

Chapter II

THE SAMPLE

(17)

Contents

	Page
Reasons for the use of sample data.....	19
Some cost comparisons for sample and complete tabulations.....	19
The sampling unit.....	19
Stratification of the sample.....	19
Selection of sample farms.....	19
Geographic distribution of sample farms for Morrow County, Ohio.....	19
Map—Morrow County, Ohio—Distribution of farmsteads for 2-percent sample, compared with that for all farms, April 1, 1940.....	20
Table IV.—Number of 2-percent sample farms compared with all farms by Minor Civil Divisions, for Morrow County, Ohio: Census of 1940.....	21
Method of expanding the sample.....	21
Precision of data.....	21
"Recorded totals" shown in bold face type.....	22

Chapter II.—THE SAMPLE

Reasons for the use of sample data.—The 1940 Farm and Ranch Schedule provided the basic data necessary for making tabulations for all farms of the value of farm products by color and tenure of farm operator. The decision to use sample data as a basis for obtaining these statistics was based primarily upon cost considerations. In addition there was the possibility of securing experience which would be valuable in preparing plans for future censuses, either on a complete or on a sample basis.

Some cost comparisons for sample and complete tabulations.—The cost of pulling and coding the 2-percent sample cards and making one tabulation run *by States* has been estimated, for this work, at roughly one-sixth to one-eighth of the cost of making a single tabulation run *by counties* for all farms. In both cases, the cost of reproducing the codes from one punch card to another has been included; and the cost for the sample tabulation includes the complete tabulation for the farms with "\$10,000 and over" total value of farm products. The cost of coding and tabulating these large-income farms exceeded the cost of pulling, coding, and tabulating the 2-percent sample punch cards for farms with less than \$10,000 total value of farm products. The cost of making an additional run of the sample punch cards on a *State basis*, after the sample punch cards have been drawn and coded, was estimated in the neighborhood of 1 percent of the cost of making a single tabulation run *by counties* for all farms. These comparisons relate only to costs for machine tabulation and do not include costs of preparing the data for publication. The technical and clerical costs of processing these sample data have proved to be very much higher, on a per unit basis, than the corresponding costs for mass processing of complete tabulations.

The sampling unit.—Most studies of sampling techniques for proposed sample censuses of agriculture have centered about the problem of size-of-sampling unit; i.e., the problem of balancing enumeration difficulties and costs for small sampling units, such as individual farms, against the loss of information for large sampling units, such as clusters of farms. All of the 1940 Census data for one farm for a group of related items, such as value of farm products, were on a single punch card. Consequently, it was feasible to use the smallest available sampling unit, that is, the individual farm.

Stratification of the sample.—According to the 1940 Census of Agriculture, the number of farms in the United States varies from approximately 135,000 for the New England Division to nearly 1,100,000 for the West North Central Division. By tenure the number of farm operators for the United States ranges from approximately 36,000 managers to more than 3,000,000 full owners. It was recognized that it would be desirable to vary the sampling ratio of number of farms for each geographic area for which statistics were desired; likewise, it would be desirable to vary the sampling ratio of number of farm

operators by color and tenure groups. Administrative considerations made such procedures impractical; consequently, the following alternative was substituted:

The punch cards were sorted into two primary strata: Farms with less than \$10,000 and farms with \$10,000 or more total value of products. The farms with "\$10,000 and over" total value of farm products represented only 1 percent of all farms, but accounted for approximately 17 percent of the gross farm income for the United States in 1939. The punch cards for all farms with "\$10,000 and over" total value of farm products were sorted out and tabulated, i.e., a 100-percent sample was used.

For farms with less than \$10,000 total value of products a 2-percent sample was selected on a county basis; that is, the secondary stratification of the "Under \$10,000" strata was geographic.

Selection of sample farms.—The 2-percent sample of farms "Under \$10,000" was selected by machine by sorting out all punch cards with serial numbers ending in 15 and 65. The same terminal digit "5" was selected to reduce the number of cards handled in the subsequent sort; with two different terminal digits the number of cards handled in the subsequent sort would have been doubled, i.e., 1,200,000 compared with 600,000. The "15" was selected to reduce the number of counties excluded from the sample. There were only 22 counties in the United States, chiefly independent cities in Virginia, which reported less than 15 farms at the time of the 1940 Census and were, therefore, excluded from this 2-percent sample. The serial numbers had been placed on the individual farm schedules prior to the time that the data were transferred from the schedules to the punch cards. Since, for the most part, the schedules were in order of enumeration, there was no reason to infer any relationship between the characteristics of the farms and their serial numbers.

