consuming operation because of the many documents which had to be carefully reviewed during the process and the precision required in correcting discrepancies between the numbering of blocks during enumeration and the block numbers on the maps. It was found that for the most part the assignment of block numbers during the enumeration had been properly done. It took approximately 90 man-days to complete the block check operation.

TRANSCRIBING INFORMATION FROM SUPPLEMENTAL ENUMERATION FORMS

A number of supplemental forms were used before and during the Puerto Rico census enumeration in order to simplify the work of the enumerator and also to enumerate (1) persons away from home on the census date or not found at home after repeated visits, (2) persons living in military installations, and (3) persons who were members of the crews of vessels docked at the various ports of the island on the census date. All of the information reported on these supplemental forms was transcribed to the appropriate questionnaires.

Advance Census Reports

Questionnaires called Advance Census Reports were distributed to households by mailmen in advance of the census, to be filled in and given to the enumerator when he visited the household during the enumeration. Form 60PH-6PR (see appendix) was mailed to households in the three "block cities" and form 60PH-5PR to households elsewhere on the island. Usually, the enumerator transcribed the information from these Advance Census Reports to the regular census questionnaires. However, some people, for example, persons who expected to be away from home during the enumeration, mailed their Advance Census Reports to the census field office in their district; for most of these, transcription was done in the field offices. Some Advance Census Reports were not received by the field offices until after the corresponding enumeration books had been sent to the Operations Office. These forms were sent to the Operations Office for subsequent transcription.

Individual Census Reports

The Individual Census Reports (form 60PH-10PR) were supplied to the enumerator and were completed by him whenever he encountered a visitor in a private home who was fairly certain that no one at his home address would supply the information for him. These forms were sent to the census field office of the area in which the visitor lived, where the information was to be transcribed to the appropriate enumeration book, provided that a search of the enumeration book indicated that he had not been enumerated. While it was originally intended that the name search and any required transcription be done in the field offices, this phase of the work was actually carried out in the Operations Office.

Reports for Military and Maritime Personnel

Persons residing in military installations in Puerto Rico and crews of vessels docked in ports of the island were enumerated on a questionnaire called the Report for Military and Maritime Personnel, form 60PH-13PR. After the enumeration, these questionnaires also were sent to the Operations Office for transcription. For 75 percent of the forms, the information reported was transcribed to the 100-percent questionnaires, and for the remaining 25 percent, to the sample questionnaire. The crews of vessels were allocated to the ED's which included the port areas at which the vessels were docked on the census date, April 1, 1960.

CODING

After completion of the block check and supplemental forms operations, it was possible to begin editing and coding of schedule content. As mentioned earlier, the major portion of the questions on both the basic and the sample questionnaires provided for precoded answers, i.e., the possible responses were listed on the questionnaire and the enumerator simply marked a coded check box to indicate the reply of the respondent. An example of this is the item on relationship to the head of the household, shown below. Marking any one of the first three entries automatically supplied the proper code. (Subsequent entries for this item were coded by the clerks since various answers were possible.)

- Head 1
- Wife of head 2
- Son or daughter of head 3
- Other (specify)

Precoding was used to the maximum extent possible because of the gains in efficiency it introduced. Codes were assigned manually to all items which could not be adapted to precoding.

The purpose of the codes, whether assigned by precoding or by manual coding, was to put the entries made by the enumerators on the questionnaires into a form which could be transferred into punched entries on a punchcard.

The manual coding instructions also called for examining certain population and housing entries and correcting them if they were inconsistent with respect to other entries shown. An illustration of such an inconsistency would be a wife reported as head of the household with her husband being reported as wife. (A more elaborate edit for errors and inconsistencies was performed, using mechanical equipment, prior to the tabulation operation.) The manual coding operations were divided into two major parts, "general" coding and industry and occupation coding.

General Coding

The manual coding of all questionnaire items requiring the assignment of a code except the industry and occupation items was referred to as general coding. The general items which were coded or examined were the following:

100-Percent Questionnaire

<u>Coded</u>	<u>Examined</u>
P-3 Relationship	P-4 Sex
P-7 Place of birth	P-6 Marital status
	H-3 Number of persons or vacant
	H-12 Tenure
	H-13 Vacancy status
	H-16 Rent

Sample Questionnaire

<u>Coded</u>	<u>Examined</u>
P-3 Relationship	P-4 Sex
P-6 Marital status	P-18 Born before or after 1946
P-7 Place of birth	P-23 When last worked
P-10 Migration	H-3 Number of persons or vacant
P-25 Place of employment	H-12 Tenure

P-29 Individual income	H-13 Vacancy status
P-30 for 1959 and the	H-16 Contract rent
P-31 total of that income—also total family income	H-33 and totaling of gross rent

In addition (on both 100-percent and sample questionnaires), group quarters were coded to differentiate between institutions, such as jails, hospitals, etc., and other types of group quarters such as boarding houses, military installations, etc. The persons enumerated in group quarters were coded to indicate their status as inmates, residents of institutions, lodgers, etc.

The specific codes to be assigned for each of the items to be coded or examined, and a detailed explanation of the coding scheme and its use, were furnished in an instruction manual for coding. To simplify the work of the coders, a separate code card was prepared which provided in a convenient summary form a list of the items to be coded and the codes which could be assigned (see appendix). Some entries, however, could not be coded by the clerks without first being submitted for professional review. For these a problem referral slip, form No. 60-50-5, was designed. When the coder encountered entries which he could not code, he described the problem on this form and referred it to a more highly trained Technical Assistant for decision. Unusual entries and consistent patterns of errors were submitted to the appropriate subject-matter personnel in Washington for study so that proper remedial action could be taken.

Training of the first group of general coders started on May 23, 1960, and as the flow of work into the coding operation increased, additional groups of clerks were hired and trained. A 3-day training and exercise period was found to be sufficient. Questionnaires were assigned to coders on an ED basis. The 100-percent questionnaires for an ED were coded first, after which the sample questionnaires were coded. Each household was coded as a unit, i.e., all coding required for one household was completed before going on to the next. An exception to this rule was made when a coder encountered problem entries which he was unable to code. If there were a substantial number of problems, the entire ED was referred to the Technical Assistant. If there were only a few problems, the coder was instructed to fill out problem referral slips for the uncoded entries and complete the remainder of the ED. All coders' work was verified in order to control the quality of the work (see "Quality Control," below).

A total of 30 persons, including the supervisor and two Technical Assistants, were employed in the general coding operation. Questionnaires were coded at a rate of 1,900 enumerated persons per man-day. The operation required 1,459 man-days, and was completed the week ending October 7, 1960.

Industry and Occupation Coding

Because special training is required to translate into meaningful codes a respondent's replies relating to his occupation and industry, the coding of these items is normally handled as a separate operation. It was so handled in Puerto Rico.

A 3-digit code was assigned for each industry entry and a 3-digit code for each occupational entry. Codes were assigned by reference to an alphabetical index of occupations and industries, and to company name lists

which listed the industry codes which had been assigned to the major industrial establishments in Puerto Rico. These were supplemented by an instruction manual which provided the coders with the rules to be followed in determining the proper code for industry and occupation entries. Problem referral slips were also used by the industry and occupation coders when they encountered entries which required referral to the Technical Assistants.

Since the industry and occupation items appeared only in the sample questionnaires, the sample pamphlets were separated from the 100-percent enumeration books upon completion of the general coding operation, and were placed in a new set of ED portfolios. To edit and code industry and occupation items in these questionnaires, nine clerks were hired and trained. A supervisor was selected from among the best of the general coders. Training was started on June 16, 1960, and required approximately 4 days. A subject-matter specialist was assigned from Washington headquarters of the Bureau to act as the Technical Assistant to solve the more difficult problems of this operation.

The clerks coded industry and occupation entries for about 1,600 enumerated persons in the sample per day. It took approximately 622 man-days to complete the operation. The fact that the coding was done in Puerto Rico made it possible for the Technical Assistant to obtain information directly from local government agencies and private industrial and business enterprises when necessary in order to determine proper classifications for industries and occupations for which the alphabetical index of occupations and industries did not provide specific entries and codes.

Industry and occupation coding was verified using a scientifically formulated system of quality control. (See "Quality Control," below.)

KEYPUNCHING

The keypunching operation involved the transfer of the codes assigned to the various items on the questionnaires to IBM punchcards by punching holes in the appropriate positions on the punchcard. A housing card was prepared for each housing unit; a population card was prepared for each person.

The Procedure

To achieve the objectives of the tabulation program, three different kinds of punchcards were required—one for the population data and two for the housing data (see appendix). The two housing punchcards were later consolidated into one. The reason that two different punchcards were required for the housing information arose from the fact that the arrangement of the housing questions on the 100-percent questionnaire was different from the arrangement on the sample questionnaire. (The difference had been decided upon in order to make each type of questionnaire a more effective instrument of enumeration.) As a result, it was found to be simpler and more economical to punch the 100-percent housing data from the 100-percent questionnaires to a housing card specifically designed for this purpose and then, as a subsequent operation, to reproduce the data (rearranging them as required) into the format used for punching housing data from the sample questionnaires. (After checking for correctness of the

reproduction operation, the original housing deck of cards was destroyed.) The 100-percent population questions were arranged in the same sequence on both the 100-percent and the sample questionnaires, so only one type of punchcard was needed for the population data.