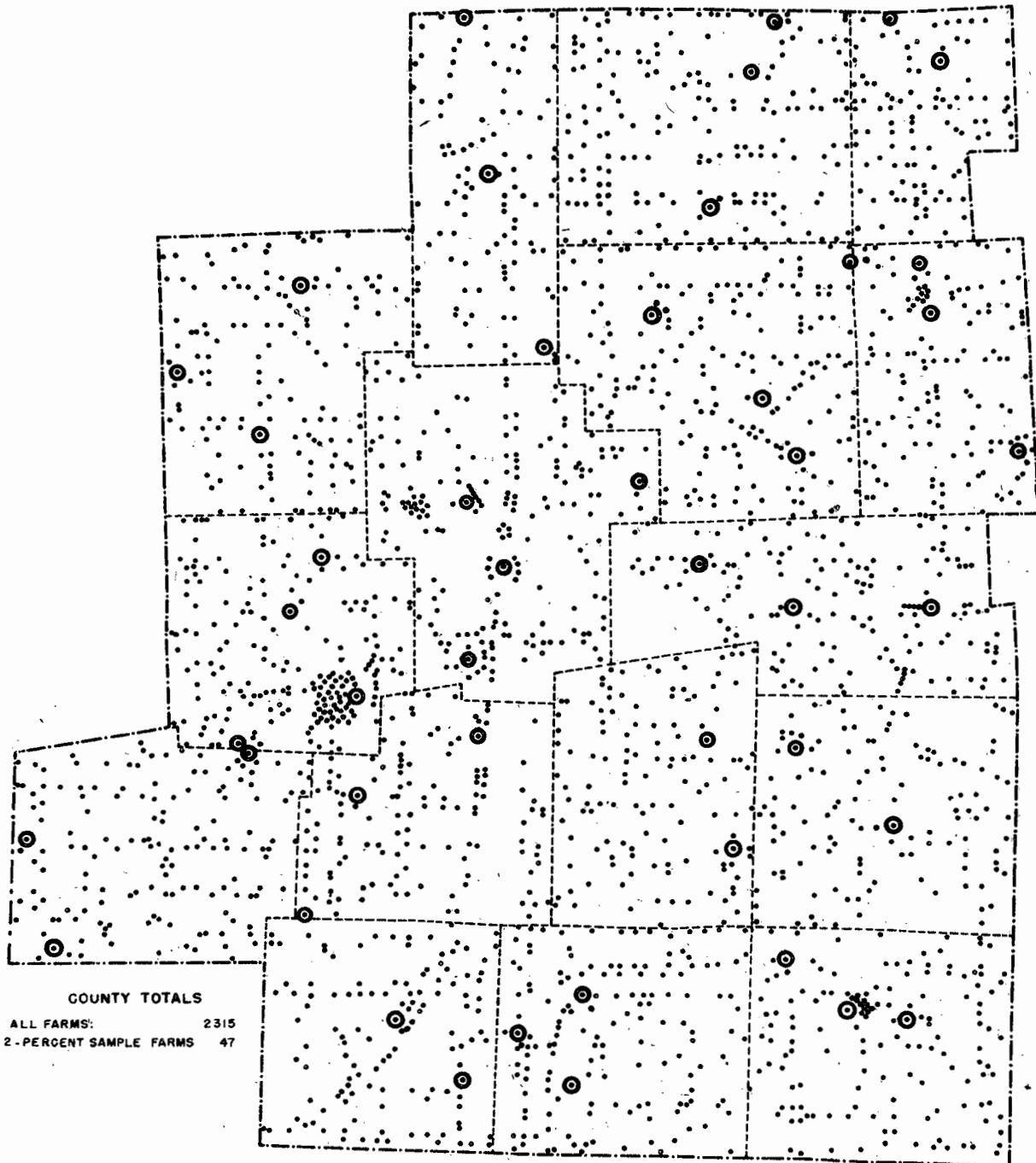
Geographic distribution of sample farms for Morrow County, Ohio.—The geographic distribution of the farmsteads, for the 2-percent sample farms compared with all farms, for Morrow County, Ohio, is illustrated in the accompanying dot map. The numbers of these farms are summarized by minor civil divisions in table IV. Morrow County, Ohio, was chosen for this illustration because, at the time this study was made, it was the only county for which a 1940 Census farmstead-location map was readily available. Although the machine selection of the sample farms was done by counties, the individual farm schedules had been arranged by minor civil divisions within each county prior to numbering. Consequently, the geographic distribution of the sample farms approaches a stratification by minor civil divisions within each county. To a considerable extent, the original order of enumeration was preserved in numbering the schedules. For this reason, the 2-percent sample also approximates a selection along the enumerator's route of every 50th farm in the "Under \$10,000" category.

VALUE OF FARM PRODUCTS

MORROW COUNTY, OHIO

DISTRIBUTION OF FARMSTEADS FOR 2-PERCENT SAMPLE, COMPARED WITH THAT FOR ALL FARMS

APRIL 1, 1940



COUNTY TOTALS
ALL FARMS: 2315
2-PERCENT SAMPLE FARMS 47

LEGEND

- = FARMSTEAD LOCATION
- = FARMSTEAD LOCATION FOR 2-PERCENT SAMPLE FARM
- - - = MINOR CIVIL DIVISION (TOWNSHIP) LINE

TABLE IV.—NUMBER OF 2-PERCENT SAMPLE FARMS COMPARED WITH ALL FARMS BY MINOR CIVIL DIVISIONS, FOR MORROW COUNTY, OHIO: CENSUS OF 1940

MINOR CIVIL DIVISION (TOWNSHIP)	NUMBER		Sample ratio
	All farms	2-percent sample farms	
Bennington.....	153	3	0.020
Canaan.....	141	3	.021
Cardington.....	194	4	.021
Chester.....	122	2	.016
Congress.....	186	4	.022
Franklin.....	149	3	.020
Gilead.....	194	4	.021
Harmony.....	125	2	.016
Lincoln.....	128	3	.023
North Bloomfield.....	167	3	.018
Perry.....	129	3	.023
Peru.....	124	2	.016
South Bloomfield.....	147	3	.020
Troy.....	101	2	.020
Washington.....	107	3	.028
Westfield.....	147	3	.020
County total.....	2,315	47	0.020

Method of expanding the sample.—Since complete tabulations were made for the "\$10,000 and over" strata, or subgroup, it was necessary to expand to a 100-percent level the 2-percent sample data for the "Under \$10,000" subgroup. By so doing, the figures for the two subgroups could be combined to secure the desired totals for color and tenure groups. Furthermore, the sampling was from a finite population, 1940 Census recorded totals for all farms (operators) being available for practically all individual items sampled,¹ such as number of unclassified farms, value of field crops sold or traded, number of vegetable farms, number of farms in the \$750-\$999 value group, etc. It was known that many users of the data would wish to compare these value figures with the complete tabulations of other items by color and tenure of farm operator. For this reason it was necessary to adjust the expanded 2-percent sample data to meet the recorded totals for all farms, i.e., all color and tenure groups. The procedure followed in expanding and adjusting the 2-percent sample was as follows:

As has been indicated in chapter I, recorded 1940 Census totals were available by States and geographic divisions for the number of farm operators in each color-tenure group. The 2-percent sample data for specified color-tenure groups were expanded to a 100-percent level by multiplying by the reciprocal of the sampling ratio, based on the number of all farms, for each color-tenure group. Another expansion factor, which could have been used, was the reciprocal of the sampling ratio based on number of all farms for the total sample, that is, for all color-tenure groups combined. Obviously this would have given less precision than that obtained by the use of sampling ratios for the individual color-tenure groups. This would be particularly true for those color-tenure groups for which the population numbers of farms are relatively small and the sampling ratios quite variable.

For the data shown in the State tables the expansions were made for two major-tenure groups in the North and West, (1) owners and managers and (2) all tenants; and for four major color-tenure groups in the South, (1) white owners and managers, (2) white tenants, (3) nonwhite owners and managers, and (4) nonwhite tenants. The figures shown in the geographic division tables were expanded in like manner for each of the respective

¹ Recorded totals were not available for number of farms reporting crops sold or traded, and/or forest products sold. See tables 3 and 11.

subtenure groups, viz, (1) full owners, (2) part owners, and (3) managers; and (4) cash tenants, (5) share-cash tenants, (6) share tenants, (7) croppers (South only), and (8) other tenants. In the South separate expansions were made for white and non-white operators for each of these subtenure groups.

Because of the small population, numbers of farms (operators) involved, and the relatively large sampling errors expected, the figures for certain States were combined before the sample data were expanded. These combinations were as follows: New England States; Delaware, Maryland, and the District of Columbia; New Mexico and Arizona; and Utah and Nevada.