Each keypunch operator was provided with an instruction manual which set forth in detail the manner in which each item on the questionnaire was to be punched, the code symbols used, and the punchcard columns which corresponded to the questionnaire items. The first 14 columns on all three types of punchcards were used only for geographic identification which was duplicated by machine from the ED Master Card (see below). The remaining items on the card were punched directly from the questionnaires.

The Operation

Training of the keypunch operators began the first week of June 1960. The first 3 or 4 days were dedicated to special training for manual dexterity since, in general, the persons hired for this operation had had no keypunching experience. Approximately 10 days were spent on practice punching.

The first assignment made to the keypunch unit after completion of training was that of punching the ED geographic identification into an ED Master Card from the coded Minor Civil Division Sheets. Then punchcards for the 100-percent questionnaires were punched and, finally, after completion of the industry and occupation coding, punchcards for the sample questionnaires were punched. Material for an ED was maintained in a work unit during the punching and verification of punching operations.

The punching of the census information was carried out in four separate punching operations. During peak production all four operations were being conducted as follows:

- a. The 100-percent population information from the 100-percent questionnaires.
- b. The 100-percent housing information from the 100-percent questionnaires.
- c. The 100-percent and sample population information from the sample questionnaires.
- d. The 100-percent and sample housing information from the sample questionnaires.

The separation of the punching into these four operations simplified appreciably the work which had to be done by the keypunch operator since he could concern himself with only one kind of questionnaire and one kind of punchcard. The punching was further simplified by the use of the ED Master Card. By placing the ED Master Card in the reading station of the keypunch machine (type 024), the operator could duplicate automatically the first 14 columns, the geographic identification, on each card. Each punched card in its turn then served as an ED Master Card for automatically reproducing the geographic identification into the next card for the same ED.

In addition, an experienced general coder (called a contact coder because he was in daily contact with the cardpunching operation) was assigned to the operation to answer questions raised by the punchers regarding the correct codes to be punched for improper, illegible, or missing entries for some questionnaire items.

Productive punching was started in mid-June of 1960 and the entire operation, including the verification of the punching, was completed by January 27, 1961. Punching the 100-percent cards (i.e., those containing entries for only 100-percent items, for 75 percent of the population and housing units) began June 15, 1960, and was completed November 1, 1960. Punching the population and housing sample cards (100-percent and sample information) began July 15, 1960, and was completed January 27, 1961. Thirty persons were hired, trained, and participated in part or all of the card punching. This number includes two persons who were supervisors, and the contact coder.

A total of 2,349,544 population cards were punched, 621,000 of which were sample cards. The 100-percent population cards were punched at an average rate of approximately 2,300 enumerated persons per man-day, and the sample population cards were punched at an average rate of 800 enumerated persons in the sample per man-day. A total of 521,959 housing cards were punched, 134,000 of which were sample cards. The 100-percent housing cards were punched at an average rate of approximately 1,700 enumerated housing units per man-day, and the sample housing cards were punched at an average rate of 750 per man-day.

The significant difference in the rates of production between the punching of the 100-percent cards and the punching of the sample cards can be attributed to the many additional columns required for punching the sample information (45 more for a total of 80 columns on the population card; 32 more for a total of 74 columns on the housing card), plus the fact that codes for some of the sample questions were somewhat more difficult to punch. In addition, the sample questionnaire was more difficult for the keypunch operator to handle since it was in booklet form whereas the 100-percent questionnaire was designed in such a way that all the 100-percent items for a housing unit or a person could be obtained from one line.

The entire cardpunching operation required a total of 3,288 man-days.

The punched cards were verified by standard quality control procedures. (See "Quality Control," below.)

QUALITY CONTROL

Quality of the coding and punching operations was maintained through a system of verification designed to establish high and controlled levels of quality. The system located coders and punchers who might require additional instruction or training because their error rates were too high. The system also provided for removal of coders and punchers who failed to meet qualification standards, but all were able to qualify although a few did require additional training.

All errors noted during the verification operations were corrected.