For the same reason the State figures on sources of income, shown in the tables, were restricted to the following combinations of items: Farm products used by farm households, farm products sold or traded, livestock and livestock products sold or traded, crops sold or traded and forest products sold, livestock sold or traded, and livestock products sold or traded. The State data on major source of income presented in the tables were likewise restricted to the combinations of major-source groups corresponding to these combined sources of income.

For each item, such as number of field crop farms, the expanded sample data by color-tenure groups were scaled by a proportionate adjustment to equal the 1940 Census recorded total for that item. The geographic division totals of the State expansions of the sample data also provided pseudoestablished totals; and the geographic division expansions of sample data were scaled to equal these pseudoestablished totals by similar proportionate adjustments. For example, the geographic division total of the State figures (expanded sample data) on number of field-crop farms for owners and managers constituted a pseudoestablished total; and the geographic division expansions of the sample data on number of field-crop farms for full owners, part owners, and managers were adjusted proportionately to equal this previously established total.

The tabulations presented in this monograph involve the following principal cross-classifications: (1) Value of all farm products by source of income and color and tenure of farm operator; (2) number of farms by major source of income and color and tenure of farm operator; (3) value of all farm products for farms by major source of income and color and tenure of farm operator; and (4) number of farms by total value of farm products and color and tenure of farm operator. For these cross-classifications both the column totals and the line totals were either recorded or pseudoestablished totals. In such cases, the expanded sample figures for all cells of the cross-classifications were adjusted to equal both sets of marginal totals by a succession of horizontal and vertical proportionate adjustments.

For cross-classifications (1) and (3) the line totals, i.e., total value of farm products or value of all farm products for all classified farms for each color-tenure group, were pseudoestablished totals. For States, these pseudoestablished totals were based upon averages of the three adjusted totals derived from the expansions of the sample data by source of income, by major-source groups, and by value groups. For geographic divisions these pseudoestablished totals were based upon averages of the two adjusted totals derived from the expansions of the sample data by source of income and by major-source groups.

Precision of data.—As previously mentioned, the expanded 2-percent sample data were adjusted to the 1940 Census recorded totals for all color-tenure groups. This was done for the convenience of users of the statistics. Because of these adjustments, it was not feasible to round the figures, based on sample data, to indicate their approximate statistical precision.

It is to be expected that the relative sampling errors will be considerably smaller for color-tenure groups for which the population number of farms is generally large, such as full owners, than for color-tenure groups which have few farms, like managers. A similar difference in relative sampling errors is to be expected as between geographic divisions with large numbers of farms and geographic divisions with fewer farms. Likewise, the relative sampling errors will differ as between items, depending, in the case of the farm counts, upon differences in the proportion of farms reporting specified characteristics, and for quantitative items, upon differences in the variability of the items.

For these reasons, in the tables in this monograph, figures are marked with asterisks when the coefficient of variation, as indicated by sample data, was approximately 10 percent or more.¹ In using these figures marked with asterisks, it should be kept in mind that the sampling error for any such individual figure may be greatly in excess of 10 percent (either way). For the figures not marked with asterisks, it is unlikely that the values shown are in error by more than about 10 percent, and very unlikely that they are in error by more than approximately 20 percent. In evaluating the reliability of the figures, the consistency in the pattern of relationships should be taken

¹ For technical notes on the formulas used in this phase of the work, see Appendix.

into account as well as the indicated precision for individual cells.

"Recorded totals" shown in bold face type.—In the tables all 1940 Census recorded totals are shown in bold face type. The term "recorded totals," as here used, means totals for all farms, i.e., complete tabulations. Not all such recorded totals have been published in previous 1940 Census Reports on Agriculture. These "recorded totals" include (1) the column, or item, totals, i.e., totals of all color-tenure groups for each item, broken down by the "Under \$10,000" and "\$10,000 and over" subgroups; (2) the line or color-tenure group numbers of all farms; and (3) all figures for the "\$10,000 and over" subgroup in each color-tenure group. The data under (1) and (2) were published earlier in chapters X and III, respectively, of the 1940 Census, General Report on Agriculture. The statistics under (3) were tabulated especially for this report and are presented for the first time in this monograph.

The figures in roman type in the tables are based, entirely or in part, on sample data. As previously explained, the statistics for the "Under \$10,000" subgroup of each color-tenure group are the expanded and adjusted 2-percent sample data. The totals for each color-tenure group were secured by combining these expanded and adjusted sample data for the "Under \$10,000" subgroup with the complete tabulations for the "\$10,000 and over" subgroup.