The successful application of the system required that the work of a coder or keypunch operator be verified in the same order in which it had been produced, and that the time lag between coding or punching and verification be held to a minimum so that the personnel could be informed of their errors and the corrective action required be determined as quickly as possible.

Verification of General Coding

The work done by a general coder was verified on a 100-percent basis until it was determined that his error rate was low enough to qualify him for sample verification--i.e., to have only a portion of his work verified. Each general coder had to code the entries for a minimum of three ED's with a total error rate of no more than 4 percent for 100-percent data and no more than 5 percent for sample data before his work could qualify for sample verification. If he did not qualify after having coded entries for 10 ED's, he was given additional training. The error rate was based on the number of items in error per 100 enumerated persons. The coding of both the 100-percent data and the sample data were taken into consideration in determining the acceptability of a general coder's work, and the coder had to qualify in coding both types of data in order to qualify in either. The sample selected for verification of general coding was one in every 40 households in the 100-percent questionnaires and one in every 20 households in the sample questionnaires. That is, for general coding, the samples used established a verification level of 2.5 percent for the basic questionnaires and 5 percent for the sample questionnaires. After having qualified for sample verification, the coder had to continue to maintain an acceptable quality of work. This was assured through a system of process control for the coder and acceptance sampling for the individual ED's. For the coder, this consisted of a comparison of the quality control of his coding, based on a sample from three ED's, with a set of standards to determine if the coder was doing acceptable work or if he should be required to requalify, be retrained, or be removed from the operation. A comparison of the error rate found for the individual ED's with another set of standards was made to determine if the quality of the coding for the individual ED was acceptable or if the ED should be 100-percent verified.

The quality of the work of the verifier was also controlled. A system of preverification was used: After completion of general coding for an ED, but before it was assigned for verification, a preverifier reviewed the coding of a sample of the data and noted errors (but did not correct them). The preverification sample was determined using a table of random numbers to obtain a selection of one in 400 of the households enumerated in the basic enumeration book and one in 200 of the households enumerated on sample questionnaires. The preverification process also provided an independent estimate of the coding error rate as well as a means for determining whether the verifier or preverifier was finding a sufficient number of the existing errors. The determination as to whether both the preverifier and verifier were finding enough of the existing errors was based on the comparison of the amount of work reviewed and the number of errors found by each during a 3-week work period. Special mathematical tests were then applied to the results of this comparison in order to establish which one of the two was finding the most errors. Whenever the results of the tests indicated that either the preverifier or verifier was not finding enough of the errors, he was so informed, was reminded of the importance of the verification operation and, if necessary, was given additional training. Continued failure to meet the established standards resulted in removal from the operation.

Verification of Industry and Occupation Coding

The procedures employed in the verification of industry and occupation coding were identical to those used in the verification of general coding, except that different qualification levels were used because of (1) the limited number of items to be coded, (2) the large number of possible responses to the items, (3) the relatively small percent of persons enumerated in sample questionnaires who were in the labor force (an estimated 30 percent), and (4) the fact that only the sample questionnaires were involved.

Initial work of an industry and occupation coder was verified on a 100-percent basis as in the case of verification of the work of a general coder. However, an industry and occupation coder was not allowed to qualify until after he had coded entries for 10 ED's with acceptable error rate, and he was expected to qualify by the time he had coded entries for 30 ED's. A coder with an error rate of 5 percent or less qualified for sample verification. The error rate was based on the number of items in error per 100 persons enumerated. As in the case of the general coding, failure to maintain an acceptable level of quality required 100-percent verification of the coder's work until he either qualified again for sample verification, was retrained, or was removed from the operation. The coding for individual ED's was subjected to the acceptance sampling procedure.

A system of preverification was also employed for the evaluation of the verifier's work quality. One household in every 200 was selected for preverification. The quality of the work of the preverifier and verifier was tested in the same manner as that of the verification of general coding.

Verification of Card Punching

The work of a card puncher, like that of a coder, was verified on a 100-percent basis until his error rate was low enough to qualify him for sample verification. A card puncher was not allowed to qualify for sample verification, however, until after he had punched 1,500 cards, but was expected to qualify by the completion of 6,000 punched cards. (This same formula was used for each of the card types punched.) After once having qualified for sample verification, the puncher had to continue to produce work which would enable him to maintain an error rate of no more than 4 percent or his work was again verified on a 100-percent basis until he had requalified. Requalification had to be achieved before he had punched 1,200 more cards or the operator was subject to retraining or removal from the operations. The sample plan employed in the cardpunching operation called for verification of 5 percent of the cards punched by the operator. In practice this was achieved by verifying all the cards punched for one household in every 20. Since each card type had been punched separately, verification was handled separately for each card type.

The error rates were stated in terms of the number of punched cards containing one or more errors per 100 cards verified (rather than in terms of the number of incorrect punches). All cards discovered to have errors were, of course, corrected by punching new cards.

To maintain a control on the quality of the verifier's work, a system of reverification on a sample basis was employed. The procedure followed was similar to that used in the preverification phase of the coding operations.

The reverification sample for each card type was determined using a random number to start and selecting one household of every 200 thereafter. A card containing the punched data for a housing unit and all the cards pertaining to the population of a particular household were selected for reverification after the cards had been verified. The reverification process, like the preverification process in the coding operations, provided an independent estimate of the punching error rate as well as a means for determining whether the verifier or reverifier were finding a sufficient number of existing errors.

OTHER PRETABULATION OPERATIONS

With completion of the card punching and verifying, the remaining pretabulation operations began. These consisted of a machine check of the data on the punchcards for the purposes of (1) sorting the cards into groups representing particular combinations of data and gang-punching a unique code (referred to as the recode) in the cards to represent each group, and (2) locating and correcting inconsistencies in the data. Recodes are used because they make for simpler tabulation procedures inasmuch as they combine the information found in different columns on the punchcards, i.e., the information obtained in replies to more than one question. Recodes were established for the following items: employment status, number of persons per room, value-income ratio, gross rent as percent of income, and condition and plumbing of the housing unit.

Employment-Status Recode

The purpose of the employment-status recode was to indicate whether or not the individual concerned was in the labor force and, if so, whether he was employed, unemployed, etc. The recode was determined for each person 14 years old and over, by reference to the entries punched in column 45, "worked last week"; columns 46-47, "hours of work"; column 48, "looking for work or on lay off"; column 49, "had a job"; column 50, "date last worked"; columns 54-56, "occupation"; column 57, "class of worker"; and column 59, "place of work." Based on the combinations of entries found, a code was punched as indicated below:

<u>Employment status of persons</u>	<u>Recode</u>
At work	1
With a job but not at work	2
Unemployed	3
In the armed forces	4
Not in the labor force	5

A complete description of all possible "employment status" characteristics and the recodes assigned is included in instructions. The number of possible combinations is so vast that the 101 Statistical Tabulator lacks capacity to handle all cases. The residue were referred to specially trained coders who determined the proper recode.

Number-of-Persons-Per-Room Recode

The purpose of the number-of-persons-per-room recode was to assign to each housing unit a code which represented the relationship between the number of persons and

the number of rooms in each housing unit. Five classifications were established as follows:

<u>Average no. of persons per room</u>	<u>Recode</u>
0.75 or less	1
0.76 to 1.00	2
1.01 to 1.50	3
1.51 to 2.00	4
2.01 or more	5

Value-Income Ratio and Gross Rent as Percent of Income

The purpose of the value-income ratio recode and the gross-rent-as-percent-of-income recode was to assign to each housing unit a code which represented the relationship between the income in 1959 of the head of the household and either the value of his home if he were the owner or the gross rent if he were the renter. The following classifications were established:

<u>a. Value-income ratio</u>	<u>Recode</u>
Value of house less than 1.50 times 1959 income	1
Value of house 1.51 to 1.99 times 1959 income	2
Value of house 2.00 to 2.49 times 1959 income	3
Value of house 2.50 to 2.99 times 1959 income	4
Value of house 3.00 to 3.99 times 1959 income	5
Value of house 4.00 or more times 1959 income	6
<u>b. Gross rent as percent of income</u>	<u>Recode</u>
Gross annual rent less than 10 percent of 1959 income	1
Gross annual rent 10 to 14 percent of 1959 income	2
Gross annual rent 15 to 19 percent of 1959 income	3
Gross annual rent 20 to 24 percent of 1959 income	4
Gross annual rent 25 to 34 percent of 1959 income	5
Gross annual rent 35 or more percent of 1959 income	6

Condition-and-Plumbing Recode

Data on condition of housing units were combined with data on plumbing facilities to provide a single measure of housing quality, and codes representing specific levels of quality were assigned. A total of 11 different categories were established.

Mechanical Edit of the Remaining Data Fields

The recoding operation made it possible to correct for inconsistencies in data in the fields included in the recoding operation because the recodes assigned assumed all possible combinations of entries for the items involved. For the remaining items, determination of consistency was accomplished by means of a "mechanical edit" of the

punched cards. That is, the data were checked—using an IBM 101 machine—to locate impossible or inconsistent codes, for example, marital status entries of "divorced" for persons under 14 years of age, occupied housing units coded as vacant, etc. The number of mechanical edits that were carried out were voluminous and covered all data fields.

In practice, control panels of the 101 machine were wired for the mechanical edits in such a way that all the cards containing a common type of error would be sorted into a particular machine pocket. These cards were then corrected³ as required. For example, marital status codes of "married," "widowed," "divorced," and "separated" for persons under 14 years of age were always corrected by the assignment of the code for "never married." In other cases, reference to the original questionnaire had to be made in order to determine the appropriate entry.

When the corrected cards had been punched, they also were passed through the machine for mechanical editing, to make certain that no new errors had been introduced in the correction procedure and that the corrections themselves had been properly made. Because of the number of items included in the censuses and the many relationships existing between them, a total of 14 mechanical edit boards, each covering a different set of data, were required to complete the operation.

INFLATING THE SAMPLE

Before the cards punched from sample household could be tabulated, they were weighted, using a ratio-estimation technique; that is, the number of punched cards in the sample file was adjusted so that the tabulations obtained from the sample file would yield, area for area for a pre-determined set of characteristics, the same total counts that the 100-percent cards would produce.⁴ The procedures followed made it possible to assign a uniform weight of four to each card—a distinct advantage in the tabulation operations as it meant that the results of the sample tabulations had only to be multiplied by four to provide equivalent 100-percent totals.

In practice, the adjustment of the sample card file was accomplished by duplicating (more generally referred to as replicating) or eliminating selected sample cards or, occasionally, by the conversion of sample cards to 100-percent cards. Ratio estimates were handled on a "weighting area" basis; a weighting area was usually equivalent to a municipio, with the larger urban areas and municipios being further subdivided.

In total, 105 weighting areas were established. Of the 76 municipios, 69 were designated as separate weighting areas; 4 other municipios were divided into 2 weighting areas each, the urban and rural areas being separated; 1 municipio was divided into 3 weighting areas; and the 3 largest municipios, San Juan, Ponce and Mayaguez, in which the three cities bearing those same names were located, were divided along barrio and subbarrio boundaries into 16, 6, and 3 weighting areas respectively.

For each weighting area, the sample cards were sorted

³Punched cards, of course, are not actually corrected. Instead, new cards containing only correct data are punched and the rejected cards are destroyed.

⁴The sample cards could, of course, have been tabulated by multiplying by 4. However, processing the sample as was done provided more precise statistics, on the average, in the sample tabulations.

into the specified ratio-estimate groups and counts were obtained for each group to permit comparison with a similar set of counts which had earlier been obtained from the 100-percent cards. The ratio-estimate groups established were as follows:

A. Population cards

1. Males
 - a. Heads of households
 - b. Other males 14 years of age and over
 - c. Other males under 14 years of age
2. Females
 - a. Heads of households
 - b. Other females 14 years of age and over
 - c. Other females under 14 years of age

B. Housing cards

1. Owner-occupied housing units
2. Renter-occupied housing units
3. Vacant units

These counts, both 100-percent and sample (plus several balancing totals) were posted to a Ratio Estimate Control Record, and formed the basis for computation of the required weighting of the sample file.

The detailed instructions for the procedures and a reproduction of the Ratio Estimate Control Record are contained in the appendix.

The entire ratio-estimating operation had to be carried out with great care in order to insure accuracy of result as any errors would have automatically caused the tabulations from the sample to disagree among each other as well as to yield differing results from the 100-percent tabulations.

TABULATION

After completion of the mechanical edit and weighting operations, the punched cards were ready for tabulation, that is, the population and housing information punched in the cards could now be summarized to provide a wide range of population and housing information for the Commonwealth of Puerto Rico as a whole and for its various geographic subdivisions.

Tabulation Plan

The tabulation program was designed to produce the statistical data required to meet the needs of the publication program as it had been developed jointly by the professional staffs of the U. S. Bureau of the Census and the Puerto Rico Planning Board. In practice, publication requirements are presented in the form of "table outlines" which indicate both the various characteristics which are desired and the form in which the data are to be published. From these, the specifications for the tabulation processes are developed. These specifications describe how the punched cards were to be classified (grouped) and tabulated (counted).

The publication program for Puerto Rico required the tabulation of a considerable amount of data, and the tabulation program was designed to provide the information

with maximum efficiency and economy. In addition, consideration also had to be given to the priority of publications to ensure meeting publication deadlines and to have available the staff and equipment needed to complete the job. This required an analysis of the entire program in order to estimate workloads (that is, the required number of times the applicable population and housing punched cards would have to be passed through the tabulating equipment), the need for multiple shift operations, the use of summary cards, etc.

Although it had been planned originally to process all the Puerto Rican census in Puerto Rico, it quickly became evident that the volume of tabulation output required made this impossible if publication schedules were to be met. It was, therefore, decided to process the series PC-1(D) and HC(2) tabulations in the Washington headquarters of the Census Bureau, utilizing the Bureau's own tabulating equipment. A duplicate set of sample population cards was reproduced in Puerto Rico and shipped to Washington for this purpose; this made it possible to tabulate in both locations simultaneously.

A different set of tabulations was required for each of the publication series. Of these, the major ones are listed below in order of processing, together with a brief explanation of their purpose, the punched card file required, and the type of equipment utilized.

Tabulation in Puerto Rico

The tabulation process carried out in Puerto Rico as a whole employed a peak of 12 persons, took 10 months to complete, and required approximately 135 card passes, as follows:

100-percent population deck (2,349,544 cards):	31 card passes
100-percent housing deck (521,868 cards):	60 card passes
25-percent sample population deck (587,386):	79 card passes
25-percent sample housing deck (130,469):	65 card passes

The machines used in carrying out these operations were:

Type	Number
024 Keypunch	20
056 Verifier	15
083 Sorter	2
077 Collator	1
514 Reproducer	2
101 Electronic Statistical Machine	4
402 Accounting Machine	1
490 Census Processor Sta- tistical Machine	1
604 Calculating Punch (IBM Service Bureau)	1

Report PC(1)-53A, Number of Inhabitants. This report was obtained from the tabulations required for the PC(1)-

53B report (see below). It provides counts of the resident population for each separately identifiable political and statistical area in Puerto Rico (traced areas excepted) and shows their relationship to earlier censuses.

Report PC(1)-53B, Characteristics of the Population. This report presents statistics of the resident population by nativity, age, sex, marital status, and relationship to head of household for political subdivisions and statistical areas (except tracts) in Puerto Rico. Nativity and country of birth, State of birth of the population born in the United States, and citizenship of the foreign-born by age and sex are presented for the Commonwealth.

The data were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Sorting	Sample and 100-percent population cards using only the 100-percent data portion of the sample card	083	To separate the cards into the smallest geographic areas (urban places and barrios by residence) for which data were to be shown
Tabulation	Same as above	101	To provide counts of the 100-percent characteristics for each such area

Report PC(1)-53C, General Social and Economic Characteristics of the Population. This report provides statistics, based on a 25-percent sample, on the general social and economic characteristics of the resident population for each of the municipios and its urban and rural parts and for each urban place, urbanized area, and SMSA.

The data for this report were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Ratio estimation	Sample and 100-percent population cards	101 (also manual clerical operations)	To establish the proper number of sample cards
Sorting	Sample population cards	083	To separate the cards for each urban place and municipio by residence
Tabulation	Same as above	101	To provide counts of the 25-percent characteristics for each such area

Population and Housing PHC(1) Reports, for Census Tracts. These reports provide statistics, based on both the 100-percent and the 25-percent sample data, on selected social and economic characteristics of the resident population and characteristics of housing units by census

tracts for the three SMSA's in Puerto Rico: Mayaguez, Ponce, and San Juan.

The data for the three reports were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Sorting	100-percent and sample population and housing cards	083	To sort the cards by tract number within each of the three tracted areas
Tabulation	Same as above	101	To provide counts of the 100-percent and 25-percent characteristics for each census tract

Report HC(1)-53, Housing, Puerto Rico. This report presents basic statistics on the characteristics of housing units for Puerto Rico as a whole, by inside and outside SMSA's and by urban and rural areas, and for each SMSA, urbanized area, municipio, and place of 1,000 inhabitants or more.

The data were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Ratio estimation	Sample and 100-percent housing cards, using only the 100-percent data portion of the sample card	101 (also manual clerical operations)	To establish the proper number of sample cards
Sorting	Same as above	083	To separate the cards for each municipio and each place of 1,000 or more population
Tabulation	Sample and 100-percent housing cards	101	To provide data on the characteristics for each such area

Housing Series HC(3) Reports, City Block Statistics. A report for each of the major cities in the island, Mayaguez, Ponce, and San Juan, contained statistics on housing characteristics for each block and totals by census tracts within the specified city.

The data for these reports were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Sorting	Sample and 100-percent housing cards, using only 100-percent portion of sample cards	083	To sort the cards by block number within each tract for each of the three block cities
Tabulation	Same as above	402	To provide counts of the 100-percent characteristics for each block, including rent and value
Computation of average rent, value, and number of bedrooms	Same as above	604	To provide the required averages for each block

Tabulation in Washington

The tabulation process carried out in Washington, D.C., employed a peak of eight persons, took 6 months to complete, and required approximately 118 cardpasses distributed as follows:

25-percent sample housing deck (77,963 cards for 3 SMSA's):	28 card passes
25-percent sample population deck (587,386 cards):	90 card passes

The machines used in carrying out these operations were:

<u>Type</u>	<u>Number</u>
083 Sorter	2
490 Census Tabulator	3
407 Accounting Machine	1

Reports on Metropolitan Housing. These reports present detailed cross-tabulations of housing information based on the 25-percent sample for two SMSA's: Ponce and San Juan.

The data for these two reports were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Sorting	Sample housing cards	083	To separate the cards for each SMSA and central city and to group the cards within each area by selected housing characteristics
Tabulation	Same as above	490	To provide tabulations showing relationships between various housing and household characteristics

Report PC(1)-53D, Detailed Characteristics. This report presents detailed cross-tabulations based on the 25-percent sample data on social and economic characteristics of the population for Puerto Rico and selected areas.

The data for the report were obtained by means of the following mechanical operations:

<u>Operation</u>	<u>Cards Utilized</u>	<u>Type of Equipment</u>	<u>Purpose</u>
Sorting	Sample population cards	083	To separate the cards for each SMSA, city of 100,000 or more population, and the balance of the Commonwealth and to group the cards within each area by selected population characteristics
Tabulation	Same as above	490	To provide cross-tabulations showing relationships between various 25-percent characteristics of the population

Tabulation Control

As each tabulation run was produced, it was checked for accuracy, that is, for agreement with previously determined control totals and for absence of internal inconsistencies.⁵ Generally, the machine can be wired to check for inconsistencies in distributions provided that the distributions are either complete (that is, they equal the total number of cards being tabulated) or agree, in total, with other distributions based on the same universe. Where this is not possible, the distributions must be manually checked for correctness.

Obtaining agreement with previously established control totals was always a manual operation, and various types of worksheets were designed for this purpose. As each tabulation was produced, key totals were posted to the reconciliation sheet. If these were not identical to the totals obtained in previous tabulations for the same universe or characteristic, the error was located and corrected. Very small differences, however, were ignored (since no tabulation process can be expected to produce perfect results) and were manually adjusted in the table preparation operations.

TABLE PREPARATION

Table preparation, usually referred to at the Census Bureau as "result work" or "final results," follows tabulation. The end result of this operation is tabulated data arranged in table format ready for typing and printing for publication. (Typing and preparing the tables for publication was not handled in the Puerto Rican Opera-

⁵ Standard operating procedures always include the use of "test decks" of punched cards designed to ensure that the control panel of the tabulating machine has been wired according to specifications and that correctly tabulated results will be obtained. These same test decks, since they contain known totals, are also used to guard against machine malfunction. If a test deck tabulation (which is normally run several times a day) fails to come up with this predetermined result, the tabulation process is suspended until the cause of the error is located and corrected.

tions Office. This function was carried out centrally in Washington for all phases of the 1960 population and housing censuses.)

Preparation of Table Format Worksheets

Since it is normally not efficient (and may, in fact, be impossible) to produce a tabulation on conventional equipment in the exact format required for publication, some further manipulation of the data is required. This, for Puerto Rico, involved:

1. Manual addition of detail for small geographic areas in order to obtain totals for the larger geographic areas.
2. Posting the tabulated data to "table format worksheets." In the same operation for the sample tabulations, each number was multiplied by 4 in order to produce from the 25-percent sample the equivalent 100-percent estimates.
3. Adjustment of minor inconsistencies to ensure agreement with control totals and internal agreement within the distributions being shown.
4. Calculation of percentages, medians, ratios, etc., as required.
5. Verification of all data posted to the worksheet and calculations made therefrom.

Detailed instructions were prepared covering each phase of the operations for each tabulation and table worksheet. They indicated the source of the data, the key totals which were to be checked against control counts, the fields and characteristics to be added, the consistency checks to be applied, the computational techniques to be used, etc.

Table preparation operations began in October of 1960 and continued through completion of the sample tabulations for report PC(i)-53C in September of 1961. Approximately 1,180 man-days were used with a peak employment of 14 persons.